

# The *tcl-3* proto-oncogene altered by chromosomal translocation in T-cell leukemia codes for a homeobox protein

Ming Lu, Zhiyuan Gong<sup>1</sup>, Weifang Shen<sup>2</sup> and Anthony D.Ho

Northeastern Ontario Regional Cancer Centre and Department of Medicine, University of Ottawa, 41 Ramsey Lake Rd, Sudbury, Ontario, Canada P3E 5J1, <sup>1</sup>Banting Institute, University of Toronto, 100 College St, Toronto, Ontario, Canada M5G 1L5 and <sup>2</sup>Department of Medicine, University of California at Davis, Davis, CA 94553, USA

Communicated by T.Graf

**The t(10;14)(q24;q11) chromosomal translocation found in malignant cells of 5–10% of patients with T-cell acute lymphoblastic leukemia (T-ALL) involves the T-cell receptor  $\delta$  chain gene on chromosome 14 and a breakpoint cluster region on chromosome 10. The candidate proto-oncogene *tcl-3*, thought to be involved in the pathogenesis of t(10;14) T-ALL, was cloned and found to be elevated in expression in leukemic cells harboring the t(10;14) translocation. Sequence analysis revealed that *tcl-3* is a new homeobox-containing gene. Comparison of the *tcl-3* cDNA and its 5' genomic sequences with DNA sequences from the t(10;14) translocation breakpoints showed that this gene is structurally altered in four patients with t(10;14)(q24;q11) T-ALL. These findings suggest that homeobox-containing genes that normally act as transcription factors may contribute to T-cell leukemogenesis when abnormally expressed.**

**Key words:** chromosomal translocation/gene rearrangement/homeobox/oncogene/T-cell leukemia

## Introduction

It has been well established that non-random chromosomal translocations constitute an important mechanism for neoplastic transformation of hematopoietic cells. The best studied examples are translocations involving the *c-myc* gene in Burkitt's lymphomas, the *bcl-2* gene in non-Hodgkin's follicular lymphomas and the *c-abl* gene in chronic myelogenous leukemias as well as in some acute leukemias. Many T-cell leukemias carry chromosomal translocations (Hecht *et al.*, 1984; Ueshima *et al.*, 1984; Williams *et al.*, 1984; Raimondi *et al.*, 1988) and some of these translocation breakpoints have recently been characterized (for reviews see Boehm and Rabbitts, 1989; Tycko and Sklar, 1990). The genes identified near the chromosomal translocation breakpoints in leukemic T-cells include *ttg*, a zinc-finger protein associated with the t(11;14)(p15;q11) translocation (McGuire *et al.*, 1989), and *lyl-1* and *scl* (*tal-1/tcl-5*), proteins each with a helix-loop-helix DNA binding domain associated with the t(7;19)(q35;p13) and t(1;14)(p32;q11) translocations (Mellentin *et al.*, 1989; Begley *et al.*, 1989; Chen *et al.*, 1990), respectively.

The acquired chromosomal translocation t(10;14)(q24;q11) is a primary chromosome change found in the

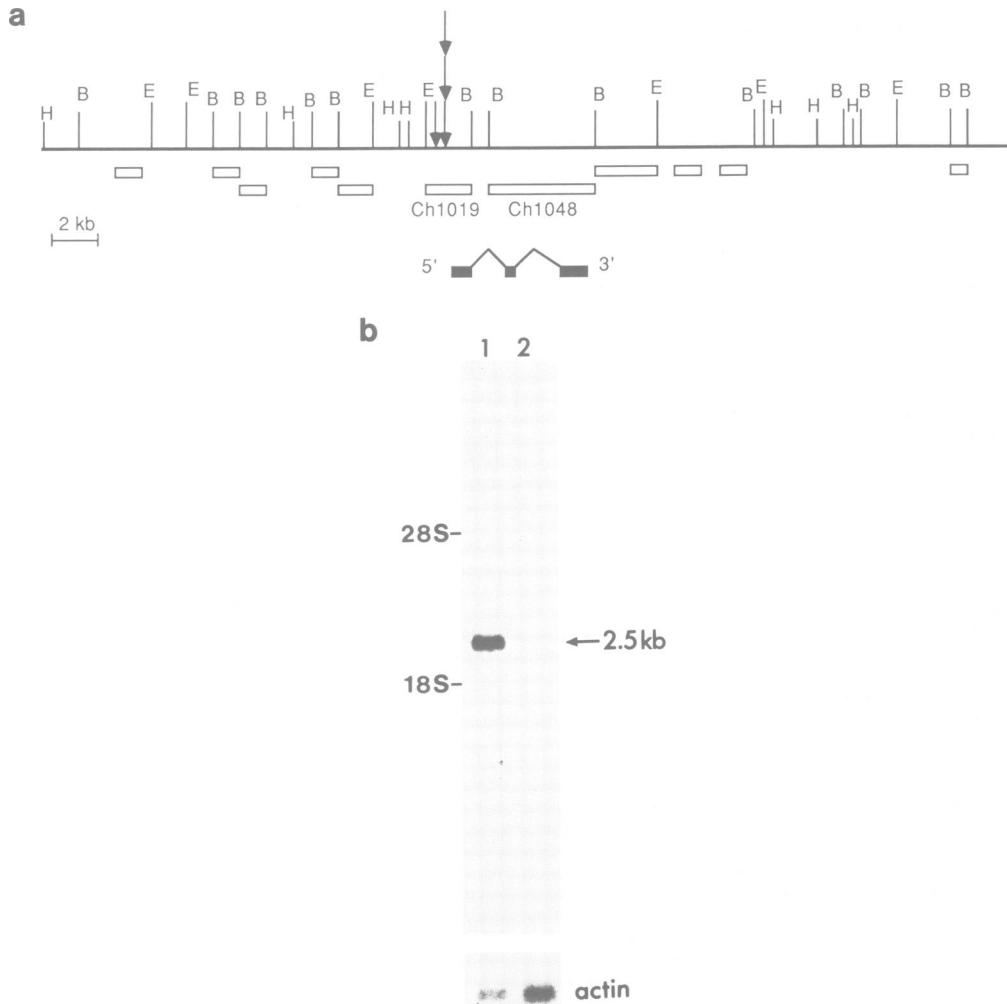
malignant cells of 5–10% of patients with T-cell acute lymphoblastic leukemia (T-ALL). Molecular cloning and sequence analysis revealed that the t(10;14) chromosomal translocation breakpoints occurred in the T-cell receptor  $\delta$  chain gene (TCRD) on chromosome 14 and in a 9 kb DNA region on chromosome 10 (Kagan *et al.*, 1989; Lu *et al.*, 1990; Zutter *et al.*, 1990). The tight clustering of t(10;14) translocation breakpoints on chromosome 10 suggests that the translocation may result in dysregulation of a proto-oncogene in 10q24, termed *tcl-3*, via illegitimate physiological recombination with TCRD. However, no candidate proto-oncogene has been identified thus far.

