Leukemia (2001) 15, 1906–1913

www.nature.com/leu

Molecular characterization of genomic *AML1-ETO* fusions in childhood leukemia

Z Xiao[1], MF Greaves[1], P Buffler[2], MT Smith[3], MR Segal[4], BM Dicks[5], JK Wiencke[5] and JL Wiemels[5]

[1]*Leukaemia Research Fund Centre, Institute of Cancer Research, London, UK;* [2]*Division of Public Health Biology and Epidemiology, University of California, Berkeley, CA, USA;* [3]*Division of Environmental Health Sciences, University of California, Berkeley, CA, USA;* [4]*Division of Biostatistics, University of California, San Francisco, CA, USA; and* [5]*Laboratory for Molecular Epidemiology, Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA*

T(8;21) *AML1(CBFA2)-ETO(MTG8)* is the most common chromosomal translocation in acute myeloid leukemia (AML) in both children and adults. We sought to understand the structure and gain insight into the fusion process between *AML1* and *ETO* by sequencing genomic fusions in 17 primary childhood AMLs and two cell lines with t(8;21). Reciprocal translocations were sequenced for seven of the 19 samples. We assumed a null hypothesis that the translocation breakpoints would be evenly distributed along the intronic breakpoint cluster regions. Testing for multimodality via smoothed bootstrap statistical methods suggested, however, the presence of two separate cluster regions within both the *AML1* and *ETO* breakpoint cluster regions. *ETO* breakpoints were predominantly located in intron 1B in a defined cluster 5′ of exon 1A (scan statistic *P* value = 0.00001). All patients with available RNA expressed an *AML1-ETO* mRNA fusion between exon 5 of *AML1* and exon 2 of *ETO*. Since the structural restraints for the fusion protein of *AML1-ETO* exclude exon 1A, we reason that *ETO* intron 1B harbors a structural feature with propensity for breakage and/or recombination. Chromosomal breakpoints displayed evidence of fusion by a non-homologous end joining process, with microhomologies and nontemplate nucleotides at some fusion junctions. Breakpoints in general displayed similar complexity of duplications, deletions, and insertions to other common pediatric leukemia translocations (*TEL-AML1, MLL-AF4, PML-RARA, CBFB-MYH11*) that we and others have analyzed. *Leukemia* (2001) **15,** 1906–1913.
Keywords: translocation; acute myeloid leukemia; AML1; ETO; cluster

Introduction

Chromosomal translocations resulting in specific fusion genes are a hallmark of the leukemias. While studies of the fusion proteins have yielded extensive information on the biology and clinicopathology of the leukemias, less is understood about the structure and etiology of chromosomal fusions themselves. Genomic mapping and sequencing of chromosomal fusion junctions of specific types has led to speculations on the possible mechanisms of formation of these fusion junctions. The most explicit causal mechanism appears to be the aberrant involvement of the V(D)J recombinase *RAG* genes in the formation of chromosomal fusions of *TCR* and *IG* genes with various oncogenes, and occasionally between introns of genes not involved in *TCR* or *IG* rearrangement in the lymphomas and some leukemias (eg Refs 1–5). Clinical and molecular reports have implicated multiagent chemotherapy including topoisomerase II-inhibiting drugs as a putative cause of several leukemia translocation types, including *MLL* rearrangements, *NUP98-EOS* and various fusions involving *AML1*.[6–12] Progress in characterizing the majority of leukemia translocations has been slower due to dispersed nature of the chromosomal fusions and difficulties in sequencing translocations because of large intronic regions; however, large-scale sequencing efforts have recently been undertaken.[13–17]

The only pediatric translocation to date with hypothesized etiologies are the *MLL* fusions in infant leukemia, in which topoisomerase II inhibition by environmental and dietary agents may play a role.[18–20] The most common rearrangement in childhood leukemias is t(12;21)*TEL-AML1*, and we and others have noted structural features of the introns that may contribute to genomic breakage and refusion, including unstable repeat sequences and signs of non-homologous end joining repair.[14,21–24]

T(8;21) (*AML1-ETO*) is a common recurrent translocation (~12% overall) in both childhood and adult AML. This translocation is primarily a *de novo*, or idiopathic event, by the lack of identified causal exposure in most of the cases. T(8;21) occasionally occurs in leukemias associated with prior exposure to cancer chemotherapy of both alkylator and topoisomerase II inhibitor activity,[8,25–29] and has also been reported in leukemias associated with certain occupational exposures.[30–33] *AML1* is occasionally fused to other partners in therapy-related leukemias, including *EVI-1/MDS/EAP*, *MTG16*, and other partners in chromosomes 1, 12, 14 and others.[9,34,35] *AML1-ETO* fusion produces a chimeric oncoprotein consisting of the *runt*-homology domain of *AML1* on chromosome 21 fused to nearly the entire gene of *ETO* on chromosome 8. *AML1* and *ETO* are both involved in transcriptional regulation of genes in hematopoietic precursor cells (reviewed in Ref. 36).

In an effort to explore the nature of *de novo AML1-ETO* translocations, we sequenced the genomic fusion between chromosomes 8 and 21 in a series of 17 primary childhood leukemias and two cell lines that were derived from adult leukemias. *AML1-ETO* fusions were previously demonstrated by Tighe and Calabi, using Southern blotting, to predominantly harbor breakpoints in intron 1B of *ETO*.[37,38] Using genomic sequencing and two complementary statistical approaches we confirmed and further demarcated this particular feature as well as demonstrating evidence of clustering of *AML1* breakpoints. We also note other features indicating a complex breakage-refusion process that is consistent with the 'damage-repair' model of chromosomal translocation.

Materials and methods

Patients and cell lines

Patients were derived from two case–control epidemiology studies of childhood leukemia, the UK Childhood Cancer Study and the Northern California Childhood Leukemia Study. These are population-based epidemiology studies and therefore we anticipate no bias in patient selection that would

Correspondence: JL Wiemels, Laboratory for Molecular Epidemiology, Department of Epidemiology and Biostatistics, University of California, San Francisco, 94143–0560, CA, USA; Fax: (415) 476–6014
Received 21 May 2001; accepted 9 August 2001

influence breakpoint characteristics. Biological samples were obtained before treatment, by consent, and consisted of either blood or bone marrow. T(8;21) translocations were identified in diagnostic karyotypes. For some patients, RNA was isolated from mononuclear blast cells (purified by Ficoll density-differential centrifugation), using phenol-guanidium isothiocyanate procedures. Expression of *AML1-ETO* was confirmed by reverse transcription-PCR (RT-PCR) using standard primers and techniques.[39] High molecular weight DNA was isolated using conventional SDS-proteinase K digestion DNA isolation methods. All studies were reviewed and passed by the Institutional Review Boards at the institutions of the authors, and at all of the hospitals where patient material was derived.

