

COBL is a novel hotspot for IKZF1 deletions in childhood acute lymphoblastic leukemia

Bruno Almeida Lopes¹, Claus Meyer², Thayana Conceição Barbosa¹, Udo zur Stadt³, Martin Horstmann^{3,4,5}, Nicola C. Venn⁶, Susan Heatley^{7,8}, Deborah L. White^{7,8}, Rosemary Sutton⁶, Maria S. Pombo-de-Oliveira¹, Rolf Marschalek², Mariana Emerenciano¹

¹Pediatric Hematology-Oncology Program, Research Center, Instituto Nacional de Câncer, Rio de Janeiro, RJ, Brazil

²Diagnostic Center of Acute Leukemia/Institute of Pharmaceutical Biology/ZAFES, Goethe-University of Frankfurt, Biocenter, Germany

³Center for Diagnostics, University Medical Center Hamburg Eppendorf, Hamburg, Germany

⁴Research Institute Children's Cancer Center, Hamburg, Germany

⁵Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

⁶Children's Cancer Institute, Lowy Cancer Research Centre UNSW, Sydney, New South Wales, Australia

⁷South Australian Health and Medical Research Institute (SAHMRI), Adelaide, South Australia, Australia

⁸Discipline of Medicine, University of Adelaide, Adelaide, South Australia, Australia

Correspondence to: Mariana Emerenciano, **email:** memerenciano@inca.gov.br

Keywords: acute lymphoblastic leukemia, COBL, IKZF1, RAG, relapse

Received: May 03, 2016

Accepted: June 30, 2016

Published: July 13, 2016

ABSTRACT

IKZF1 deletion (Δ IKZF1) is an important predictor of relapse in childhood B-cell precursor acute lymphoblastic leukemia. Because of its clinical importance, we previously mapped breakpoints of intragenic deletions and developed a multiplex PCR assay to detect recurrent intragenic Δ IKZF1. Since the multiplex PCR was not able to detect complete deletions (*IKZF1* Δ 1-8), which account for ~30% of all Δ IKZF1, we aimed at investigating the genomic scenery of *IKZF1* Δ 1-8. Six samples of cases with *IKZF1* Δ 1-8 were analyzed by microarray assay, which identified monosomy 7, isochromosome 7q, and large interstitial deletions presenting breakpoints within *COBL* gene. Then, we established a multiplex ligation-probe amplification (MLPA) assay and screened copy number alterations within chromosome 7 in 43 diagnostic samples with *IKZF1* Δ 1-8. Our results revealed that monosomy and large interstitial deletions within chromosome 7 are the main causes of *IKZF1* Δ 1-8. Detailed analysis using long distance inverse PCR showed that six patients (16%) had large interstitial deletions starting within intronic regions of *COBL* at diagnosis, which is ~611 Kb downstream of *IKZF1*, suggesting that *COBL* is a hotspot for Δ IKZF1. We also investigated a series of 25 intragenic deletions (Δ 2-8, Δ 3-8 or Δ 4-8) and 24 relapsed samples, and found one *IKZF1*-*COBL* tail-to-tail fusion, thus supporting that *COBL* is a novel hotspot for Δ IKZF1. Finally, using RIC score methodology, we show that breakpoint sequences of *IKZF1* Δ 1-8 are not analog to RAG-recognition sites, suggesting a different mechanism of error promotion than that suggested for intragenic Δ IKZF1.

INTRODUCTION

Several genetic mutations are usually necessary for the onset of cancer [1]. In case of childhood acute lymphoblastic leukemia (ALL) only few mutations are required [2]. In most cases, products deriving from

chromosomal translocations are the oncogenic initiating or driving lesions and have been characterized at the molecular and functional level during the last decades. In addition to the widely known gene fusions, genomic technologies - SNP arrays or whole genome sequencing - also allowed the identification of complementing genetic

alterations that are important contributors of hematological malignancies. One of the identified alterations concerns the *IKZF1* gene, located at 7p12, encoding the transcription factor Ikaros, that is essentially involved in the development of the B-cell lineage [3].

The biology of *IKZF1* is complex because this gene consists of 8 exons, and encodes 11 different splice variants [4]. Five of these isoforms are translated into proteins that are acting as transcriptional activators (1, 2, 2a, 3, 3a), while six isoforms result in dominant-negative versions of Ikaros (4, 4a, 5, 6, 7, 8). The Ikaros protein composes a regulatory complex involving hematopoietic transcription factors (E2A, EBF1-3, and PAX5) with the distinct function of driving lymphoid development [5], which is impaired by “dominant-negative” isoforms.

In 2007, the first study demonstrating the importance of *IKZF1* deletions ($\Delta IKZF1$) in leukemia was published [2]. Genome wide SNP array analyses of 242 patients revealed that several genes involved in B-lineage development are mutated in B-cell precursor ALL (BCP-ALL) patients. These initial findings were confirmed by a contemporary study that investigated 40 leukemia patients [6], which identified again recurrent submicroscopic deletions in several genes linked to B-lineage development (*PAX5*, *EBF1*, *TCF3*, *IKZF1* and others). Thereafter, clinical studies revealed that leukemia patients with $\Delta IKZF1$ should be classified as high-risk, as the presence of a deletion was shown to be an independent prognostic factor for event-free survival in pediatric BCP-ALL [7]. Of note, a subsequent international study showed that all types of $\Delta IKZF1$ are associated with unfavorable prognosis [8]. The clinical effects of $\Delta IKZF1$ and the results from basic research have been thoroughly reviewed [9].

Due to the clinical importance of $\Delta IKZF1$, multiplex ligation-probe amplification (MLPA) assays have been developed to detect such alterations in BCP-ALL. As $\Delta IKZF1$ are predictive of relapse and given the lack of sensitivity of MLPA assays, scientific efforts focused on novel diagnostic tools for minimal residual disease investigation. In 2011, a real-time quantitative PCR assay was developed to detect the most common microdeletion in ALL *IKZF1* Δ 3-6 (now known as *IKZF1* Δ 4-7), which results in the Ik6 transcript [10]. This was followed in 2013 by two studies that developed multiplex PCR assays to detect recurrent intragenic deletions of *IKZF1*, such as Δ 2-3, Δ 2-7, Δ 2-8, Δ 4-7, and Δ 4-8 [11, 12]. Although these methodologies present a higher sensitivity of at least 10^{-2} , they are not able to detect complete deletions (*IKZF1* Δ 1-8), which represent ~30% of all deletions. Therefore, the aim of this study was to unravel the genomic landscape of *IKZF1* Δ 1-8.

