REGULAR ARTICLE

blood advances

Classification and genetics of pediatric B-other acute lymphoblastic leukemia by targeted RNA sequencing

Natacha Azussa Migita,[1,2] Patrícia Yoshioka Jotta,[1] Natália Paiva do Nascimento,[1] Victor Sande Vasconcelos,[1,2] Gabriel Lopes Centoducatte,[1,2] Katlin Brauer Massirer,[3] Amilcar Cardoso de Azevedo,[1] Silvia Regina Brandalise,[1] and José Andrés Yunes[1,4]

[1]Centro Infantil Boldrini, Campinas, Brazil; [2]Graduate Program in Genetics and Molecular Biology, Biology Institute, State University of Campinas, Campinas, Brazil; [3]Center for Medicinal Chemistry, Center for Molecular Biology and Genetic Engineering, State University of Campinas, Campinas, Brazil; and [4]Genetics Department, Faculty of Medical Sciences, State University of Campinas, Campinas, Brazil

Key Points

- Targeted RNA sequencing with a commercially available kit can be used to identify the novel subgroups of B-other ALL, except for iAMP21.
- Differential gene-expression analysis of targeted RNA-sequencing data identify *DUX4*-r, a subgroup not defined by fusions or mutations.

Acute lymphoblastic leukemia (ALL) can be classified into different subgroups based on recurrent genetic alterations. Here, targeted RNA sequencing was used to identify the novel subgroups of ALL in 144 B-other and 40 "classical" ALL samples. The classical *TCF3-PBX1*, *ETV6-RUNX1*, *KMT2A*-rearranged, and *BCR-ABL1*, and novel *P2RY8-CRLF2, ABL-*, *JAK2-*, *ZNF384-*, *MEF2D-*, and *NUTM1*-fusions were easily identified by fusion transcript analysis. *IGH-CRLF2* and *IGH-EPOR* were found by abnormally high levels of expression of *CRLF2* or *EPOR*. *DUX4*-rearranged was identified by the unusual expression of *DUX4* genes and an alternative exon of *ERG*, or by clustering analysis of gene expression. *PAX5*-driven ALL, including fusions, intragenic amplifications, and mutations were identified by single-nucleotide variant analysis and manual inspection using the IGV software. Exon junction analysis allowed detection of some intragenic *ERG* and *IKZF1* deletions. *CRLF2*-high associated with initial white blood cell (WBC) counts of $\geq 50 \times 10^3/\mu L$ and *GATA3* risk alleles (rs3781093 and rs3824662), whereas *ABL*/*JAK2*/*EPOR*-fusions associated with high WBC counts, National Cancer Institute's high-risk classification, and IKZF1del. *ZNF384*-fusions associated with CALLA-negativity and *NUTM1*-fusions in infants. In conclusion, targeted RNA sequencing further classified 66.7% (96 of 144) B-other ALL cases. All BCP-ALL subgroups, except for iAMP21, hyperdiploid and hypodiploid cases, were identified. Curiously, we observed higher frequencies of females within B-rest ALLs and males in *PAX5*-driven cases.

Introduction

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy, accounting for 25% of pediatric cancers.[1] Over the last decades, advances in ALL treatment and supportive care have dramatically increased the survival rate of pediatric ALL to between ~80% and 90%.[2-4] Contemporary protocols rely on sophisticated risk stratification of patients, which includes genetic features[5,6] and minimal residual disease levels.[7,8] In this context, the genetic risk factors include recurrent chromosomal rearrangements, aneuploidy, and cooperating mutations involved in leukemogenesis.

Submitted 18 October 2022; accepted 14 February 2023; prepublished online on *Blood Advances* First Edition 27 February 2023. https://doi.org/10.1182/bloodadvances.2022009179.

A subset of genomic data for representative cases of each of the alteration described in our study were submitted on Zenodo: https://doi.org/10.5281/zenodo.7559347.

Data are available on request from the corresponding author, José Andrés Yunes (andres@boldrini.org.br).

The full-text version of this article contains a data supplement.

Approximately 70% of B-cell precursor ALL (BCP-ALL) can be classified by standard genetic analyses into well-known subtypes, based on the presence of t(9;22) *BCR-ABL1* (Ph-positive), rearrangements of *KMT2A (MLL)* at 11q23, t(1;19) *TCF3 (E2A)-PBX1*, t(17;19) *TCF3-HLF*, t(12;21) *ETV6-RUNX1*, high-hyperdiploidy (HeH), or hypodiploidy.[9,10] Until some years ago, the remaining 30% of pediatric BCP-ALL were called "B-other", because this portion did not have any of the aforementioned characteristic chromosomal aberrations at diagnosis, and the underlying driver events were unknown.[9] After extensive genetic studies, some B-other ALL cases could be grouped into common; nonoverlapping genetic subgroups, such as Ph-like, iAMP21, *DUX4*-cluster, *ZNF384*-rearranged, *MEF2D*-rearranged, and *NUTM1*-rearranged; and *PAX5*-driven.[11-13] Identifying these new biomarkers are important because some serve as novel therapeutic targets and/or provide prognostic information.[14]

Current ALL diagnostic strategies include karyotyping or DNA index by cytometry to detect high-hyperdiploidy, and reverse transcription polymerase chain reaction (RT-PCR) or fluorescence in situ hybridization (FISH) for "classical" fusions. Patients at high risk who are negative for this initial testing are eligible for RT-qPCR, FISH, Sanger and/or next generation sequencing (NGS) to detect other actionable/prognostic biomarkers, like *CRLF2*-high, *JAK1*/*JAK2* mutations, and *ABL*-class fusions.[15,16] This complex workflow might be simplified by NGS methods and extended to all patients.

Here, we present a study on 144 pediatric B-other ALL cases in conjunction with 40 BCP-ALL cases of the classical genetic groups, to assess the feasibility and benefits of introducing a commercial panel of targeted RNA sequencing into the routine diagnostics. Moreover, we present, to the best of our knowledge, for the first time, the frequencies of these novel ALL subtypes in a South American population and explore their associations with clinico-biological features at presentation.

Materials and methods

Patients and study design

This study included 144 of 177 consecutive cases of B-other ALL treated in a single institution and 40 cases of classical ALL subtypes included as controls: high-hyperdiploidy (n = 3), *TCF3-PBX1* (n = 3), *ETV6-RUNX1* (n = 24), *BCR-ABL1* (n = 4), and *KMT2A*-rearrangements (n = 6). Patients (≤18 years) were diagnosed and treated at the Boldrini Children's Center in accordance with the Brazilian GBTLI ALL-1999[17] and -2009[18] protocols. General demographic and genetic information of patients enrolled in these protocols is summarized in supplemental Table 1. Leukemia diagnosis was established based on morphology, immunophenotype, cytogenetics, RT-PCR, and multiplex ligation probe-dependent amplification (MLPA, using kits from MRC Holland) as part of routine clinical diagnostics. B-other classification was established by a lack of t(12;21), t(1;19), t(9:22), and *KMT2A*-rearrangements and by having ≥44 and <51 chromosomes (or DNA index of <1.16). The study was approved by the institutional ethical committee (CAAE51991015.0.0000.5376). All patients, parents, or guardians provided informed consent for sample collection and research in accordance with the Declaration of Helsinki.

Library preparation and targeted RNA sequencing

RNA was obtained from bone marrow mononuclear cells. RNA integrity (DV200 > 50%) was checked on an Agilent 2100 Bioanalyzer (Agilent Technologies). Targeted RNA sequencing of 1385 cancer-related genes was done using the TruSight RNA Pan-Cancer kit (Illumina) per the manufacturer's instruction. Approximately 3 million reads per sample (2 × 76 cycles) were sequenced in MiSeq equipment.