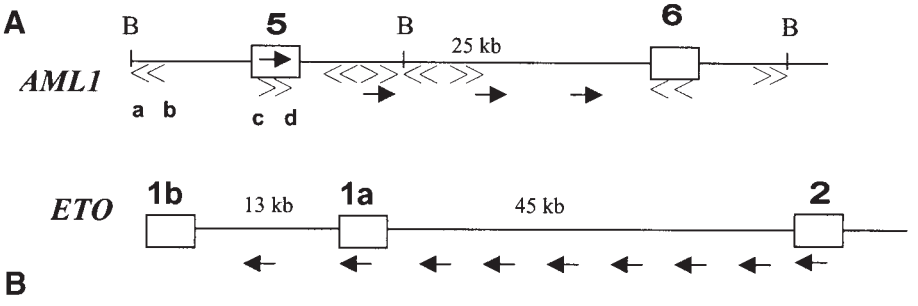
The Kasumi[40] and SKNO-1[41] cell lines were kindly provided by N Kamada (Research Institute for Nuclear Medicine and Biology, Hiroshima, Japan) and S Matozaki (Hyogo Medical Center for Adults, Japan), respectively. DNA and RNA were isolated as above. A final patient sample previously sequenced was included in the clustering analysis.[42]

Sequencing of genomic fusion of AML1-ETO

Genomic fusions of *AML1* intron 5 and *ETO* intron 1 or 2 were performed by long distance inverse PCR (LDI-PCR) especially when DNA was limiting, and by conventional long distance PCR techniques. Since the *AML1* genomic breakpoint region is smaller than *ETO*, the *AML1* region was used as the anchor for LDI-PCR. *Bam*H1 digestion yields two large fragments of this region, allowing for four sets of inverse primers (Figure 1). Inverse PCR was performed exactly as described,[22] but using only a single enzyme, *Bam*H1, which creates two fragments from *AML1* exon 5. This allowed for four primer sets to identify the *AML1-ETO* fusion (AMLinv1 and AMLinv3 primer set) or the *ETO-AML1* reciprocal (AMLinv2 and AMLinv4 primer sets). Six pmoles of each primer in 60 mM Tris-SO_4 (pH 9.1), 18 mM $(NH_4)_2SO_4$, 1.7 mM $MgSO_4$, 200 μM each deoxynucleotidyl transferase (dNTP), and 0.5 μl Elongase enzyme (Life Technologies, Carlsbad, CA, USA) in a 50 μl reaction was subjected to 35 cycles (94°C for 1 min, 60°C for 1 min, and 72°C for 1 min) followed by an extension at 72°C for 7 min. PCR products were checked on agarose gels.

For conventional LD-PCR, nine primers at approximately 6 kb intervals were designed for *ETO*, and four primers for *AML1* (Figure 1). To limit the number of reactions, two *AML1* primers (AML 6 and 18 in one reaction, and 12 and 24 in the other) were combined with each *ETO* primer, to allow for a total of 18 reactions for each patient. The three primers (6 pmoles each) were combined in a PCR reaction with 200 ng DNA with Elongase enzyme (Life Technologies) and the same reaction conditions as described above. PCR cycling for all LDI and LD reactions consisted of 40 cycles of 92°C, 30 s, 68°C, 20 min.

Amplification bands with suspected rearrangements (based on size) were excised from agarose gels, cleaned (Qiagen Gel Purification Kit, Valencia, CA, USA), and sequenced using the same primers as were used in the PCR reaction. Upon confirming the presence of *AML1* and *ETO* sequence at the distal ends of the PCR product, additional PCR primers were synthesized at 1 kb intervals on the *AML1* gene only. An additional round of PCR with the single *ETO* primer (or the *AML1* inverse PCR primer proximal to the *Bam*H1 site) and each new 1 kb interval *AML1* primer was used to narrow down the location of the breakpoint. Usually, only one additional sequencing reaction with the *AML1* primer most proximal to the breakpoint revealed the breakpoint sequence.



821A-6 GTCTTCACAAACCCACCGCAAGTCG	AML-inv1A CCTACAACTCTAGCACCTCAGTGAC [b]
821A-12 GCTGGTCACAGGTGCCAGCAGTTG	AML-inv1A1 GTCCCAAATCAAGGTGGCAGCAG [a]
821A-18 CTGAGTCTCAGTGCCAGTTTCCCAGGA	AML-inv1B GCTTCACTCTGACCATCACTGT [c]
821A-24 CTGATGAAACAGCTGGAGGCAGAGGG	AML-inv1B1 CCGCAAGTCGCCACCTACCACAGAG [d]
ETO-1B CACTTACTGAGCACCCTCAGTGCAAGTC	AML-inv2A GGTTTTGCGAGGGTCCAGACGG
ETO-1A GCACATGTGGCCTTGAACCCAGCC	AML-inv2A1 CTCTTCTCTGCTCCCACATTGG
ETO-6 AGCCATGGATACCCACCAGTCTCACC	AML-inv2B GCCACACCAGTGAGGTCAGGATC
ETO-13 CAGAAACAGATTTCCAACTGGACACTGCC	AML-inv2B1 CTGTCTGCATTTAGGCTTCTG
ETO-20 CCAGACAATTCACCTGGACCTCAGTCTCTC	AML-inv3A CTCGTCCCATTGCCTGGTTCAACAG
ETO-27 CCAGGAAAACCATTGGGGAACCTAGAAC	AML-inv3A1 GCTCAAGTTCACACCCACTGGCC
ETO-34 GGAACCGAGAGGTGCATGGTGATGTTA	AML-inv3B ACATGCACTCAGGTGGTACATGAG
ETO-39 CCTCCACATCTAGAAAGCAGGTCTACAGCAC	AML-inv3B1 GAAATTCTCTGGCTGTAAGTGGCTGC
ETO-46 GGTGAGTCTGGCATTGTGGAGTGCTTC	AML-inv4A GTGGAGTGGTTCAGGGAGGCACG
	AML-inv4A1 GCCGCAGCTGCTCCAGTTCACTGAG
	AML-inv4B CAGCTAAGTGGCCTGTCTGTGGAG

Figure 1 (A) Location scheme of PCR primers to amplify *AML1-ETO* fusion junctions. Numbers refer to exons. (B) refers to *Bam*H1 sites in *AML1*. Closed arrows are LD-PCR primers, and open-ended arrows refer to LDI-PCR primers (lower case a, b, c, and d refer to 'inv1' set of primers in (B); the same orientation holds for the three other sets). (B) Primer sequences for LD-PCR (left) and LDI-PCR (right).

Statistical analysis

Using GenBank-derived genomic sequence, maps of breakpoint locations were constructed. Clustering of breakpoints was assessed using both Silverman's (1981) smoothed bootstrap procedure and scan statistic approximations, these being either large deviation[43] or moment[44] based. The approaches are complementary in that the smoothed bootstrap procedure provides a formal test for the number of clusters (equivalently modes) that is global (ie dependent on all the data) in nature, whereas the scan statistic is tuned to assessing the significance of a target (local) cluster and is unaffected by the data density/sparsity outside that target. Detailed description of methods are presented elsewhere.[45]

Silverman's smoothed bootstrap procedure uses (Gaussian) kernel density estimation. A critical bandwidth is defined such that any decrease in this bandwidth results in the corresponding density estimate gaining a mode. The density estimate associated with the critical bandwidth is called the critical density. Smoothed bootstrap resampling from this critical density is then performed a large number of times. For each replicate sample, a kernel density estimate, using the critical bandwidth, is constructed and the number of modes determined. The proportion of times this number exceeds the number of modes in the critical density gives the *P* value for whether an additional mode is required.

