Genomic organization of TEL: the human ETS-variant gene 6.

M Baens, P Peeters, C Guo, et al.

Genome Res. 1996 6: 404-413
Access the most recent version at doi:10.1101/gr.6.5.404

RESEARCH

Genomic Organization of *TEL*: The Human *ETS*-variant Gene *6*

Mathijs Baens, Pieter Peeters, Caiying Guo, Jeroen Aerssens, and Peter Marynen[1]

Human Genome Laboratory, Center for Human Genetics, Flanders Interuniversity Institute for Biotechnology, University of Leuven, Leuven, Belgium

We have constructed a detailed map of the genomic region containing the *ETS*-variant gene *6* (*ETV6*), involved in translocations and deletions associated with hematologic malignancies. Thirty-eight cosmids were characterized belonging to two contigs spanning 340 kb, and an *Eco*RI restriction map was developed. The gap between the two contigs, 2 kb in size, was closed by PCR. The contigs contain the complete coding sequence and the 5′ and 3′ UTRs of *ETV6*. Eight exons accounting for the *ETV6* cDNA sequence were identified. The helix–loop–helix (HLH) motif is coded by exons 3 and 4, whereas exons 6–8 code for the *ETS* DNA-binding domain. All introns show consensus 5′ donor and 3′ acceptor splice sites. Introns 1 and 2 span 100 and 82 kb, respectively, and introns 3–7 range from 15 to 1.3 kb. An alternative exon 1 (exon 1B) is localized in intron 2. The 5′ end of the *ETV6* gene is associated with a CpG island characterized by the presence of four *Not*I, four *Sac*II, and three *Bss*HII recognition sites and several *SP1*- and *AP2*-binding motifs. Alternative polyadenylation at the 3′ end of the *ETV6* gene generates the three transcripts of 6200, 4300, and 2400 nucleotides, respectively. The *ETV6* gene spans 240 kb and is flanked at its 5′ and 3′ end by *D12S1697* and *D12S98*, respectively. The markers *D12S1095* and *D12S89* are located in the first intron. Two new DNA polymorphisms were identified in the *ETV6* gene, which will be useful for the analysis of loss of heterozygosity reported for the *ETV6* gene in leukemia.

The *ETS*-variant gene *6* (*ETV6*, previously named *TEL*) was cloned by Golub et al. (1994) by virtue of its involvement in the t(5;12)(q33;p13) found in a chronic myelomonocytic leukemia (CMML). The *ETV6* gene encodes a protein with a putative helix–loop–helix (HLH) dimerization domain at the amino terminus, conserved in a subset of *ETS* family members, and an *ETS* DNA-binding domain at the carboxyl terminus. In the t(5;12) a 5′ *ETV6–PDGFRB* 3′ fusion occurs linking the HLH domain of *ETV6* to the transmembrane and tyrosine kinase domains of *PDGFRB*. The reciprocal fusion transcript is not expressed (Golub et al. 1994).

ETV6 appears to be involved in different leukemia-associated translocations. In rare cases of childhood acute lymphoblastic leukemia (cALL) with t(9;12)(q34;p13) the HLH domain is fused to the tyrosine kinase domain of *ABL*, which then shows increased tyrosine kinase activity (Papadopoulos et al. 1995). In myeloproliferative disorders with t(12;22)(p13;q11), resulting in the fusion of *ETV6* to the *MN1* gene on 22q11, expression of both chimeric transcripts is reported (Buijs et al. 1995). It is suggested that the 5′ MN1–ETV6 3′ fusion protein is important for the transformation of early myeloid precursors. This fusion protein contains almost the entire MN1 protein linked to the carboxy-terminal part of the ETV6 protein, providing an *ETS* DNA-binding domain, and might act as an altered transcription factor. *ETV6* is also implicated in the pathogenesis of cALL through its fusion to the *AML1* gene on 21q22 (Golub et al. 1995; Romana et al. 1995a). The chimeric transcript encoding the *ETV6* HLH domain fused to the *AML1 runt* and *trans*-activation domains is predicted to have oncogenic potential, as the reciprocal 5′ *AML1–ETV6* 3′ transcript is expressed only in a subset of patients (Raynaud et al. 1996). Remarkable is the systematic deletion of the normal *ETV6* allele in patients with 5′ *ETV6–AML1* 3′ fusions, which results in the loss of wild-type *ETV6* function in the leukemic cells (Golub et al. 1995; Romana et al. 1995a). Independently, Stegmaier et al. (1995) and Cavé et al. (1995) detected frequent loss of

[1]**Corresponding author.**
E-MAIL Peter.Marynen@Med.KULeuven.ac.be; FAX 32-16-34.59.97.

6:404–413 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

heterozygosity (LOH) of the *ETV6* region in cALL. It is now clear that these cases represent a cryptic t(12;21) and a 5′ *ETV6–AML1* 3′ fusion (Raynaud et al. 1995).

Fluorescence in situ hybridization (FISH) analysis of 12p abnormalities revealed that *ETV6*, together with *CDKN1B*, is frequently deleted in hematopoietic malignancies (Sato et al. 1995; Höglund et al. 1996; Wlodarska et al. in press), which has led to the speculation that *ETV6* might also have tumor-suppressor activity.

The *ETV6* gene is therefore implicated in the pathogenesis of a number of distinct hematologic malignancies through the fusion of either its HLH or its *ETS* DNA-binding domain to different translocation partners, which is sometimes associated with loss of the second *ETV6* allele. In the majority of the cases, the t(12;21) and the rearrangement involving chromosomes 9 and 12 are not detectable by classic cytogenetics but can be detected by FISH (Romana et al. 1995b). In addition, genomic *ETV6* probes are needed to analyze the extent of *ETV6* deletions. To develop tools to investigate the promiscuous involvement of the *ETV6* gene in leukemia we have constructed a detailed physical map of the *ETV6* gene, characterized its genomic structure, and generated additional polymorphic markers within the *ETV6* gene.