Sequencing data analysis

Data analysis was performed using *Illumina BaseSpace* apps: RNA-seq Alignment version 1.1 and version 2.0, and DRAGEN RNA Pipeline version 3.5, with standard settings. Reads were mapped to the GRCh37/hg19 reference genome. Fusion detection was also performed using CICERO.[19] Fusions were considered true positive if reported as "PASS" by at least 2 software. *IGH*, *ERG,* and *CRLF2* fusions were considered true positive even by a single software (usually DRAGEN RNA or CICERO). Predicted fusions were manually curated on ProteinPaint/FusionEditor (https://proteinpaint.stjude.org/FusionEditor/), excluding fusions caused by misalignment of reads to pseudogenes. All rearrangements were manually reviewed using the integrative genomics viewer (IGV).[20] Detection of *PAX5* amplifications, *CRLF2*-rearrangements, and *ERG* and *IKZF1* intragenic deletions were manually curated by analyzing "soft-clipped" and exon junction reads in IGV. Strelka variant caller was used for single-nucleotide variant (SNV) and small indel identification. Salmon was used for gene-expression quantification in transcripts per million (TPM). Strelka and Salmon were run in the RNA-seq Alignment version 2.0 platform. A subset of genomic data for representative cases of each of the alterations described in our study was submitted on Zenodo under URL https://doi.org/10.5281/zenodo.7559347.

Hierarchical clustering analysis (HCA) and t-SNE plot

Individual gene-expression values (TPM) were log2 transformed and then subjected to Euclidean hierarchical clustering analysis (HCA) with a method of Ward. Different probe sets used for ALL clustering analysis were selected from the literature and genes matching with the 1385 targeted panel were used for HCA. The final lists of genes used in clustering analysis are provided in supplemental Table 2. For t-distributed stochastic neighbor embedding (t-SNE) analysis, the log2 TPM expression values were filtered using the *varFilter* function on the *genefilter* R package (version 1.74.0), with a cutoff set at 0.5 and with interquartile range (IQR) as the base function. The resulting 731 most variable genes were used for t-SNE. The t-SNE plot was generated using the *tsne* function on the *M3C* R package (version 1.14.0)[21] using a perplexity score of 10 (set.seed(1238) perplex = 10).

GATA3 single-nucleotide polymorphism genotyping and validation of fusion transcripts

The rs3781093 and rs3824662 *GATA3* polymorphisms were genotyped by a 2-step PCR (primers in supplemental Table 3) followed by Sanger sequencing.[22] For *GATA3* allele frequencies in non-ALL controls, we used the ABraOM (Brazilian Genomic Variants) database, consisting of 1171 admixed individuals from São Paulo (Brazil).[23] The validation of fusion transcripts was done by RT-PCR and/or FISH, by routine laboratory procedures.

Statistical analysis

Possible associations between genetic abnormalities and demographic or clinical data were estimated using a 2-sided Fisher exact test with Monte Carlo simulations (based on 10 000 sampled tables with starting seed of 2 000 000). Analyses were performed using the SPSS Statistics, WinSTAT, GraphPad Prism, and/or R (www.r-project.org) software.

Results

In total, ~32% (177 of 561) of non–Down syndrome (DS) BCP-ALL cases treated at the Boldrini Children's Center (Brazil) from 1999 to 2017 could not be classified in any of the classical molecular subtypes of ALL and were provisionally classified as B-other ALLs (supplemental Table 1). Of these, 144 cases had frozen samples available and were selected for the study. The median age at diagnosis was 7.38 years, males were predominant (55.6%), and 86% were standard risk (SR) at diagnosis based on the National Cancer Institute (NCI) risk classification (Table 1). Of the 144 cases, 7 had intrachromosomal amplification of chromosome 21 (iAMP21) by FISH and MLPA (supplemental Figure 1). We kept iAMP21 as part of B-other ALLs, because it was still a provisional entity when this work began.[10]

Identifying most subtype-defining fusion transcripts

Except for the high-hyperdiploid cases, all other classical subtypes of ALL included as controls were correctly detected by the Pan-Cancer panel. Fusion transcripts were found in 32.6% (47 of 144) of B-other samples investigated (Figure 1). All the recently described B-other fusion transcripts were detected, including *ABL*-class (n = 7, 4.9%), *JAK2*-rearranged (n = 3, 2.1%), *P2RY8-CRLF2* (n = 6, 4.2%; 1 of which was only found by manually searching soft-clipped reads in the IGV software), *ZNF384*-rearranged (n = 9, 6.3%), *MEF2D*-rearranged (n = 4, 2.8%), *NUTM1*-rearranged (n = 4, 2.8%), and *ETV6*-rearranged (n = 2, 1.4%) (supplemental Figure 2-3). Commonly missed by many popular fusion detection tools, we found 1 *IGH-EPOR* fusion but only by using CICERO (supplemental Table 4). *IGH-CRLF2* cases could not be detected in fusion transcript analysis but were later identified by gene-expression analysis (further sections).

Table 1. Clinical characteristics of 144 pediatric patients with B-other ALL registered at the Boldrini Children's Center who were included in this study

	B-other features
Total of B-other analyzed	144
Treatment protocol	
GBTLI LLA-1999	86
GBTLI LLA-2009	58
Age at diagnosis (y)*	
Median	7.38
Range	0.28-17.85 (average 8.36)
Sex, n (%)	
Male	80 (55.6)
Female	64 (44.4)
WBC count at diagnosis (x 10^9/L)	
Median	20.9
Range	1.3-459.0
WBC count at risk	
<50 × 10^9/L	97
≥50 × 10^9/L	47
NCI risk at diagnosis, n (%)	
Standard risk	124 (86.1)
High risk	20 (13.9)

*One patient with outlier age (21 years) was registered and treated within the GBLTI-1999.

We found novel gene partners for *JAK2* (*TOP2B-JAK2*), *ZNF384* (*NCOA3-ZNF384*, *SPI1-ZNF384*), *MEF2D* (*MEF2D-PYGO2*), *NUTM1* (*KAT6A-NUTM1*), *ETV6* (*ETV6-MANSC1*), and *PAX5* (*PAX5-ZBED3*) in the B-other cohort. Other novel in-frame fusions are shown in Figure 1C; supplemental Table 4, a subset of which were validated by FISH, RT-PCR, or MLPA (supplemental Figure 2).

Identifying *PAX5* fusions and mutations, and the analysis of soft-clipped sequences to detect *PAX5* intragenic amplifications

There are 2 subtypes of B-ALL defined by *PAX5* alterations: PAX5-altered (PAX5alt) and *PAX5* p.Pro80Arg mutated (PAX5 P80R). The former comprises diverse *PAX5* rearrangements, including fusions, focal intragenic amplifications, and mutations (other than P80R). Fusion transcript analysis resulted in the identification of 11 cases with *PAX5* fusions, 4 out-of-frame (supplemental Table 4). Two out-of-frame *PAX5*-rearranged cases (*PAX5-ZCCHC7*, *PAX5-MLLT3*) also harbored a *P2RY8-CRLF2* fusion. One of the *ETV6-RUNX1* controls also carried a *PAX5-ZCCHC7* fusion. To detect *PAX5* intragenic amplification, soft-clipped sequences were manually inspected in the IGV software, as previously described.[13] Four patients with B-other ALL showed altered *PAX5* transcripts indicative of intragenic exons 2-5 amplification (supplemental Figure 4), which was confirmed by MLPA (data not shown). SNVs in the *PAX5* coding and exon/intron regions were found in 12 cases, 5 of which corresponding to p.P80R. The variant allele frequency of *PAX5* P80R for these patients was 44%, 49%, 72%, 99%, and 100%, respectively (messenger RNA level). Loss of the other *PAX5* allele is a hallmark of *PAX5* P80R.[12] The first patient harbored a frame-shift mutation in the other *PAX5* allele whereas the last 2 patients had a single *PAX5* loss by MLPA (data not shown). In total, we identified 13 PAX5alt (9%) and 5 *PAX5* P80R (3.5%) cases.

