View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Dadun, University of Navarra

Oncogene (2003) 22, 1400–1410
© 2003 Nature Publishing Group All rights reserved 0950-9232/03 $25.00
www.nature.com/onc

A t(11;15) fuses *MLL* to two different genes, *AF15q14* and a novel gene *MPFYVE* on chromosome 15

Vandana Chinwalla[1], Andy Chien[1], Maria Odero[1,2], Mary Beth Neilly[1], Nancy J Zeleznik-Le[1,3] and Janet D Rowley[1,*]

[1]*Section of Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, IL, USA*

The *mixed lineage leukemia* gene (*MLL*, also known as *HRX*, *ALL-1* and *Htrx*) located at 11q23 is involved in translocations with over 40 different chromosomal bands in a variety of leukemia subtypes. Here we report our analysis of a rare but recurring translocation, t(11;15)(q23;q14). This translocation has been described in a small subset of cases with both acute myeloblastic leukemia and ALL. Recent studies have shown that *MLL* is fused to *AF15q14* in the t(11;15). Here we analyse a sample from another patient with this translocation and confirm the presence of an *MLL–AF15q14* fusion. However, we have also identified and cloned another fusion transcript from the same patient sample. In this fusion transcript, *MLL* is fused to a novel gene, *MLL partner containing FYVE domain* (*MPFYVE*). Both *MLL–AF15q14* and *MLL–MPFYVE* are in-frame fusion transcripts with the potential to code for novel fusion proteins. *MPFYVE* is also located on chromosome 15, approximately 170 kb telomeric to *AF15q14*. MPFYVE contains a highly conserved motif, the FYVE domain which, in other proteins, has been shown to bind to phosphotidyl-inositol-3 phosphate (PtdIns(3)P). The MLL–MPFYVE fusion may be functionally important in the leukemia process in at least some patients containing this translocation.

Oncogene (2003) **22,** 1400–1410. doi:10.1038/sj.onc.1206273

Keywords: t(11;15)(q23;q15); *MLL*; *AF15q14*; *MPFYVE*; FYVE domain; leukemia

Introduction

Chromosomal translocations involving chromosomal band 11q23 are the single most common cytogenetic abnormality in infants with acute leukemia, regardless of the phenotype (Kaneko *et al*., 1988; Gibbons *et al*., 1990; Martinez-Climent *et al*., 1995; DiMartino and Cleary, 1999). The *mixed lineage leukemia* gene (*MLL*, also known as *HRX, ALL-1* and *Htrx*) located at 11q23 is involved in translocations with over 40 different chromosomal bands in a variety of leukemia subtypes (Rowley, 1999). All translocation breaks in *MLL* occur in an 8.3 kb *Bam*HI genomic segment named the breakpoint cluster region (BCR), encompassing exons 8–14 (numbering according to Nilson *et al*., 1996; Thirman *et al*., 1993).

MLL is a large protein with an estimated molecular mass of 430 kDa. The amino-terminus of MLL contains three AT-hook motifs, a transcriptional repression domain and a region of homology to mammalian DNA methyltransferase, all of which are retained in each MLL fusion protein found to date (Gu *et al*., 1992; Tkachuk *et al*., 1992; Zeleznik-Le *et al*., 1994). The carboxy-terminal half of the MLL protein contains the plant homeobox domain (PHD) and the SET domain (Su(var)3-9, enhancer of zeste and trithorax), which are highly homologous to similar domains found in the *Drosophila* protein, trithorax (trx). The trx protein is a positive regulator of homeotic genes that are required for normal embryonic and adult development of *Drosophila* (Breen and Harte, 1993). The heterozygous *Mll* mutant mice show abnormal expression of some of the *Hox* genes, providing further evidence that *MLL* is the mammalian homolog of *trx* (Yu *et al*., 1995). In addition, the expression of many *Hox* genes is not properly maintained in embryonic fibroblasts derived from *Mll* homozygous mutant mice (Hanson *et al*., 1999).

The mechanism of action of the MLL fusion proteins in the development of leukemia is unclear. At present, 30 different partner genes to which *MLL* is fused by translocations have been cloned as an initial step to elucidate the mechanism of leukemogenesis caused by the 11q23 translocations (Ayton and Cleary, 2001). The partner proteins do not share any distinct common sequence motif and are functionally diverse, ranging from potential nuclear transcription factors to cytoplasmic structural proteins.

Present evidence indicates that the fusion proteins formed by translocations involving the *MLL* gene play an important role in leukemogenesis. *MLL* translocations always yield in-frame fusion mRNA where the 5′ portion of *MLL* is fused to different

*Correspondence: JD Rowley, Interim Deputy Dean for Science, Blum-Riese Distinguished Service, Professor of Medicine, of Molecular Genetics and Cell Biology and of Human Genetics, Section of Hematology/Oncology, University of Chicago, 5841 S. Maryland, Chicago, IL 60637, USA;
E-mail: jrowley@medicine.bsd.uchicago.edu

[2]Current address: Department of Genetics, University of Navarra, Pamplona, Spain

[3]Current address: Department of Medicine, Cardinal Bernardin Cancer Center, Loyola University, Maywood, IL, USA
Received 12 August 2002; revised 18 November 2002; accepted 22 November 2002

Table 1 Primer names and sequences

Primer name	*Primer sequence*
MLL primers	
MLL 5A	5′AAG CCC GTC GAG GAA AAG AGT GAA GAA GGG AAT GTC TCG
MLL 5NP	5′ GAT AAG CTT CCA GGA AGT CAA GCA AGC AGG 3′
MLL 5UNP	5′GCC TGA ATC CAA ACA GGC CAC CAC T 3′
MLL F4	5′ AAA AAA TTT AGG CTT GGC AAG GCG 3′
MPFYVE primers	
B1	5′ GCC TGT GGG ACT GCC CTT CCG GGA GGA CC 3′
B2	5′ GGA GTC GTG CTA GGC GCT CAG CAA TCA TCT GG 3
NB1	5′ GGT GTC TTC AGT GCT CTT GGC CTG CAC GTG C 3′
NB2	5′ CTT GTC GTG TCA GTC CCT GGC TCT GGG AAG 3′
NG 32T	5′ GAG GAA GTG CTT AAT AAA TGG AGG CTA TTG 3′
NG 406B	5′ GGA GAT CCC CCT CTC ACC CGC CCC TGC AGG 3′
AF15 primers	
AF15 6104B	5′ TCT GTC GGG GTT TCT GTA TCA AGA TGT GGA 3′
AF15 7421B	5′ TGC TTG GTT GTT CCA ATG GTG GTC CAA GGG 3′
AF15 6084B	5′ CTC CCT TAT TGT TAT TCT CCC ATC TTG 3′
AF15 7366B	5′ CAT TCT TCA GGT AGT TGT TCT CCA GTG 3′
AF15 T	5′ GAC TGT TTT CTA TTC TAG TTG TAA TGA TGC 3′
AF15 P2	5′ TCT GAT GTG GTC ATA AGA TTT TGG AGC 3′
AF15 NP2	5′ TTC TTT TTC GGG TGT TGG TGC AGC TGC 3′

partner genes, giving rise to novel fusion genes and proteins (Rowley, 1993; Thirman *et al*., 1993; Bernard and Berger, 1995; Waring and Cleary, 1997). In a murine bone marrow immortalization system, the *MLL* partner genes *ENL, ELL* and *CBP* were critical for immortalization and leukemogenesis (Lavau *et al*., 1997, 2000a, b). Thus, it is clear that the partner genes contribute to the leukemogenic ability of the MLL fusions but their exact role is still not determined. Most of the common *MLL* translocations have been cloned. Studying the rare but recurrent translocations such as t(11;15) may offer some insight regarding the role of *MLL* partner genes.