RESULTS

Construction of a Cosmid Contig for the *ETV6* Gene

To isolate the *ETV6* gene, first an *ETV6* cDNA fragment (bp 19–1567; Golub et al. 1994) and a 170-bp probe for alternative exon 1B (Romana et al. 1995a) were hybridized to the arrayed LL12NCO1 library (Montgomery et al. 1993). Twenty-four cosmids were isolated. Restriction analysis and Southern hybridizations with the *ETV6* cDNA revealed three nonoverlapping cosmid contigs. Contig 1 (179A6, 29F2, 51E2, 168G1, 118A6, and 15A4/15B5) contained exon 1 of *ETV6* (see below); contig 2 (50F4 and 95H4/97D12) contained exon 2, and contig 3 (258D11, 171H6, 242E1, 2G8, 159A7/158H3, 189A3, 45E12, 163E7, 184C4, 54D5, 24G12, 88A9, and 148B6) contained exons 3–8 and the alternative exon 1B.

A walking strategy was then applied to close the gaps between the three contigs. The ends of cosmids that flank the gaps were rescued by vectorette PCR and used as probes to screen the LL12NCO1 library. The cosmids isolated with 15A4T3 (67C6/68F8 and 145A11) and 50F4T7 (67C6/68F8, 6C11/7F1, and 205H1) did overlap and joined the first and second contigs. Southern hybridizations with the rescued T7 and T3 ends of the isolated cosmids confirmed the overlaps (Fig. 1). Cosmid 145A11, however, was very unstable, which might explain the underrepresentation of this region in the LL12NCO1 library. Restriction analysis (*Eco*RI) of a 10-kb fragment (TELLRP1), generated by long-range PCR and spanning the region of minimal coverage in the first intron, confirmed the restriction patterns of cosmid clones 67C6 and 145A11 (data not shown).

Five additional cosmids were isolated with the T7 end of 171H6 that extended the contig 25 kb farther into the second intron (Fig. 1). Subsequent attempts to close the second gap using the T7 and T3 ends of 95H4 and 132B11, respectively, were hampered by the absence of the intervening sequence in the LL12NCO1 library. Direct visual hybridization (DIRVISH) experiments with cosmids 50F4 and 219D7, however, revealed that the gap was <15 kb (data not shown). Long-range PCR was therefore used to amplify the intervening region. A 2100-bp PCR product (TELLRP2) was obtained using both human DNA and DNA of YAC 958B8, containing the *ETV6* gene, as the template DNA with primers derived from sequences of both cosmid end fragments. No amplification product was obtained with DNA of the cosmid pools of the LL12NCO1 library, suggesting the absence of this region in the library. To demonstrate that TELLRP2 closed the gap between the two cosmid contigs, both ends were sequenced and shown to overlap with the sequences of the cosmid end fragments used for primer construction. Furthermore, hybridization of genomic Southern blots with the T7 and T3 end probes of 95H4 and 132B11 detected *Eco*RI restriction fragments of 3.4 and 2.8 kb, respectively, as is expected by adding the length of the *Eco*RI end fragments of the cosmids to the length of the *Eco*RI end fragments of TELLRP2 (data not shown). Finally, the T3 end fragment of cosmid 148B6 was used for a walking experiment and yielded three cosmids (136B11, 167A6, and 244E8) extending an additional 40 kb toward the centromere (Fig. 1).

The analysis of the *Eco*RI restriction patterns of all isolated cosmids, combined with the analy-