Identifying some *ERG* deletions by exon junction track analysis

Five B-other ALLs had a fusion between *ERG* and LINC01423, likely resulting from an intrachromosomal deletion at the *ERG* locus (ERGdel). When *ERG* soft-clipped sequences were manually searched in IGV, we were able to detect another 7 cases with *ERG-LINC01423* fusions, totaling 12 cases. This *ERG* fusions seemed characteristic of some ERGdel and were easily visualized in a Sashimi plot[24] as splice junction reads connecting *ERG* exon 4 (or exon 11 in few cases) to LINC01423 (supplemental Figure 5). One of these *ERG-LINC01423* cases harbored a *NUP214-ABL1* fusion whereas all the other 11 were later shown to belong to

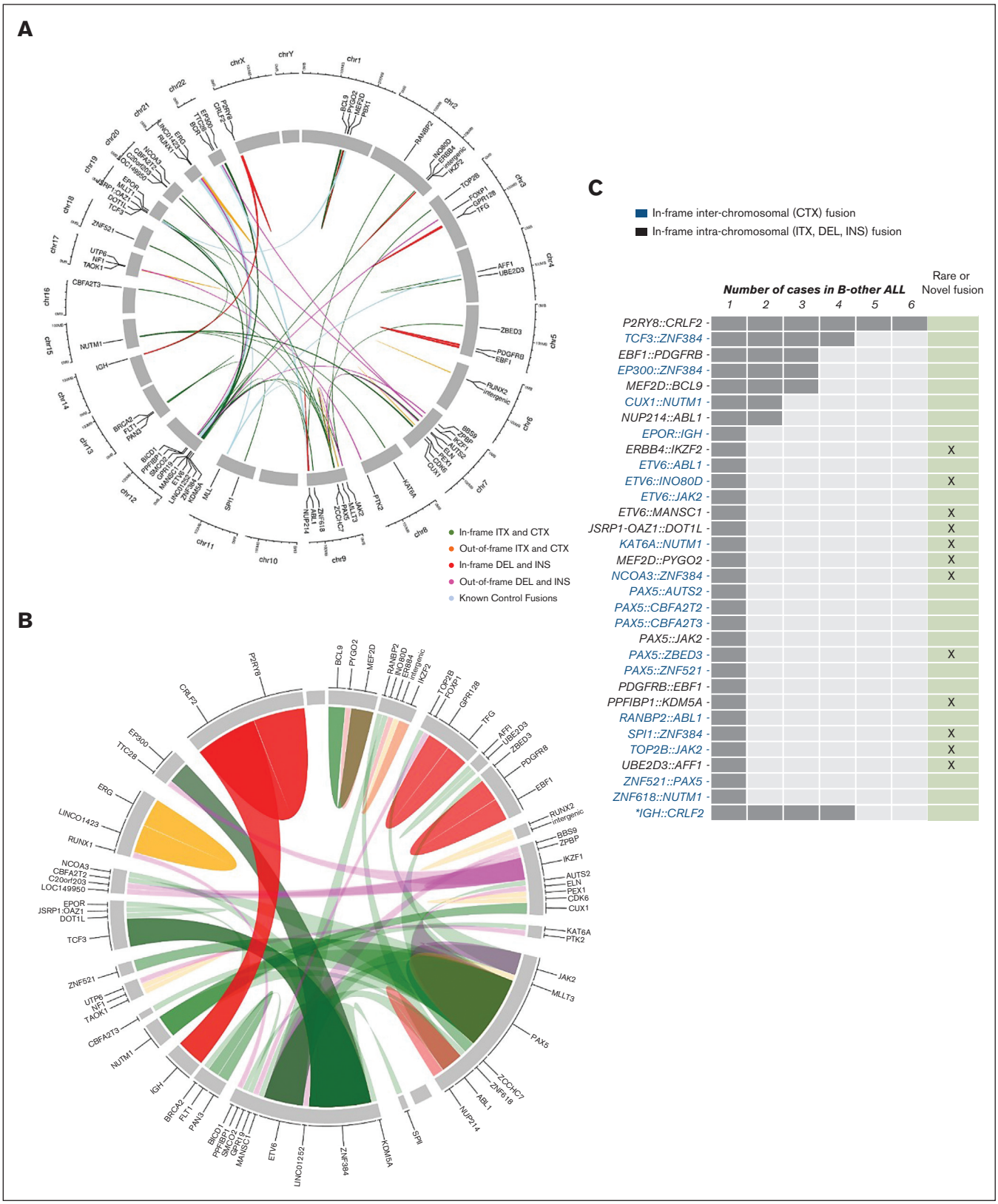

Figure 1. Overview of gene fusions discovered by targeted RNA sequencing. (A) Circos plot showing the gene fusions schematically organized into chromosome position, both for classical and B-other ALL cases. Genes are arranged based on their genomic position in the chromosome (clockwise from chromosome 1 to 22, followed by X and Y). The line links the 2 partner genes in a fusion, and the colored legend shows the rearrangement type. All classical gene fusions are shown as light blue ribbons. (B) Zoomed-in view of fusion genes to show in more detail the rearrangements found specifically in the 144 B-other–classified cases. Ribbon widths are proportional to the frequency

DUX4-rearranged (see further sections). However, when MLPA was used to investigate *ERG* deletions in the entire cohort (supplemental Table 5), only 9 of 12 *ERG-LINC01423* were confirmed. Moreover, MLPA detected 7 additional cases (5 belonging to *DUX4*-rearranged and 2 to other groups; supplemental Figure 5), not found by Pan-Cancer RNA sequencing. Therefore, two-thirds of all *ERG* deletions could be found by Pan-Cancer RNA sequencing. Interestingly, MLPA analysis revealed that *ERG-LINC01423* was associated with *ERG* exons 5-11 deletions, an alteration found exclusively in *DUX4*-rearranged cases.

Identifying some *IKZF1* deletions by exon junction track analysis

MLPA analysis revealed that 30% (41 of 136) of cases bore different *IKZF1* deletions (IKZF1del). Having this information, we investigated whether IKZF1del could be identified by splice junctions track analysis of soft-clipped sequences in the IGV software. The *IKZF1* gene has a relatively complex pattern of alternative splicing, making Sashimi plots hard to interpret. Nevertheless, by knowing beforehand which *IKZF1* exons were deleted, we were able to find certain exon-exon connecting arcs, with at least 4 reads, unique to intragenic IKZF1del cases (Figure 2; supplemental Figure 6A). For example, the second most common type of IKZF1del, that is, exon 4-7 deletions (Δ4-7), were characterized by >200 reads spanning exon 3-8 junctions (Figure 2B; supplemental Figure 6B). The same 3-8 junctions, but with less reads, were seen in a case with Δ5-6. *IKZF1* Δ4-8 showed junction reads spanning *IKZF1* exon 3 to downstream genes, *FIGNL1* or *DDC* (Figure 2C; supplemental Figure 6C).

We noticed that junctions between exons 1-5 or 1-8 and junctions between *IKZF1* exon 3 and *FIGNL1* did not occur in nondeleted samples, being exclusive to IKZF1del cases. Therefore, we believe these 3 junctions can be trusted to discriminate some IKZF1del cases. In summary, only the following IKZF1del isoforms could be realistically detected: Δ1-3, Δ1-7, Δ2-3, Δ2-7, Δ4-7, Δ4-8, and Δ5-6, which represent approximately half of IKZF1del cases. Neither whole gene *IKZF1* deletions (Δ1-8), which account for ~40% of IKZF1del cases, nor isoforms Δ1 and Δ1-6 could be detected by splice junction read analysis (supplemental Table 6). Representative Sashimi plots for all different IKZF1del are presented in supplemental Figure 6.