We report here the isolation of a novel homeobox gene from the t(10;14)(q24;q11) breakpoint cluster region. This gene is overexpressed in a cell line harboring the t(10;14)(q24;q11) chromosomal translocation. Sequence analysis revealed that this gene shares homology in its homeodomain with a number of other homeobox-containing genes that code for sequence-specific DNA binding proteins. Moreover, our data show that this homeobox gene is structurally altered in four patients with T-ALL, indicating that the putative proto-oncogene *tcl-3* was cloned. These findings suggest that a homeobox gene that normally functions as a transcription factor may play an important role in leukemogenesis.

## Results

In previous studies we and others have shown that the t(10;14) chromosomal translocation breakpoints are clustered on chromosome 10 (Kagan *et al.*, 1989; Lu *et al.*, 1990; Zutter *et al.*, 1990). To identify the putative *tcl-3* gene, which may have a role in the pathogenesis of t(10;14)(q24;q11) T-ALL, chromosomal walking was conducted and a genomic DNA region from 10q24 of ~45 kb cloned (Figure 1a). From this region, 11 single-copy sequences were isolated and used as probes to search for a transcription unit dysregulated by the t(10;14) translocation. The probe Ch1048 identified a 2.5 kb mRNA in a leukemic cell line designated SIL, which harbors the t(10;14) translocation (Figure 1b). This transcript, however, was not detectable in a similar pre-T cell line DND41 that does not contain the t(10;14) translocation (Figure 1b). Northern analysis was extended to RNA isolated from human T-cell lines (SKW3 and Jurkat), myeloid cell line (HL60), B-cell lines (Nalm6, Hoon, Daudi and Cordick), human thymus and tonsil cells, and bone marrow cells from three patients with T-ALL that do not carry the t(10;14) translocation. No transcript was identified using Ch1048 as probe (data not shown).

Screening of a SIL cDNA library with probe Ch1048 resulted in the isolation of nine overlapping cDNA clones. Sequence analysis yielded a composite cDNA sequence of 2137 bp (Figure 2). This cDNA terminates in a stretch of adenosine residues separated by 18 bp from a consensus AATAAA polyadenylation signal (Figure 2). A single major open reading frame of 1234 bp was found ending with a



**Fig. 1.** A physical map of the *tcl-3* locus, the t(10;14) translocation breakpoints in 10q24, detection of the *tcl-3* transcript as well as its alignment with respect to the genomic region. (a) Repeat-free DNA fragments are shown in open boxes. Exons are indicated by filled boxes. The four t(10;14) translocation breakpoints are indicated by arrows. They are clustered in a 112 bp DNA stretch within the DNA fragment Ch1019 (see Figure 4). Vertical bars indicate restriction sites. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hind*III. (b) Northern blot analysis showing elevated expression of the putative proto-oncogene *tcl-3* in the leukemic cell line, SIL, containing the t(10;14)(q24;q11) translocation. The 4.8 kb *Bam*HI genomic fragment, Ch1048, identified a 2.5 kb mRNA in 1  $\mu$ g of poly(A)<sup>+</sup> RNA isolated from SIL (lane 1). The transcript was not detectable in 2  $\mu$ g of poly(A)<sup>+</sup> RNA isolated from the pre-T cell line DND41 that does not carry the t(10;14) translocation (lane 2). The filter was deprobed and rehybridized to  $\beta$ -actin. Size markers (kb) were estimated with RNA ladders (Gibco/BRL).