The scan statistic is based on the maximal number of breakpoints occurring within an interval of prescribed length. The scan statistic null distribution is based on the maximum number of breakpoints in the interval assuming a uniform breakpoint density, this reflecting an absence of clustering. Again, a detailed description of this methodology is presented elsewhere.[45]

Results

All of the 19 available patient samples and cell lines were successfully sequenced for *AML1-ETO* using the combined approaches of LD- and LDI-PCR (six were sequenced by both approaches). One breakpoint (patient 3) mapped to chromosome 11 rather than chromosome 8, indicating that this breakpoint may be a three-way translocation. For seven patients, amplification of the reciprocal *ETO-AML1* by inverse PCR (using the inv 2A/B or inv 4A/B combinations, see Figure 1) were successful. For the remainder, breakpoints of the reciprocal may have been outside of the introns targeted, or simply yielded PCR products out of the practical range of the inverse PCR assay.

Of the 19 total patients and cell lines available, RNA was obtained from nine. RT-PCR analysis of these nine revealed a single band indicating a fusion between exon 5 of *AML1* and exon 2 of *ETO*. It was therefore intriguing that 13 of the 19 breakpoints, plus one additional previously sequenced *AML1-ETO* fusion,[42] were positioned prior to an alternative exon 1 of *ETO* (exon 1A) which was not found in any mRNAs tested (Figure 2), confirming an earlier observation using Southern blotting.[37,38] mRNA sequences of all patients tested were indistinguishable with regards to mRNA sequence to Kasumi and SKNO-1 cell lines, which both had breakpoints within intron 2 of *ETO* (Figure 2).

AML1-ETO fusions displayed a high degree of complexity proximal to the breakpoint; a similar or higher complexity than another breakpoint that we have analyzed extensively, *TEL-AML1*. For example, seven of 27 (26%) *AML1-ETO* fusions displayed nontemplate *N*-nucleotides, and another 14 (52%) displayed microhomologies at breakpoints (Table 1). Only six (22%) lacked either feature. *TEL-AML1* fusions that we have sequenced (data from Refs 14 and 22) demonstrated *N*-nucleotides in only 1/27 breakpoints (4%), but a similar rate of microhomologies, at 14 of 27, (52%). Of the seven breakpoints in which *AML1-ETO* and its reciprocal were sequenced, deletions of 5 to 18 099 bp were present in eight of 14 breakpoints and duplications of 1 to 1512 bp were apparent in the other six breakpoints (Table 2). These large and variable additions/deletions of DNA indicate that the translocations were not exactly reciprocal, but that DNA at the proximal ends of breakpoints was extensively processed prior to fusion. A single fusion (breakpoint No. 18) demonstrated an internal tandem duplication on the same derivative chromosome (Figure 3), a feature not previously observed in *TEL-AML1* translocations. Despite the presence of nontemplate nucleotides in seven breakpoints, other features of V(D)J recombination were absent from breakpoints, including consensus heptamer–nonamer RSS sequences or strong clustering around cryptic RSS sequences. In addition, the duplications present in reciprocal translocations are not compatible with typical V(D)J-mediated recombination mechanisms (Figure 4).

A statistical test of clustering, the scan statistic, was applied to both *AML1* and *ETO* genomic regions for the 20 total breakpoints. Scan statistics are based on the number of events occurring within a window along the intronic sequence. A marginally significant cluster was apparent on the 3′ end of *AML1* exon 5 ($P = 0.09$, right mode on Figure 5a). The 3′ region of *AML1* did not harbor a significant cluster ($P = 0.83$, left mode, Figure 5a). *ETO*, on the other hand, demonstrated a highly significant cluster ($P = 0.00001$) 5′ to exon 1A (Figure 2a and 5b). This held true even when evaluating significance without using prescribed interval lengths. To evaluate the overall clumping of breakpoints in the region, we applied a bandwidth test with significance evaluation by smoothed bootstrap after Silverman.[46] This procedure is more global in nature than the scan statistic, and is used to assess the number of modes, or bumps, in the breakpoint density. This procedure rejected the null hypothesis of a single cluster for both *AML1* ($P < 0.0001$ at a critical bandwidth of 27 826 bp) and *ETO* ($P = 0.001$ at a critical bandwidth of 47 078 bp) but did not reject a null of two clusters (null hypothesis for two clusters for *AML1*, $P = 0.73$ at a bandwidth of 6013; *ETO*, $P = 0.31$ at a bandwidth of 14 420). The data then suggest the existence of two modes on each gene as depicted in Figure 5. This is an indication of overall clumping, or clustering, of the breakpoints, but not a formal assessment of the significance of individual clusters as provided by the scan statistic.

Discussion

AML constitutes about 20% of all childhood leukemias, and t(8;21) occurs in 12–20% of these cases. While t(8;21) is generally a marker of a more favorable prognosis in adults, this is not true for children, with only a 30% long-term remission rate.[47–49] Ultimately, prevention of this and other leukemia subtypes is a goal, one that can only be reached by better understanding the disease's causes.

Apart from infant leukemias with *MLL* translocations, there are no known and few hypothesized etiologies for childhood acute myeloid leukemia (AML). In the current report, we describe methods to rapidly sequence and analyze t(8;21) fusions in children for the purpose of addressing the question