RESULTS

Clinical and genetic characteristics of BCP-ALL with *IKZF1* complete deletions

Patients with *IKZF1* Δ 1-8 were mainly male (60.5%), aged 1-9 years old (72.1%) at diagnosis, and only 30.2% had $WBC \geq 50 \times 10^9/L$. The predominant immunophenotype was c-ALL (65.1%).

Copy number alterations (CNAs) in genes frequently altered in BCP-ALL were compared between patients with or without *IKZF1* deletions in order to discover possible concomitant alterations associated with *IKZF1* Δ 1-8. The results showed that *IKZF1* Δ 1-8 deletions extended to its surrounding genes, such as *ZPBP* (deleted in 85.7% of cases), *FIGNL1* (90.5%) and *DDC* (90.5%). On the other hand, deletions of *EBF1* and *BTG1* were respectively absent or rarely (4.1%) identified in patients with *IKZF1* Δ 1-8 (Supplementary Table S1).

Genome-wide CNAs in samples with complete deletion of *IKZF1* at diagnosis

Six random DNA samples of BCP-ALL with *IKZF1* Δ 1-8 were selected for the microarray assay. The details of each sample are presented in Supplementary Table S2. Microarray analysis revealed that samples with *IKZF1* Δ 1-8 presented large deletions within chromosome 7 ($n = 5$) and partial loss of 9p arm ($n = 3$), as shown in Supplementary Figure S1. Three sorts of alterations on chromosome 7 were found for samples with *IKZF1* Δ 1-8: monosomy 7 ($n = 2$), large interstitial deletions ($n = 3$), and isochromosome 7q ($n = 1$) (Figure 1A). Interestingly, two patients (S35 and S36) with large interstitial deletions presented breakpoints within *COBL* intron 5 (Figure 1B) downstream of *IKZF1*. On the other hand, there was no indication of a genetic hotspot upstream of *IKZF1*, as 50% of cases presented loss of chromosome 7 short arm, and the remaining patients had remarkably variable breakpoints at the 7p telomeric side.

Screening CNAs on chromosome 7 with the customized MLPA

Because the microarray data showed that a diverse spectrum of alterations within chromosome 7 promotes *IKZF1* Δ 1-8, we designed two customized MLPAs to identify such CNAs in the whole series of patients included in this study. The MLPA probes were distributed throughout chromosome 7, and most of them were placed within *COBL*. The main characteristics and localization of the MLPA probes are described in Supplementary Table S4, and probe sequences are available upon request.

First, the custom MLPAs were validated with five samples previously analyzed by microarray. The comparison of the results produced by both techniques is illustrated in Supplementary Figure S2. In sum, the MLPA

results were concordant with microarray experiments. As expected, patient S22 (with monosomy 7 by microarray) presented monoallelic deletion for all of the probes tested on chromosome 7, while patients S35 and S36 (with *COBL* intron 5 rearrangements) presented deletion of a series of probes within *COBL*, thus correctly indicating the breakpoint was localizing within intron 5. Although our MLPA did not detect all of the expected CNAs (e.g. patient S24 bears an interstitial deletion within chromosome 7 and patient S34 has isochromosome 7q), the analysis of whole probe set contributed to the correct interpretation of results. Thus, we confirmed that our custom MLPAs were able to detect distinct CNAs within chromosome 7.

Subsequently, the custom MLPAs were used to investigate CNAs for chromosome 7 in 43 BCP-ALL pediatric patients with *IKZF1* Δ 1-8. The MLPA results could not be interpreted for seven samples, which were excluded from further analysis. Five sorts of alterations were identified on chromosome 7: monosomy 7 ($n = 7$), isochromosome 7q ($n = 4$), 7p loss ($n = 4$), large interstitial deletions within the

7p arm ($n = 10$), *IKZF1* Δ 1-8 with breakpoints within *COBL* ($n = 6$), and *IKZF1* complete deletion without involvement of surrounding regions ($n = 5$) (Figure 2).

Breakpoint sequences of *IKZF1* Δ 1-8

We designed four long-distance inverse PCRs (LDI-PCRs) as well as two multiplexed long-distance PCRs (MP-PCRs) assays in order to find the sequence of the breakpoints predicted by both microarray and MLPA results. The sequences of breakpoints for *COBL* rearrangements are summarized in Supplementary Figure S3. Patient S10 presented a large interstitial deletion on the short arm of chromosome 7, comprising the whole *IKZF1* until *COBL* intron 5 and a subsequent \sim 1.1 Mb inversion. Patients S35 and S36 presented interstitial deletions of 18.8 and 1.7 Mb within 7p arm, respectively. The deletions fused *COBL* intron 5 to non-coding regions of the short arm of chromosome 7: 7p14.3-*COBL* (S35) and 7p12-*COBL* (S36).

