Leukemia Research 33 (2009) 1082–1088

Contents lists available at ScienceDirect

Leukemia Research

journal homepage: www.elsevier.com/locate/leukres

# Cluster analysis of genomic *ETV6–RUNX1* (*TEL–AML1*) fusion sites in childhood acute lymphoblastic leukemia

H. von Goessel[a], U. Jacobs[a], S. Semper[a], M. Krumbholz[a], T. Langer[a], T. Keller[b], A. Schrauder[c], V.H.J. van der Velden[d], J.J.M. van Dongen[d], J. Harbott[e], E.R. Panzer-Grümayer[f], M. Schrappe[c], W. Rascher[a], M. Metzler[a,*]

[a] *Department of Pediatrics, University of Erlangen, Loschgestrasse 15, 91054 Erlangen, Germany*
[b] *ACOMED Statistik, Statistical Analyses, Leipzig, Germany*
[c] *Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany*
[d] *Department of Immunology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands*
[e] *Department of Pediatric Hematology and Oncology, University of Giessen, Giessen, Germany*
[f] *Children's Cancer Research Institute and St. Anna Kinderspital, Vienna, Austria*

ARTICLE INFO

*Article history:*
Received 3 September 2008
Received in revised form 18 October 2008
Accepted 2 November 2008
Available online 10 December 2008

*Keywords:*
Childhood ALL
*ETV6–RUNX1*
*TEL–AML1*
Chromosomal translocations
Genomic fusion sites
Cluster analysis

ABSTRACT

Fusion between *ETV6* and *RUNX1* defines the largest genetic subgroup in childhood ALL. The genomic fusion site, unique to individual patients and specific for the malignant clone, represents an ideal molecular marker for quantification of minimal residual disease. Sequencing of DNA breakpoints has been difficult due to the extended size of the respective breakpoint cluster regions. We therefore evaluated a specially designed multiplex long-range PCR assay in 65 diagnostic bone marrow samples for its suitability in routine use. Resulting fusion sites and breakpoints derived from previous studies were subject to cluster analysis to identify potential sequence motifs involved in translocation initiation.



## 1. Introduction

In childhood acute lymphoblastic leukemia (ALL), fusion between *ETV6* (*TEL*) and *RUNX1* (*AML1*) as a result of t(12;21)(p13;q22) occurs in 20–25% of cases and represents the largest genetically defined subgroup [1]. Genomic fusion sites are restricted to specific breakpoint cluster regions (BCR) in both genes. In *ETV6* chromosomal breakpoints occur almost exclusively within the 15 kb intron 5; the *RUNX1* BCR includes the large 155 kb intron 1 and the 5.5 kb intron 2. The resultant fusion gene codes for chimeric proteins comprising the amino-terminal portion of ETV6, including its HLH domain, and nearly complete RUNX1 with both its transactivation domains and the DNA- and protein-binding Runt homology domain [2]. The fusion protein is suggested to function as transcriptional repressor, presumably by inhibition of normal RUNX1 target genes [3].

High concordance rates in identical twins and *ETV6–RUNX1* rearrangements detectable in neonatal blood spots of individuals affected by *ETV6–RUNX1* positive ALL later in childhood supported the hypothesis that the translocations occur already in utero [4–7]. In contrast to other subgroups of leukemia such as *MLL* associated translocations, the *ETV6–RUNX1* rearrangement seems to establish pre-leukemic clones with extended latency and lower frequency for development of leukemic disease [8,9]. These aspects and the detection of *ETV6–RUNX1* fusion transcripts in normal cord blood [10] have been interpreted as indication that additional secondary and postnatal genetic events are necessary to transform a preleukemic clone harbouring the *ETV6–RUNX1* fusion to overt leukemia.

Although the generation of *ETV6–RUNX1* rearrangements in hematopoietic progenitor cells seems to be relatively common events, the underlying mechanism causing these gene rearrangements remains largely elusive. The fingerprint of mechanisms potentially involved in *ETV6–RUNX1* translocation may be recognizable by localization and sequence features of the genomic fusion sites. Clustering of breakpoints at defined "hot spots" has led to specific DNA motifs critical to initiation of chromosomal translocation exemplified by therapy-related leukemia associated with *MLL* rearrangements [11–13] or *E2A-PBX1* (*TCF3-PBX1*) translocation positive B-cell progenitor ALL [14].