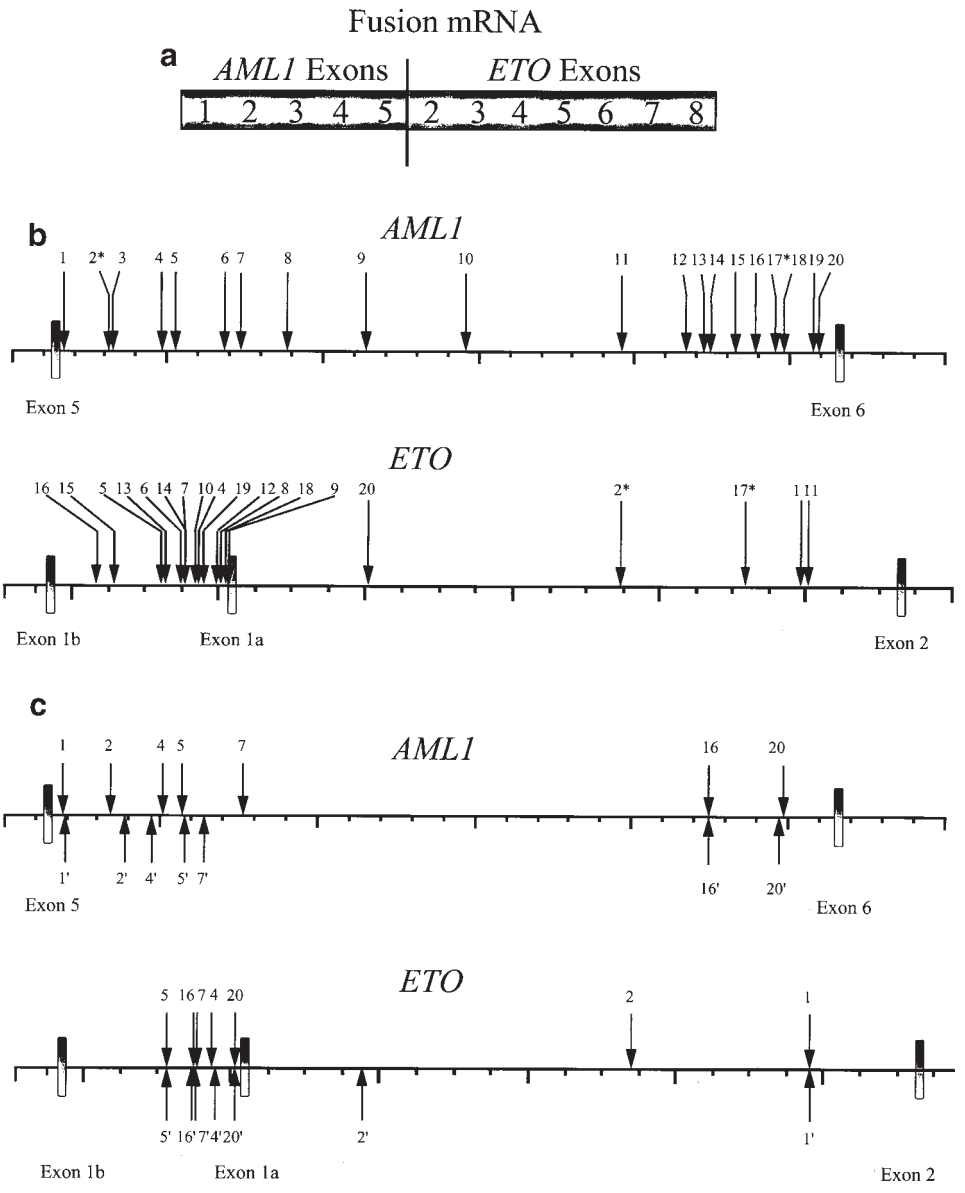


Figure 2 Map of 18 *AML1-ETO* breakpoints. (a) All patients tested exhibited the AML1 exon 5-ETO exon 2 fusion mRNA. (b) Scale map of *AML1-ETO* breakpoint positions along *AML1* and *ETO* intronic regions. Breakpoints 1, 5, 6, 7, 8, 11, 18, and 20 are from the UK Childhood Cancer Study; breakpoints 3, 4, 9, 10, 12, 13, 15, 16, and 19 from the Northern California Childhood Leukemia study; breakpoints 2 and 17 from Kasumi and SKNO-1 cell lines, respectively, and breakpoint 14 is patient KH.[42] (C) Map of breakpoints of the der8 (*AML1-ETO*) and reciprocal der21 (*ETO-AML1*) for those patients in which both were sequenced. *AML1-ETO* are indicated with arrows above the introns, and *ETO-AML1* below.

of what caused this fusion. These data, when combined with large-scale epidemiology studies of childhood leukemia and efforts to define the temporal origin of breakpoint fusions currently in progress, should ultimately result in a defined etiology of the disease and suggest preventative strategies.

We begin with the knowledge from clinical and functional studies that t(8;21) *AML1-ETO* constitutes a fusion protein between exon 5 of *AML1* and exon 2 of *ETO*. Functional studies have implicated that both the *AML1* and *ETO* protein subunits are necessary for the oncogenic function of the chimeric gene,[50,51] despite reports of the occurrence of alternative spliced mRNAs which are present at much lower relative quantity to the *AML1* exon 5–*ETO* exon 2 fusion in leukemic blasts.[38,52,53] These studies of aberrant transcripts have all utilized non-quantitative nested PCR or PCR-Southern blotting to detect such transcripts, and do not indicate their relative abundance to the *AML1* exon 5–*ETO* exon 2 transcript. Seven of our patient samples with available RNA and two cell lines expressed the common transcript exclusively after a single round of PCR. We previously found in a screen of 40 patients that the *AML1* exon 5–*ETO* exon 2 fusion was the exclusive detectable product after a single round of PCR followed by Southern blotting and hybridization of the product.[54]

Additional work on the pathophysiologic relevance of the fusion protein has indicated that the *ETO* domain is necessary for the transcriptional alterations triggered by the fusion protein, making the contribution of aberrant out-of-frame transcripts uncertain. It is therefore peculiar that all but two of the childhood leukemia *AML1-ETO* fusions were within intron 1B of *ETO*, and furthermore that 12 of the breakpoints were

Table 1 Twenty t(8;21) *AML-ETO* breakpoints demonstrating fusion *AML1-ETO* (der8) and *ETO-AML1* (der21)

No.	*Derivative chromosome*	*Translocation fusion sequence*	*AML1 breakpoint*[a]	*ETO breakpoint*[b]
1	Der 8	TCCACTTGGGGCTGGTACA**C**TTTGCTTTTTTCTGCCTGTT	21 808 391	201 552
	Der 21	ATGATAATCTCTCATTACTcCTCAAAGCTTGTTTCTGTGC	21 808 257	201 557
2	Der 8	CTTCTTTATGAGTGAAAAGCTTGAGAACACTTTCCTGTAT	21 806 837	213 882
	Der 21	AAGAGGTTATTAAACCTC**TT**GGCAGTCAGAGGGGAAAAAA	21 806 318	231 982
3	Der 8	AATGTAACCCTGCATTAAACGGTCAGGGTGCCATCTATCT	21 806 718	317 899[c]
4	Der 8	CAGCTCGTCTGTTTTTCA**T**CTCAGATGCACACACATTCAG	21 805 110	242 555
	Der 21	TGGCTTTTCAAAAATAC**TCTCT**CTATCCATACCATACTGT	21 805 444	242 118
5	Der 8	AGGATATTTGAGAGTTTTtTCAACCAGAGTTAAAAGTCAA	21 804 678	245 196
	Der 21	AGAGGACATTGTAC**TTCAA**aa**AGAGT**CTCACTCTGTTGCC	21 804 649	245 193
6	Der 8	TATAGCAAACCTATTATAT**A**GGGTCTCCTCTAAGAGCAGG	21 803 139	243 778
7	Der 8	TAAATAGTGTTATTTGCC**CA**GCTCCTTTGAATTAACTCAC	21 802 586	243 434
	Der 21	GCAGGCCAGAGCAGATGG**GG**GCAGCCTTTACCTGTCTGGT	21 804 099	243 488
8	Der 8	TGTTTTGTTTTGTTTTGTTCctGGACTGCTGAGAAGAAAAT	21 801 165	241 170
9	Der 8	CATCAACTTCTATGTGTAAAGGCAAAATAAGGGTTAACCT	21 798 572	240 543
10	Der 8	GTAGGAGCAATTGTTCTTTgggaTTGCATAAATGCATTC	21 795 400	242 822
11	Der 8	CCATACACAGGTGGACCC**AG**ACATTAATTTTAATGTTTAT	21 790 374	201 060
12	Der 8	CACAGAAATGTGTATTTCTACAAGTATAGTACTAAAAAAA	21 788 384	242 181
13	Der 8	TATATAGTTTATATTATATAATTATTAACTTACTTTAACT	21 787 860	245 009
14	Der 8	GCCAACATCTTCTGAAATACGGCAGGCCAGAGCAGATGGG	21 787 517	243 507
	Der 21	ACTATTAATTCTCTTGTATagGACTATATAAAATTCTTAT	21 787 500	243 520
15	Der 8	AACTCTGCTTAAGATTAAG**A**AGTGTTATGTGTATCCTCAG	21 786 798	248 006
16	Der 8	GTTTTCTCTGAGAGTTATGCATGCAGACTGAAGGGCATTT	21 786 102	248 371
17	Der 8	ACTCGTAATGCCAGGCAT**TA**ATCGATTGCCATAGATTTAT	21 785 538	208 350
18	Der 8	GCATGCTGAGCCTGCGTGAgagTAAAAATGTTTGTAGGCT[d]	21 785 181	240 713
	Der 21	AACTGGCACAATATACAAcctCCAAGAACAGGCCTCCCAC	21 785 220	240 722
19	Der 8	ACTCGTAATGCCAGGCAT**TA**ATCGATTGCCATAGATTTAT	21 784 223	242 270
20	Der 8	GGGGCCCTACCACTAAGC**TG**TAATTTATCCTGATTATACT	21 784 026	231 096