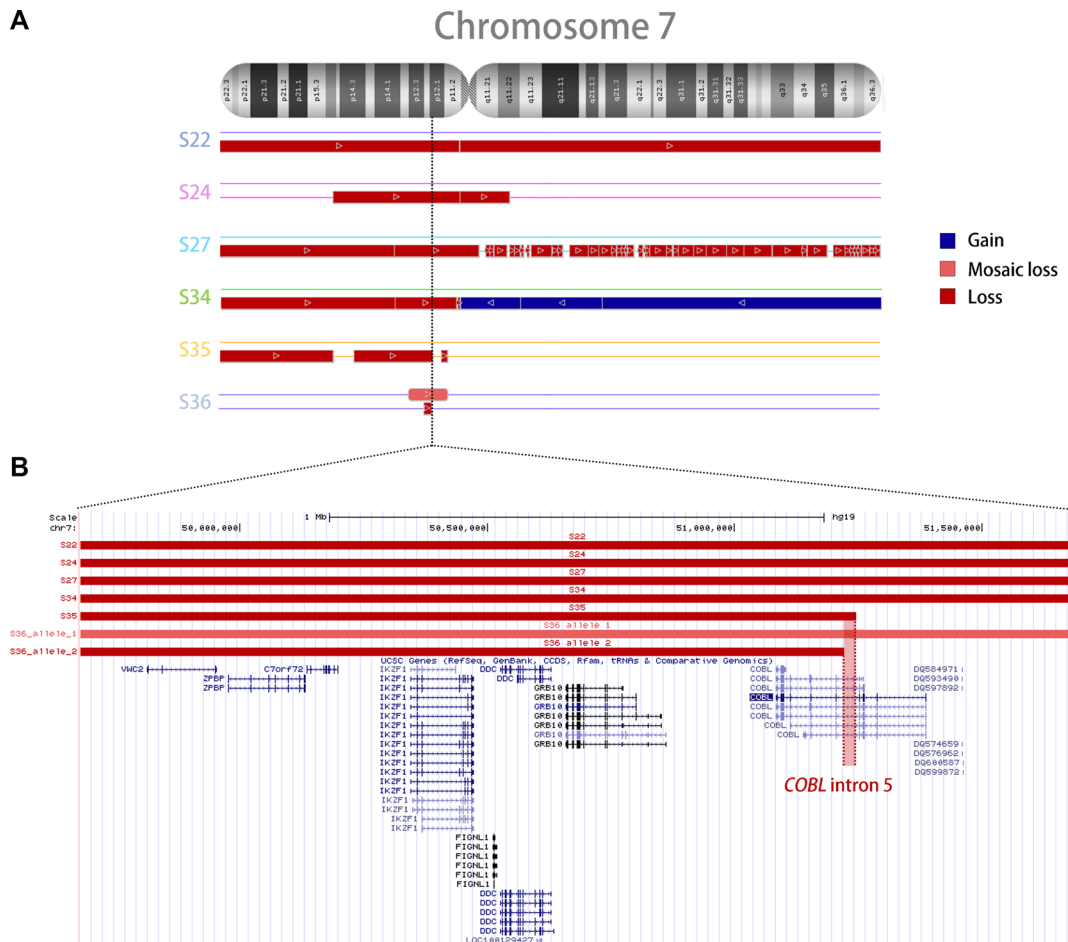


Figure 1: Copy number alterations in the chromosome 7 of six samples (S22, S24, S27, S34, S35, and S36) with complete deletion of *IKZF1*. (A) The figure shows chromosome 7 CNAs, where red and blue lines indicate deletions and amplifications, respectively. (B) UCSC Genome Browser closer view of the region containing *IKZF1* and its surrounding genes, showing complete deletions of *IKZF1* and breakpoint within *COBL* intron 5 in two samples (S35 and S36). Genomic coordinates were standardized to the GRCh37 (hg19) assembly of the human genome.

Characteristics of patients with breakpoints in *COBL*

COBL rearrangements were found in six patients with *IKZF1* Δ1-8 using our customized MLPA and LDI-PCR. We also investigated *IKZF1-COBL* fusions in 25 newly diagnosed ALL samples with *IKZF1* 3'-end deletions (i.e. retention of *IKZF1* exon 1 and a heterozygous deletion from exons 2, 3 or 4 through to exon 8). None of these 25 samples had *IKZF1-COBL* fusions. In addition, we detected a single patient with *IKZF1-COBL* fusion when RNA sequencing was performed on 24 relapsed BCP-ALL samples. Our initial MLPA analysis did not find this deletion, but closer inspection showed ratios of 0.84–0.93 for exons 4–8, consistent with 15–30% of cells in the sample having both *IKZF1-COBL* and *CDNK2A* microdeletions. This patient had an mRNA fusion between *IKZF1* exon 3 and a cryptic exon located in intron 5 of the *COBL* gene, resulting in an arbitrary fusion protein (Supplementary Figure S4). We successfully sequenced the breakpoints of this patient and three others

(S10, S35, S36, and S48) (Supplementary Figure S3). Due to PCR size limitations, the breakpoint region of the remaining three samples with *COBL* rearrangements by MLPA (S03, S05, and S29) could not be confirmed by Sanger sequencing. Patient and laboratory characteristics of cases with *COBL*-rearrangements as well as their clinical data are described in Table 1.

Identification of possible mechanisms underlying the deletions occurrence

After sequencing the breakpoints, we explored which mechanism would be involved in the generation of the interstitial deletions and gene fusions found in this study. First, we investigated RAG1/2 recombination signal sequences (RSSs) nearby the breakpoints. However, this analysis showed that most of the breakpoint sequences had RIC scores for 12RSSs and 23RSSs below the critical threshold (-38.81 and -58.45, respectively), and were not associated with functional RSSs. Only the patient S48 with *IKZF1-COBL* fusion presented significant RSS