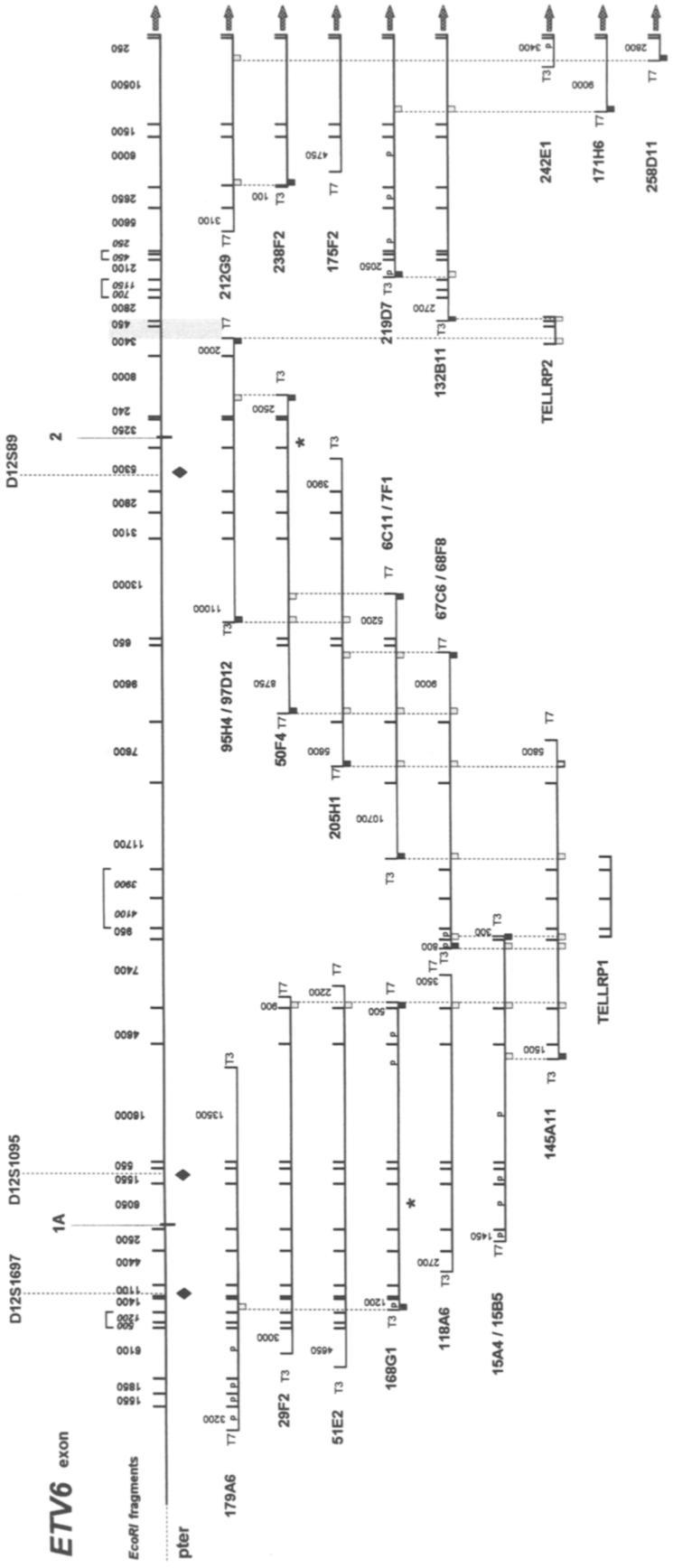

Figure 1 *(Continued on facing page.)*

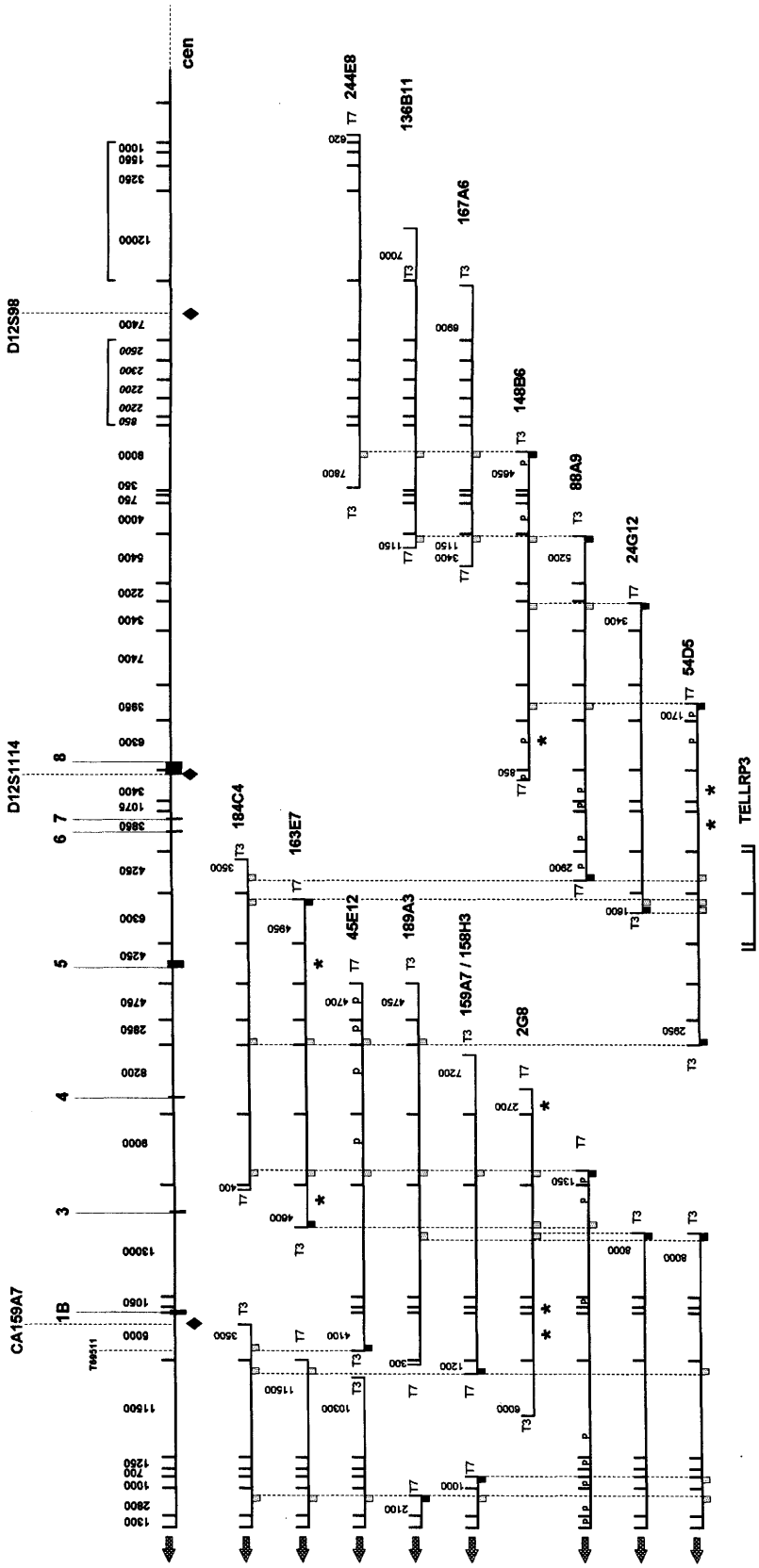

Figure 1 Genomic structure of the *ETV6* gene. Two cosmid contigs covering the *ETV6* gene are shown. All cosmids were derived from the LL12NCO1 chromosome 12 cosmid library, and the corresponding addresses are given. Both contigs are linked by a PCR clone (TELLRP2). The region absent in the LL12NCO1 library is indicated by the shaded area. The positions of the eight *ETV6* exons and six chromosome 12 STSs are indicated above the *Eco*RI restriction map. The size of the *Eco*RI fragments is indicated in bp. When the relative order of consecutive fragments is not known, they are indicated with a bracket. The T7 and T3 ends of the cosmid inserts are shown. (■) Ends rescued by vectorette PCR; (□) *Eco*RI fragments detected with the end clones by Southern analysis; (◆) the position of the STSs present in the contigs. An asterisk (*) marks the fragments subcloned to determine the exon/intron boundaries. (p) Partial digestion and hybridization with T7 and T3 primers was used to determine the order of the restriction fragments.

sis of partial *Eco*RI digests for 9 cosmids and Southern hybridizations with 34 rescued T3 or T7 ends of selected cosmids (Fig. 1) allowed us to construct a detailed *Eco*RI restriction map of the region harboring the *ETV6* gene. Double digests of cosmids using *Eco*RI in combination with several rare cutting restriction enzymes (*Not*I, *Sac*II, *Mlu*I, *Sfi*I, and *Sal*I) allowed us to integrate them into this map and identified a CpG island with four *Not*I and four *Sac*II sites at the 5′ end of the *ETV6* gene (Fig. 2B).