Here we report on our analysis of a rare but recurring translocation, t(11;15)(q23;q14), from an 11-year-old boy with acute myeloblastic leukemia (AML-M2). This translocation has been described in a small subset of cases of both AML and ALL (Raimondi *et al*., 1989; Hunger *et al*., 1993; Thirman *et al*., 1993; Rubnitz *et al*., 1994; Hernandez *et al*., 1995; Hayette *et al*., 2000). *MLL* involvement in this case was confirmed previously by Southern hybridization (Thirman *et al*., 1993). Recent studies have shown that *MLL* is fused to *AF15q14* in t(11;15)(q23;q14) (Hayette *et al*., 2000). Here, we confirm the presence of the *MLL–AF15q14* fusion in our patient sample. In addition to the *MLL–AF15q14* fusion transcript, we have cloned another fusion transcript from the same patient sample in which *MLL* is fused to a novel gene, *MLL partner containing FYVE domain* (*MPFYVE*).

Results

An 11-year-old boy with *de novo* AML-M2 had a complex karyotype of 46,XY(18 cells)/47,XY, del(4)(q2?3q2?8),dic(6;17)(q27;p11),del(7)(p13p21), −15,t(11;15)(q23;q14),t(12;12)(q13;q23), + mar1, + mar2, + mar3(3 cells)/NCA: similar to clone 1, −9,−18, −21, −dic(6;17)(1 cell). *MLL* involvement in the t(11;15)(q23;q14) was shown by a Southern blot analysis using genomic DNA from the bone marrow cells collected at the time of relapse (Thirman *et al*., 1993).

At the time of initiation of this study, cloning of the *MLL* partner gene *AF15q14* was reported in an abstract (Kuefer *et al*., 1997). We received partial sequence information of *AF15q14* from Dr Steve Morris (St Jude Hospital, TN, USA). After performing nested PCR with the *MLL* top primer, *MLL* 5NP, and an *AF15q14* bottom primer, *AF15* NP2, a distinct 400 bp amplified product was obtained (data not shown). Sequencing this product revealed that the *MLL* exon 8 sequences were fused to 260 bp of unknown sequence (Figure 2). This new sequence failed to match any known gene in the database including what was known of the *AF15q14* sequence at the time. Analysis of the 400 bp chimeric cDNA amplified product showed that an open reading frame was maintained after the *MLL* sequence, thus potentially encoding a novel fusion protein. The breakpoint in *MLL* results in the loss of the final 8 nt from the 3′ end of *MLL* exon 8.

To confirm that this 400 bp amplified product was not an artefact, we directly amplified the fusion from patient material using the *MLL* exon 8 primers, and either bottom primer B1 or B2 designed using the sequence derived from the 400 bp product (Table 1, Figures 1 and 2). After nested PCR, we obtained products of the expected sizes of 360 and 240 bp, respectively (Figure 3a). The sequence of the PCR products matched perfectly the sequence of the 400 bp product obtained earlier, confirming that the 400 bp PCR product cloned initially was derived from the chimeric cDNA and was not a PCR artefact.

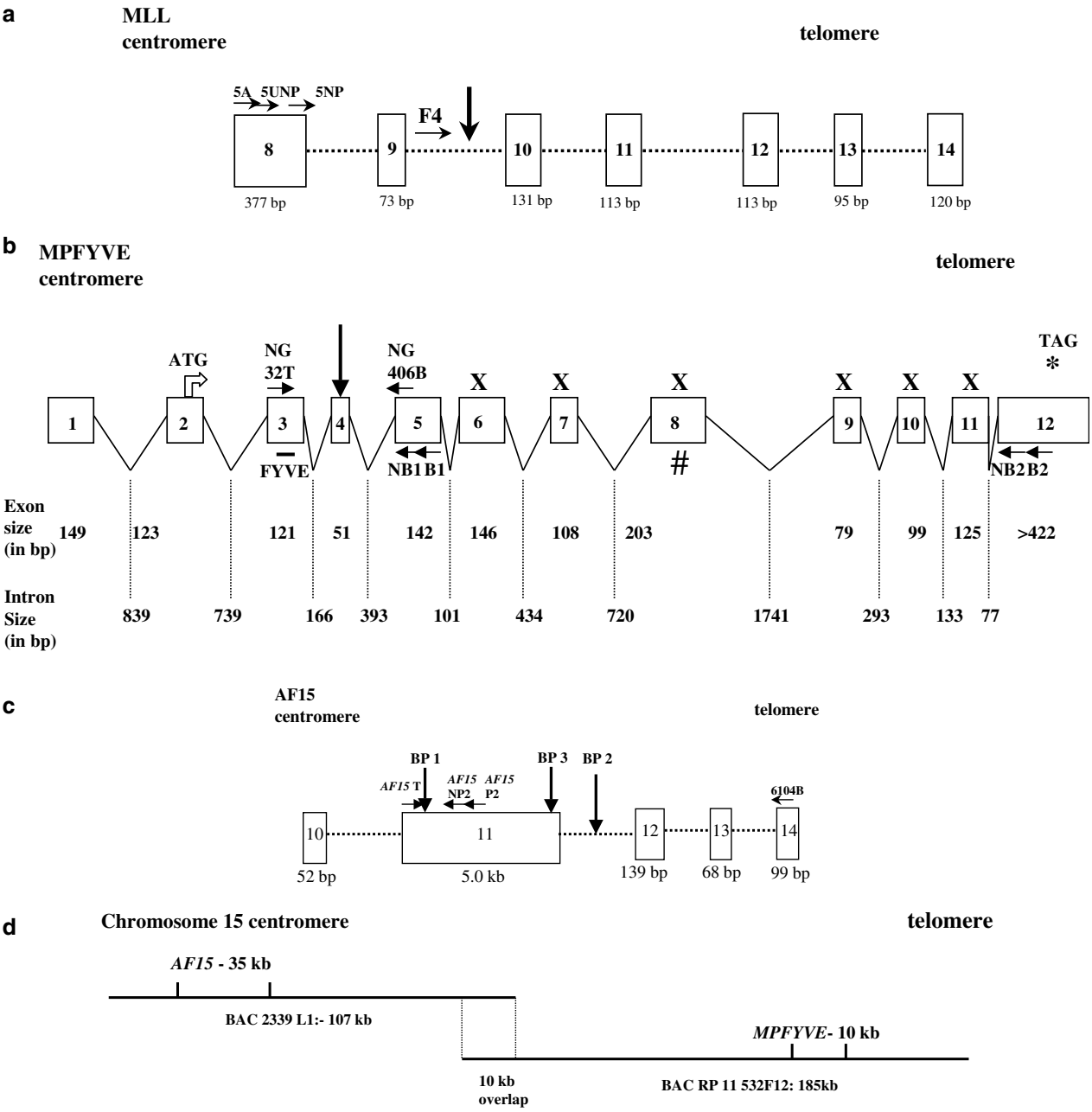

Figure 1 Genomic structure of *MLL, MPFYVE* and *AF15* breakpoint regions. (**a**) Schematic representation of the *MLL* genomic breakpoint region in the t(11;15). The long arrow indicates the chromosomal break. The locations of PCR primers are indicated as horizontal arrows. Intron and exon sizes are indicated. (**b**) Genomic structure of MPFYVE. 'X' represents exons absent in the fusion mRNA. '#' represents the alternatively spliced exon between the two ESTs T74136 and AI191114. (**c**) Genomic structure of the breakpoint region of *AF15q14*. The three known breaks in *AF15q14* BP1 (Kuefer *et al*., this issue), BP2 (case studied here) and BP3 (Hayette *et al*., 2000) are indicated by long horizontal arrows. Primer 7421B is located in exon 27 (the last exon) and is not shown here. (**d**) Orientation of the two BACs containing the *MPFYVE* and *AF15q14* genes on chromosome 15

BLAST search of the EST database identified a set of ESTs that were homologous to the novel sequence fused to *MLL*. We sequenced two ESTs, T74136 (1.2 kb) and AI191114 (2 kb), to identify the cDNA sequence of the gene fused to *MLL*. The sequence comparison of the two ESTs and the fusion cDNA revealed that they were alternative splice products of the same gene (Figure 1b). RT–PCR using RNA isolated from the cell line U937 and different primers within the EST sequence gave a variety of alternatively spliced products including one similar to the fusion transcript (data not shown).