Identifying *CRLF2*-high and *IGH-EPOR* by gene overexpression

In addition to identifying recurrent fusion genes, targeted RNA sequencing simultaneously measures the expression of all captured genes within each sample. Ranking of *CRLF2* expression values (log2 TPM+1) in all samples resulted in a sigmoid curve showing a group of 11 top outliers with *CRLF2* levels 3 times higher than the median (Figure 3A). As expected, all *P2RY8-CRLF2* cases found by fusion transcript analysis were in the *CRLF2*-high group (supplemental Figure 7A-7B). Four of the 5 remaining outliers were found to bear an *IGH-CRLF2* rearrangement by FISH (supplemental Figure 7C). Unfortunately, there was no material for FISH analysis of the last case, which, however, had *CRLF2* p.F232C mutation commonly seen in *CRLF2*-high cases.[25] The frequency of *CRLF2* rearrangements was consistent with the 1:1 proportion of translocations (*IGH-CRLF2*) to deletions (*P2RY8-CRLF2*) usually found in non-DS pediatric ALL.[26] Of note, patients in the *CRLF2*-high group also had *JAK2* (p.R683, n = 3; p.A658S, n = 1; p.L925P, n = 1) or *CRLF2* (p.F232C, n = 3) mutations that are known to be associated with induction of ligand-independent activation of downstream signaling pathways.[27,28] A similar approach was used to analyze *EPOR* expression. A single case of *EPOR*-high was found (Figure 3B), which corresponded to the *IGH-EPOR* detected by fusion transcript analysis.

Identifying *DUX4*-rearrangements by read depth pattern or gene-expression profiling

DUX4-rearranged ALL originates from translocation of a truncated copy of *DUX4* into the *IGH* or *ERG* (less common) locus, resulting in a chimeric transcript formed by most of the *DUX4* coding sequence plus noncoding regions of *IGH* or *ERG*.[29] We could not detect *DUX4*-rearrangements by fusion transcript analysis. However, by simply analyzing read coverage over *DUX4 loci* in chromosomes 4 and 10, and the *ERG* locus in chromosome 21, we differentiated a total of 35 cases showing high *DUX4* gene family expression, of which 27 expressed also an alternative isoform of *ERG* (ERGalt) with an extra exon 6 (Figure 4). As expected, 11 of 12 *ERG-LINC01423* cases (see earlier results) were among the 35 cases in the *DUX4*-rearranged subgroup.[30]

Because the subtypes of ALL are characterized by distinct gene-expression profiles, we investigated whether unsupervised clustering analysis would differentiate the subtypes of BCP-ALL. Pan-Cancer genes matching genes previously used for ALL clustering (supplemental Table 2) were used for HCA. As shown in Figure 5A, using 124 genes of the 864 selected by Liu et al (2016),[11] there was a clustering tendency for some ALL subtypes, particularly *DUX4*-, *ZNF384*-, and *MEF2D*-rearranged ALL. Some groups were better clustered when using a different gene set, although segregation was never perfect (supplemental Figure 8). Curiously, HCA based on 45 genes differentially expressed by Ph-like ALL (supplemental Table 2) resulted in a group of 9 of 11 ALL cases carrying ABL or JAK/STAT-activating fusions and 5 of 11 *CRLF2*-rearranged cases (2 with *P2RY8-CRLF2*, 2 with *IGH-CRLF2*, and 1 with *CRLF2* F232C mutation) (supplemental Figure 8A). The fact that half of *CRLF2*-high and almost all ABL and JAK/STAT-activating rearrangements were in the same group was consistent with Ph-like ALL.[14]

Importantly, when gene-expression profiles were analyzed by t-SNE, *DUX4*-rearranged and *ETV6-RUNX1* cases were clearly segregated (Figure 5B). Besides, *ZNF384*-rearranged, ABL or JAK/STAT-activating fusions, iAMP21, and *PAX5*-driven cases were somehow grouped but mostly intermingled, which may be a consequence of the low number of cases included in the analysis.

Figure 1 (continued) of a specific gene in a fusion event. (C) Number of in-frame fusion genes found in 47 of 144 B-other ALL cases. The bars indicate the number of patients in which a given fusion gene was observed. Rare and novel fusion genes discovered in this study are highlighted. §*PAX5-ZCCHC7* was found in a patient that also has a functional *ETV6-RUNX1* fusion. **IGH-CRLF2* were detected by high *CRLF2* expression and confirmed by FISH. CTX, interchromosome; DEL, deletion; INS, insertion; ITX, intrachromsome.

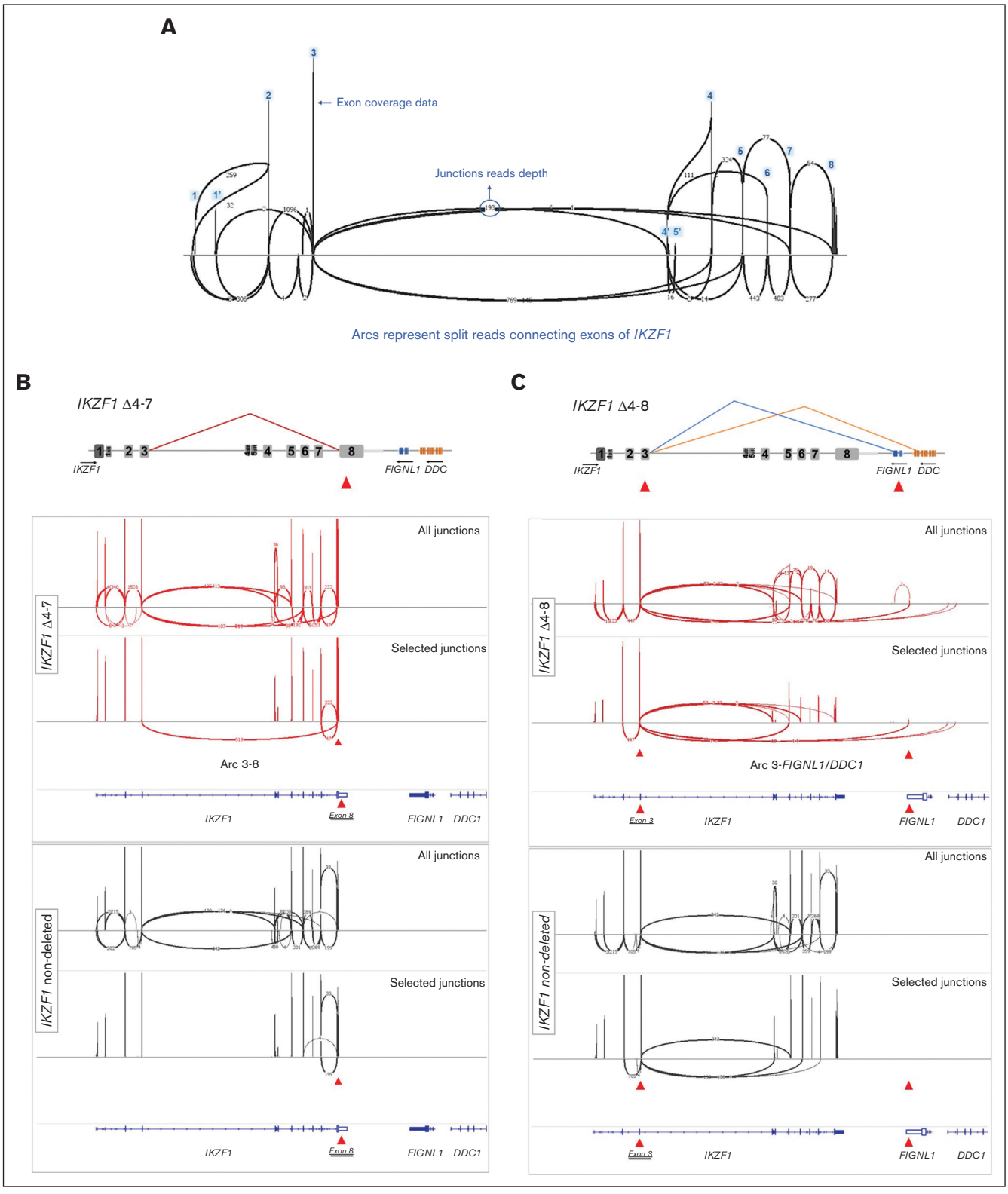

Figure 2. Identification of representative *IKZF1* deleted cases by splice junction track analysis. (A) Schematic representation of Sashimi plot for 1 *IKZF1* nondeleted case displaying all expected arcs connecting *IKZF1* exons. The vertical peaks represent coverage depth on *IKZF1* exons. Arcs represent splice junctions that connect exons, and the numbers on curved lines show the junction read depth that split across the junction. (B) Schematic view and Sashimi plots obtained in the IGV software for a representative *IKZF1* Δ4-7 case (red) displaying abnormal splice junction arcs connecting exon 3-8, which it is not seem in the *IKZF1* nondeleted case (black). The upper panel displays all exon-exon junctions. The lower panel display selected exon junctions, which distinguish a particular *IKZF1* deletion. This is accomplished in IGV by clicking on the

In summary, after identifying all the group-defining genetic alterations, a total of 96 of 144 B-other cases could be further classified into 1 of the modern subgroups of ALL. For the remaining 48 cases (33.3%) (B-"rest"), no new recurrent biomarker could be found.