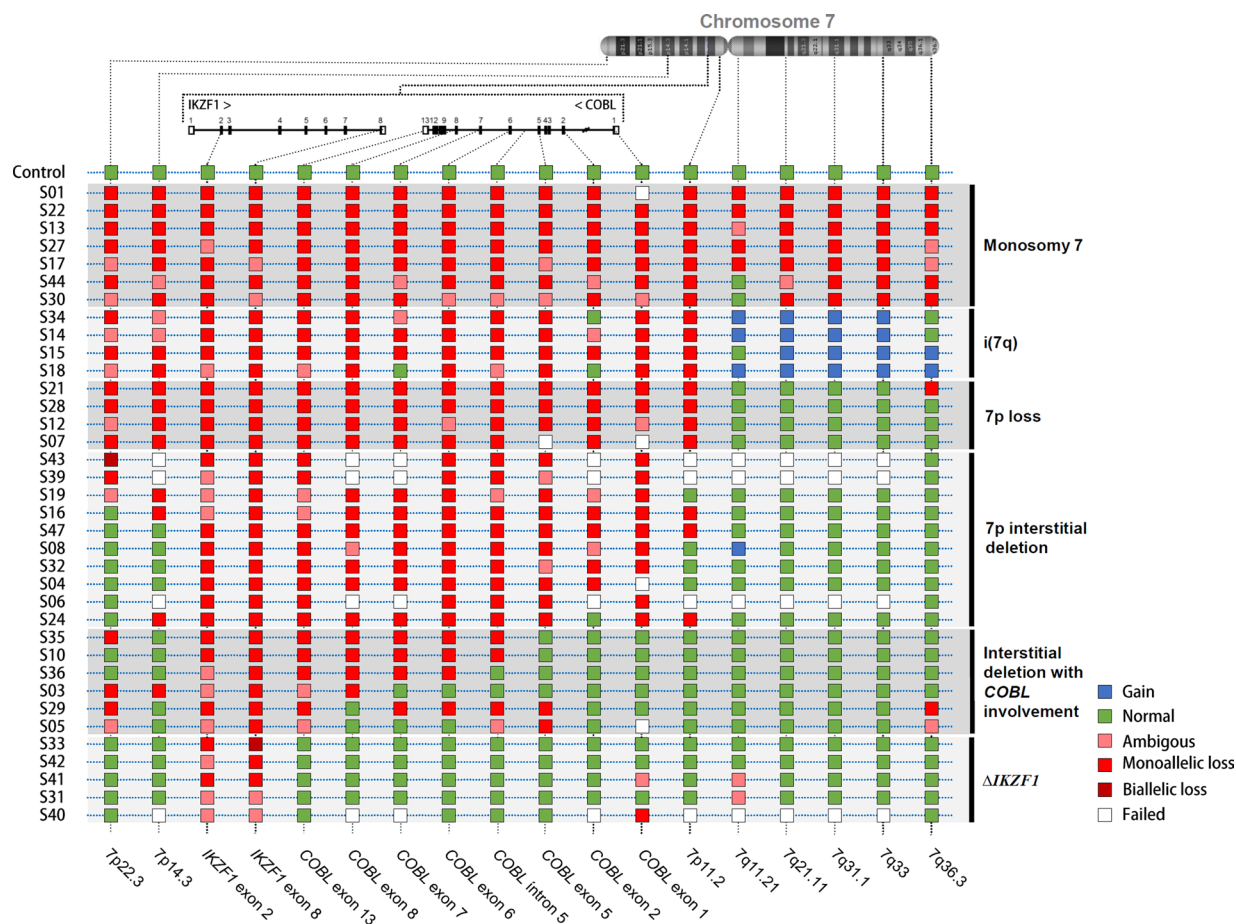


Figure 2: Chromosome 7 CNAs of pediatric BCP-ALL with *IKZF1* Δ1-8. The CNA results for samples with *IKZF1* Δ1-8 are displayed in each row, while vertical lines specify the localization of MLPA probes within chromosome 7. The colored squares represent regions with amplification (blue), without CNAs (green), and monoallelic deleted (red) or biallelic deletions (dark red). Ambiguous results (pink square) represents borderline peak ratios. The samples were clustered into six groups based on CNAs for chromosome 7: monosomy 7, isochromosome (7q), 7p loss, 7p interstitial deletion, interstitial deletions in the chromosome 7, and complete deletion of *IKZF1* only.

Table 1: Demographic, laboratory and clinical characteristics of *IKZF1* deleted cases with *COBL* involvement

Characteristics	Patient identification						
	S03	S05	S10	S29	S35	S36	S48
Age (yrs)	1 [†]	5	5	16	1 [§]	5	15
Sex	M	F	M	M	F	F	M
Laboratory							
WBC (×10 ⁹ /L)	5.0	5.7	16.4	55.0	459.6	7.5	NA
Immunophenotype	c-ALL	c-ALL	Pre-B	c-ALL	c-ALL	c-ALL	Pre-B
ALL subtype	NA	NA	NA	High hyperdiploid	<i>ETV6-RUNX1</i>	<i>ETV6-RUNX1</i>	B-other
<i>CDKN2A/B</i> status	Deleted	wt	NA	Deleted	Deleted	wt	Deleted
<i>PAX5</i> status	Deleted	Deleted	NA	Deleted	Deleted	wt	wt
Clinical data							
Clinical trial	COALL97	COALL92	COALL97	GBTLI-93	None [¶]	GBTLI-93	UKALLR3
CNS disease	No	No	No	No	NA	No	No
MRD (D33)	Negative	Negative	Negative	NA	-	NA	2E-2
CR (D33)	Yes	Yes	Yes	NA	-	NA	Yes
NCI risk group	LR	LR	LR	HR	HR	LR	HR
Relapse	NA	Yes	Yes	Yes	No	No	Yes
Time to relapse (yrs)	NA	6.5	9.0	2	-	-	3.3
Outcome	NA	Alive	Alive	Dead	Dead	Alive	Dead
Follow-up (mo)	NA	96	47	13	0.5	69	8

c-ALL, common acute lymphoblastic leukemia; CNS, central nervous system; CR, complete morphological remission; F, female; HR, high risk; LR, low risk; M, male; MRD, minimal residual disease; mo, months; NA, not available; NCI, national cancer institute; SR, standard risk; wt, wild-type; yrs, years.

[†]13 months-old at diagnosis.

[§]20 months-old at diagnosis.

[¶]The child died before any chemotherapy treatment.

sequences at the proximity of the breakpoint site. Then, we compared the RIC scores of complete deletions and intragenic deletions of *IKZF1* from our previous report [12]. As summarized in the Figure 3A–3C, RSSs and RIC scores were significantly different between these groups. Intragenic deletions, including *IKZF1-COBL* fusions, presented RAG analog sequences, while complete deletions did not present functional RSSs, supporting the hypothesis that another mechanism may be associated with complete deletions of *IKZF1*. Moreover, data analyses from publicly available tracks from the ENCODE consortium showed that the active chromatin regions associated with the presence H3K4me1 and DNase hypersensitive sites were located at the promoter region, as well as intron 5 and intron 7 of *COBL* gene (Figure 3D).

DISCUSSION

Nearly 30% of all Δ *IKZF1* in pediatric BCP-ALL comprise complete deletions of *IKZF1*, termed *IKZF1* Δ 1-8. Here, we developed novel methods to characterize

these patient cases in more detail. Our results showed that most of the *IKZF1* Δ 1-8 derive from large interstitial deletions within chromosome 7, spanning genes either upstream (*ZBPB* and *C7orf72*, ~210 Kb upstream of *IKZF1*) or downstream (*FIGNL1* and *DDC*, 39 and 53 Kb downstream, respectively) of *IKZF1*. Interestingly, *EBF1* and *BTG1* deletions were found in patients with intragenic *IKZF1* deletions, but they were rarely deleted in the cohort of *IKZF1* Δ 1-8 patients, corroborating with a previous report showing that *BTG1* and *EBF1* deletions co-occur in ALL [13]. Our data suggest that *IKZF1* Δ 1-8 promotes the development of a leukemogenic process that is independent of alterations in *EBF1* and *BTG1*. Furthermore, a recent study found that mutual *BTG1* and *IKZF1* deletions cooperatively increased the incidence of relapse in pediatric BCP-ALL cases. However, the patients were not stratified based on *IKZF1* deletion subgroups [14]. Therefore, *BTG1* deletions may play a synergistic role with intragenic *IKZF1* deletions, but the same might not be true for *IKZF1* Δ 1-8.

Contrary to intragenic deletions, *IKZF1* Δ1-8 is characterized by larger chromosomal deletions on chromosome 7. After a first screening round with microarrays in order to identify CNAs within chromosome 7, we developed a novel MLPA assay and performed detailed analyses to better characterize such alterations. In this study, children with BCP-ALL presented *IKZF1* Δ1-8 due to monosomy 7 (7/36, 19%) or large interstitial deletions that occurred on chromosome 7 (16/36, 44%). Alterations such as isochromosome 7q, 7p loss, and *IKZF1* Δ1-8 without involvement of surrounding regions were also found, but at much lower frequency. Earlier studies have also associated *IKZF1* Δ1-8 to monosomy 7 and interstitial deletions on chromosome 7 [11, 15, 16].