Structure of the *ETV6* Gene

Hybridizations of oligonucleotides derived from the *ETV6* cDNA to Southern blots of *Eco*RI-digested cosmids identified exon-containing fragments. These bands were isolated, cloned in pGEM-3Z, and partially sequenced. Comparison of the genomic and cDNA sequences revealed the exon organization and the exon–intron boundaries of the *ETV6* gene. Eight exons were identified (Fig. 2A). The sequences at the boundaries are reported in Table 1: All of the splice junctions showed the canonical GT/AG dinucleotides (Breathnach et al. 1978). This revealed further that the HLH domain is coded by exons 3 and 4, whereas the *ETS* DNA-binding domain is present in exons 6, 7, and 8 (Fig. 2A). The alternative exon 1B was found upstream of exon 3 (Fig. 1). A $(CA)_n$ repeat was identified immediately upstream of exon 1B (Fig. 3; GenBank accession no. U45431). Furthermore, sequencing the ends of the 5000-bp *Eco*RI subclone containing exon 1B revealed the presence of a sequence identical to a short 298-bp expressed sequence tag (EST) (accession no. T69511).

The partial cDNA for *ETV6*, published by Golub et al. (1994), contains 57 bp of exon 1A, which was sequenced in a 6000-bp *Eco*RI subclone of 168G1. Additional sequences of the 5′-untranslated region (UTR) of the *ETV6* cDNA were obtained following a single-strand ligation to single-stranded cDNA (SLIC) protocol (Dumas Milne Edwards et al. 1991). Amplification products of ~70 and 140 bp were obtained using an antisense primer derived from exon 1A (bp 12–30). Sequence analysis of eight clones identified three different products containing 34, 39, and 108 bp, respectively, upstream of position +1, which were colinear with the genomic sequence (Fig. 4).

Northern analysis with the *ETV6* cDNA detects three transcripts of 2400, 4300, and 6200 nucleotides, respectively (Golub et al. 1994). To investigate the origin of the different transcripts, we first screened the GenBank data base with the 1580-bp *ETV6* cDNA. Three overlapping

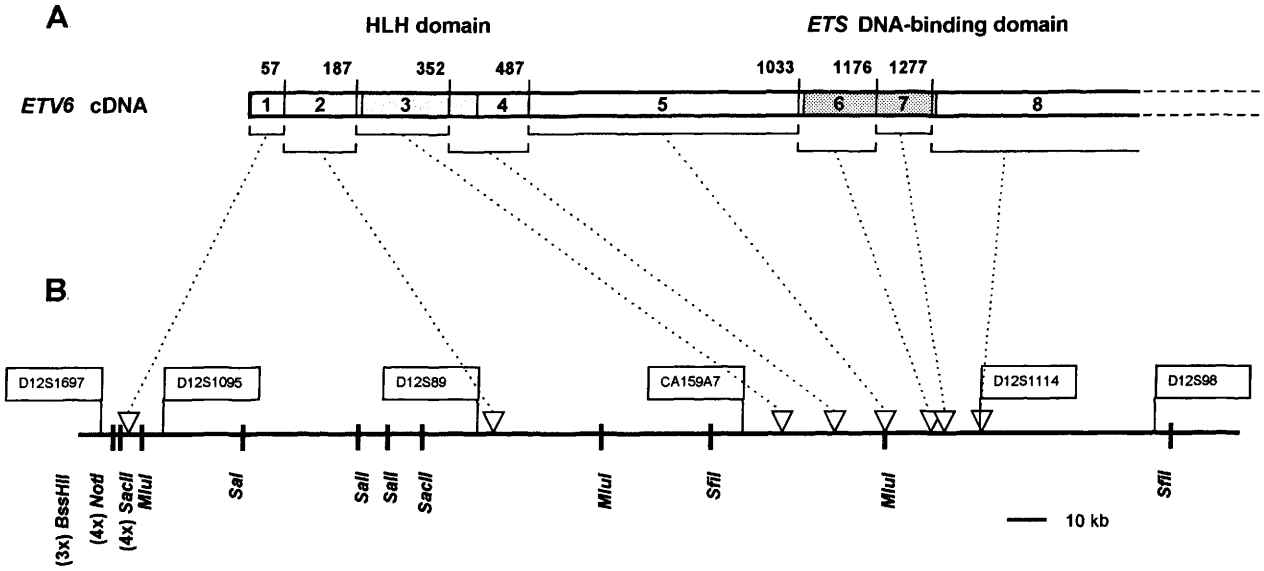

Figure 2 (*A*) The exon structure of the *ETV6* gene. The position of the last base of each exon is indicated above the *ETV6* gene according to the numbering of the sequence reported by Golub et al. (1994). The positions of the putative HLH domain and the *ETS* DNA-binding domain are indicated. (*B*) Physical map of the *ETV6* gene. The map was determined by restriction digests of the cosmid clones with *Not*I, *Sac*II, *Sfi*I, *Mlu*I, and *Sal*I. (▽) The exons. Also indicated are the positions of the chromosome 12 STSs. The scale in kb is indicated.

Table 1. Exon and Intron Junctions of the Human *ETV6* Gene

Exon		Sequence at exon/intron junctions			
no.	nucleotide position in cDNA[a]	5′ splice donor	intron	size (kb)	3′ splice acceptor
1A	1 - 57	CATTAAG gtaaaa	I	100	tatttttaacag CAGGAAC
2	58 - 187	CACCTGC gtgagt	II	82	tctctccaacag GCTTGCA
3	188 - 352	CATTCAG gtgaga	III	14.1	ttttctttccag GTGATGT
4	353 - 487	GAAGAAG gtactg	IV	15.4	tctgctccacag ATAACTG
5	488 - 1033	ATAGCAG gtgagt	V	14.2	tttcctctgtag ACTGTAG
6	1034 - 1176	CCATAAG gtaaaa	VI	1.3	tttgaacaaacag AACAGAA
7	1177 - 1277	TGTTCAG gtagca	VII	5.2	cttcttccaaag GTTTATG
8	1278 -				
1B		CTGTAGG gtcaga			

[a]The numbering is based on the human *ETV6* cDNA sequence (Golub et al. 1994).