TGA termination codon at nucleotide 1255. Examination of this sequence showed an in-frame ATG codon at nucleotide 265 in a favorable context (AGCATGG) for translation initiation (Kozak, 1989). The open reading frame codes for a 330 amino acid polypeptide, extended from nucleotide 265 at the 5' end to the termination codon TGA at nucleotide 1255. An upstream in-frame TGA termination codon was also identified 5' of the methionine codon at nucleotide 265. A primer extension analysis identified two major sites for transcription initiation. They were located 52–53 and 234 nucleotides upstream from the 5' end of the cDNA, respectively. The size of the cDNA, therefore, corresponds closely to the size of the mRNA estimated from Northern blots, given an average of ~200 residues in the poly(A) tail.

To investigate the possible function of the *tcl-3* gene product, we searched the GenBank database for homologous sequences. Significant homology between amino acids encoded by nucleotides 865–1044 of *tcl-3* and the 60 amino acids that comprise the homeoboxes encoded by other distinct

genes was observed (Figure 3). The *tcl-3* gene product contains the four invariant amino acids, tryptophan, phenylalanine, asparagine and arginine at positions 48, 49, 51 and 53, respectively (Figure 3). These four amino acids are present in all homeodomains with the exception of the yeast genes (Scott *et al.*, 1989). Among the eight additional amino acids that are highly conserved (Scott *et al.*, 1989), seven are found in *tcl-3* at positions 5, 12, 16, 20, 45, 55 and 58, respectively (Figure 3). The homeodomain for *tcl-3* showed the highest amino acid homology with k8 (50% identical). The homeobox gene k8 has been implicated in playing a role in development of the hematopoietic system (Kongsuwan *et al.*, 1988). No significant homology with other proteins was identified in regions outside the homeodomain, suggesting that *tcl-3* is a new member of the homeobox-containing gene family.

To search for possible structural changes in the *tcl-3* gene, DNA sequences of the t(10;14) translocation breakpoints were compared with the *tcl-3* cDNA and genomic DNA