Our results also revealed seven patients with BCP-ALL bearing large interstitial deletions that all started within intronic regions of the *COBL*, which is localized

~611 Kb downstream of *IKZF1*. The frequency of *COBL* rearrangements varied for complete (16.7%) and intragenic (0%) deletions of *IKZF1* at diagnosis, and were found in 4.2% of cases at relapse. *COBL* is an actin nucleator and contains three copies of the WH2 (WASP homology 2) actin-binding domain, thus promoting actin polymerization [17]. It has important roles for neuronal morphogenesis and regulation of microvillar length. Its alterations have been associated to neuronal disorders (autism spectrum disorders) [18], and autoimmune diseases (eg. type 1 diabetes) [19]. In 2013, Meyer *et al.* characterized Δ*IKZF1* in pediatric BCP-ALL patients, and described for the first time an *IKZF1-COBL* tail-to-tail fusion as a consequence of an ~800 kb interstitial deletion between *IKZF1* intron 1 and *COBL* intron 5 [12]. In the present study, we also found an *IKZF1-COBL* fusion in a relapsed sample. In 2015, Baughn *et al.* found one patient

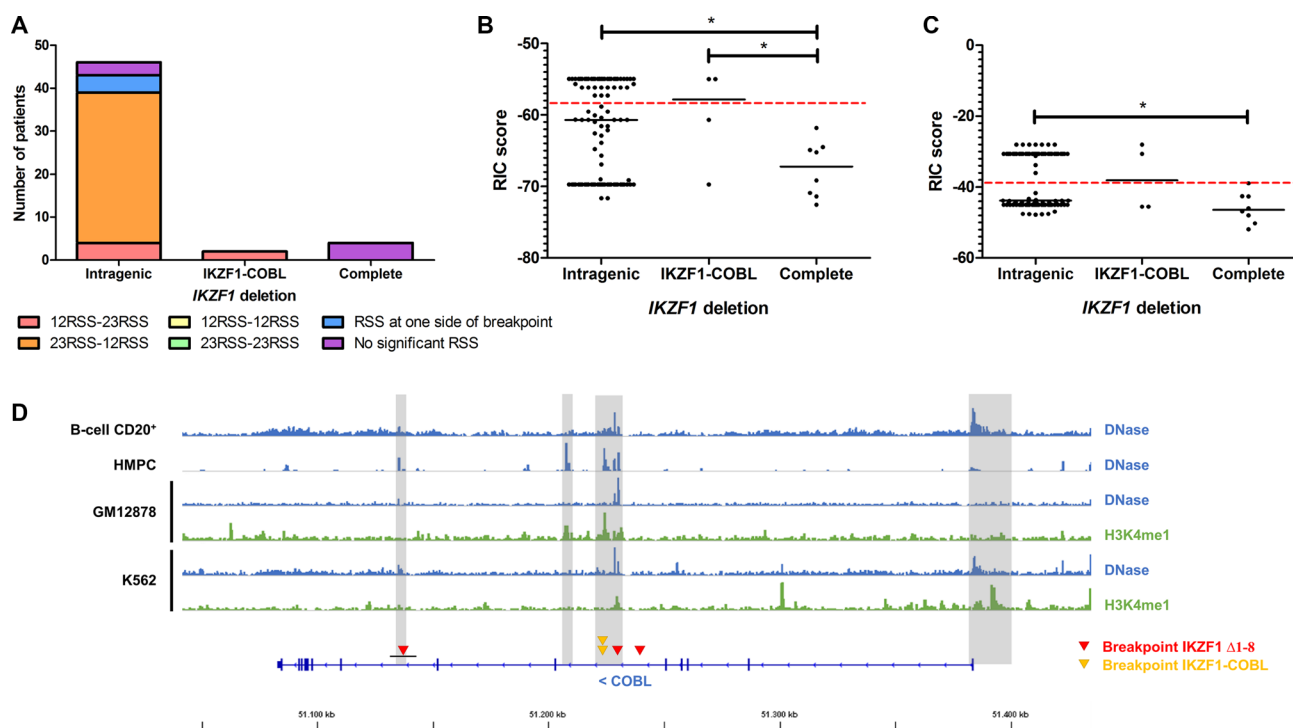


Figure 3: Comparison of RSSs at the breakpoint sequences of intragenic and complete deletions of *IKZF1* in childhood BCP-ALL. (A) The figure illustrates the RSS combination at each side of the breakpoints for intragenic and complete deletions of *IKZF1*, showing that intragenic deletions, which includes *IKZF1-COBL* fusions, present 23RSS-12RSS and 12RSS-23RSS at the breakpoint region, while complete deletions of *IKZF1* do not present RAG analogous sequences. We also compared RAG-analogue sequences for (B) 12-spaced RSSs, and (C) 23-spaced RSSs using the RIC score methodology. The greater RIC scores indicates higher similarity to RSSs. Dashed red lines show threshold scores associated to functional RSSs, and dots represent RSS-analogous sequences for each patient. Because recombination events generally occurs between 12-RSS and 23-RSS, each side of breakpoint region is limited to one statistically significant and another sequence not significant for 12/23-RSSs. Intragenic deletions were collected from our previous study [12], and complete deletions from the current study. The results show that complete deletions of *IKZF1* do not present statistically significant RSS-analogous sequences (* *p*-value < 0.05). (D) Integrative Genomics Viewer (IGV) visualization of DNase-seq and ChIP-seq (target: H3K4me1) data retrieved from ENCODE database showing accessible chromatin sites within *COBL* for human primary CD20-positive B-cells, hematopoietic precursor-cells (HMPC), as well as B lymphoblastoid (GM12878) and chronic myeloid leukemia (K562) cell lines. The grey squares highlight the regions associated to open chromatin at the promoter, as well as intron 5 and intron 7 of *COBL* gene, located on the antisense strand of 7p12.1. Colored triangles illustrate the position of the breakpoints within *COBL* leading to *IKZF1* complete deletions (red) and *IKZF1-COBL* fusions (yellow). The black horizontal line at *COBL* intron 7 defines the range of breakpoint region for patient S03 detected by MLPA. The scale bar refers to human GRCh37/hg19 genome assembly.