EST AI 191114 contained almost the full-length cDNA, but it is an incompletely processed transcript containing intronic sequence at its 5′ end. By comparing the genomic sequence of the region available in the database with our cDNA sequence (GenBank accession number AF 445414), we predicted that *MPFYVE* contains 12 exons and spans 10 kb of genomic DNA

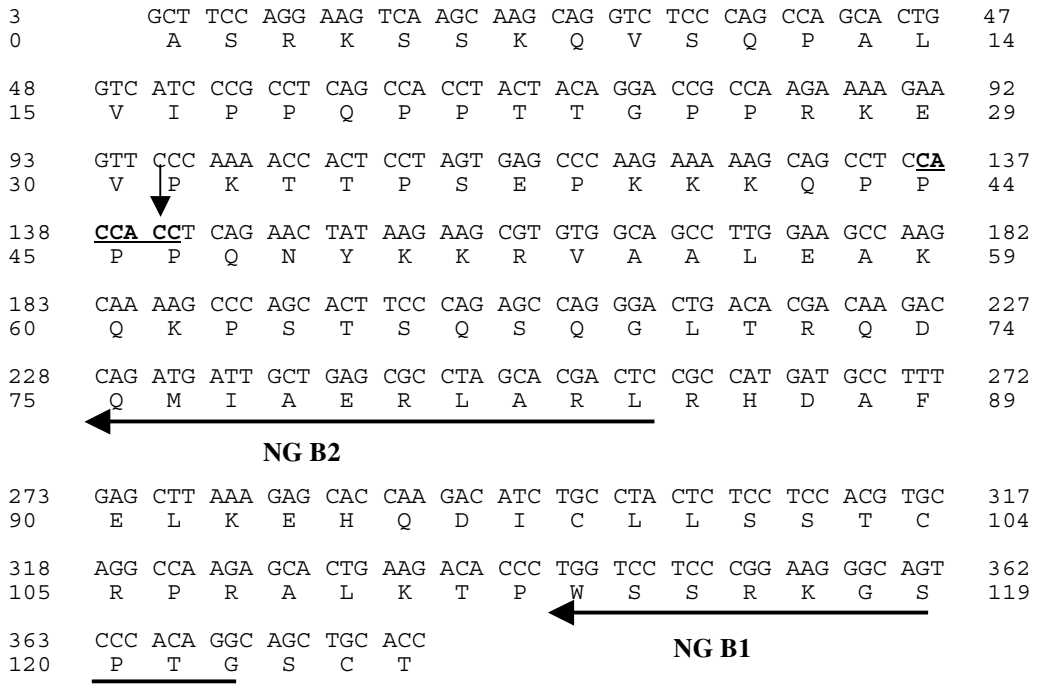

Figure 2 Sequence of the chimeric *MLL–MPFYVE* cDNA. Sequence of the 400 bp chimeric cDNA cloned from the t(11;15) patient is shown. The vertical arrow indicates the fusion junction between *MLL* and *MPFYVE*. Primers used in RT–PCR are indicated. *MLL* and *MPFYVE* contain seven identical nucleotides at the breakpoint (underlined and bold)

(Figure 1b). The first two exons of 149 and 123 nt, respectively, were not present in our longest EST clone and were obtained by the Panorama program (available through the University of Texas, Southwestern Medical center internet site, http://www.atlas.swmed.edu) using the genomic sequence from the database. The exons predicted by gene scan and the rest of the exons predicted by us, based on the sequence comparison, matched perfectly.

MPFYVE and *AF15q14* share only 15 bases of homology at the 3′ end of the 30 bp long *AF15* NP2 primer. This 15-base identity with the *AF15* NP2 primer explains the amplification of the 400 bp RT–PCR product obtained initially.

To evaluate the expression pattern of *MPFYVE*, we probed a multiple tissue Northern blot with DNA amplified from the 3′ end of the EST T74136. A major transcript of 1.35 kb was expressed in heart (H), skeletal muscle (S), kidney (K) and liver (L) (Figure 4a). Minor transcripts of 2.4 and 4.0 kb were also detected. A Northern blot containing total RNA from two different hematopoietic cell lines probed with the same probe showed multiple transcripts ranging from 1.35 to 4.0 kb (Figure 4b). The size of transcripts detected with the *MPFYVE* probes is different from that of *AF15q14* (8 and 9.5 kb), indicating that these are two different genes. Probes from EST AI191114 and the RT–PCR products obtained earlier gave the same expression pattern and transcript sizes (data not shown).

As an independent confirmation of *MLL–MPFYVE* fusion transcript expression, we used total RNA isolated from patient material for RNase protection analysis. The *MLL–MPFYVE* fusion fragment, originally isolated by RT–PCR, was protected in the presence of patient RNA, confirming the existence of the *MLL–MPFYE* transcript in our patient (Figure 4d, lane 1). We also confirmed expression of normal *MPFYVE* in this same sample (Figure 4d, lane 1). The human *β-actin* probe was used as a control for the quality of the patient RNA (lane 4).

MPFYVE encodes a predicted protein of 48 kDa (Figure 5a). The translation start codon is in the second exon. The predicted amino-acid sequence showed no striking homology to any of the known proteins in the database but showed the presence of an FYVE domain near its amino-terminus within exon 3. The FYVE domain was detected using the NCBI's Conserved Domain (CD) database search. We named the new gene that is fused to *MLL* as *MLL* partner containing a FYVE domain (*MPFYVE*).

The predicted MLL–MPFYVE fusion protein contains the AT-hooks and the repression domain of MLL with P^{1334} of MLL fused to P^{73} of MPFYVE (Figure 5b). The predicted MLL–MPFYVE fusion protein does not retain the FYVE domain of MPFYVE. The *MLL–MPFYVE* fusion transcript isolated from the patient material is an alternatively spliced form of the *MPFYVE* gene and is capable of encoding a 1435 aa protein with a predicted molecular mass of 160 kDa.

To determine whether or not we could detect a fusion of *MLL* and *AF15q14* in our patient, we performed RT–PCR using the other downstream primers of *AF15q14*, *AF15* 6104B and *AF15* 7421B and *MLL* exon 8 top