Associations of demographic, clinical features, and IKFZ1del and *GATA3* polymorphisms with the novel subgroups of B-other ALL

After demonstrating that the Pan-Cancer assay was able to identify several of the new subgroups of B-other ALL, we searched for possible associations between these subgroups and some of the patients' features at presentation (Figure 6A). After testing for gender biases, we observed that girls were significantly enriched in the B-rest subgroup ($P = .049$), that is, 48 B-other cases without any known genetic marker. *PAX5*-driven cases tended to occur more frequently in boys (Fisher exact test; $P = .043$; Monte Carlo; $P = .071$) (Figure 6B).

To increase the statistical power, *ABL*-, *JAK2*-, and *EPOR*-fusion ALLs were analyzed as a single group named AJ2E-fusion. Both AJ2E ($P = .007$) and *CRLF2*-high ($P = .042$), which are characterized by an activated kinase signature, were significantly associated with higher white blood cell (WBC) counts at diagnosis. Moreover, the AJ2E-fusion group was associated with NCI high-risk classification ($P = .038$) and with higher frequency of IKZF1-del ($P < .0001$) (Figure 6D; supplemental Table 7).

In total, 8 of the 144 B-other ALLs had CALLA (CD10)-negative immunophenotype, which were significantly enriched in the *ZNF384*-rearranged group ($P < 10^{-7}$). Noteworthy, 2 of 4 *NUTM1*-rearranged cases were infants, confirming the notion that *NUTM1*-rearrangements are rare but recurrent in infants, particularly among those who lack *KMT2A*-rearrangements (Figure 6D; supplemental Table 7).[31,32] No other associations with clinical and demographic data were observed (supplemental Table 7).

Independent genome-wide association studies in different ethnic populations have identified several risk *loci* for ALL susceptibility, including *ARID5B, IKZF1, CEBPE, PIP4K2A, CDKN2A,* and *GATA3*. Some of these polymorphism were associated with distinct ALL subtypes.[22,33] Notably, 2 inherited genetic variations in *GATA3*, single-nucleotide polymorphisms rs3824662 and rs3781093, are strongly associated with leukemia treatment outcomes and are particularly relevant in patients of Hispanic ancestry.[34] The susceptibility related to *GATA3* polymorphisms appears to be affected by cytokine and kinase-activating alterations, as in cases of Ph-like ALL.[35] We found the A risk allele at rs3824662 and the C risk allele at rs3781093 to be over-represented in most B-other ALL subgroups, particularly in *CRLF2*-high (Figure 6C). Interestingly, B-rest had *GATA3* allele frequencies similar to that of healthy individuals.[22,23]

Discussion

The laboratory methods used in the clinical diagnosis of genetic alterations in BCP-ALL include karyotyping, FISH, and RT-PCR. These methods restrict the number of genetic loci that can simultaneously be analyzed and are limited to the classical subtypes of ALL. NGS RNA sequencing circumvents this problem by providing the analysis of several *loci* at once and is particularly useful because it allows the identification of fusion transcripts derived from recurrent chromosomal translocations that are the hallmarks of ALL.[36-46] However, in the current diagnostics of ALL, NGS is used in addition to previous methods in intricate workflows. Our results suggest that targeted RNA sequencing might substitute FISH and RT-PCR methods, simplifying the current diagnostic strategy for ALL, while providing identification of both classical and modern subgroups of ALL, except for ploidy and copy number alterations. Despite the relatively low number of positive controls analyzed, fusion detection by targeted RNA sequencing was highly consistent with results obtained by cytogenetic and RT-PCR for *BCR-ABL1*, *ETV6-RUNX1*, *TCF3-PBX1, KMT2A-AFF1, KMT2A-MLLT1*, and *KMT2A-MLLT3*. Except for iAMP21 cases, targeted RNA sequencing allowed detection of all other novel subgroups of B-other ALL, which includes *CRLF2*-high, *ABL*-fusion, *JAK2*-fusion, *EPOR*-fusion, *ZNF384*-rearranged, *NUTM1*-rearranged, *MEF2D*-rearranged, *ETV6*-fusion, *PAX5*-altered, *PAX5* P80R, and *DUX4*-rearranged (including segregation of cases with *ERG*alt).

Targeted RNA sequencing takes longer than total RNA sequencing but is less expensive, provides a deeper coverage of transcripts, and is easier to analyze. Moreover, targeted RNA sequencing is applicable to lower quantities/quality of RNA. We successfully used the TruSight Pan-Cancer kit for the analysis of bone marrow smears (data not shown). The advantage of using a commercial sequencing assay, with global kit distribution, is that analytical performance has been evaluated by the manufacturer and by independent researchers. The TruSight series of sequencing panels ensure deep, uniform coverage, which is 1 of the most important parameters for highly accurate variant calling. In hematologic malignancies, the TruSight Pan-Cancer was shown to detect 100% of fusions transcripts.[42,43] Regarding SNV, concordance between Pan-Cancer RNA sequencing and exome analysis was 86%, except for nonsense mutations likely subjected to messenger RNA decay.[42] Serial dilution of *BCR-ABL1*$^{+}$ RNA into normal control RNA revealed a detection limit ranging from 1% to 10%.[47]

The TruSight RNA Pan-Cancer kit has been tested with success within the diagnostic workflow of the international AIEOP-BFM ALL 2017 trial in a cohort of 117 BCP-ALL cases.[43] All classical group-defining fusion transcripts were identified as well as some of the rearrangements involving *CRLF2, PAX5, JAK2, PDGFRB*, *ETV6, ZNF384, MEF2D,* and *NUTM1* genes, commonly found in B-other ALL. However, the number of B-other cases included in this previous study (n = 36) was relatively small. In this manuscript, we used the Pan-Cancer panel to analyze 144 B-other ALL in conjunction with 40 BCP-ALL of the classical groups as controls. Besides confirming previous observations on the utility of this targeted RNA-sequencing panel in identifying group-defining fusion transcript, we also explored SNVs, exon junction tracks, and gene-expression data (see workflow in supplemental Figure 9), to show

Figure 2 (continued) specific exon, as indicated by the red arrowhead. (C) Schematic view and Sashimi plots for a representative *IKZF1* Δ4-8 case (red) displaying the abnormal splice junctions connecting exon 3 to downstream genes *FIGNL1* and *DDC*. Note that the abnormal splice junction arcs are only seen in *IKZF1* deleted cases. Results for the whole group of patients are summarized in supplemental Figure 6 and supplemental Table 6. Genomic coordinates plotted on the last track of each panel correspond to RefSeq, Hg19.