with BCP-ALL and normal karyotype that presented a ~917 Kb interstitial deletion within chromosome 7, leading to *IKZF1* Δ 1-8. Again, the breakpoint was located within *COBL* intron 6 [20]. In 2011, Flach *et al.* described one patient with an evolution from myelodysplastic syndrome to acute myeloid leukemia after accumulation of 7p12.1–12.2 deletion ranging from *IKZF1* to *COBL* [21]. In 2016, Duployez *et al.* reported a patient with myeloproliferative neoplasm who progressed to blast crisis upon acquisition of biallelic *IKZF1* deletions, as well as *EBF1* and *CDKN2A/B* deletions. In this case, *IKZF1* deletion involved its surrounding genes, from *VWC2* until *COBL* [22]. Furthermore, Gonzalez-Gonzalez *et al.* used a SNP-array to identify a ~941 Kb amplification between *IKZF1* and *COBL* intron 2 in a patient with metastatic colorectal cancer [23]. These data show that *COBL* rearrangements are recurrently found in *IKZF1* Δ 1-8, and are also found in cases with intragenic deletions of *IKZF1* (*IKZF1-COBL* fusions), suggesting a relationship between genes located at 7p12.1 (*IKZF1*, *DDC*, *GRB10*, and *COBL*) and cancer. Interestingly, breakpoints within *COBL* were also found in autism spectrum disorders [18]. Although our data supported by these aforementioned studies suggest that *COBL* is a downstream hotspot for Δ *IKZF1*, the breakpoint sites varied considerably at the telomeric side of deletions. This finding has been concordant in both SNP array and MLPA screenings. For that reason, it was not possible to investigate any breakpoint hotspot upstream of *IKZF1* and, consequently, it is not feasible to include detection of *IKZF1* Δ 1-8 in the multiplex PCR panels previously published.

Several studies have suggested that RAG-recognition errors might promote intragenic Δ *IKZF1*, based on the identification of RSS-analogue sequences in the vicinity of the identified breakpoints [24, 25]. However, the association between RAG recombination and *IKZF1* Δ 1-8 is still unclear. Using the RIC score methodology, we have demonstrated that breakpoint sequences of *IKZF1* Δ 1-8 investigated were not similar to RSSs. Also, *IKZF1* Δ 1-8 with *COBL* rearrangements did not present additional nucleotides at the breakpoint sites, suggesting that complete deletions of *IKZF1* are not attributed to neither aberrant RAG activity nor terminal deoxynucleotidyl transferase (TdT) involvement. On the other hand, RSSs analogous sequences and additional nucleotides were found for patients with intragenic deletions of *IKZF1*, including patients with *IKZF1-COBL*. Therefore, analyses of breakpoint sequences reveal that intragenic deletions are possibly mediated by RAG-recombination events, while the remaining large interstitial deletions and monosomy 7 leading to *IKZF1* Δ 1-8 are the result of other mechanism that caused chromosome instability. Such mechanisms could involve genomic hotspots due to the changes in the architecture of chromosomes, or, other mechanisms that are associated with DNA double-strand breakage. It is noteworthy that the breakpoints found in our study

positioned within accessible chromatin regions, therefore, it is plausible that such area is more susceptible to double-strand DNA breaks.

In conclusion, we demonstrate that monosomy 7 and large interstitial deletions within chromosome 7 are the main causes of complete deletions of *IKZF1*. *COBL* rearrangements were recurrently found in these patients, showing that *COBL* represents a genetic hotspot for Δ *IKZF1*. Both cases with *IKZF1-COBL* had breakpoints within one base pair in *COBL* intron 5, so screening of new patient sets with 3'-*IKZF1* deletions may reveal similar patients with rare RAG initiated deletions. Further investigation of *COBL* rearrangements are needed to better characterize its role in BCP-ALL, and to answer the question whether the deletion of *COBL* or other genes localizing between *IKZF1* and *COBL* could be important for leukemogenesis and prognosis. For this purpose, we developed a customized MLPA assay for the evaluation of CNAs within *COBL*.

MATERIALS AND METHODS

Subjects

First, forty-three diagnostic samples of children with *IKZF1* Δ 1-8 were selected, being 24 patients enrolled in a Brazilian previously published study [26] and 19 patients registered in the German CoALL [12]. Briefly, *IKZF1* Δ 1-8 were analyzed by MLPA (SALSA MLPA P335-A3-B2 probe mix and/or SALSA MLPA P202-B1, MRC Holland, Amsterdam, The Netherlands), according to the manufacturer's recommendations. Based on the data obtained in this first screening, we also investigated two additional series of patients: (i) with lack of *IKZF1* exon 8, namely Δ 2–8, Δ 3–8 or Δ 4–8 ($n = 25$), being 13 Brazilian and 12 Australian samples (ANZCHOG ALL8 or AIEOP-BFM ALL2009 trials), which were identified by MLPA analysis of 399 and 568 new diagnosis samples, respectively, and (ii) 24 relapse samples investigated by both MLPA and sequencing analysis. In accordance with the Declaration of Helsinki, clinical data collection (e.g. gender, age at diagnosis, white blood cell (WBC) count at diagnosis, and ALL subtype) and laboratory procedures have been evaluated and approved by the Ethics Committees of Instituto Nacional de Câncer-INCA (#33243214.7.0000.5274) and the Sydney Children's Hospital Network LNR.13.SCHN.367.

Microarray assay

The microarray analysis was performed using the CytoScan HD Array according to the manufacturer protocol (Affymetrix, Inc., Santa Clara, CA, USA). Briefly, 250 ng of genomic DNA from six patients were digested with *NspI*, and then amplified with Titanium Taq PCR Kit (Clontech Laboratories, Inc., Mountain View, CA). After fragmentation

and labeling, the DNA was hybridized to the microarray for 16 hours, washed on the GeneChip Fluidics Station 450, stained with Affymetrix GeneChip Stain Reagents, and scanned on the GeneChip Scanner 3000 7G (Affymetrix, Inc., Santa Clara, CA, USA). Data were analyzed using Chromosome Analysis Suite software version 3.0 (Affymetrix, Inc., Santa Clara, CA, USA) based on the GRCh37/hg19 build of the Human Genome Assembly.

Multiplexed long-distance PCR

The breakpoints of interstitial deletions or gene fusion indicated by microarray were confirmed by MP-PCRs in order to analyze the breakpoints at the nucleotide level. The reaction was performed with a set of ten primers flanking a region of ~20 Kb surrounding each breakpoint. Amplification was performed with PCR Extender System (5Prime, Germany) and the primers listed in Supplementary Table S3.

Customized multiplex ligation-dependent probe amplification

Two in-house customized MLPA assays were designed to investigate CNAs within chromosome 7, with a special focus on *COBL*. The design of the probes was based on the manual “Designing synthetic MLPA probes”, version 14 (MRC-Holland); probe details are described in Supplementary Table S4. The validation was performed by a comparison of CNA data between microarray and customized MLPA. In brief, 100 ng of genomic DNA were denatured and hybridized overnight with the customized probes. Then, the probes were ligated and amplified with SALSA MLPA EK1 reagents (MRC Holland, The Netherlands). The fragments were separated by ABI 3,500 Genetic Analyzer (Applied Biosystems, EUA), and analyzed with GeneMarker v1.85 (SoftGenetics), where the relative copy numbers are normalized according to the peaks observed in controls.