CGGGCTTTTTCGCCTGGTGATTGATGTCCCAGAGTCAACAGCGAGCGAGCCGGAGCGGGGAAGCAGAAGCCA	75
***	
GAGAGGGGAAGAATACGGCGCCCCCTCTCCCTCCCCCTCCCTTCTACTTTAGCCTTTCTGCGCACTTCGCT T	150
CCAAGTCTCCGCGCAGCCAGGAGCCGCTGTGCTCCAGCCCTGCTAGCTGCCCGGAGCCGAGCGCAGCGA	225
GCGCCCGCGCCCGGGCCCCCGGTGGGGCCAGGGCCAGCATGGAGCACCTGGGTCCGACCCACCTCCACCCGGGT	300
M E H L G P H H L H P G	12
CACGCAGAGCCATTAGCTTCGGCATCGACCAGATCCTCAACAGCCCGGACCAGGGTGGCTGCATGGGACCCGCC	375
H A E P I S F G I D Q I L N S P D Q G G C M G P A	37
TCGCGCTCCAGGACGGAGAATACGGCCTTGGCTGCTTGGTTCGGAGGCGCCTACCTTACGGCGCGGGGGCTCC	450
S R L Q D G E Y G L G C L V G G A Y T Y G G G G S	62
GCGCCCGCAGCGGGGTGGAGGAGCGGGGGCTATGGTACTGGAGGTCCCGCGGGCCCCGAGGGCCCGCAGGC	525
A A A T G A G G A G A Y G T G G P G G P G G P A G	87
GGCGCGCGCCTGCAGCATGGGTCTCTGACCGCTCTACAACGTGAACATGGCCTTGGCAGGCGGGCCCCGGT	600
G G G A C S M G P L T G S Y N V N M A L A G G P G	112
CCTGGCGCGCGCGCGCAGCAGCGCGGTGCCGGGGCACTAGCGCTGCCGGGGTAATCCGGGTGCCGGCACAC	675
P G G G G S S G G A G A L S A A G V I R V P A H	137
AGCCCGCTCCCGGAGCTGTGGCCACCCCGACCCCTGGCCACCGGCTTGGCCACCGTGCCTCTGTGCTGCC	750
R P L A G A V A H P Q P L A T G L P T V P S V P A	162
ATGCCGGCGCTCAACAACCTCACTGGCCTCACCTTCCCCTGGATGGAGAGTAACCGCAGATACACAAGGACAGG	825
M P G V N N L T G L T F P W M E S N R R Y T K D R	187
TTCACAGGTACCCCTATCAGAACCGGACGCCCCCAAGAAGAAGCCGCGCACGTCTTACACCGCTGCAG	900
F T G H P Y Q N R T P P K <u>K K K P R T S F T R L Q</u>	212
ATCTGCGAGCTGGAGAAGCGCTTCCACCGCCAGAAGTACCTGGCCTCGGCCGAGCGCGCCGCCCTGGCCAGGCG	975
<u>L C E L E K R F H R O K Y L A S A E R A A L A K A</u>	237
CTCAAAATGACCGATGCGCAGGTCAAAAACCTGGTTCAGAACCGGGGACAAAGTGGAGACGGCAGACTGCGGAG	1050
<u>L K M T D A Q V K T W F O N R R T K W R R O T A E</u>	262
GAACGGGAGGCCGAGAGGCAGCAAGCGAACCCGATCTCTGCGAGTTGCAGCAGGAGGCTTCCAGAAGAGCCTG	1125
E R E A E R Q Q A N R I L L Q L Q Q E A F Q K S L	287
GCACAGCGCTGCCCGCTGACCCCTGTGCGTGCACAACCTCGTCTCTTCG CCCTGCAGAATCTGCAGCCGTGG	1200
A Q P L P A D P L C V H N S S L F A L Q N L Q P W	312
TCTGACGACTCGACAAAATCACTAGCGTACGTCGGTGGCGTGGCCTGCGAGTGAGCCTGCCAATTCTGCCCT	1275
S D D S T K I T S V T S V A S A C E ***	337
GTGGGACCCAGGCCACTCAGGGTCACTGAGGCTGAGACCCAGGACTCCTCCCCACCTCCTGGCCTCAGAC	1350
TGCACCCAGGAGGGGAACACTGCCCTCGCAGCCCCAAAGGGCCCCACATTTGTGCCGACACTGTTCTCCTTCGG	1425
TGGAAGAGCTCAAGGACAAGGACACGCGCCCCCTCCAGAGGCGTCCCGCACCTGTCTGAACGTTAAGAAAT	1500
CTGTTTTTGTATTATTTATTTAATTTAATTTTAACTGGGATTCAGAGAAAGGCAAGGGAGGTAAGGGAGGAG	1575
GAGCTTCTGGGGTCCCCAGGGCTGTCATCTGAATTTGCCCTGGGAAACCCCTTCTCTGTGACCCACTTCTCATCA	1650
CACACATGGAACCCATAGTCCACACACAGGTGGTGTACTGTCCCTCTGGTGTACCCAGACCCACACATG	1725
GGCATCTATGGAGAGTGTCAACCAGACAGAGGGTACAGTGTTTACACTTTGGACCTTACGATCAGGCACAGGT	1800
CAGGGTGACACAGACTCATCTGAACAGCATGGCACTCCCTCCAGCACAAACACAAGGTATGGCCACACTGTG	1875
ACACACTACCCACACACAACAGCCAACAGCTACAACAGCCTCACTTGGTCTGCCAGGCCCCACCCACACATCCC	1950
AGCCCAATCCAGGTACGCACAGACAGGTTTTACATAAATGCAGCCATTTCTCCAGAACCCATTTGAGGGGTGG	2025
GGGGGTGTAATTTATGCACITATAAGGTGTTTTCTGTGTAACCATTTTATAAAGTGCTTGTGTAATTTATGTGAAAA	2100
AATAAATAAAGCCTCAAAAGCCTCCGGAAAAA	2175
=====	

**Fig. 2.** Nucleotide and predicted amino acid sequence of cDNAs containing the entire coding and 3' untranslated region of *tcl-3*. Sequences are displayed 5' to 3'. The in-frame termination codons both upstream and downstream from the coding sequence are designated by three asterisks. Underline indicates homeodomain; double underline indicates polyadenylation signal.

sequences on chromosome 10. One new and three previously isolated translocation breakpoints were mapped upstream from the 5' in-frame termination codon TGA, suggesting that mutational events occurred in the 5' regulatory region of the *tcl-3* gene as a result of the chromosomal translocation (Figure 4). Thus, these data provide strong evidence that the cloned cDNA represents the candidate *tcl-3* proto-oncogene.

**Discussion**

The t(10;14) chromosomal translocation is one of the most frequently occurring chromosomal abnormalities found in patients with T-ALL. Its tight association with T-ALL suggests that a proto-oncogene is dysregulated by the chromosomal translocation and the dysregulated gene directly contributes to T-cell leukemia. In this study we have isolated