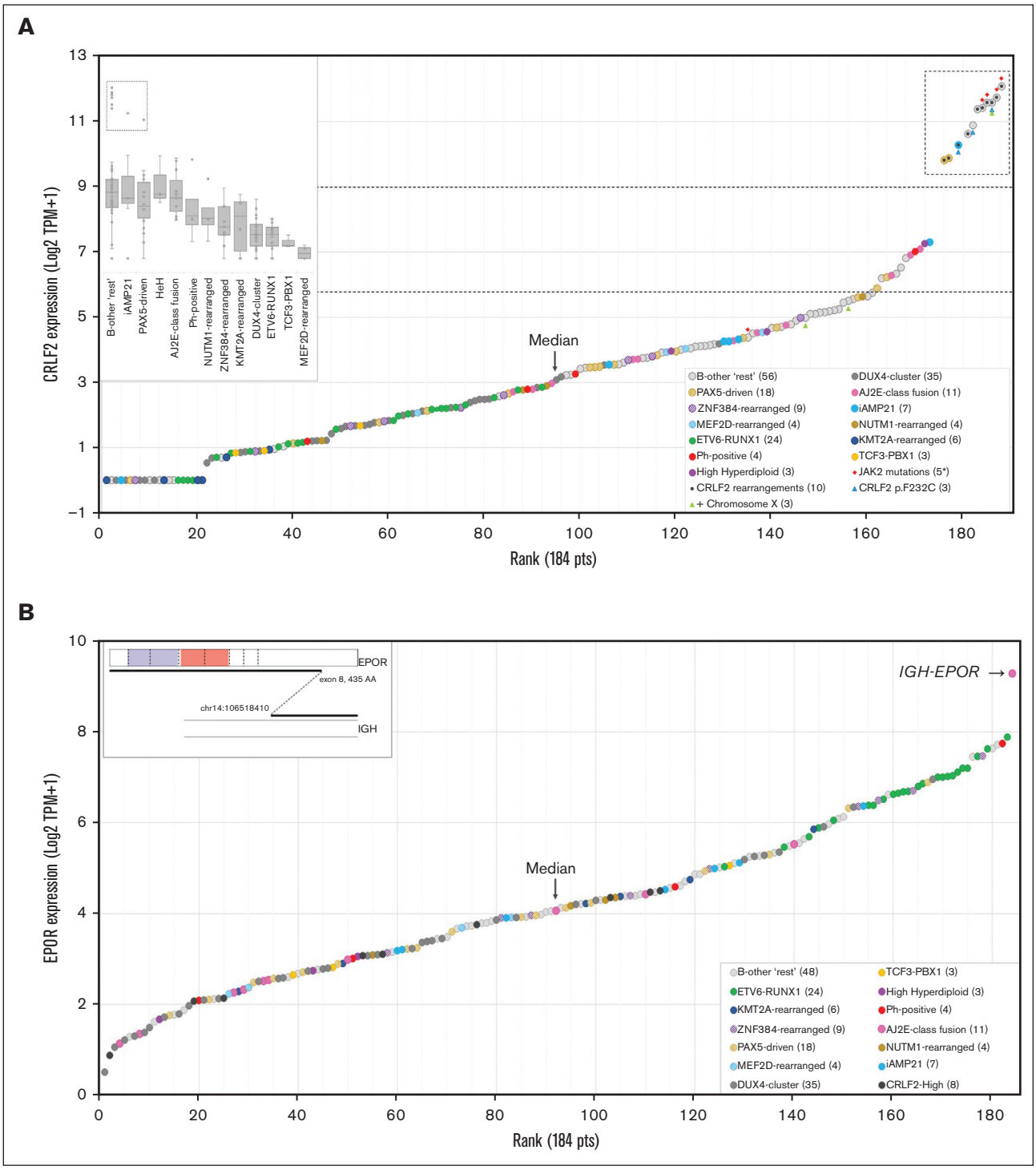

Figure 3. Overexpression of *CRLF2* and *EPOR*. (A) Normalized expression (Log2 TPM+1) of *CRLF2* in ascending order. Data from 184 BCP-ALLs (144 B-other and 40 classical subtypes of ALL) are shown. Symbols represent different genetic alterations. The dashed box highlights samples with *CRLF2* expression three-fold higher than the median, which correspond to samples with *CRLF2*-rearrangements (*P2RY8-CRLF2*, *IGH-CRLF2)* or *CRLF2* p.F232C. (B) Normalized rank-ordered expression (Log2 TPM+1) of *EPOR*. The "outlier" case corresponds to the single *IGH-EPOR* found by fusion transcript analysis. AJ2E, group of *ABL*-, *JAK2*-, and *EPOR*-fusion cases.

for the first time its utility in the identification of some other molecular markers of ALL, particularly *PAX5* P80R, *IGH-CRLF2*, intragenic ERGdel and IKZF1del, and *DUX4*-rearrangements. The frequencies of these genetic alterations in the Brazilian cohort of B-other ALLs (Figure 7) seemed compatible to those found by total RNA sequencing in a recent European study of unselected consecutive B-other cases,[13] except that we found a much higher number of fusions resulting in ABL and JAK/STAT kinase activation

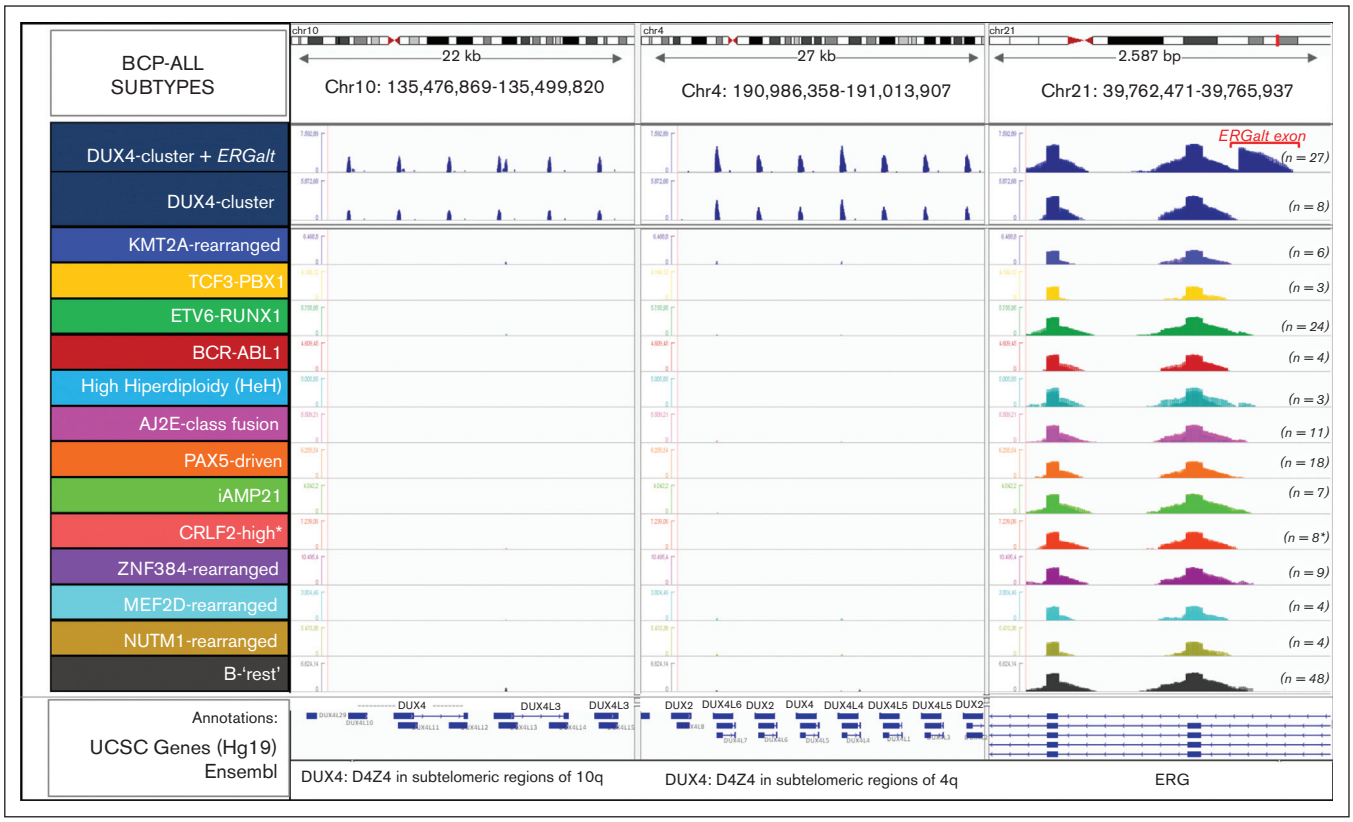

Figure 4. Identification of *DUX4*-rearranged ALL by the differential expression profile of the *DUX4* gene clusters and *ERG* exons. The plot shows read depth coverage on 3 different chromosomal regions, that is, *DUX4* multifamily gene array in the telomeric regions of chromosomes 4 and 10, and *ERG* gene on chromosome 21 aligned on the human Hg19 genome. *DUX4* gene is exclusively expressed in the *DUX4*-rearranged group, and is associated, in most cases (27 of 35), with expression of an alternative exon 6 of *ERG* (ERGalt). The location of the ERGalt exon is shown in red. Screenshot of the IGV browser. *Here, 2 cases grouped within the *PAX5*-driven group and 1 in the iAMP21 group has an outlier expression of *CRLF2*. AJ2E, group of *ABL*-, *JAK2*-, and *EPOR*-fusion cases.