* Corresponding author. Tel.: +49 9131 85 33118; fax: +49 9131 85 33113.
*E-mail address:* Markus.Metzler@uk-erlangen.de (M. Metzler).

0145-2126/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.leukres.2008.11.001

Whereas the *ETV6–RUNX1* fusion transcript detection by RT-PCR is well established for routine application in pediatric ALL to aid subgroup classification and minimal residual disease (MRD) assessment [15], only a limited number of genomic *ETV6–RUNX1* breakpoints has yet been sequenced and characterized. However, the *ETV6–RUNX1* DNA fusion sequence, unique to each individual patient's leukemic cell clone, has proven a valuable tool for MRD quantification and analysis of clonal evolution in relapsed *ETV6–RUNX1* leukemia [16,17].

In the present study, we applied an optimized multiplex long-range PCR (MLR-PCR) protocol for genomic fusion site sequencing in a series of 65 pediatric *ETV6–RUNX1* leukemia patients to test its suitability for routine application and performed scan statistics on the breakpoint distributions in combination with previously published fusion sites to evaluate sequence motifs and regions associated with chromosomal rearrangements for their relevance in *ETV6–RUNX1* translocation initiation.

## 2. Materials and methods

### 2.1. Patients and samples

DNA was extracted from leukemic blasts of 65 pediatric ALL patients enrolled to the Austrian and German ALL-BFM 2000 trial or Dutch DCOG-ALL9 trial and prescreened for *ETV6–RUNX1* translocation by RT-PCR or FISH. All patients gave written informed consent, in accordance with the Declaration of Helsinki.

Previously published *ETV6–RUNX1* fusion sequences for comparison with present data were derived from Ford et al. ($n = 1$) [5], Romana et al. ($n = 2$) [18], Thandla et al. ($n = 4$) [19], Wiemels and Greaves ($n = 9$) [20], Wiemels et al. ($n = 11$) [21], Andersen et al. ($n = 9$) [22], Maia et al. ($n = 1$) [7], and McHale et al. ($n = 14$) [23].

### 2.2. Genomic ETV6–RUNX1 fusion site analysis

Breakpoint-spanning DNA fragments were amplified by two-round multiplex long-range PCR with minor modifications to previously published protocols [16,17]. Four separate nested primer sets, each with 1 *ETV6* sense (393 or 7904), and 8 (intron 1) or 10 (intron 1 and 2) *RUNX1* antisense primers (set A or B) were used with the Expand Long Template PCR System (Roche) according to the manufacturer's instructions (Fig. 1A). Positions and nucleotide sequences of all primers are shown in Supplementary Table 1. Four separate PCR reactions were chosen because combination of all primers in one PCR reaction or additional *ETV6* sense primers led to a significantly higher rate of unspecific products, despite extensive primer and PCR parameter optimization. The first run products of the positive PCR reactions (e.g. lanes 1 and 3 in Fig. 1B) were used as template for another second round long-range PCR, but with separate *RUNX* antisense primers per reaction (Fig. 1C). The *ETV6* sense primer closer to the genomic fusion site is defined by the presence of the shorter product resulting from initial MLR-PCR. To make the *ETV6–RUNX1* fusion site accessible to direct sequencing from both directions, the previous product (Fig. 1C, lane7) is shortened by running a third second round PCR using the first round MLR-PCR product as template and additional *ETV6* sense primers (Fig. 1D). The shortest product is finally directly sequenced using the respective PCR primers. If the *ETV6–RUNX1* primer set failed to generate a specific PRC product, the reciprocal *RUNX1–ETV6* fusion site was attempted to amplify using accordingly orientated primers listed in supplementary Table 1. Breakpoint junction sequences obtained by MLR-PCR were confirmed by a patient-specific breakpoint PCR using a fresh aliquot of the original genomic DNA.

Alignments of patient-specific fusion sequences were performed using the BLAST2 tool (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). The *ETV6* BCR reference sequence for further analyses was downloaded from the ENSEMBL database (release 38) starting from the first nucleotide of exon 5 to the last nucleotide of exon 6 (15,164 bp). The *RUNX1* BCR reference was defined as the sequence from the first nucleotide of exon 1 to the last nucleotide of exon 3 (162,502 bp). Breakpoint distribution and density within the BCR was visualized using GraphPad Prism software.