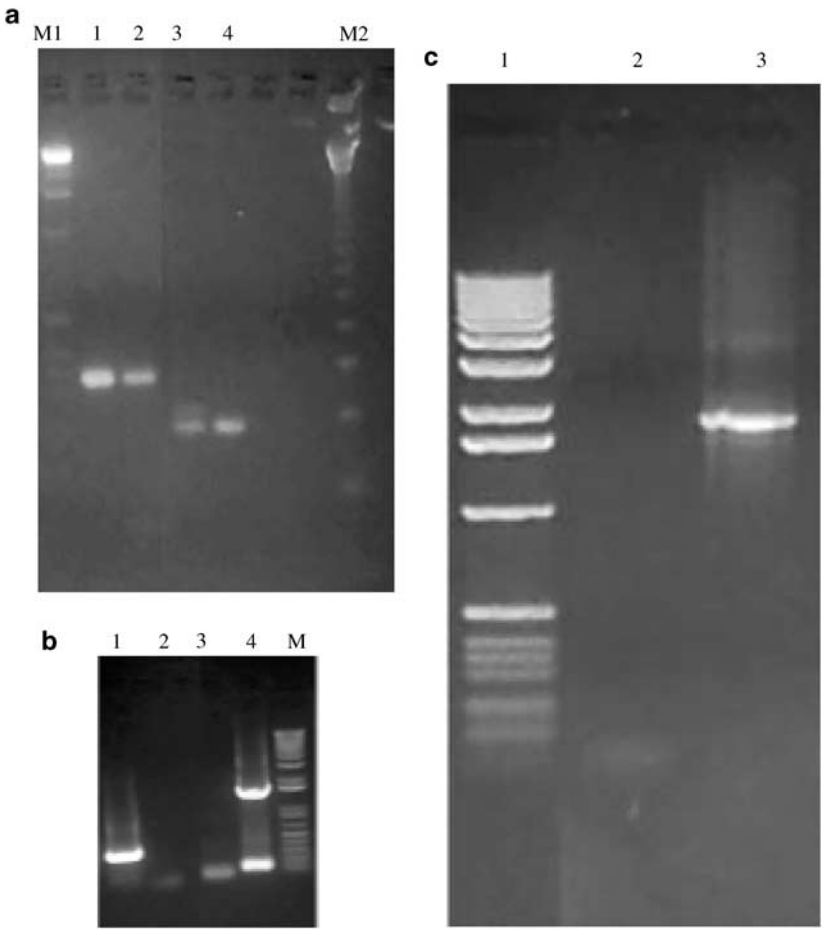

Figure 3 RT–PCR showing the presence of both *MLL–MPFYVE* and *MLL–AF15q14* chimeric transcripts in the t(11;15) patient sample, and PCR of patient sample genomic breakpoint. (**a**) After nested PCR using *MLL* 5NP and the *MPFYVE* bottom primers, either NB1 (lanes 1 and 2) or NB2 (lanes 3 and 4) products of 360 and 240 bp, respectively, were obtained. Either 1 : 30 (lanes 1 and 3) or 1 : 300 (lanes 2 and 4) dilution of the first round PCR-amplified DNA was used as a template in the second round. M1 and M2 represent the 1 and 100 bp ladder (Gibco-BRL molecular weight standards), respectively. (**b**) RT–PCR of total RNA from the patient's bone marrow sample using *MLL* top primers and *AF15q14* bottom primers. After nested PCR with *MLL* 5NP and either *AF15q14* 6084B or 7366B, 500 bp (lane 1) and 1.8 kb (lane 4) products were obtained, respectively. Lanes 2 and 3 represent no DNA template negative controls. The marker is 1 kb plus ladder (Gibco-BRL) (lane M). (**c**) A 1.8 kb product was obtained (lane 3) using *MLL* intron 6 top primer (*MLL* F4) and *AF15q14* 6084B bottom primer and genomic DNA isolated from the patient material. Marker (lane 1, 1 kb ladder) and no template DNA PCR negative control (lane 2) are shown

primers (Table 1 and Figure 1). Surprisingly, on the second round of amplification, we obtained 500 and 1.8 kb products, respectively (Figure 3b). Cloning and sequencing these products showed that the *MLL* exon 9 is fused to exon 12 of *AF15q14*. The size difference of the RT–PCR products obtained with the two bottom primers reflected the distance between the sites where they annealed, approximately 1.3 kb. The chimeric cDNA maintains the open reading frame, thus encoding a putative MLL–AF15q14 fusion protein. Translation of this chimeric cDNA sequence showed that the amino-terminus of MLL up to K^{1362} is fused to the carboxy-terminus of the AF15q14 starting at the amino acid I^{1819} (Figure 5c). The entire *MLL–AF15q14* coding sequence is 5658 bp and encodes a 1886 aa chimeric protein with a predicted molecular mass of 208 kDa.

The genomic breakpoint of the t(11;15) from our patient was cloned to identify the exact break in *MLL* and *AF15q14*. Long-range PCR using genomic DNA isolated from the bone marrow cells from our patient and the *MLL* intron 9 top primer (*MLL* F4) and *AF15q14* bottom primer (*AF15* 6084B) amplified a 1.8 kb band that was sequenced to obtain the exact sequence of the fusion junction between *MLL* and *AF15q14* (Figure 3c). Genomic PCR using *MLL* top primers and *MPFYVE* bottom primers did not amplify any fragment, indicating that there is only one genomic breakpoint. The two different fusions we see here could be due to alternative intragenic splicing between *AF15q14* and *MPFYVE*.

The sequence of the BACs containing *AF15q14* and *MPFYVE* is available in the database (AC022405 and AC012476, respectively). The position of the two genes within each BAC is known, as is the fact that these two BACs overlap each other by 10 kb. We could therefore estimate that the distance between the two genes is

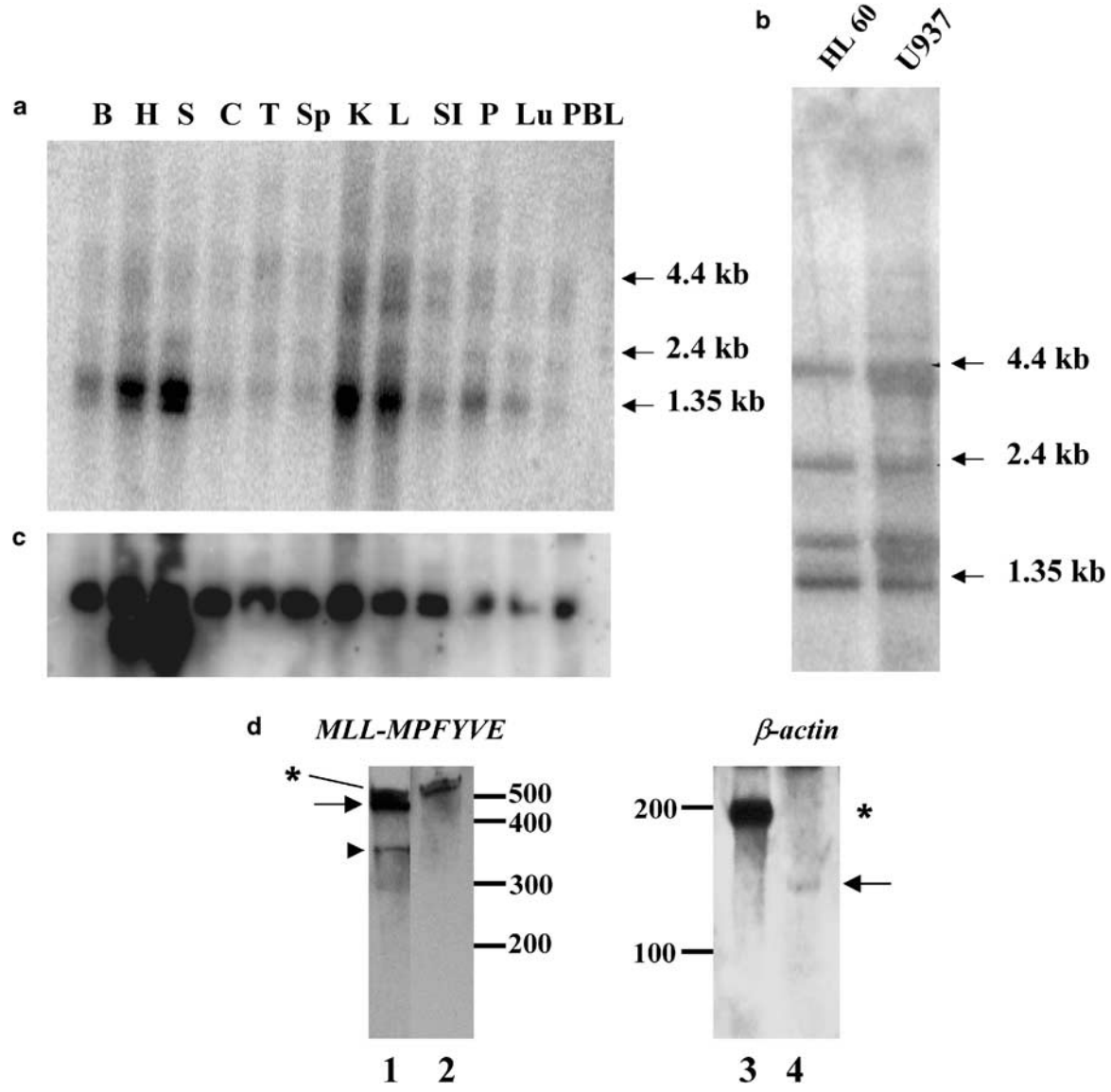

Figure 4 Expression of *MPFYVE* and *MLL–MPFYVE* RNA. (**a**) Multiple tissue Northern blot (Clontech Inc.) was probed with EST T74136. A major transcript of 1.35 kb is expressed abundantly in heart, skeletal muscle, kidney and liver. Minor transcripts are detected at 2.4 and 4.0 kb. (**b**) Northern blot containing total RNA from the two hematopoietic cell lines, HL60 and U937, indicates that there are similar multiple transcripts encoded by *MPFYVE* ranging from 1.35 to 4.0 kb. (**c**) The same blot as in (**a**) was rehybridized with an actin probe. (**d**) RNase protection assay. Patient RNA was hybridized with antisense RNA probe from either *MLL–MPFYVE* clone (lane 1) or human *β-actin* (lane 4). The full length *MLL–MPFYVE* probe (506 nt) is indicated by (*). The fusion *MLL–MPFYVE* (416 nt) and normal *MPFYVE* (350 nt) protected messages are indicated by arrow and arrowhead, respectively (lane 1). Full-length (188 nt) and protected (127 nt) human *β-actin* probe are indicated by (*) and arrow, respectively (lanes 3 and 4). Lanes 2 and 3 are undigested probes

approximately 170 kb (Figure 1d). By fluorescence *in situ* hybridization (FISH) using these BACs, we have determined that *AF15q14* is centromeric to *MPFYVE* on chromosome 15 (Figure 6b).