(11 of 144 compared with 1 of 110). We believe that this discrepancy is not a consequence of sampling bias because our inclusion rate was of 86.1% (supplemental Table 1). The discrepancy might relate to methodological differences: targeted RNA sequencing vs total RNA sequencing, read depth, and/or bioinformatics tools. Another possible reason for the increased fraction of ABL- and JAK-STAT-activating fusions in the South American cohort could be related to ancestry.[48]

A series of novel fusion partner genes were identified: *TOP2B* for *JAK2*, *KAT6A* and *ZNF618* for *NUTM1*, *MANSC1* for *ETV6*, *NCOA3* and *SPI1* for *ZNF384*, and *PYGO2* for *MEF2D*. Although the Pan-Cancer targets only 1385 genes, new in-frame fusion genes (*JSRP1*; *OAZ1-DOT1L, UBE2D3-AFF1, PPFIBP1-KDM5A*, and *ERBB4-IKZF2*) were identified in ALL cases with no other known driver alteration, deserving further studies. All these genes were already described as participating in fusion transcripts in the FusionGDB2 database (https://compbio.uth.edu/FusionGDB2/). However, to the best of our knowledge, the combinations found here have not yet been reported, except for *IKZF2-ERBB4* fusion that was found in T-cell lymphomas[49] and ovarian tumors.[50]

Importantly, some of the subgroup-defining fusion genes could not be found by fusion transcript analysis. Fortunately, the identification of these cases was still possible by means of analyzing gene-expression levels and read coverage. Patients with *IGH-CRLF2* or *IGH-EPOR* translocations express exceptionally high levels of *CRLF2* or *EPOR*, respectively. *NUTM1*-fusion cases also had high levels of expression of *NUTM1*, a gene that is normally not expressed in ALL (supplemental Figure 3F). Likewise, in comparison with other ALLs, *DUX4*-rearranged ALL showed a significantly higher pattern of read coverage along the *DUX4* gene cluster. The *DUX4* subgroup is enriched in *ERG* deregulation, that is, aberrant *ERG* expression, expression of an alternative splicing isoform of *ERG* (*ERG*alt), or *ERG* deletions.[30] ERGalt was identified in 27 of 35 cases of *DUX4*-rearranged ALL. Finally, *DUX4*-rearranged ALL formed an independent group in hierarchical clustering analysis, and both *DUX4*-rearranged and *ETV6-RUNX1* ALL were easily segregated in t-SNE analysis. HCA was tentatively used to identify Ph-like ALL but with ambiguous results. Because targeted RNA sequencing identified all the *ABL*-class and JAK/STAT-activating gene alterations, which is decisive in guiding the choice of an alternative drug use, there is probably no need for Ph-like classification.

Focal IKZF1del are frequently found in B-other ALL. Although IKZF1del does not define a separated subgroup of ALL, identification of IKZF1del is of high relevance as a prognostic factor, being associated with worse outcome.[51,52] Here, we showed that certain exon-exon junctions captured by targeted RNA sequencing

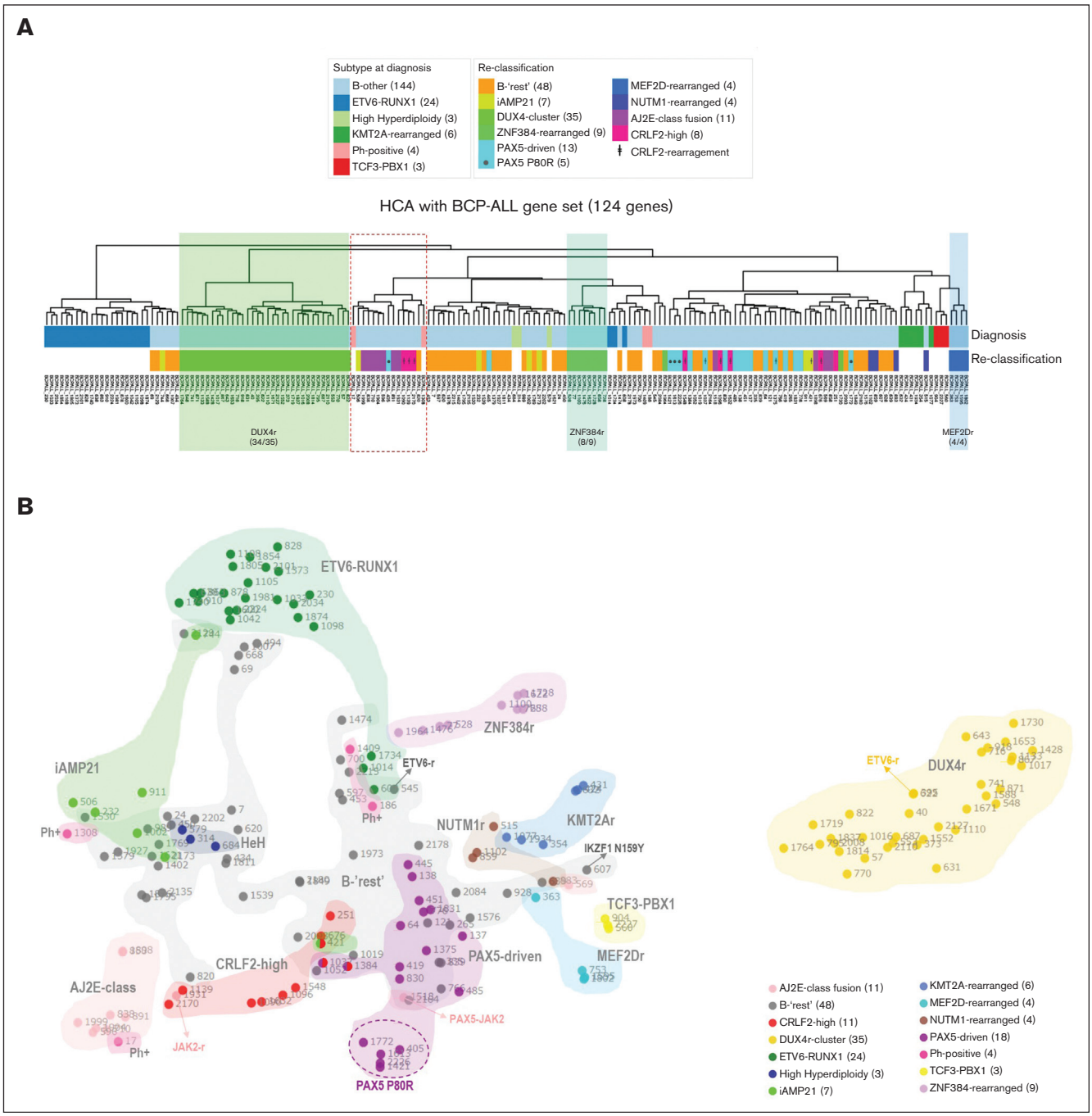

Figure 5. Unsupervised clustering analyses segregate specific subgroups of patients with shared gene-expression patterns. (A) HCA of the 144 B-other and 40 classical BCP-ALL samples based on the expression of 124 genes subtracted from a list of 864 genes selected by Liu et al (2016)[11] for ALL classification. Resulting groups/column dendrograms are shown (Euclidean and ward D). The list of genes is presented in supplemental Table 2. Genetic annotations are split into 3 lanes: the first lane shows the initial classification into the classical subtypes; the second lane shows the subgroup classification among B-other ALLs, as provided by the present study; and the third lane shows patients' identification. (B) Two-dimensional t-distributed stochastic neighbor embedding (t-SNE) plot of the 144 B-other and 40 classical BCP-ALL samples based on the 731 most variably expressed genes. Each dot represents a sample colored by subgroup. The 731 genes were selected and processed by the t-SNE algorithm with a perplexity score of 10. Of note, the 2 in-frame *ETV6*-fusion cases did not cluster together in gene-expression profiling analyses and thus were not classified in a separate subgroup of B-other ALL. AJ2E, group of *ABL*-, *JAK2*-, and *EPOR*-fusion cases.