Organism	Class	Gene	200	260	Identities
human		<i>tcl-3</i>	KKKPRTSFTRLQICELEKRFHRQKYLASAERAALAKALKMTDAQVKTWFQNRRTKWRQT		
			* * * * *	O * * * * *	
human	ANTP	<i>cl1</i>	R-RG-QTY--Y-TL----E--YNR--TRRR-IEI-H--CL-ER-I-I-----M--KKEN		28
human	DFD	<i>Hul</i>	G-RA--AY--Y-TL----E--FNR--TRRR-IEI-H--CLSER-I-I-----M--KKDN		28
mouse	labial	<i>hox1.6</i>	PNAV--N--TK-LT----E--FN--TR-R-VEI-AS-QLNET---I-----M--QKKRE		28
mouse	en	<i>en1</i>	D-R---A--AE-LQR-KAE-QANR-ITEQR-QT--QE-SLNES-I-I----K-A-IKKA-		23
human	<i>hox2.4</i>	<i>huhox2.4</i>	RRRG-QTYS-Y-TL----E-LFNP--TRKR-IEVSH--GL-ER---I-----M--KKEN		25
human	<i>hox1.5</i>	<i>c13+1</i>	S-RV--AY-SA-LV----E--FNR--CRPR-VEM-NL-NL-ER-I-I-----M-YKKDQ		26
human	Abd B	<i>huhox2.5</i>	SR-K-CPY-KY-TL----E-LFNM--TRDR-HEV-RL-NLSER---I-----M-MKKMN		24
human	POU	<i>Oct-1</i>	RR-K---IETNIRVA---S-LENQKPT-E-ITMI-DQ-N-EKEVIRV--C---Q-EK-IN		20
fly	prd	<i>prd</i>	QRRC--T-SAS-LD---RA-E-TQ-PDIYT-EE--QRTNL-E-RIQV--S---ARL-K-H		22
frog	eve	<i>xhox-3</i>	MRRY--A---E--AR---E-Y-EN-VSRPR-CE-----NLPETTI-V-----M-DK--R		28
human		<i>prl1</i>	-RRNFKQATEILN-YFYSHLSNP-PSEEAKEE---KCGI-VS--SN--G-K-IRYKKN		13
human		<i>k8</i>	ARRL--AY-NT-LL----E--FN--CRPR-VEI-AL-DL-ER---V-----M-HK---		30
nematode		<i>mab5</i>	S-RT-QTYS-S-TL----E--YH--TRKR-QEISET-HL-ER---I-----M-HKKEA		26
nematode		<i>JML1001</i>	MRRA-----YE-LVRV-NK-LTSR--SVV--LN--IQ-QLSET---I-----KKNH		26
nematode		<i>mec3</i>	RRG---TIKQN-LDV-NEM-SNTPKPSKHA--K--LETGLSMRVIQV-----S-E--LK		19
fly		<i>bcd</i>	PRRT--T--SS--A---QH-LQGR--TAPRL-D-SAK-ALGT---I--K---RRHKI-S		25
fly		<i>cad</i>	-D-Y-VVY-DF-RL---EYCTSR-ITIRKSE--QT-SLSER---I-----A-E-TSN		24
fly		<i>cut</i>	S--Q-VL-SEE-KEA-RLA-ALDP-PNYGTIEF--NE-GLATRTITN--H-H-MRLKQ-V		16
fly		<i>H2.0</i>	RSWS-AV-SN--RKG--IQ-QQ---ITKPD-RK--AR-NL-----V-----M---HTR		29
fly		<i>ro</i>	QR-Q--T-STE-TLR--VE---NE-ISRSR-FE--ET-RL-E--I-I-----A-DK-IE		27
fly		<i>zen1</i>	V-LK--A--SV-LV---NE-KSNM--YRTR-IEI-QR-SLCER---I-----M-FKKDI		25
fly		<i>zen2</i>	S-RS--A-SS--LI---RE--LN----RTR-IEISQR-AL-ER---I-----M-LKKS-		29

Fig. 3. Comparison of the homeodomain of *tcl-3* with other eukaryotic homeobox-containing genes. The class designation is taken from Scott *et al.* (1989). The 11 highly conserved amino acids found in *tcl-3* are marked by asterisks. The four invariant amino acids among non-yeast homeodomains are at positions 48, 49, 51 and 53. The amino acid identities of *tcl-3* with the other homeodomains are indicated by a dash and tabulated to the right of each line.