cDNAs were identified (accession nos. T79177, T79261, and T91395) extending 371 bp toward the 3′ end of the gene and indicated that an *Eco*RI restriction site flanked the 3′ end of the published sequence (bp 1580). Sequencing the ends of the 3400-bp *Eco*RI fragment, which hybridized with exon 8, and its flanking 6300-bp fragment confirmed that exon 8 extends in the latter over the *Eco*RI site. A detailed restriction map was constructed for this 6300-bp *Eco*RI fragment, and different fragments were then used for Northern analysis (Fig. 5). A cDNA probe containing the coding region of *ETV6* detects the three described transcripts, probes A and B detect the signals of 4300 and 6200 nucleotides, and probe E hybridizes only with the longest transcript. Probes C and D do not yield a signal upon Northern analysis, indicating that this 6300-bp *Eco*RI fragment delimits the *ETV6* gene and that the reported transcripts are the result of differential polyadenylation in the 3′ region.

gccaggggattagtaacctttacaactctaatactagggaatatattaataactggtgag
aaaaaagattacagctttttattgataggtctataaaagccatttccgtatgaatgtcc
tcagaaaggaggaataaggggatgagatagaatagaatagagagaagtctgtctcacaca
cacacacacacacacacacacacacacacacacacacaaatgcaacacaattatctttaaat
agaagttgcttctccaagtttgtttttttttgtttgtttgtttcatgaggcgtcagcaac
gctgttaaaatgatttcatcctgcctgTGTGTCACAGCAGTCTGCCCCGCGAAATGCTGA
ATCAGCAGCAGCCCAGACAGCCTTTCGGTGGAGGGTAGATTCGACACCCAAAGGCAACCA
ACCGGGCCTTTGAAAGCACTTGGAAGCACTCTCCAAAGAGCCAGGGAAGAGGACAAGGAA
GAAGGGGGTGCAGTGGAATTCTCCTTAAAAAACAGCATGTTCTAGAAAAGCCATGCTTTG
GAGGTGCTGTAGGgtcagatttg

Figure 3 Characterization of the 5′-upstream region of exon 1B. The nucleotide sequence of exon 1B is shown in uppercase letters, the internal *Eco*RI site (box) and the 5′ splice donor (italics) are indicated. The primers used for amplification of the $(CA)_n$ repeat are marked by arrows. The sequence was submitted to GenBank under accession no. U45431.

The *ETV6* Promoter Region

Three 5′-end cDNAs of different length were isolated applying a SLIC protocol (Fig. 4). Sequence analysis revealed the absence of additional 5′ exons and suggests that transcription starts in the genomic region preceding exon 1A (Fig. 4). To identify promoter elements, we determined 3 kb of sequence 5′ to exon 1A (GenBank accession no. U45432). Sequence analysis revealed a high GC content, especially in the 1500 bp upstream of exon 1A, with the presence of four *Not*I, four *Sac*II, and three *Bss*HII recognition sequences. Southern analysis with a single-copy probe flanking exon 1A showed the *Not*I and *Sac*II sites to

```
-420 TTCTGCACTGAAACTCTCAAGATCAATGAGCAAAGAGCTTTCTCAGTTCTGCCTTTCAGTTTCTCTCTTC  -351

-350 CAGGAAGGAAAACATTCGAGAGAGCGAGGGAGAGCCGCGGGAGGGCGGGGGCGGGGGCGCCGGCTGCGGG  -281

-280 TGGGAGGAGAGACCGGGAGGCCGGCCGGGCTGCGTCCCGGGTCCCCGCGCGCGCCGCGACCTGCAGACC  -211

-210 CCGCCGCCGCGCTCGGGCCCGTCTCCCACGCCCCCGCCGCCCCGCGCGCCCAACTCCGCCGGCCGCCCCG  -141

-140 CCCCGCCCCGCGCGCTCCAGACCCCCGGGGCGGCTGCCGGGAGAGATGCTGGAAGAAACTTCTTAAATGA   -71
                                  *

-70 CCGCGTCTGGCTGGCCGTGGAGCCTTTCTGGGTTGGGGAGAGGAAAGGAAAGTGGAAAAACCTGAGAACT    -1
                                  * *

+1  TCCTGATCTCTCTCGCTGTGAGACATGTCTGAGACTCCTGCTCAGTGTAGCATTAAGgtaaaaatcttct  +70
    <-------- primer 256
```

Figure 4 Characterization of the 5′ end of the *ETV6* gene. The nucleotide sequence of the 5′-flanking region of *ETV6* is shown. The sequence is numbered relative to the start of the *ETV6* cDNA (Golub et al. 1994), which is marked as +1. The 5′ ends of the cDNAs isolated by SLIC are indicated with asterisks. Consensus sequences factor-binding sites, identified by GCG software, are underlined (*SP1*) or boxed (*AP2*). The sequence has been submitted to GenBank/EMBL under accession no. U45432.

be unmethylated (P. Peeters, unpubl.). No TATA or CAAT boxes could be identified, as has been described for the promoters of different housekeeping genes. A sequence analysis shows the presence of numerous *SP1* and *AP2* recognition sequences as can be expected for a CpG island (Fig. 4).

Integration of the *ETV6* Gene in the Physical Map

Krauter et al. (1995) presented a sequence-tagged site (STS)-based yeast artificial chromosome (YAC) contig map of human chromosome 12. The presence of markers flanking *D12S1114* (exon 8 of *ETV6*) was evaluated to integrate the *ETV6* gene in the physical map of the short arm of chromosome 12. The *ETV6* gene spans 240 kb and is flanked by *D12S1697* (AFM333wb5), 7 kb upstream of exon 1A, and *D12S98*, ~50 kb downstream of exon 8. *D12S1095* and *D12S89* are present in the first intron immediately downstream (6 kb) and upstream (<7 kb) of exons 1 and 2, respectively. Our map determines the relative order of *D12S1697* and *D12S1095*, which remained undefined in the YAC-based map (Krauter el al. 1995).

New Polymorphic Markers in the *ETV6* Gene

A $(CA)_n$ repeat was identified immediately upstream of exon 1B by partially sequencing the exon 1B subclone (Fig. 3). PCR analysis using 32 unrelated individuals showed this markers to be polymorphic: Fifteen alleles were identified with a heterozygosity of 0.7.