The karyotype of this patient is complex and contains a deleted chromosome 15 and three marker chromosomes. There is no normal chromosome 15. Our FISH analysis with the *AF15q14* BAC showed that it hybridizes to three different chromosomes, the der(11), der(15), which is the deleted chromosome 15 and a marker chromosome; the probe is duplicated on the marker chromosome. Thus, there appears to be a duplication of *AF15q14*. The whole chromosome paint 15 (WCP15) confirmed the presence of chromosome 15 sequences on the marker chromosome to which the *AF15q14* BAC hybridizes (data not shown). Hybridization of the *AF15q14* probe to der(11), deleted chromosome 15 and the marker chromosome with chromosome 15 sequences, indicates that it contains the breakpoint of t(11;15) (Figure 6a). The *MLL* probe also gave a split signal between the der(11) and der(15) chromosomes, confirming the involvement of *MLL* in this case; the critical 5′ *MLL*-3′*AF15q14* fusion is on the der(11) chromosome. Different *MPFYVE* containing BACs did

a MVKADCPVPITDLPDSSGKLQYGCKNCGRAFCSGCLSFSAAVPRTGNTQQKVCKQCHE
VLTRGSSANASKWSP*PQNYKKRVAALEAKQKPSTSQSQGLTRQDQMIAERLARLR*QENKP
KLVPSQAEIEARLAALKDERQGSIPSTQEMEARLAALQGRVLPSQTPQPAHHTPDTRTQ
AQQTQDLLTQLAAEVAIDESWKGGGPAASLQNDLNQGGPGSTNSKRQANWSLEEEKS
RLLAEAALELREENTRQERILALAKRLAMLRGQDPERVTLQDYRLPDSDDDEDEETAI
QRVLQQLTEEASLDEASGFNIPAEQASRPWTQPRGAEPEAQDVDPRPEAEEEELPWCCI
CNEDATLRCAGCDGDLFCARCFREG**HDAFELKEHQDICLLSSTCRPRALKTPWSSR**
KGSPTGSGTHFWAQPQDVRWESLSGSTDDG*

b **MLL-MPFYVE**

MLL ——— | ——— **MPFYVE**

PKTTPSEPKKKQPPP^{1334} P^{73}QNYKKRVAALEAKQK

c **MLL-AF15**

MLL ——— | ———**AF15**

APRPSIPVKQKPKEK^{1362} I^{1819}FDHHTEEDIDKSAN

Figure 5 (**a**) Predicted amino-acid sequence of MPFYVE, MLL–MPFYVE and MLL–AF15q14. The predicted molecular mass of MPFYVE is 48 kDa. The MPFYVE protein contains an FYVE domain near its amino-terminus (underlined). The part of MPFYVE present in MLL–MPFYVE is a product of an alternatively spliced mRNA where the first half (italicized) and second half (bold) are joined together by splicing. (**b, c**) Predicted amino-acid sequence of the fusion junctions of MLL–MPFYVE (**b**) and MLL–AF15q14 (**c**) are shown

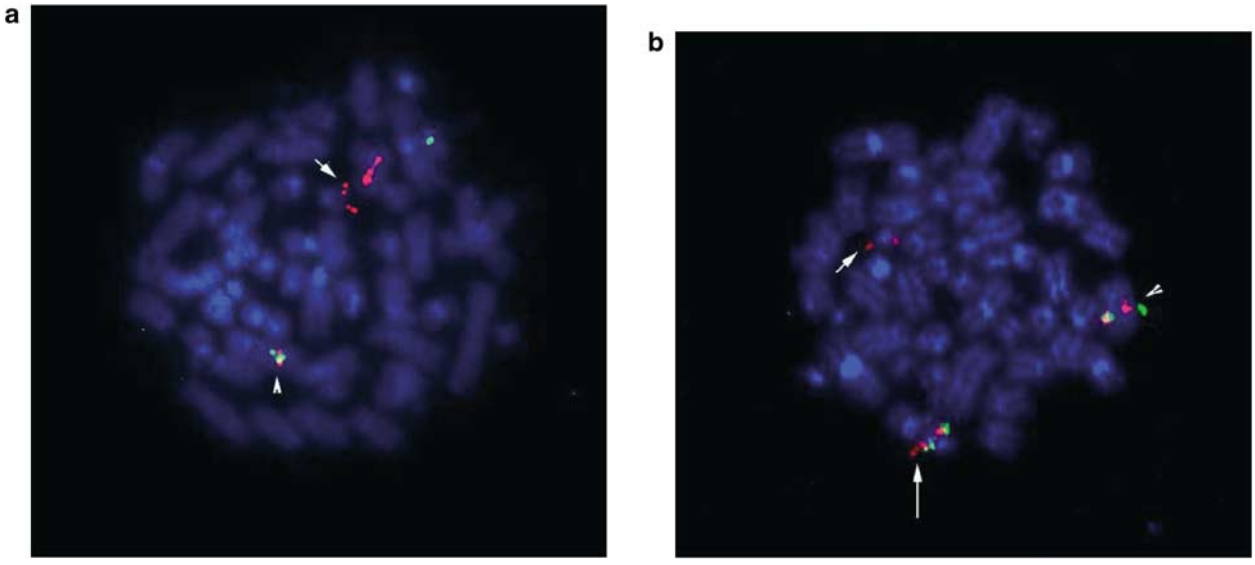

Figure 6 FISH analysis of chromosomes from the t(11;15) patient sample. (**a**) FISH analysis using *MLL* and *MPFYVE* BACs as probes. The *MLL* probe (red) showed three signals, on normal 11, der(11) (arrowhead) and der(15) (arrow). *MPFYVE* BACs showed only two signals (green): marker chromosome which is a normal chromosome 15 and der(11) where the entire BAC was translocated, telomeric to *MLL*. This indicates that the genomic break is 5′ of *MPFYVE*. (**b**) FISH analysis using *AF15q14* and *MPFYVE* BACs as probes. The *AF15q14* BAC (red) showed three signals, on der(11) (arrowhead), der(15) (short arrow) and a marker chromosome, which contains chromosome 15 sequences. The *AF15q14* probe is duplicated on the marker chromosome as seen by the presence of two red signals (long arrow). The *MPFYVE* probe (green) labels the der(11) and has a single signal on the marker chromosome

not show a split signal; instead the entire *MPFYVE* signal was translocated to the der(11), indicating that *MPFYVE* is telomeric to *AF15q14* (Figure 6b). Moreover, the *MPFYVE* containing BACs only gave one signal on the marker chromosome, indicating that the *MPFYVE* gene is not duplicated. The importance of the *AF15q14* duplication is not currently understood.