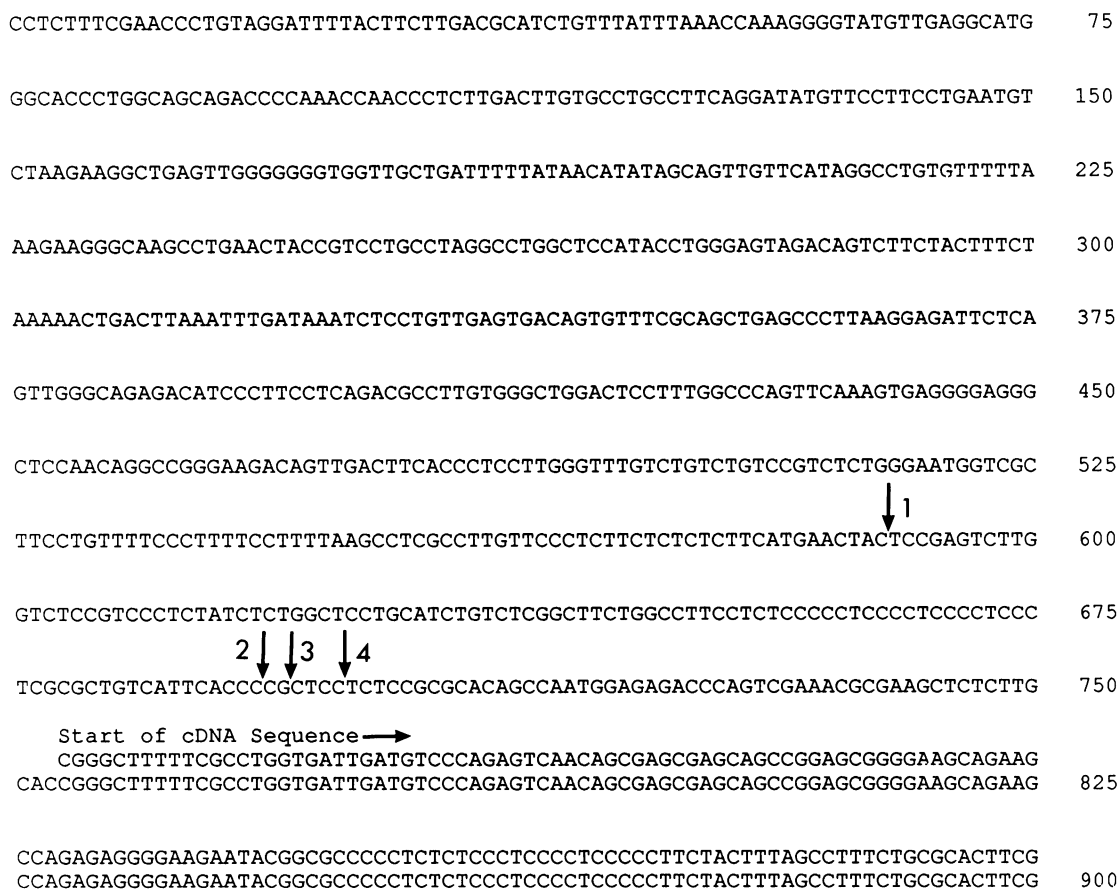


Fig. 4. Upstream genomic nucleotide sequences of *tcl-3* on chromosome 10. Vertical arrows with numbers indicate the t(10;14) chromosomal translocation breakpoints cloned from a new patient BA (arrow 1) (this study); patient BO (arrow 2) (Lu *et al.*, 1990); patients DW and JM (arrows 3 and 4) (Kagan *et al.*, 1989).

and characterized the gene as a novel homeobox gene from the t(10;14) breakpoint cluster region. Thus, for the first time, we provide evidence that a homeobox gene may be directly involved in the pathogenesis of T-cell acute leukemia.

Homeobox-containing genes were first described as *Drosophila* homeotic proteins that are important in controlling the developmental fate of embryonic cells (reviewed in Scott *et al.*, 1989; Gehring *et al.*, 1990). These genes all share a highly conserved 180 bp DNA stretch called homeobox which predicted a helix-turn-helix secondary protein structure similar to the *Hin* recombinase DNA binding domain (Affolter *et al.*, 1991). Recently, homeodomains were also identified in documented mammalian transcription factors such as *Pit-1*, *Oct-1* and *Oct-2* (Bonder *et al.*, 1988; Clerc *et al.*, 1988; Ingraham *et al.*, 1988; Ko *et al.*, 1988; Muller *et al.*, 1988; Schneidreit *et al.*, 1988; Sturm *et al.*, 1988). Sequence analysis revealed that the homeodomain of the *tcl-3* gene contains the hallmark of non-yeast homeobox genes, the four invariant amino acids at positions 48, 49, 51 and 53, respectively (Figure 3). Among the eight additional amino acids that are highly conserved in the homeoboxes encoded by distinct genes, seven of them are identical in *tcl-3* (Figure 3). They are at positions 5, 12, 16, 20, 45, 55 and 58, respectively. Although the amino acid at position 40 in the *tcl-3* homeodomain, methionine, differs from the highly conserved amino acid, leucine, it is identical to the amino acid at the same position in a distinct class of homeobox genes that contain the POU domain (Figure 3). The POU domain homeobox genes are known to be transcription factors. By analogy the *tcl-3* gene could also act as a transcription factor by binding DNA and modifying the transcriptional activity of specific target genes. Other transcription factors such as *jun*, *fos* and *myc* can act as oncoproteins.

Our data support the notion that this new homeobox gene is the candidate proto-oncogene designated *tcl-3*. First, this gene is elevated in expression in SIL cells harboring the t(10;14) chromosomal translocation (Figure 1b). Overexpression of *tcl-3* in a patient with t(10;14) T-ALL was observed independently by Zutter *et al.* (1990). No expression of this gene was detected by Northern blotting in a number of myeloid B-, and T-cell lines, and in leukemic T-cells that do not carry the t(10;14) translocation (Zutter *et al.*, 1990; this study). Second, this gene is structurally altered by the t(10;14) chromosomal translocation in four individual patients with t(10;14) T-ALL. The breakpoints were mapped in the 5' regulatory region of the gene (Figure 4), resulting in head-to-tail juxtaposition of *tcl-3* with the TCRD. The translocation event may bring a promoter or enhancer elements from the TCRD locus to the immediate vicinity of the *tcl-3* coding region. The new regulatory elements presumably are capable of directing gene expression in T-cells, providing a possible explanation for the elevated expression of *tcl-3*. Alternatively, the elevation in expression could be explained as a result of loss of a tissue-specific regulatory element or a silencer element from the *tcl-3* locus. Possibilities that cannot be excluded by our data include differential expression of the gene in distinct types of T-cells, potential somatic mutations or both.