When long-range PCR was applied to generate a probe for the genomic region between exons 5 and 6 (TELLRP3; Fig. 1), *Eco*RI restriction analysis of the amplification product (10 kb) revealed a discrepancy with the established map of *ETV6*. A 5600-bp *Eco*RI fragment was visualized instead of the expected 6300-bp fragment. Southern analysis using a single-copy *Xba*I fragment

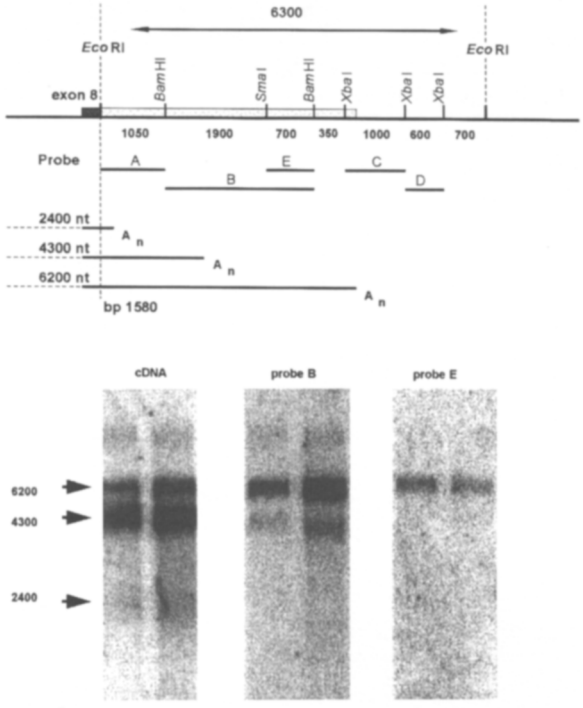

Figure 5 Northern blot analysis of poly(A)$^+$ RNA from human fibroblasts using an *ETV6* cDNA probe containing the coding region and probes derived from the 6300-bp *Eco*RI subclone of cosmid 148B6 containing the 3′ UTR of the *ETV6* gene. The sizes of the *ETV6* transcripts are indicated. The restriction map of the subclone (*left*) shows the positions and the sizes of probes A, B, C, D, and E and the restriction enzymes required for their purification.

(2000 bp) derived from the *Eco*RI fragment revealed the existence of two alleles of 5600 and 6300 bp, respectively. One hundred thirty-four chromosomes were analyzed, and the frequency of both alleles was determined to be 0.44 (5600 bp) and 0.56 (6300 bp).

DISCUSSION

In this study a 340-kb contig covering the entire *ETV6* gene locus was constructed using 33 unique cosmid clones and one PCR amplification product. A restriction map for *Eco*RI, *Not*I, *Sac*II, *Mlu*I, *Sfi*I, and *Sal*I was then generated, and six chromosome 12p STSs were localized within this region. Overall, a fourfold coverage of the region was obtained. Regions with lower coverage were shown to be unstable or absent in the LL12NCO1 library and were analyzed further using long-range PCR. A minimum tiling path (15A4, 67C6, 6C11, 95H4, TELLRP2, 132B11, 242E1, 184C4, 88A9, 136B11) covers 300 kb of the *ETV6* gene locus and has only 45 kb of overlap (15%). Thus, these cosmids can serve as a resource for determining the complete sequence of the *ETV6* gene locus.

The *ETV6* gene is a member of the *ETS* family of transcription factors (Wasylyk et al. 1993). The highest homology is found with the subfamily containing *ETS1* and *ETS2*. The members of this subfamily encode a protein with a HLH domain at the amino terminus and an *ETS* DNA-binding domain at the carboxyl terminus. *ETS1* also consists of eight exons with a very large first intron and a long (6 kb) 3′ UTR (Jorcyk et al. 1991).

Sequence analysis of the putative promoter region of *ETV6* further revealed the presence of recognition sequences for *SP1* and *AP2* factor-binding sites and the absence of a CAAT or TATA box. The same characteristics were reported for the promoter region of *ETS1*. Promoters lacking a TATA box are found for many housekeeping genes, which correlates with the expression of *ETV6* in all tissues examined (Golub et al. 1994). The presence of an unmethylated CpG island with several recognition sequences for transcription factors immediately upstream of exon 1A of *ETV6* and the analysis of the 5′-end cDNAs for the gene isolated by means of a SLIC protocol strongly suggests transcription initiation in this region. Primer extension experiments or a RNase protection assay need to be performed to identify the transcription start sites.

Northern analysis with *ETV6* cDNA consistently detects three transcripts of 6200, 4300, and 2400 nucleotides, respectively. Consecutive probes generated for the 6.3-kb region flanking the published sequences of exon 8 progressively detect only the longer transcripts. In all cases, the positions of the different probes according to the restriction map of the region coincide with the lengths of the detected transcripts. Although no sequence data are available, these experiments strongly suggest the presence of alternative polyadenylation sites that give rise to the three transcripts described. However, the existence of an additional (small) intron in the 6300-bp *Eco*RI fragment cannot be excluded. In the cell lines examined by us, the expression of the smallest transcript is consistently low. Romana et al. (1995a) reported variation of the relative levels of the three transcripts in different cell lines.

The alternative exon 1B (Romana et al. 1995a) was identified 12 kb upstream of exon 3. No in-frame ATG start codon was identified in the 223-bp sequence of exon 1B. Translational initiation at the third in-frame ATG (nucleotides 289–291, exon 3) would generate a shortened protein that partially lacks the putative HLH motif. Interestingly, the transcription start site at exon 1B is proximal to the class I breakpoint found in the t(12;22) cases (Buys et al. 1995). The expression of the shortened ETV6 protein could therefore be unaffected by this translocation. Sequence analysis of exon 1B in the cosmid clones confirmed the existence of a putative 5′ splice donor site. The 5′ upstream region of exon 1B, however, contains a $(CA)_n$ polymorphic repeat. The promoter activity of this region needs to be analyzed further.

The detection of an EST in intron 2 in intriguing in view of its involvement in myeloproliferative disorders with class I breakpoints. The absence of an open reading frame in EST T69511 suggests that the clone is derived from the 3′ UTR region of a gene in intron 2 of *ETV6* or is a genomic contaminant present in the cDNA library. The latter is supported by the fact that no additional ESTs with this sequence are present in the Human cDNA Database (HCD) [The Institute for Genomic Research (TIGR) data base; Adams et al. 1995].