Long distance inverse PCR

The LDI-PCR was used to analyze breakpoints within *COBL* intron 5. The technique was previously described for *KMT2A* rearrangements detection, and basically consists of seven steps: (1) DNA digestion with restriction enzymes, (2) religation of the ends to form circular DNA, (3) amplification of the circular DNA of interest (4) Agarose gel electrophoresis to separate derivative bands from wild type bands (5) gel extraction of derivative bands (6) sequencing of derivative bands (7) via BLAST alignment of the identified sequence with the human genome [27]. The primer sequences are listed in Supplementary Table S3.

Identification of mechanisms leading to *IKZF1* deletions

RAG1/2 RSSs were investigated along the breakpoint sequences. The “RSS database” searches RSS sequences, consisting of a heptamer (5'-CACAGTG-3') and a nonamer (5'-ACAAAAACC-3') separated by either 12 or 23 nucleotides (12RSS and 23RSS), and classifies the sequences based on the “RIC score”, which estimates the similarity between the sequence of interest and the RSS consensus sequence [28]. 12RSSs and 23RSSs greater than -38.81 and -58.45, respectively, were attributed as possibly functional. In addition, DNase-seq and ChIP-seq data were retrieved from ENCODE and visualized with Integrative Genomics Viewer (IGV) version 2.3.77 to assess chromatin structure of *COBL*.

Statistical analysis

This study compared clinical-demographic characteristics and CNAs between samples according to *IKZF1* status (*IKZF1* Δ1-8 vs. wild-type or intragenic deletions) with Fisher's exact test using SPSS Statistics 18 (IBM, EUA). For the analysis of RAG consensus sequences at breakpoints, we used an unpaired t test to compare RIC scores among *IKZF1* deletion subgroups. GraphPad Prism 5 (GraphPad Software, Inc., California, USA) software) was used for this analysis. *P*-values < 0.05 were interpreted as statistically significant.

ACKNOWLEDGMENTS

We are grateful to the children and their parents for participating in the study. We thank the Pediatric Hematology-Oncology Program staff (AJ Faro, BKS Souza, BAA Gonçalves, CFG Andrade, C Zampier, ETG Pina, G Vasconcelos), J Merkens, and all physicians who contributed with clinical and demographical data (Adriana Vanessa Santini, Alejandro Mauricio Arancibia, Cibelli Navarro, Dora Marcia Santana Alencar, Everaldo Ruiz Junior, Fernando de Almeida Werneck, Gustavo Ribeiro Neves, Imarui Costa, Isis Maria Quezado Magalhaes, Jozina Maria de Andrade Agareno, Lilian Maria Burlacchini de Carvalho, Marcelo dos Santos Souza, Patricia Brito, Renata Silva de Carvalho Gurgel, Rosania Maria Basegio, Silvia Rocha Caetano Agra, Teresa Cristina Cardoso Fonseca, Terezinha de Jesus Marques Salles). We acknowledge parents and children, and the Australian and New Zealand Children's Haematology and Oncology Group hospital staff and oncologists especially Professor Glenn Marshall, Dr Frank Alvaro, Dr Luciano Dalla Pozza, Dr Heather Tapp and Dr Chris Fraser.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This investigation was supported by the Brazilian National Council of Technological and Scientific Development (CNPq#447385/2014-3), by the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ#E-26/110.533/2014), and Instituto Nacional de Câncer (INCA). ME has been supported by CNPq (PQ-2014#304142/2014-0) and FAPERJ (JCNE#2015-2017) research scholarships. RM is being supported by the DFG grant Ma 1876/11-1. RS and DW acknowledge funding support from the National Health and Medical Research Council in Australia.

Author's contributions

BAL and ME wrote the manuscript. BAL, CM, TCB, US and NV performed and analyzed laboratory data. MH, SH, DLW and RS contributed with clinical and demographical data. MSPO and RM contributed to the writings and critical analysis of the data. CM and ME contributed to the conception of the study, writings and critical analysis of the data. All authors contributed with revision of the final version of the manuscript.

REFERENCES

1. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013; 499:214–218.
2. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, Su X, Pui CH, Relling MV, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*. 2007; 446:758–764.
3. Georgopoulos K, Moore DD, Derfler B. Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. *Science*. 1992; 258:808–812.
4. Rebollo A, Schmitt C. Ikaros, Aiolos, Helios: transcription regulators and lymphoid malignancies. *Immunol Cell Biol*. 2003; 81:171–175.
5. Ramirez J, Lukin K, Hagman J. From hematopoietic progenitors to B cells: mechanisms of lineage restriction and commitment. *Curr Opin Immunol*. 2010; 22:177–184.
6. Kuiper RP, Schoenmakers EF, van Reijmersdal SV, Hehir-Kwa JY, van Kessel AG, van Leeuwen FN, Hoogerbrugge PM. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions

affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia*. 2007; 21:1258–1266.

7. van der Veer A, Waanders E, Pieters R, Willemse ME, Van Reijmersdal SV, Russell LJ, Harrison CJ, Evans WE, van der Velden VH, Hoogerbrugge PM, Van Leeuwen F, Escherich G, Horstmann MA, et al. Independent prognostic value of BCR-ABL1-like signature and IKZF1 deletion, but not high CRLF2 expression, in children with B-cell precursor ALL. *Blood*. 2013; 122:2622–2629.
8. Boer JM, van der Veer A, Rizopoulos D, Fiocco M, Sonneveld E, de Groot-Kruseman HA, Kuiper RP, Hoogerbrugge P, Horstmann M, Zaliouva M, Palmi C, Trka J, Fronkova E, et al. Prognostic value of rare IKZF1 deletion in childhood B-cell precursor acute lymphoblastic leukemia: an international collaborative study. *Leukemia*. 2015; 30:32–38.
9. Jia M, Wang ZJ, Li JY, Yang SL, Zhao HZ, Cheng YP, Luo ZB, Tang YM. The impact of IKZF1 deletion on the prognosis of acute lymphoblastic leukemia: an updated meta-analysis. *Cancer Biomark*. 2014; 14:493–503.
10. Venn NC, van der Velden VH, de Bie M, Waanders E, Giles JE, Law T, Kuiper RP, de Haas V, Mullighan CG, Haber M, Marshall GM, Md N, van Dongen JJ, et al. Highly sensitive MRD tests for ALL based on the IKZF1 Delta3-6 microdeletion. *Leukemia*. 2012; 26:1414–1416.
11. Caye A, Beldjord K, Mass-Malo K, Drunat S, Soulier J, Gandemer V, Baruchel A, Bertrand Y, Cave H, Clappier E. Breakpoint-specific multiplex polymerase chain reaction allows the detection of IKZF1 intragenic deletions and minimal residual disease monitoring in B-cell precursor acute lymphoblastic leukemia. *Haematologica*. 2013; 98:597–601.
12. Meyer C, Zur Stadt U, Escherich G, Hofmann J, Binato R, Barbosa TC, Emerenciano M, Pombo-de-Oliveira MS, Horstmann M, Marschalek R. Refinement of IKZF1 recombination hotspots in pediatric BCP-ALL patients. *Am J Blood Res*. 2013; 3:165–173.
13. Waanders E, Scheijen B, van der Meer LT, van Reijmersdal SV, van Emst L, Kroeze Y, Sonneveld E, Hoogerbrugge PM, van Kessel AG, van Leeuwen FN, Kuiper RP. The origin and nature of tightly clustered BTG1 deletions in precursor B-cell acute lymphoblastic leukemia support a model of multiclonal evolution. *PLoS Genet*. 2012; 8:e1002533.
14. Scheijen B, Tijchon E, van Ingen Schenau D, Marke R, van Emst L, Van Der Meer LT, Kuiper RP, Boer JM, Pieters R, Hoogerbrugge PM, Den Boer ML, van Leeuwen FN. Tumor Suppressors BTG1 and IKZF1 Cooperate during Mouse Leukemia Development and Impact Relapse Rate in Childhood Acute Lymphoblastic Leukemia. *Blood*. 2015; 126:905–905.
15. Heerema NA, Harbott J, Galimberti S, Camitta BM, Gaynon PS, Janka-Schaub G, Kamps W, Basso G, Pui CH, Schrappe M, Auclerc MF, Carroll AJ, Conter V, et al. Secondary cytogenetic aberrations in childhood