[a]GenBank accession: NT_011512.3.
[b]GenBank accession: NT_008075.4.
[c]Chromosome 11, GenBank accession: NT_009115.3.
[d]See Figure 3 for complete fusion structure.
Key of patient numbers is in caption to Figure 2. *AML1* sequences are underlined, *ETO* are not. Microhomologies (nucleotides at the breakpoint that are the same on both *AML1* and *ETO*) are noted in bold, and non-template sequences in lower case letters. Locations of breakpoints are referenced to locations in GenBank sequences.

Table 2 Duplications (+) and deletions (–) at *AML1-ETO* breakpoints and their reciprocals

	AML	*ETO*
1	–48	–4
2	–518	–18 099
4	+334	+438
5	–28	+4
7	+1512	–53
16	+18	–13
20	+38	+8

located within a range of 3245 bp, resulting in highly significant clustering by statistical analysis (Results and Figure 5b). This clustering, as well as evidence of clumping of breakpoints at the 5′ and 3′ ends of *AML1* intron 5, could indicate regions of DNA or chromatin structure particularly susceptible to recombination in comparison to the surrounding sequence. The lack of specific recombination site sequences or DNA interspersed repeats in proximity to the breakpoints argues against the presence of 'hotspots' such as V(D)J recombinase site sequences. This complexity of the breakpoints is shared by *MLL-AF4*, *TEL-AML1*, *PML-RARA*, and *CBFB-MYH11*, translocations which occur in pediatric ALL as well as AML.[14–16,22,55–57] These translocations like *AML-ETO* in children are not associated with etiologic causal agents in the vast

Figure 3 Fine structure of *AML1-ETO* fusion from patient No. 18. An internal tandem duplication (underlined) is shown at the fusion. Non-template nucleotides are in lower case letters, and diagonal hash marks indicate a gap in sequence between the duplicated region of *AML1*.

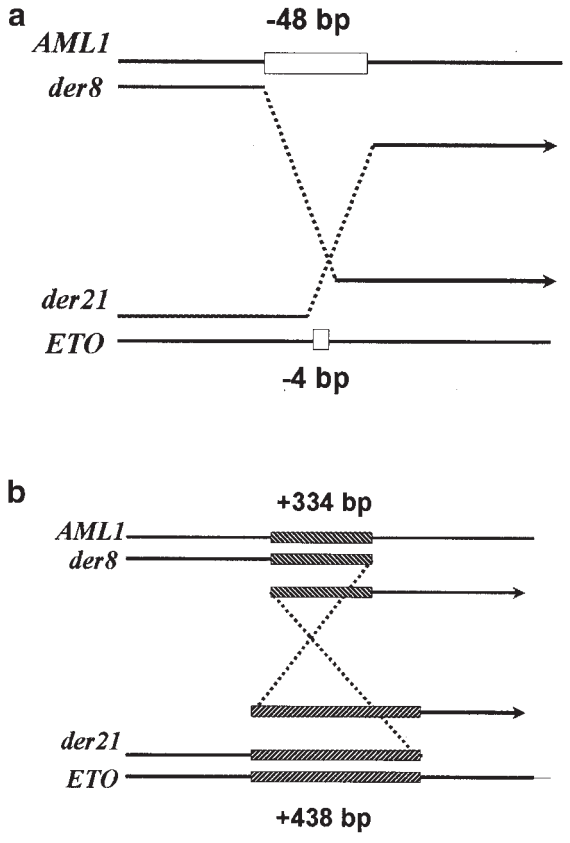


Figure 4 Schematic diagrams indicate duplications (striped boxes) and deletions (hollow boxes), and demonstrate pictorially the relationship between the der 8 (*AML1-ETO*), der 21 (*ETO-AML1*) and the parent chromosomes containing *AML1* and *ETO*. We represent only two patients, as an example, in which information regarding both reciprocals is known, information about the other five are in Table 2. (a) Patient No. 1; (b) patient No. 4.

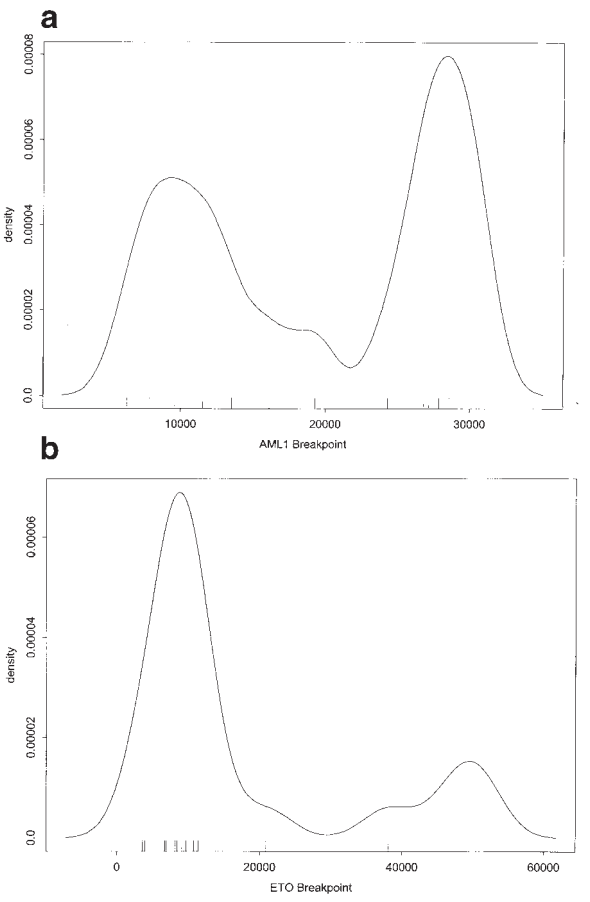


Figure 5 Silverman smoothed bootstrap for the determination of the numbers of cluster 'modes'. The curves are drawn according to breakpoint density, and smoothing based on bootstrap resampling (see Methods). Breakpoint density maps visually suggest two or three cluster modes for both *AML1* and *ETO*. The corresponding statistic for both regions rejects the null hypothesis of one single cluster, but fails to reject, in a separate test, the null hypothesis of two clusters (see Results), thus providing evidence for two separate cluster regions within both (a) *AML1* and (b) *ETO* introns.

majority of cases. *TEL-AML1* is known to occur *in utero* in most patients tested,[58–60] and the temporal origin of *AML1-ETO* in childhood leukemia is currently under scrutiny. Future studies on *AML1-ETO* breakpoint etiology will benefit from a comparison of the *de novo* children's breakpoints described in the current report with those in adult and therapy-related leukemias.