The dysregulation of *tcl-3* may occur during early T-cell differentiation and therefore represents a major genetic alteration in T-cell leukemogenesis. The exact role of *tcl-3*

in neoplastic transformation of T lymphocytes, however, is not clear. It is generally believed that full leukemic transformation requires defects in both growth and differentiation (Sawyers *et al.*, 1991). For instance, overexpression of IL-3 or GM-CSF caused a myeloproliferative syndrome, but not leukemia, suggesting that leukemogenesis may require additional events that would block normal differentiation (Johnson *et al.*, 1988; Chang *et al.*, 1989; Wong *et al.*, 1989). Co-expression of IL-3 and a homeobox gene *Hox-2.4* resulted in murine myeloid leukemia, implying that homeobox genes may play a role in regulating key differentiation processes (Perkins *et al.*, 1990). Recently, a homeobox-containing gene, *prl* (now named *pbx*), has been demonstrated to be associated with the t(1;19) translocation in pre-B leukemia (Kamps *et al.*, 1990; Nourse *et al.*, 1990). The 3' coding region of this gene including the homeodomain was found to be fused to the amino terminus of the transcription factor, *E2A*. Subsequent work showed that the resultant *E2A-pbx* fusion proteins could transform NIH-3T3 cells (Kamps *et al.*, 1991). Thus, some homeobox-containing genes, acting by binding DNA and modulating the transcription activity of specific target genes, may constitute a new family of oncogenes which contribute importantly to tumor formation of hematopoietic origin.

## Materials and methods

### Chromosomal walking

A human placental genomic library was screened using a previously isolated DNA fragment, Ch1019, as probe. Positive  $\lambda$  phage clones with human inserts were subjected to plaque purification, DNA isolation, restriction mapping and subcloning according to standard recombinant DNA procedures (Sambrook *et al.*, 1989). End DNA fragments from the genomic clones that are free of repetitive sequences were isolated and used in the next round of library screening. Fifteen overlapping genomic clones covering a DNA region of 43 kb on chromosome 10 were obtained.

### Northern blotting

Total RNA was isolated from the cell line SIL as described (Sambrook *et al.*, 1989). Poly(A)<sup>+</sup> RNA was selected twice through an oligo(dT) column (Pharmacia). The RNA was electrophoresed on a 1.4% gel containing 2.2 M formaldehyde and transferred to nylon membranes (GeneScreen Plus, DuPont). The membranes were hybridized with <sup>32</sup>P-labeled probes made by random priming method (Feinberg and Vogelstein, 1984).

### cDNA synthesis

Synthesis of the first cDNA strand was accomplished using SuperScript RNase H<sup>-</sup> reverse transcriptase (BRL). Second strand synthesis was performed using DNA polymerase I, RNase H and *Escherichia coli* DNA ligase. The double-stranded cDNA was blunted using T4 DNA polymerase and subsequently ligated with *EcoRI-NorI* adaptors. The cDNA was phosphorylated with T4 polynucleotide kinase and ligated into  $\lambda$  Zap II (Stratagene). The ligation mixture was packaged (Gigapack Gold, Stratagene) and plated on the host strain XL1-Blue. The library was screened with Ch1048.

### DNA sequence analysis

Hybridizing cDNA clones identified with Ch1048 were excised *in vivo* with helper phage R408. Nested sets of deletion mutants were made using unidirectional deletion by ExoIII (BRL). Double-stranded DNAs were purified and subjected to sequence analysis using the dideoxy termination method (Sanger *et al.*, 1977) and a modified *Taq* DNA polymerase (Promega). The predicted polypeptide sequences encoded by cDNAs were compared with GenBank databases using the IFIND program and the FASTDB program. The genomic DNA fragment Ch1019 was subcloned into Bluescript II SK (Stratagene) and used as template for determination of nucleotide sequences. DNA sequences were determined by direct sequencing with either the T3 or T7 primers, or custom-made primers.

**Primer extension analysis**

End-labeled oligonucleotide (10 fmol) was annealed with 1 µg of poly(A)<sup>+</sup> RNA at 42°C for 5 h in 50 mM PIPES, 0.4 mM NaCl and 1 mM EDTA (pH 6.4). Extension was carried out using 200 U of SuperScript RNase H<sup>-</sup> reverse transcriptase for 1 h at 40°C. The reaction mixture was extracted with phenol/chloroform, ethanol precipitated, and analyzed on a 6% denaturing polyacrylamide gel.

**Acknowledgements**

We thank Dr Corey Largman for critical reading of the manuscript. This work was supported by the Ontario Ministry of Northern Development and Mines. The GenBank accession number for the combined genomic and cDNA sequences reported in this paper is M62626.