LOH at the *ETV6* gene locus was reported is association with t(12;21) in cALL. Two new DNA polymorphisms were detected within the *ETV6* gene that will aid the analysis of LOH. A polymorphic $(CA)_n$ repeat was identified upstream of alternative exon 1B. The second polymorphism was identified by performing long-range PCR be-

tween exons 5 and 6. PCR generated an amplification product 700 bp shorter than expected from the map data. Restriction analysis showed that the polymorphism originated from a deletion/insertion in the 5600/6300 *Eco*RI fragment. In patients showing t(12;21)(p13;q22), the translocation breakpoint occurs in the fifth intron of the *ETV6* gene. Therefore, when analyzing cALL samples for potential rearrangements of *ETV6* by Southern hybridization, the occurrence of a size polymorphism in intron 5 needs to be taken into consideration.

METHODS

Isolation and Characterization of Cosmid Clones

Cosmid clones were derived from the arrayed human chromosome 12 library LL12NCO1 (Montgomery et al. 1993) by colony hybridization using ^{32}P-labeled probes. A 1549-bp *ETV6* cDNA probe was generated by PCR using primers derived from the published cDNA sequence (Golub et al. 1994) and liver cDNA as template. A probe for alternative exon 1 (exon 1B) (Romana et al. 1995a) was a gift from R. Berger (Institut de Génétique Moléculaire, INSERM U301, Paris, France). Cosmid end fragments were rescued by ligating a vectorette unit (Riley et al. 1990) to *Rsa*I-digested cosmids, followed by PCR with the T3 or T7 vector primers together with a primer complementary to the vectorette. *Eco*RI-digested cosmid DNA was electrophoresed in 0.7% agarose gels and blotted onto Hybond-N$^+$ membranes (Amersham) using 0.4 N NaOH. Exon-containing fragments were identified using oligonucleotides derived from the *ETV6* cDNA. An *Eco*RI restriction map was generated from the *Eco*RI digestion patterns of overlapping clones, by probing Southern blots of *Eco*RI cosmid digests with rescued T3 or T7 end-fragments of selected cosmids (Fig. 1), and by hybridizing Southern blots of partial *Eco*RI digests of *Sfi*I-digested clones with ^{32}P-labeled T3 and T7 oligonucleotides.

DIRVISH Analysis

Cosmid DNA was labeled by nick translation with biotin-11–dUTP (Sigma) or digoxigenin-16–dUTP (Boehringer Mannheim) using a commercially available kit (Life Technologies) and used as probe to construct a high-resolution visual map of stretched DNA by fluorescence hybridization (DIRVISH). DNA stretching and hybridization were performed as described by Parra and Windle (1993). The signals were visualized by digital imaging microscopy using a cooled charge-coupled device camera (Photometrics Ltd., Tucson, AZ) and Smart Capture software (Imagenetics, Stuttgart, Germany).

PCR and SLIC Analysis

Long-range PCR was performed with *Taq* and *Pwo* DNA polymerases according to the recommendations of the manufacturer (Boehringer Mannheim).

TELLRP1 was generated using primers 404 (5′-ggagtggtacagagatgcatacagcaactt-3′) and 471 (5′-gtaacttgcccgaagttcatgattcattg-3′), derived from the T3 end sequences of cosmid 15A4 and 6C11, respectively. TELLRP2 was generated using primers 402 (5′-ctttctgctataccatgcggccacctgaaa-3′) and 403 (5′-accattccaatgcaatctaattgcaaagtg-3′) derived from the T3 and T7 end sequences of cosmid 132B11 and 95H4, respectively.

TELLRP3 was generated using primers 400 (5′-gccaacacatccgcattgcagagaaacat-3′) and 401 (5′-ggtggttgaggtcagtgtattgctgtgaat-3′), derived from sequences of subclones containing exons 5 and 6/7, respectively.

The 5′ end of the *ETV6* cDNA was cloned following the SLIC protocol (Dumas Milne Edwards et al. 1991). Human brain cDNAs with adaptor sequences ligated to their 5′ ends (Clontech) were amplified using an antisense primer derived from exon 3 (5′-ctgaatgaggagatcgatagcg-3′) and the adaptor. Heminested amplification was performed with 1/100 of the first reaction product using the adaptor primer and an internal *ETV6* primer (exon 1A; 5′-agacatgtctcacagcgag-3′). After phosphorylation, fragments were cloned in pUC18, linearized with *Sma*I, and sequenced.

DNA Sequencing and Analysis

The exon–intron boundaries were sequenced from genomic *Eco*RI fragments subcloned in pGEM-3Z (Fig. 1; fragments marked with an asterisk). All primers were derived from the *ETV6* cDNA. The 5′-upstream region was sequenced by primer walking. Nucleotide sequencing was carried out by dideoxynucleotide chain termination with FITC–dATP or fluorescently labeled primers and analyzed on an ALF sequencer (Pharmacia). Cosmid end fragments were rescued by vectorette PCR and sequenced using a solid-phase approach with a biotinylated vectorette primer and FITC-labeled T3 or T7 primers. Data bases were searched using the BLASTN algorithm (Altschul et al. 1990). The sequence of the promoter region was analyzed with the GCG software.

ACKNOWLEDGMENTS

These investigations have been supported by grant G.0153.96 from the Nationaal Fonds voor Wetenschappelijk Onderzoek of Belgium, by the Interuniversity Network for Fundamental Research sponsored by the Belgian government (1991–1996), and by the Flemish government project VIS/94/16. P.M. is an Onderzoeksdirecteur of the Nationaal Fonds voor Wetenschappelijk Onderzoek, Belgium. We thank the IWT (vlaams Instituut voor de bevordering van het Wetenschappelijk-Technologisch onderzoek in de industrie) for a scholarship to P.P., Hilde Braeken, and Riet Somers for their expert technical assistance, and Karel Rondou for the photography. The sequence data described in this paper have been submitted to the GenBank data library under accession nos. U45431–U45432.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Adams, M.D., A.R. Kerlavage, R.D. Fleischmann, R.A. Fuldner, C.J. Bult, N.H. Lee, E.F. Kirkness, K.G. Weinstock, J.D. Gocayne, O. White, et al. 1995. Initial assessment of human gene diversity and expression pattern based upon 83 million nucleotides of cDNA sequence. *Nature* suppl. **377:** 3–174.