Discussion

Cloning the chimeric cDNA and genomic breakpoint junction formed by the t(11;15) from this patient reveals a complex situation. We have identified two in-frame fusions in which the 5′ portion of *MLL* is fused to a novel gene, *MPFYVE*, and a recently identified gene,

a Alignment of the predicted FYVE finger domain of MPFYVE with closely related FYVE domains present in other proteins

```
                          10        20        30        40        50        60        70
                 ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*.....
                   W# D -       C% C   F@ %%     R+HHCR CG ##C   CS    % %            %R#C  C#  #
consensus     1 PHWVPDEEVS---NCMRCGKPFT-LT---KRRHHCRACGRIFCS-SCSSKTVPLPPMG-----ERPVRVCDSCYDLLNK 66
MPFYVE       66 cfrgtmvkad---cpvpitdlpdssg---klqygCKNCGRAFCS-GCLSFSAAVPRTG-----NTQQKVCKQCHEVLTR 132
1HYI A        1 RKWAEDNEVQ---NCMACGKGFS-VT---VRRHHCRQCGNIFCA-ECSAKNALTPSS------KKPVRVCDACFNDLQG 65
gi 2388904  168 PDWTDSEV------CLRCRTPFT-FT---NRKHHCRNCGGVFCN-QCSSKTLSLPHLg----iNQPVRVCDSCYSLRTK 231
gi 1089781  158 PDWVDAEE------CHRCRVQFG-VV---TRKHHCRACGQIFCG-KCSSKYSTIPKFg----iEKEVRVCEPCYEQLNK 221
gi 1326382  191 PEWADGPE------CYRCRSVFS-VF---TRKHHCRACGQIFCD-KCSSRELALPQFg----iEKEVRVCETCYEKKVA 254
gi 3880102  858 DHWVQDVTRQ---RCDDCEHKFT-LA---DRKHHCRNCGQIFCS-TCSRFESHITRMn----iSRPVRVCRKCFQRLQQ 924
gi 1706789  725 PTPIREKEVT---MCMRCQEPFNsIT---KRRHHCKACGHVVCG-KCSEFRARLVYDn-----NRSNRVCTDCYVALHG 791
gi 7436516  536 PYWIPDSE------CPNC-MLFT-II---TRRHHCRACGRVLCG-SCCNEKAFLEYLqeegkkLQAVRVCKPCSAMLAR 602
gi 1743430  341 VRWQWEDDVE---NCSGCAVSVA-KI---KPRPRCLHCCKIFCS-TCVQHTVPSGPT------RRPANVCQVCHTLLNR 405
```

b FYVE-related fingers

```
                          10        20        30        40        50        60        70
                 ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*.....

MPFYVE       66 cfrgtmvkad---cpvpitdlpdss---gklqygCKNCGRAFCS-GCLSFSAAVPRTG------NTQQKVCKQCHEVLTR132
RnRIM       131 EHKDDAP-----TCGICHKTKFAD-----GCGHICSYCRTKFCA-RCGGRVSLR-----SNNEDKVVMWVCNLCRKQQE 193
HsNOC2       83 MRRNVMGNGLS-QCLLCGEV-LGFLG---SSSVFCKDCRKKVCT-KCGIEASPG---------QKRPLWLCKICSEQRE 146
CeYPT7      162 MRRRATGNGVT-HCLLCHTE-FGLLASK-SYAAMCVDCRKYVCQRNCGVETTDV----NQTTGKVETVFLCKICSEARE 233
HsBassoon_B 378 AKPKTMPKERA-ICPLCQAE-LNVGSKSPANYNTCTTCRLQVCN-LCGFNPTPH--------LVEKTEWLCLNCQTKEL 445
HsBassoon_B  84 PQIAPLPSST--LCPICKTSDLTSTPSQ-PNFNTCTQCHNKVCN-QCGFNPNPH--------LTQVKEWLCLNCQMQRA 150
MmRabphilin  82 MRKNVAGDGVN-RCILCGEQ-LGLMG---SACVVCEDCKKNVCT-KCGEVTSNN---------RPHPVWLCKICLEQRE 145
```

Figure 7 FYVE domain alignment. (**a**) Alignment of the predicted FYVE finger domain of MPFYVE with closely related FYVE domains present in other proteins. The alignment shown here is the result of the CD-Browser analysis (NCBI database). This figure is modelled after the FYVE alignment published previously (Stenmark and Aasland, 1999). The FYVE domain of MPFYVE contains only six cysteines and is missing the first two cysteines present in the consensus. R + HH is absent in MPFYVE, which makes it more similar to the FYVE-related fingers. The absence of 'R + HH' is shown here by highlighting the existing residues 'lqyg' in bold. The weakly conserved hydrophobic residue at position 67 is not conserved in MPFYVE (numbering specific to this figure). Arginine to lysine at position 68 is a conservative change and therefore may not change the structure of the finger. # strongly conserved hydrophobic; − strongly conserved acidic; + strongly conserved basic; % weakly conserved hydrophobic. (**b**) FYVE-related fingers. The MPFYVE sequence is aligned with the known FYVE-related fingers described by Stenmark and Aasland (1999). Like other FYVE-related fingers, MPFYVE lacks the R + HH motif, but still contains the glycine residue (bold) at a conserved position 39 (numbering specific to this figure) similar to the FYVE domain

AF15q14. The two putative fusion proteins retain the AT-hooks, DNA methyltransferase homology and the repression domains of MLL. In the *MLL–MPFYVE* fusion, *MLL* exon 8 (minus 8 nt from the 3′ end of the exon) is fused to *MPFYVE* exon 4. Since the site of fusion of *MLL* with *MPFYVE* is within an exon, it could be a cryptic splice donor site/acceptor site enabling the fusion to maintain its open reading frame. Usage of cryptic splice donor and acceptor has been well documented in *MLL* fusions (Caldas *et al*., 1998; Divoky *et al*., 2000). Although canonical splice donor/acceptor sequences are not present at the fusion junction, we are confident that this is the real junction sequence. We have independently PCR amplified this junction using different primer pairs and sequenced the PCR products both directly and after subcloning. Further data supporting the existence of *MLL–MPFYVE* fusion RNA come from the RNase protection assay (Figure 4d). Another interesting observation of the fusion RNA is that the 7 nt at the junction between *MLL* and *MPFYVE* are identical. The importance of this 7 bp homology at the point of fusion is unclear. Furthermore, the genomic break in *MLL* is in intron 9, but the *MLL–MPFYVE* fusion did not contain exon 9. This is likely owing to alternative splicing of *MLL* exon 9, which has been previously described.

MPFYVE encodes a small protein containing 438 amino acids. The MPFYVE protein does not share extensive homology to any previously identified proteins but contains a highly conserved FYVE domain near its amino-terminus. The FYVE domain is a zinc-finger motif, which is conserved among many yeast, *Caenorhabditis elegans* and mammalian proteins (Figure 7). It was named 'FYVE' after four proteins (**F**ab1p, **Y**OTB, **V**ac1p and **E**EA1) containing this motif (Stenmark *et al*. 1996). The FYVE domain contains eight conserved cysteines, which coordinate two Zn^{2+} ions (Schwabe and Klug, 1994; Stenmark *et al*., 1996). In addition to the cysteines, the R(R/K)HHCRxCG motif surrounding a third and a fourth cysteine residue and an arginine at position 68 are also highly conserved in the family (Stenmark and Aasland, 1999). Several proteins, such as Rabphilin 3A and Rim, lack the R(R/K)HHCRxCG motif and the arginine at position 68 and are thus characterized separately as FYVE-related domain containing proteins. These proteins contain a conserved

glycine at position 39. The FYVE domain of MPFYVE resembles the FYVE-related domain as it lacks the R(R/K)HHCRxCG motif and the arginine at position 68. MPFYVE contains only six cysteines instead of eight and hence represents a variant member of this family.