Several etiologies of the adult AMLs are well defined, most notable, AML resulting from prior multiagent chemotherapy regimens, radiation, and benzene. T(8;21) is a common translocation in the AML subclass FAB-M2 (~40%), and has been associated with benzene exposure.[61] Translocation(8;21) is also an occasional cytogenetic abnormality in therapy-related leukemias, and has in particular been associated with exposure to the topoisomerase II inhibitor, doxorubicin.[8,29] In this context, the marginally significant cluster at the 3′ end of *AML1* ($P = 0.09$, Figure 5a) is worthy of notice in light of a strong *in vivo* topoisomerase II site found in this area by Stanulla and colleagues,[62] who used topoisomerase II-inhibiting drug-treated cultured cell lines in defining this hypersensitive region. A near-consensus topoisomerase II site was also found in DNA sequence analysis,[62] and one of our childhood t(8;21) breakpoints (No. 18) was found to be located 30 bp from this site, with nine of the 20 breakpoints forming a cluster to the 5′ end of the site. The potential significance of this association is unknown, as we emphasize that pediatric AML with t(8;21) has not been associated with exogenous chemical exposures in any studies to date. The sequencing of additional patient breakpoints, in particular those from adults with occupational exposures or prior exposure to chemotherapy, will provide more information about the relevance of this feature.

A final interesting feature of t(8;21) is the presence of non-template nucleotides at a significant number of the breakpoints. These nucleotides were present at a higher rate (7/25, 28%) than that which we previously observed for *TEL-AML1* breakpoints in children (1/27, 4%, see Results). It should be noted that others have observed a higher rate of *N*-nucleotides in *TEL-AML1*, being present in three of six breakpoints, or 50%.[21] The presence of non-template nucleotides is suggestive of the activity of the enzyme terminal deoxynucleotidyl transferase (*TdT*) which is normally a lymphocyte-specific enzyme, and would not be expected for a myeloid-origin leukemia such as t(8;21)$^+$ AML. Non-template nucleotides can arise from other sources, such as fragments of deletions as has been suggested for *MLL-AF4*,[56] or templated short patch DNA synthesis as suggested for *IGH-BCL2*.[2] Given the very short nature of the non-template nucleotides present here (1–4 nucleotides) we are currently unable to trace the origin of these nucleotides.

While the human genome project is essentially finished,

techniques such as inverse PCR or panhandle PCR might be thought to be supplanted by conventional long-distance PCR for genomic applications such as breakpoint sequencing. However, in our experience, inverse PCR is highly useful for samples in which DNA is limiting, or cases in which genomic breakpoint cluster regions are of vastly different sizes on partner chromosomes. In the latter case, inverse or panhandle PCR techniques can use the smaller breakpoint cluster region as the anchor, the larger region as the target. We continue to refine the molecular biologic and statistical methodology to characterize these breakpoint regions in the leukemias, and will incorporate these techniques into traditional large-scale epidemiology studies of leukemia to better understand the etiology and natural history of leukemia.

Acknowledgements

We thank the UK Childhood Cancer Study Group and individual clinicians in the UK and California for access to patient material. We thank also Susan Colman, Dr Xiaomei Ma, Yunxia Wang, and Dr Luoping Zhang for initial processing and abstracting of samples. We also thank Professors N Kamada and S Matozaki for providing cell lines. These studies were supported by a Leukaemia Research Fund Specialist Programme grant (ZX, MFG), and grants from the National Institutes of Health (PB (ES09137), MTS (ES04705), MRS (AI40906, AI39932), JLW (CA89032)).