Altschul, S.F., W.W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Breathnach, R., C. Benoist, K. O'Hare, F. Gannon, and P. Chambon. 1978. Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the intron-exon boundaries. *Proc. Natl. Acad. Sci.* **75:** 4853–4857.

Buijs, A., S. Sherr, S. van Baal, S. van Bezouw, D. van der Plas, A. Geurts van Kessel, P. Riegman, R. Lekanne Deprez, E. Zwarthoff, A. Hagemeijer, and G. Grosveld. 1995. Translocation (12;22)(p13;q11) in myeloproliferative disorders results in fusion of the *ETS*-like TEL gene on 12p13 to the *MNI* gene in 22q11. *Oncogene* **10:** 1511–1519.

Cavé, H., B. Gerard, E. Martin, C. Guidal, I. Devaux, J. Weissenbach, J. Elion, E. Vilmer, and B. Grandchamp. 1995. Loss of heterozygosity in the chromosomal region 12p12-13 is very common in childhood acute lymphoblastic leukemia and permits the precise localisation of a tumor-suppressor gene distinct from p27KIP1. *Blood* **86:** 3869–3875.

Dumas Milne Edwards, J.B., J. Delort, and J. Mallet. 1991. Oligodeoxyribonucleotide ligation to single-stranded cDNAs: A new tool for cloning 5′ ends of mRNAs and for constructing cDNA libraries by in vitro amplification. *Nucleic Acids Res.* **19:** 5227–5232.

Golub, T.R., G.F. Barker, M. Lovett, and D.G. Gilliland. 1994. Fusion of PDGF receptor β to a novel *ETS*-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell* **77:** 307–316.

Golub, T.R., G.F. Barker, S.K. Bohlander, S.W. Hiebert, D.C. Ward, P. Bray-Ward, E. Morgan, S.C. Raimondi, J.D. Rowley, and D.G. Gilliland. 1995. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci.* **92:** 4917–4921.

Höglund, M., B. Johansson, J. Pedersen-Bjergaard, P. Marynen, and F. Mitelman. 1996. Molecular characterisation of 12p abnormalities in hematologic malignancies—Deletion of *KIP1*, rearrangement of *TEL* and amplification of *CCND2*. *Blood* **87:** 324–330.

Jorcyk, C.L., D.K. Watson, G.J. Mavrothalassitis, and T.S. Papas. 1991. The human *ETS1* gene: Genomic structure, promoter characterization and alternative splicing. *Oncogene* **6:** 523–532.

Krauter, K., K. Montgomery, S. Yoon, J. Leblanc-Straceski, B. Renault, I. Marondel, V. Herdman, L. Cupelli, A. Banks, J. Lieman, J. Menninger, P. Bray-Ward, P. Nadkarni, J. Weissenbach, D. Le Paslier, P. Rigault, I. Chumakov, D. Cohen, P. Miller, D. Ward, and R. Kucherlapati. 1995. A second-generation YAC contig map of human chromosome 12. *Nature* **377:** 321–333.

Montgomery, K.T., J.M. Leblanc, P. Tsai, J.S. Mcninch, D.C. Ward, P.J. De Jong, R. Kucherlapati, and K.S. Krauter. 1993. Characterisation of two chromosome 12 cosmid libraries and development of STSs from cosmids mapped by FISH. *Genomics* **17:** 682–693.

Papadopoulos, P., S.A. Ridge, C.A. Boucher, C. Stocking, and L.M. Wiedemann. 1995. The novel activation of ABL by fusion to an *ETS*-related gene, TEL. *Cancer Res.* **55:** 34–38.

Parra, I. and B. Windle. 1993. High resolution visual mapping of stretched DNA by fluorescent hybridisation. *Nature Genet.* **5:** 17–21.

Raynaud, S.D., H. Cavé, M. Baens, C. Bastard, V. Cacheux, J. Grosgeorge, E. Vilmer, P. Marynen, and B. Grandchamp. 1996. The 12;21 translocation involving *TEL* and deletion of the other *TEL* allele: Two frequently associated alterations found in childhood Acute Lymphoblastic Leukemia. *Blood* **87:** 2891–2899.

Riley, J., R. Butler, D. Ogilvie, R. Finniear, D. Jenner, S. Powell, R. Anand, J.C. Smith, and A.F. Markham. 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18:** 2887–2890.

Romana, S.P., M. Mauchauff, M. Leconiat, I. Chumakov, D. Le Paslier, R. Berger, and O.A. Bernard. 1995a. The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. *Blood* **85:** 3662–3670.

Romana, S.P., H. Poirel, M. Leconiat, M.A. Flexor, M.M.P. Jonveaux, E.A. Macintyre, R. Berger, and O.A. Bernard. 1995b. High frequency of t(12;21) in childhood B-lineage acute lymphoblastic leukemia. *Blood* **86:** 4263–4269.

Sato, Y., Y. Suto, J. Pietenpol, T.R. Golub, D.G. Gilliland, E.M. Davis, M.M. Le Beau, J.M. Roberts, B. Vogelstein, J.D. Rowley, and S.K. Bohlander. 1995. TEL and KIP1 define the smallest region of deletions on 12p13 in hematopoietic malignancies. *Blood* **86:** 1525–1533.

Stegmaier, K., S. Pendse, G.F. Barker, P. Bray-Ward, D.C. Ward, K.T. Montgomery, K.S. Krauter, C. Reynolds, J. Sklar, M. Donelly, S.K. Bohlander, J.D. Rowley, S.E. Sallan, D.G. Gilliland, and T.R. Golub . 1995. Frequent loss of heterozygosity at the TEL gene locus in acute lymphoblastic leukemia of childhood. *Blood* **86:** 38–44.

Wasylyk, B., S.L. Hahn, and A. Giovane. 1993. The *ETS* family of transcription factors. *Eur. J. Biochem.* **211:** 7–18.

Wlodarska, I., P. Marynen, R. La Starza, C. Mecucci, and H. Van den Berghe. 1996. The *ETV6, CDKN1B* and *D12S178* loci are involved in a segment commonly deleted in various 12p aberrations in different hematological malignancies. *Cytogenet. Cell Genet.* **72:** 229–235.

Received January 17, 1996; accepted in revised form April 2, 1996.