Sequence motifs known to be associated with chromosomal translocations [24] were searched with the use of VectorNTI software. Repetitive elements were identified by the RepeatMasker program (http://www.repeatmasker.org).

### 2.3. Statistical analysis

Breakpoint density and cluster analysis by scan statistics were performed in analogy to Segal and Wiemels [25] using components of the free software environment R, Version 2.6.1 (www.r-project.org) and SaTScan software v6.1.2 [26], respectively.

Within Kernel density analysis, bandwidth selection was performed according to Sheather and Jones [27]. Clusters were identified in regions where lower limit of 95% confidence band determined by bootstrapping procedure are higher than a density function resulting from simulations at randomly distributed pseudo-breakpoints. Both bootstrapping and simulations used 10,000 permutations.

## 3. Results

### 3.1. Sequencing and alignment of genomic ETV6–RUNX1 fusion sites

Multiplex long-range PCR amplified the patient-specific *ETV6–RUNX1* fusion site in 57 of 65 tested bone marrow DNA from t(12;21) positive pediatric ALL patients. In three cases, only the reciprocal *RUNX1–ETV6* was amplifiable by the MLR-PCR assay. Alignments of the fusion sequence to *ETV6* and *RUNX1* germline sequences are shown in Supplementary Fig. 1. Twenty-six of 60 breakpoints (43%) had microhomologies at the fusion site, 20 (33%) showed clean transitions between the two contributing genes, and filler DNA was present in 14 cases (23%). These features are suggestive for activity of NHEJ repair mechanisms involved in *ETV6–RUNX1* fusion. In patient UPN 480, the *ETV6* breakpoint aligned to intron 6, outside the common BCR, and is therefore not represented in Figs. 2 and 3. In *RUNX1*, 56 of 60 breakpoints (93%) were located in intron 1; the remaining 4 breaks (7%) occurred in the intron 2 portion of the *RUNX1* BCR. DNA quality, tested by germline *ETV6* amplification, was insufficient for long-range PCR amplification in two of the five MLR-PCR negative samples. Attempts to analyze the remaining three cases by long-distance inverse PCR (LDI-PCR) also failed to amplify a breakpoint spanning product.

To compare the breakpoint localization of the current study with previously published genomic *ETV6–RUNX1* fusion sites and to combine all data for cluster analysis, a total of 51 additional sequences were derived from 8 studies with smaller sample numbers [5,7,18–20,22,23,28]. In two cases, the given sequence sections were not sufficient to identify a non-ambiguous *RUNX1* breakpoint. Therefore, only 49 were included in Kernel density and scan statistic analysis. Breakpoint distribution and density curves for the combined collection are shown in Fig. 2A.

### 3.2. Statistical analysis for breakpoint subclusters

Genomic fusion sites of 60 *ETV6–RUNX1* cases amplified by MLR-PCR showed no significant clustering neither within *ETV6* nor *RUNX1* BCR. The same applied for the 51 breakpoints pooled from 8 previous reports. A comparison between MLR-PCR and LDI-PCR derived fusion sites showed no evidence for distortion of breakpoint distribution depending on the PCR protocol applied for fusion site amplification. However, *RUNX1* breakpoint density in cohorts, both current study and pool of previous studies, showed some increase towards the BCR 3′ end (Fig. 2B).

In combination with previously reported *ETV6–RUNX1* fusion sites, a data set of 110 *ETV6* and 109 *RUNX1* genomic breakpoints was analyzed. Kernel density analysis with bandwidth 12,574 for *RUNX1* selected according to Sheather and Jones [27] detected an increased density in a region telomeric of position 104,000; significance could be shown only for a small region (Fig. 2A). This result is in concordance with findings obtained by scan statistics. A cluster centered at position 131,500 with a 30,400 bp radius (region 101,100–161,900 bp) was detected ($p = 0.006$). This region includes 65 breakpoints opposed to 41 breakpoints expected for homogeneous distribution.

For ETV6 breakpoint distribution, Kernel density analysis using a bandwidth of 1673 detected no overlap of confidence intervals, and scan statistics did not identify clusters within the *ETV6* BCR.