Most of the FYVE domain proteins are involved in endosomal trafficking of proteins, but some are also shown to be involved in other cellular processes, such as signal transduction and cytoskeletal regulation (reviewed by Stenmark and Aasland, 1999). The FYVE domain has been shown to bind to phosphotidylinositol-3 phosphate (PtdIns(3)P) in the basic pocket created by the R(R/K)HHCRxCG motif (Burd and Emr, 1998; Gaullier *et al*., 1998; Patki *et al*., 1998; Stenmark and Aasland, 1999). Proteins containing the FYVE-related domain contain fewer basic residues in this pocket area and only one of them is conserved (position 39, Figure 7). Therefore, it is predicted that these proteins bind to a less acidic or a neutral ligand. Proteins containing the FYVE-related domain play roles in regulated secretion. The MPFYVE protein contains the variant form of the FYVE-related domain and future functional tests will determine if it plays a role in regulated secretion as do the other members of the family.

The second fusion we have identified is *MLL–AF15q14* in which the *AF15q14* exon 12 is fused to *MLL* exon 9. The breakpoint in *AF15q14* reported here is different from the two published breaks in *AF15q14* (Hayette *et al*., 2000, Kuefer *et al*., this issue). The break in *AF15q14* in both cases is in exon 11. The MLL–AF15q14 fusion protein described by Hayette *et al*. contains only 85 amino acids of AF15q14, out of which 43 are encoded by exon 11. The remaining amino acids are unique to the AF15q14 sequence described by Hayette *et al*. and are not present in the sequence reported by Kuefer *et al*. (this issue). The fusion described here and by Kuefer *et al*. contains amino acids encoded by exon 12 and further 3′ of *AF15q14*. Therefore, the AF15q14 amino-acid sequence of the chimeric proteins studied here and by Kuefer *et al*. is completely different from that present in the fusion described by Hayette *et al*. The presence of two different fusion transcripts in which *MLL* is fused to two different genes is unique to the patient we studied. We analysed RNA from the t(11;15) patient sample studied by Kuefer *et al*. and could not amplify an *MLL–MPFYVE* fusion from that patient material.

In the leukemia cells we studied here, the genomic break is in intron 9 of *MLL*, and in intron 11 of *AF15q14*. There is only one genomic break on chromosome 15. The presence of two different rearranged transcripts of *MLL* has been observed in other cases where *MLL* tandem duplication is present along with the *MLL–AF9* fusion transcript (Whitman *et al*., 2001). The two different fusion transcripts we cloned here may be the result of intragenic splicing. Although such intragenic splicing has not been observed for *MLL* fusions before, it has been reported in a t(3;21) translocation (Nucifora *et al*., 1994). In the case of the t(3;21), the *AML1* gene on chromosome 21 is fused to three genes *EAP, MDS1* and *EVI1*, which are separated from each other over a distance of approximately 450 kb. The two genes *AF15* and *MPFYVE* are only 170 kb apart, with *MPFYVE* being telomeric to *AF15q14* (Figure 1d).

The presence of two different *MLL* fusion transcripts raises an important question as to which of the two transcripts (or both) are important for leukemia. Future experiments using the two fusions in retroviral transduction of murine bone marrow progenitors will be able to address this question.

Materials and methods

Patient material

Leukemic cells from an 11-year-old male with *de novo* AML-M2 had a t(11;15)(q23;q14). Cytogenetic analysis at diagnosis was unsuccessful because of inadequate material. The patient relapsed after 1 year and cytogenetic analysis of the metaphase chromosomes revealed a t(11;15)(q23;q14) as one of the cytogenetic abnormalities. This patient material was obtained with informed consent.

Cloning of the chimeric cDNA using 3′ rapid amplification of cDNA ends (RACE)

Total RNA from a frozen bone marrow sample collected at the time of relapse was extracted using Tri Reagent (Molecular Research Corp., Cincinnati, OH, USA) according to the manufacturer's guidelines. The first strand of cDNA was made with 2 μg of total RNA using the Adapter primer following the specifications outlined by the Gibco-BRL 3′ RACE kit (Gaithersburg, MD, USA). Polymerase chain reaction (PCR) was used to amplify the putative fusion transcript using the *MLL* exon 8 top primer *MLL* 5A (Table 1, Figure 1) and the abridged universal amplification primer (AUAP) provided in the kit. One-tenth of the first strand cDNA reaction was used to synthesize the second strand in a standard PCR. The nested PCR was carried out using 1/30th of the first round PCR-amplified DNA, *MLL* 5NP and *AF15* NP2 (Table 1, Figure 1).

The PCR products were resolved by electrophoresis on a 1% TAE–agarose gel, purified using Gene Clean Extraction Kit (BIO 101, Vista, CA, USA) and cloned in a TA cloning vector (Invitrogen, Carlsbad, CA, USA). Sequencing reactions used the ABI PRISM dye Terminator cycle sequencing Reaction Ready kit (Perkin-Elmer, Foster City, CA, USA), following the manufacturer's specifications. Gel electrophoresis, data collection and analyses were performed with an automated sequencer (Model 377A, Applied Biosystems, Foster City, CA, USA). Searches for sequence similarity were performed with the GenBank nucleic acid databases using the BLAST program (Adams *et al*., 1991).

Reverse transcription (RT)–PCR

One-tenth of the first strand cDNA synthesized for the 3′ RACE protocol was used as a template to amplify the chimeric cDNA product using *MLL* top primers and the gene-specific bottom primers as follows:

Amplification of the MLL–MPFYVE fusion transcript A different *MLL* exon 8 top primer (*MLL* 5UNP) and either *MPFYVE* bottom primers B1 or B2 were used in the first round PCR (Table 1, Figure 1). The nested PCR used 1/30th and 1/300th of the amplified DNA from first round, *MLL* 5NP with either NB1 or NB2. The cycling conditions used were

2 min at 94°C, 30 cycles of 30 s at 94°C, 2 min at 58°C, 2 min at 72°C and final extension at 72°C for 7 min.

Amplification of the MLL–AF15q14 fusion transcript *MLL* 5A and the *AF15* 6104B or *AF15* 7421B with an annealing temperature of 60°C were used in the first round of PCR amplification (Table 1, Figure 1). One-thirtieth of the first round PCR-amplified DNA with nested primers *MLL* 5NP and the *AF15* 6084B or 7366B were used to amplify the RT–PCR products (Table 1, Figure 1), respectively. The numbers of the *AF15* primers refer to the position in the nucleotide (nt) sequence. The RT–PCR products were purified by gel electrophoresis using 1.5% TAE–agarose, cloned into TA cloning vector and sequenced as described above.

Long-range genomic PCR

Genomic DNA from the bone marrow sample of the patient was isolated (Puregene DNA isolation kit, Gentra Systems, Minneapolis, MN, USA), following the manufacturer's instructions. Long-range PCR (TaqPlus Long PCR system, Stratagene, La Jolla, CA, USA) was performed using 300 ng genomic DNA, 1 × high-salt buffer, 200 nM of primers *MLL* F4 and *AF15* 6084B, 200 μM each of dATP, dCTP, dTTP, dGTP and 5 U of TaqPlus Long polymerase. The cycling conditions used were 2 min at 94°C, 30 cycles of 30 s at 94°C, 1 min at 60°C, 5 min at 72°C and final extension at 72°C for 7 min. The amplified fragment was purified and sequenced using the *MLL* F4 primer.

Isolation of BACs containing AF15q14 and MPFYVE

A BAC library, containing human genomic DNA (Research Genetics, Huntsville, AL, USA), was screened using the manufacturer's guidelines to isolate genomic clones containing the *AF15q14* and the *MPFYVE* genes. BAC 21B18 containing the *AF15q14* gene was isolated using primers *AF15* T and *AF15* P2. BACs 164E2 and 147G20 containing *MPFYVE* were isolated using primers NG 32 T and NG 406B (Table 1, Figure 1).

Purified BAC DNA was prepared using the Qiagen large-scale plasmid kit. DNA isolated from these BACs and the DNA from the BACs used for sequencing in the database (534F12 and 2339L15) were used as probes in fluorescence *in situ* hybridization (FISH) of patient chromosomes. A previously published probe for *MLL* from ONCOR was used.

Northern blot analysis

The human multiple tissue Northern blot was purchased from Clontech Inc. (Palo Alto, CA, USA). The prehybridization and hybridization were performed according to the manufacturer's protocol using the expressed sequence tag (EST) clones containing the *MPFYVE* sequences as probes.