- Philadelphia chromosome positive acute lymphoblastic leukemia are nonrandom and may be associated with outcome. *Leukemia*. 2004; 18:693–702.
16. Martinelli G, Iacobucci I, Storlazzi CT, Vignetti M, Paoloni F, Cilloni D, Soverini S, Vitale A, Chiaretti S, Cimino G, Papayannidis C, Paolini S, Elia L, et al. IKZF1 (Ikaros) deletions in BCR-ABL1-positive acute lymphoblastic leukemia are associated with short disease-free survival and high rate of cumulative incidence of relapse: a GIMEMA AL WP report. *J Clin Oncol*. 2009; 27:5202–5207.
 17. Winckler B, Schafer DA. Cordon-bleu: a new taste in actin nucleation. *Cell*. 2007; 131:236–238.
 18. Griswold AJ, Ma D, Cukier HN, Nations LD, Schmidt MA, Chung RH, Jaworski JM, Salyakina D, Konidari I, Whitehead PL, Wright HH, Abramson RK, Williams SM, et al. Evaluation of copy number variations reveals novel candidate genes in autism spectrum disorder-associated pathways. *Hum Mol Genet*. 2012; 21:3513–3523.
 19. Winkler C, Krumsiek J, Lempainen J, Achenbach P, Grallert H, Giannopoulou E, Bunk M, Theis FJ, Bonifacio E, Ziegler AG. A strategy for combining minor genetic susceptibility genes to improve prediction of disease in type 1 diabetes. *Genes Immun*. 2012; 13:549–555.
 20. Baughn LB, Biegel JA, South ST, Smolarek TA, Volkert S, Carroll AJ, Heerema NA, Rabin KR, Zweidler-McKay PA, Loh M, Hirsch B. Integration of cytogenomic data for furthering the characterization of pediatric B-cell acute lymphoblastic leukemia: a multi-institution, multi-platform microarray study. *Cancer Genet*. 2015; 208:1–18.
 21. Flach J, Dicker F, Schnittger S, Schindela S, Kohlmann A, Haferlach T, Kern W, Haferlach C. An accumulation of cytogenetic and molecular genetic events characterizes the progression from MDS to secondary AML: an analysis of 38 paired samples analyzed by cytogenetics, molecular mutation analysis and SNP microarray profiling. *Leukemia*. 2011; 25:713–718.
 22. Duployez N, Nibourel O, Ducourneau B, Gardel N, Boyer T, Bories C, Darre S, Coiteux V, Berthon C, Preudhomme C, Roche-Lestienne C. Acquisition of genomic events leading to lymphoblastic transformation in a rare case of myeloproliferative neoplasm with BCR-JAK2 fusion transcript. *Eur J Haematol*. 2016; 97:399–402. doi: 10.1111/ejh.12752.
 23. Gonzalez-Gonzalez M, Fontanillo C, Abad MM, Gutierrez ML, Mota I, Bengoechea O, Santos-Briz A, Blanco O, Fonseca E, Ciudad J, Fuentes M, De Las Rivas J, Alcazar JA, et al. Identification of a characteristic copy number alteration profile by high-resolution single nucleotide polymorphism arrays associated with metastatic sporadic colorectal cancer. *Cancer*. 2014; 120:1948–1959.
 24. Mullighan CG, Miller CB, Radtke I, Phillips LA, Dalton J, Ma J, White D, Hughes TP, Le Beau MM, Pui CH, Relling MV, Shurtleff SA, Downing JR. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature*. 2008; 453:110–114.
 25. Iacobucci I, Storlazzi CT, Cilloni D, Lonetti A, Ottaviani E, Soverini S, Astolfi A, Chiaretti S, Vitale A, Messa F, Impera L, Baldazzi C, D'Addabbo P, et al. Identification and molecular characterization of recurrent genomic deletions on 7p12 in the IKZF1 gene in a large cohort of BCR-ABL1-positive acute lymphoblastic leukemia patients: on behalf of Gruppo Italiano Malattie Ematologiche dell'Adulto Acute Leukemia Working Party (GIMEMA AL WP). *Blood*. 2009; 114:2159–2167.
 26. Barbosa TC, Terra-Granado E, Quezado Magalhaes IM, Neves GR, Gadelha A, Guedes Filho GE, Souza MS, Melaragno R, Emerenciano M, Pombo-de-Oliveira MS. Frequency of copy number abnormalities in common genes associated with B-cell precursor acute lymphoblastic leukemia cytogenetic subtypes in Brazilian children. *Cancer Genet*. 2015; 208:492–501.
 27. Meyer C, Schneider B, Reichel M, Angermueller S, Strehl S, Schnittger S, Schoch C, Jansen MW, van Dongen JJ, Pieters R, Haas OA, Dingermann T, Klingebiel T, et al. Diagnostic tool for the identification of MLL rearrangements including unknown partner genes. *Proc Natl Acad Sci U S A*. 2005; 102:449–454.
 28. Merelli I, Guffanti A, Fabbri M, Cocito A, Furia L, Grazini U, Bonnal RJ, Milanese L, McBlane F. RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. *Nucleic Acids Res*. 2010; 38:W262–267.