References

1 Cayuela JM, Gardie B, Sigaux F. Disruption of the multiple tumor suppressor gene MTS1/p16(INK4a)/CDKN2 by illegitimate V(D)J recombinase activity in T-cell acute lymphoblastic leukemias. *Blood* 1997; **90**: 3720–3726.
2 Jager U, Bocskor S, Le T, Mitterbauer G, Bolz I, Chott A, Kneba M, Mannhalter C, Nadel B. Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. *Blood* 2000; **95**: 3520–3529.
3 Haluska FG, Finver S, Tsujimoto Y, Croce CM. The t(8; 14) chromosomal translocation occurring in B-cell malignancies results from mistakes in V-D-J joining. *Nature* 1986; **324**: 158–161.
4 Tsujimoto Y, Gorham J, Cossman J, Jaffe E, Croce CM. The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. *Science* 1985; **229**: 1390–1393.
5 Aplan PD, Lombardi DP, Ginsberg AM, Cossman J, Bertness VL, Kirsch IR. Disruption of the human SCL locus by 'illegitimate' V-(D)-J recombinase activity. *Science* 1990; **250**: 1426–1429.
6 Pedersen-Bjergaard J, Pedersen M, Roulston D, Philip P. Different genetic pathways in leukemogenesis for patients presenting with therapy-related myelodysplasia and therapy-related acute myeloid leukemia. *Blood* 1995; **86**: 3542–3552.
7 Ahuja HG, Felix CA, Aplan PD. The t(11;20)(p15;q11) chromosomal translocation associated with therapy-related myelodysplastic syndrome results in an NUP98-TOP1 fusion. *Blood* 1999; **94**: 3258–3261.
8 Quesnel B, Kantarjian H, Bjergaard JP, Brault P, Estey E, Lai JL, Tilly H, Stoppa AM, Archimbaud E, Harousseau JL, Bauters F, Fenaux P. Therapy-related acute myeloid leukemia with t(8;21), inv(16), and t(8;16): a report on 25 cases and review of the literature. *J Clin Oncol* 1993; **11**: 2370–2379.
9 Roulston D, Espinosa R 3rd, Nucifora G, Larson RA, Le Beau MM, Rowley JD. CBFA2(AML1) translocations with novel partner chromosomes in myeloid leukemias: association with prior therapy. *Blood* 1998; **92**: 2879–2885.
10 Ahuja HG, Felix CA, Aplan PD. Potential role for DNA topoisomerase II poisons in the generation of t(11;20)(p15;q11) translocations. *Genes Chromosomes Cancer* 2000; **29**: 96–105.
11 Felix CA, Lange BJ, Hosler MR, Fertala J, Bjornsti MA. Chromosome band 11q23 translocation breakpoints are DNA topoisomerase II cleavage sites. *Cancer Res* 1995; **55**: 4287–4292.
12 Broeker PL, Super HG, Thirman MJ, Pomykala H, Yonebayashi Y, Tanabe S, Zeleznik-Le N, Rowley JD. Distribution of 11q23 breakpoints within the MLL breakpoint cluster region in *de novo* acute leukemia and in treatment-related acute myeloid leukemia: correlation with scaffold attachment regions and topoisomerase II consensus binding sites. *Blood* 1996; **87**: 1912–1922.
13 Reichel M, Gillert E, Breitenlohner I, Repp R, Greil J, Beck JD, Fey GH, Marschalek R. Rapid isolation of chromosomal breakpoints from patients with t(4;11) acute lymphoblastic leukemia: implications for basic and clinical research. *Cancer Res* 1999; **59**: 3357–3362.
14 Wiemels JL, Alexander FE, Cazzaniga G, Biondi A, Mayer SP, Greaves M. Microclustering of TEL-AML1 breakpoints in childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 2000; **29**: 219–228.
15 van der Reijden BA, Dauwerse HG, Giles RH, Jagmohan-Changur S, Wijmenga C, Liu PP, Smit B, Wessels HW, Beverstock GC, Jotterand-Bellomo M, Martinet D, Muhlematter D, Lafage-Pochitaloff M, Gabert J, Reiffers J, Bilhou-Nabera C, van Ommen GJ, Hagemeijer A, Breuning MH. Genomic acute myeloid leukemia-associated inv(16)(p13q22) breakpoints are tightly clustered. *Oncogene* 1999; **18**: 543–550.
16 Yoshida H, Naoe T, Fukutani H, Kiyoi H, Kubo K, Ohno R. Analysis of the joining sequences of the t(15;17) translocation in human acute promyelocytic leukemia: sequence non-specific recombination between the PML and RARA genes within identical short stretches. *Genes Chromosomes Cancer* 1995; **12**: 37–44.
17 Atlas M, Head D, Behm F, Schmidt E, Zeleznik-Le NH, Roe BA, Burian D, Domer PH. Cloning and sequence analysis of four t(9;11) therapy-related leukemia breakpoints. *Leukemia* 1998; **12**: 1895–1902.
18 Ross JA, Potter JD, Reaman GH, Pendergrass TW, Robison LL. Maternal exposure to potential inhibitors of DNA topoisomerase II and infant leukemia (United States): a report from the Children's Cancer Group. *Cancer Causes Control* 1996; **7**: 581–590.
19 Ross JA, Potter JD, Robison LL. Infant leukemia, topoisomerase II inhibitors, and the MLL gene. *J Natl Cancer Inst* 1994; **86**: 1678–1680.
20 Alexander FE, Patheal SL, Biondi A, Brandalise S, Cabrera ME, Chan LC, Chen Z, Cimino G, Cordoba JC, Gu LJ, Hussein H, Ishii E, Kamel AM, Labra S, Magalhaes IQ, Mizutani S, Petridou E, de Oliveira MP, Yuen P, Wiemels JL, Greaves MF. Transplacental chemical exposure and risk of infant leukemia with MLL gene fusion. *Cancer Res* 2001; **61**: 2542–2546.
21 Thandla SP, Ploski JE, Raza-Egilmez SZ, Chhalliyil PP, Block AW, de Jong PJ, Aplan PD. ETV6-AML1 translocation breakpoints cluster near a purine/pyrimidine repeat region in the ETV6 gene. *Blood* 1999; **93**: 293–299.
22 Wiemels JL, Greaves M. Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res* 1999; **59**: 4075–4082.
23 Romana S, Poirel H, Della Valle V, Mauchauffe M, Busson-Le Coniat M, Berger R, Bernard OA. Molecular analysis of chromosomal breakpoints in three examples of chromosomal translocation involving the TEL gene. *Leukemia* 1999; **13**: 1754–1759.
24 Andersen MT, Nordentoft I, Hjalgrim LL, Christiansen CL, Jakobsen VD, Hjalgrim H, Pallisgaard N, Madsen HO, Christiansen M, Ryder LP, Clausen N, Hokland P, Schmiegelow K, Melbye M, Jorgensen P. Characterization of t(12;21) breakpoint junctions in acute lymphoblastic leukemia. *Leukemia* 2001; **15**: 858–859.
25 Larson RA, Le Beau MM, Ratain MJ, Rowley JD. Balanced translocations involving chromosome bands 11q23 and 21q22 in therapy-related leukemia (letter). *Blood* 1992; **79**: 1892–1893.
26 Pedersen-Bjergaard J, Philip P. Balanced translocations involving chromosome bands 11q23 and 21q22 are highly characteristic of myelodysplasia and leukemia following therapy with cytostatic agents targeting at DNA-topoisomerase II (letter). *Blood* 1991; **78**: 1147–1148.
27 Pedersen-Bjergaard J, Daugaard G, Hansen SW, Philip P, Larsen SO, Rorth M. Increased risk of myelodysplasia and leukaemia after