RNase protection assay

The *MLL–MPFYVE* RT–PCR fragment cloned into PCR2.1 cloning vector (Invitrogen, Inc.) was used to make an antisense RNA probe using T7 RNA polymerase. The probe was synthesized by *in vitro* transcription, using the guidelines and reagents of a Maxiscript kit from Ambion Inc. (Austin, TX, USA) and biotin-labeled UTP (Roche). The RNase protection assay was performed using the procedure described in the RPA III kit (Ambion Inc.). The purified *MLL–MPFYVE* probe (264 pg) was hybridized with 20 μl of total RNA from the patient sample, overnight at 42°C. Two separate controls with the same amount of probe as test and 10 μg of yeast RNA each were set up. One of the yeast RNA tubes was treated with RNase at the end of hybridization similar to the test sample, and the other was not (intact probe). The yeast RNA sample that was treated with RNase is a control for RNase digestion. A human β-actin RNA probe (250 pg) was used to hybridize with 10 μl of patient RNA as a control for quality of RNA. After RNase digestion, the samples were run on a 5% polyacrylamide/8 M urea gel, transferred onto a nylon membrane (Ambion, Inc.), using electroblotting and 1 × TBE buffer (BioRad, Inc., CA, USA) at 80 mA for 1 h. The blot was then processed using BrightStar BioDetect, a nonisotopic detection kit from Ambion, Inc., following the manufacturer's guidelines.

Acknowledgements
We thank Dr Steve Morris and Dr Martin Kuefer for sharing sequence information of *AF15q14* and the preliminary data about the size and the expression pattern of *AF15q14* prior to publication.This study was supported by grants CA 42557 and CA 84405 (J.D.R.); CA40046 and CA78438 (N.J.Z.-L.); a postdoctoral fellowship award from Leukemia Research Foundation and Cancer Biology training grant, T32 CA09595 (VC) and by the Spastic Paralysis Foundation of Illinois, Eastern Iowa District of Kiwanis International.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick, M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B and Moreno RF. (1991). *Science*, **252,** 1651–1656.

Ayton PM and Cleary ML. (2001). *Oncogene*, **20,** 5695–5707

Bernard OA and Berger R. (1995). *Genes Chromosomes Cancer*, **13,** 75–85.

Breen TR and Harte PJ. (1993). *Development*, **117,** 119–134.

Burd CG and Emr SD. (1998). *Mol. Cell*, **2,** 157–162.

Caldas C, So CW, MacGregor A, Ford AM, McDonald B, Chan LC and Wiedemann LM. (1998). *Gene*, **208,** 167–176.

DiMartino JF and Cleary ML. (1999). *Brit. J. Hematol.*, **106,** 614–626.

Divoky V, Trka JM, Watzinger F and Lion T. (2000). *Gene*, **247,** 111–118.

Gaullier JM, Simonsen A, D'Arrigo A, Bremnes B, Aasland R and Stenmark H. (1998). *Nature*, **394,** 432–433.

Gibbons B, Katz FE, Ganly P and Chessells JM. (1990). *Br. J. Heamatol.*, **74,** 264–269.

Gu Y, Nakamura T, Alder H, Prasad R, Canaani O, Cimino G, Croce CM and Canaani E. (1992). *Cell*, **71,** 701–708.

Hanson RD, Hess JL, Yu BD, Ernst P, van Lohuizen M, Berns A, van der Lugt NM, Shashikant Ch, Ruddle FH, Seto M and Korsmeyer SJ. (1999). *Proc Natl. Acad. Sci.*, **96,** 14372–14377.

Hayette S, Tigaud I, Vanier A, Martel S, Corbo L, Charrin C, Beillard E, Deleage G, Magaud JP and Rimokh R. (2000). *Oncogene*, **19,** 4446–4450.

Hernandez JM, Mecucci C, Beverloo HB, Selleri L, Wlodarska I, Stul M, Michaux L, Verhoof G, Van Orshoven A and Gassiman JJ. (1995). *Leukemia*, **9,** 1162–1166.

Hillion J, Le Coniat M, Jonveaux P, Berger R and Bernard O. (1997). *Blood*, **90,** 3714–3719.

Hunger SP, Tkachuk DC, Amylon MD, Link MP, Carroll AJ, Welborn JL, Willman CL and Cleary ML. (1993). *Blood*, **81,** 3197–3203.

Kaneko Y, Shikano T, Maseki N, Sakurai M, Sakurai M, Takeda T, Hiyoshi Y, Mimaya J and Fujimoto T. (1988). *Leukemia*, **2,** 672–676.

Kuefer MU, Behm FG, Naeve CW, Rakestraw KM, Raimondi SC and Morris SW. (1997). *Blood*, **90** (Suppl 1), 1235.

Lavau C, Du C, Thirman M and Zeleznik-Le N. (2000a) *EMBO J*., **19,** 4655–4664.

Lavau C, Luo RT, Du C and Thirman MJ (2000b) *Proc. Natl. Acad. Sci.*, **97,** 10984–10989.

Lavau C, Szilvassy SJ, Slany R and Cleary ML. (1997). *EMBO J*., **16,** 4226–4237.

Martinez-Climent JA, Thirman MJ, Espinosa III R, Le Beau MM and Rowley JD. (1995). *Leukemia*, **9,** 1299–1304

Nilson I, Lochner K, Siegler G, Greil J, Beck JD, Fey GH and Marschalek R. (1996). *Br. J. Haematol*., **93,** 966–972.

Nucifora G, Begy CR, Kobayashi H, Roulston D, Claxton D, Pedersen-Bjergaard J, Parganas E, Ihle JN and Rowley JD. (1994). *Proc. Natl. Acad. Sci.*, **91,** 4004–4008.

Patki V, Lawe DC, Corvers S, Virbasius JV and Chawla A. (1998). *Nature*, **394,** 433–434.

Raimondi SC, Peiper SC, Kitchingman GR, Behm FG, Williams DL, Hancock ML and Mirro J Jr. (1989). *Blood*, **73,** 1627–1634.

Rowley JD. (1993) Seminars in Cancer Biology, ed. Rabbitts, T. H. (Academic, London). *Semin. Cancer Biol*., **4,** 377–385.

Rowley JD (1999). *Semin. Hematol.*, **36** (Suppl 7), 59–72.

Rubnitz JE, Link MP, Shuster JJ, Carroll AJ, Hakani N, Frankel LS, Pullen DJ and Cleary ML. (1994). *Blood*, **84,** 570–573.

Schwabe JWR and Klug A. (1994). *Nature Struct. Biol.*, **1,** 345–349.

Stenmark H and Aasland R. (1999). *J. Cell Sci*., **112,** 4175–4183.

Stenmark H, Aasland R, Toh BH and D'Arrigo A. (1996). *J. Biol. Chem*., **271,** 24048–24054.

Tkachuk DC, Kohler S and Cleary ML (1992). *Cell*, **71,** 691–700.

Thirman MJ, Gill HJ, Burnett RC, Mbangkollo D, McCabe NR, Kobayashi H, Ziemin-van der Poel S, Kaneko Y, Morgan R, Sandberg AA, Chaganti RSK, Larson RA, Le Beau MM, Diaz MO and Rowley JD. (1993). *N. Engl. J. Med.*, **329,** 909–914.

Waring PM and Cleary ML (1997). *Curr. Top. Microbiol. Immunol*., **220,** 1–23.

Whitman SP, Strout MP, Marcucci G, Freud AG, Culley LL, Zeleznik-Le NJ, Mrozek K, Theil KS, Kees UR, Bloomfield CD and Caligiuri MA. (2001). *Cancer Res.*, **61,** 59–63.

Yu BD, Hess JL, Horning SE, Brown GA and Korsmeyer SJ. (1995). *Nature*, **378,** 505–508.

Zeleznik-Le NJ, Harden AM and Rowley JD. (1994). *Proc. Natl. Acad. Sci.*, **91,** 10610–10614.