**References**

- Affolter, M., Percival-Smith, A., Müller, M., Billeter, M., Qian, Y.Q., Otting, G., Wüthrich, K. and Gehring, W.J. (1991) *Cell*, **64**, 879–880.
- Begley, C.G., Aplan, P.D., Denning, S.M., Haynes, B.F., Waldmann, T.A. and Kirsch, I.R. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 10128–10132.
- Boehm, T. and Rabbitts, T.H. (1989) *FASEB J.*, **3**, 2344–2359.
- Bonder, M., Castrillo, J.L., Theill, L.E., Deerinck, T., Ellisman, M. and Karin, M. (1988) *Cell*, **55**, 505–518.
- Chang, J.M., Metcalf, D., Lang, R.A., Gonder, R.J. and Johnson, G.R. (1989) *Blood*, **73**, 1487–1497.
- Chen, Q., Cheng, J.T., Tsai, L.H., Schneider, N., Buchanan, G., Carroll, A., Crist, W., Ozanne, B., Siciliano, M.J. and Baer, R. (1990) *EMBO J.*, **9**, 415–424.
- Clerc, R.G., Corcoran, L.M., Lebowitz, J.H., Baltimore, D. and Sharp, P.A. (1988) *Genes Dev.*, **2**, 1570–1581.
- Feinberg, A.P. and Vogelstein, B. (1984) *Analyt. Biochem.*, **137**, 266–267.
- Gehring, W.J., Müller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y.Q., Otting, G. and Wüthrich, K. (1990) *Trends Genet.*, **6**, 323–329.
- Hecht, F., Morgan, R., Hecht, B. and Smith, S. (1984) *Science*, **226**, 1445–1447.
- Ingraham, H.A., Chen, R., Mangalam, H.J., Elshotz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G. (1988) *Cell*, **55**, 519–529.
- Johnson, G.R., Gonda, T.J., Metcalf, D., Hariharan, I.K. and Cory, S. (1988) *EMBO J.*, **8**, 441–448.
- Kagan, J., Finger, L.R., Letofsky, J., Finan, J., Nowell, P.C. and Croce, C.M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 4161–4165.
- Kamps, M.P., Murre, C., Sun, X-h. and Baltimore, D. (1990) *Cell*, **60**, 547–555.
- Kamps, M.P., Look, A.T. and Baltimore, D. (1991) *Genes Dev.*, **5**, 358–368.
- Ko, H.S., Fast, P., McBride, W. and Staudt, L.M. (1988) *Cell*, **55**, 135–144.
- Kongsuwan, K., Webb, E., Housiaux, P. and Adams, J. (1988) *EMBO J.*, **7**, 2131–2138.
- Kozak, M. (1989) *J. Cell Biol.*, **108**, 229–241.
- Lu, M., Dube, I., Raimondi, S., Carroll, A., Zhao, Y., Minden, M. and Sutherland, P. (1990) *Genes Chrom. Cancer*, **2**, 217–222.
- McGuire, E.A., Hockett, R.D., Pollock, K.M., Bartholdi, M.R., O'Brien, S.J. and Korsmeyer, S.J. (1989) *Mol. Cell. Biol.*, **9**, 2124–2132.
- Mellentin, J.D., Smith, S.D. and Cleary, M.L. (1989) *Cell*, **58**, 77–83.
- Muller, M.M., Ruppert, S., Schaffner, W. and Matthias, P. (1988) *Nature*, **336**, 544–551.
- Nourse, J., Mellentin, J., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S.D. and Cleary, M.L. (1990) *Cell*, **60**, 535–545.
- Perkins, A., Kongsuwan, K., Visvader, J., Adams, J. and Cory, S. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8398–8402.
- Raimondi, S.C., Behm, F.G., Roberson, P.K., Pui, C.H., Rivera, G.K., Murphy, S.B. and Williams, D.L. (1988) *Blood*, **72**, 1560–1566.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
- Sawyers, C., Denny, C. and Witte, O. (1991) *Cell*, **64**, 337–350.
- Scheidereit, C., Cromlish, J.A., Gerster, T., Kawakami, K., Balmaceda, C.-G., Currie, R.A. and Roeder, R.G. (1988) *Nature*, **336**, 551–557.
- Scott, M., Tamkun, J. and Hartzell, G., III (1989) *Biochim. Biophys. Acta*, **989**, 24–48.

- Sturm, R.A., Das, G. and Herr, W. (1988) *Genes Dev.*, **2**, 1582–1599.
- Tycko, B. and Sklar, J. (1990) *Cancer Cells*, **2**, 1–8.
- Ueshima, Y., Rowley, J., Variakojis, D., Winter, J. and Gordon, L. (1984) *Blood*, **63**, 1028–1038.
- Williams, D., Look, A., Melvin, S., Roberson, P., Dahl, G., Flake, T. and Stass, S. (1984) *Cell*, **36**, 101–109.
- Wong, P.M.C., Chung, S., Dunbar, C.E., Bodine, D.M., Ruscetti, S.R. and Neinhuis, A.W. (1989) *Mol. Cell. Biol.*, **9**, 797–808.
- Zutter, M., Hockett, R.D., Roberts, W.M., McGuire, E.A., Bloomstone, J., Morton, C.C., Deaven, L.L., Crist, W.M., Carroll, A.J. and Korsmeyer, S.J. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 3161–3165.

Received on May 14, 1991; revised on June 27, 1991

**Note added in proof**

While this manuscript was in press, Hatano *et al.* [*Science*, **253** (1991), 79–82] reported a shorter *icl-3* cDNA with a similar predicted amino acid sequence.