etoposide, cisplatin, and bleomycin for germ-cell tumours. *Lancet* 1991; **338**: 359–363.
28 Pedersen-Bjergaard J, Sigsgaard TC, Nielsen D, Gjedde SB, Philip P, Hansen M, Larsen SO, Rorth M, Mouridsen H, Dombernowsky P. Acute monocytic or myelomonocytic leukemia with balanced chromosome translocations to band 11q23 after therapy with 4-epi-doxorubicin and cisplatin or cyclophosphamide for breast cancer. *J Clin Oncol* 1992; **10**: 1444–1451.
29 Sandoval C, Pui CH, Bowman LC, Heaton D, Hurwitz CA, Raimondi SC, Behm FG, Head DR. Secondary acute myeloid leukemia in children previously treated with alkylating agents, intercalating topoisomerase II inhibitors, and irradiation. *J Clin Oncol* 1993; **11**: 1039–1045.
30 Cuneo A, Fagioli F, Pazzi I, Tallarico A, Previati R, Piva N, Carli MG, Balboni M, Castoldi G. Morphologic, immunologic and cytogenetic studies in acute myeloid leukemia following occupational exposure to pesticides and organic solvents. *Leuk Res* 1992; **16**: 789–796.
31 Golomb HM, Alimena G, Rowley JD, Vardiman JW, Testa JR, Sovik C. Correlation of occupation and karyotype in adults with acute nonlymphocytic leukemia. *Blood* 1982; **60**: 404–411.
32 Li Y-S, Zhao Y-L, Jiang Q-P, Yang C-L. Specific chromosome changes and non-occupational exposure to potentially carcinogenic agents in acute leukemia in China. *Leuk Res* 1989; **13**: 367–376.
33 Mitelman F, Nilsson PG, Brandt L, Alimena G, Gastaldi R, Dallapiccola B. Chromosome pattern, occupation, and clinical features in patients with acute nonlymphocytic leukemia. *Cancer Genet Cytogenet* 1981; **4**: 197–214.
34 Nucifora G, Begy CR, Kobayashi H, Roulston D, Claxton D, Pedersen-Bjergaard J, Parganas E, Ihle JN, Rowley JD. Consistent intergenic splicing and production of multiple transcripts between AML1 at 21q22 and unrelated genes at 3q26 in (3;21)(q26;q22) translocations. *Proc Natl Acad Sci USA* 1994; **91**: 4004–4008.
35 Gamou T, Kitamura E, Hosoda F, Shimizu K, Shinohara K, Hayashi Y, Nagase T, Yokoyama Y, Ohki M. The partner gene of AML1 in t(16;21) myeloid malignancies is a novel member of the MTG8(ETO) family. *Blood* 1998; **91**: 4028–4037.
36 Downing JR. The AML1-ETO chimaeric transcription factor in acute myeloid leukaemia: biology and clinical significance. *Br J Haematol* 1999; **106**: 296–308.
37 Tighe JE, Calabi F. t(8;21) breakpoints are clustered between alternatively spliced exons of MTG8. *Clin Sci (Colch)* 1995; **89**: 215–218.
38 Tighe JE, Calabi F. Alternative, out-of-frame runt/MTG8 transcripts are encoded by the derivative (8) chromosome in the t(8;21) of acute myeloid leukemia M2. *Blood* 1994; **84**: 2115–2121.
39 Erickson P, Gao J, Chang KS, Look T, Whisenant E, Raimondi S, Lasher R, Trujillo J, Rowley J, Drabkin H. Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood* 1992; **80**: 1825–1831.
40 Asou H, Tashiro S, Hamamoto K, Otsuji A, Kita K, Kamada N. Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood* 1991; **77**: 2031–2036.
41 Matozaki S, Nakagawa T, Kawaguchi R, Aozaki R, Tsutsumi M, Murayama T, Koizumi T, Nishimura R, Isobe T, Chihara K. Establishment of a myeloid leukaemic cell line (SKNO-1) from a patient with t(8;21) who acquired monosomy 17 during disease progression. *Br J Haematol* 1995; **89**: 805–811.
42 Shimizu K, Miyoshi H, Kozu T, Nagata J, Enomoto K, Maseki N, Kaneko Y, Ohki M. Consistent disruption of the AML1 gene occurs within a single intron in the t(8;21) chromosomal translocation. *Cancer Res* 1992; **52**: 6945–6948.
43 Loader CR. Large deviation approximations to the distribution of scan statistics. *Adv Appl Prob* 1991; **23**: 751–771.
44 Huffer F, Lin C-T. Approximating the distribution of the scan statistic using moments of the number of clumps. *J Amer Stat Assoc* 1997; **92**: 1466–1475.
45 Segal MR, Wiemels JL. Clustering of translocation breakpoints. *J Amer Stat Assoc* 2001 (in press).
46 Silverman BW. Using kernel density estimates to investigate multimodality. *J R Stat Soc B* 1981; **43**: 97–99.
47 Sarper N, Ozbek U, Agaoglu L, Ozgen U, Eryilmaz E, Yalman N, Anak S, Devecioglu O, Gedikoglu G. Is AML1/ETO gene expression a good prognostic factor in pediatric acute myeloblastic leukemia? *Pediatr Hematol Oncol* 2000; **17**: 577–583.
48 Kalwinsky DK, Raimondi SC, Schell MJ, Mirro J Jr, Santana VM, Behm F, Dahl GV, Williams D. Prognostic importance of cytogenetic subgroups in *de novo* pediatric acute nonlymphocytic leukemia. *J Clin Oncol* 1990; **8**: 75–83.
49 Martinez-Climent JA, Lane NJ, Rubin CM, Morgan E, Johnstone HS, Mick R, Murphy SB, Vardiman JW, Larson RA, Le Beau MM, Rowley JD. Clinical and prognostic significance of chromosomal abnormalities in childhood acute myeloid leukemia *de novo*. *Leukemia* 1995; **9**: 95–101.
50 Shimada H, Ichikawa H, Nakamura S, Katsu R, Iwasa M, Kitabayashi I, Ohki M. Analysis of genes under the downstream control of the t(8;21) fusion protein AML1-MTG8: overexpression of the TIS11b (ERF-1, cMG1) gene induces myeloid cell proliferation in response to G-CSF. *Blood* 2000; **96**: 655–663.
51 Meyers S, Lenny N, Hiebert SW. The t(8;21) fusion protein interferes with AML-1B-dependent transcriptional activation. *Mol Cell Biol* 1995; **15**: 1974–1982.
52 van de Locht LT, Smetsers TF, Wittebol S, Raymakers RA, Mensink EJ. Molecular diversity in AML1/ETO fusion transcripts in patients with t(8;21) positive acute myeloid leukaemia. *Leukemia* 1994; **8**: 1780–1784.
53 Saunders MJ, Tobal K, Keeney S, Liu Yin JA. Expression of diverse AML1/MTG8 transcripts is a consistent feature in acute myeloid leukemia with t(8;21) irrespective of disease phase. *Leukemia* 1996; **10**: 1139–1142.
54 Xiao Z, Hao Y, Bian S. Acute myeloid leukemia M2b (subacute myelogenous leukemia) in China (letter). *Leuk Res* 1997; **21**: 351–352.
55 Gillert E, Leis T, Repp R, Reichel M, Hosch A, Breitenlohner I, Angermuller S, Borkhardt A, Harbott J, Lampert F, Griesinger F, Greil J, Fey GH, Marschalek R. A DNA damage repair mechanism is involved in the origin of chromosomal translocations t(4;11) in primary leukemic cells. *Oncogene* 1999; **18**: 4663–4671.
56 Reichel M, Gillert E, Nilson I, Siegler G, Greil J, Fey GH, Marschalek R. Fine structure of translocation breakpoints in leukemic blasts with chromosomal translocation t(4;11): the DNA damage-repair model of translocation. *Oncogene* 1998; **17**: 3035–3044.
57 Reichel M, Gillert E, Angermuller S, Hensel JP, Heidel F, Lode M, Leis T, Biondi A, Haas OA, Strehl S, Panzer-Grumayer ER, Griesinger F, Beck JD, Greil J, Fey GH, Uckun FM, Marschalek R. Biased distribution of chromosomal breakpoints involving the MLL gene in infants versus children and adults with t(4;11) ALL. *Oncogene* 2001; **20**: 2900–2907.
58 Ford AM, Bennett CA, Price CM, Bruin MC, Van Wering ER, Greaves M. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc Natl Acad Sci USA* 1998; **95**: 4584–4588.
59 Wiemels JL, Ford AM, Van Wering ER, Postma A, Greaves M. Protracted and variable latency of acute lymphoblastic leukemia after TEL-AML1 gene fusion *in utero*. *Blood* 1999; **94**: 1057–1062.
60 Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G, Saha V, Biondi A, Greaves MF. Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* 1999; **354**: 1499–1503.
61 Smith MT, Zhang L, Wang Y, Hayes RB, Li G, Wiemels J, Dosemeci M, Titenko-Holland N, Xi L, Kolachana P, Yin S, Rothman N. Increased translocations and aneusomy in chromosomes 8 and 21 among workers exposed to benzene. *Cancer Res* 1998; **58**: 2176–2181.
62 Stanulla M, Wang J, Chervinsky DS, Aplan PD. Topoisomerase II inhibitors induce DNA double-strand breaks at a specific site within the AML1 locus. *Leukemia* 1997; **11**: 490–496.