allowed the discrimination of some intragenic IKZF1del using the Sashimi plot tool in the IGV browser. However, this method is not straightforward. Moreover, whole *IKZF1* gene deletions (exons 1-8), which account for one-third of IKZF1del cases could not be discriminated. Thus, Pan-Cancer RNA sequencing cannot be a substitute for current MLPA or microarray methods, which are

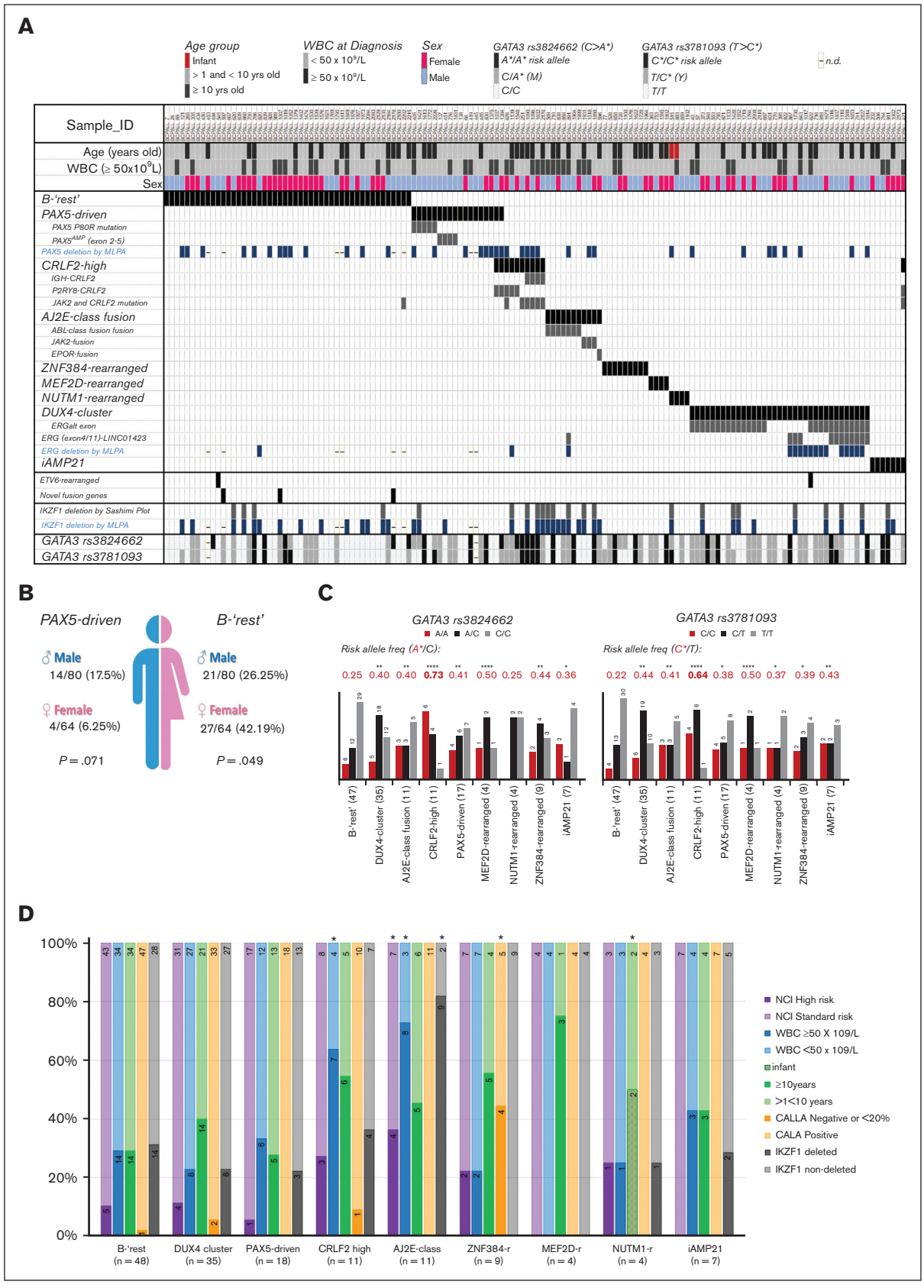
A
Age group
Infant
> 1 and < 10 yrs old
≥ 10 yrs old
WBC at Diagnosis
< 50 x 10⁹/L
≥ 50 x 10⁹/L
Sex
Female
Male
GATA3 rs3824662 (C>A*)
A*/A* risk allele
C/A* (M)
C/C
GATA3 rs3781093 (T>C*)
C*/C* risk allele
T/C* (Y)
T/T
n.d.
Sample_ID
Age (years old)
WBC (≥ 50x10⁹L)
Sex
B-'rest'
PAX5-driven
PAX5 P80R mutation
PAX5AMP (exon 2-5)
PAX5 deletion by MLPA
CRLF2-high
IGH-CRLF2
P2RY8-CRLF2
JAK2 and CRLF2 mutation
AJ2E-class fusion
ABL-class fusion fusion
JAK2-fusion
EPOR-fusion
ZNF384-rearranged
MEF2D-rearranged
NUTM1-rearranged
DUX4-cluster
ERGalt exon
ERG (exon4/11)-LINC01423
ERG deletion by MLPA
iAMP21
ETV6-rearranged
Novel fusion genes
IKZF1 deletion by Sashimi Plot
IKZF1 deletion by MLPA
GATA3 rs3824662
GATA3 rs3781093
B
PAX5-driven
B-'rest'
♂ Male
14/80 (17.5%)
♀ Female
4/64 (6.25%)
P = .071
♂ Male
21/80 (26.25%)
♀ Female
27/64 (42.19%)
P = .049
C
GATA3 rs3824662
A/A A/C C/C
Risk allele freq (A*/C):
0.25 0.40 0.40 0.73 0.41 0.50 0.25 0.44 0.36
GATA3 rs3781093
C/C C/T T/T
Risk allele freq (C*/T):
0.22 0.44 0.41 0.64 0.38 0.50 0.37 0.39 0.43
B-'rest' (47)
DUX4-cluster (35)
AJ2E-class fusion (11)
CRLF2-high (11)
PAX5-driven (17)
MEF2D-rearranged (4)
NUTM1-rearranged (4)
ZNF384-rearranged (9)
iAMP21 (7)
D
100%
80%
60%
40%
20%
0%
B-'rest (n = 48)
DUX4 cluster (n = 35)
PAX5-driven (n = 18)
CRLF2 high (n = 11)
AJ2E-class (n = 11)
ZNF384-r (n = 9)
MEF2D-r (n = 4)
NUTM1-r (n = 4)
iAMP21 (n = 7)
NCI High risk
NCI Standard risk
WBC ≥50 X 109/L
WBC <50 x 109/L
infant
≥10years
>1<10 years
CALLA Negative or <20%
CALA Positive
IKZF1 deleted
IKZF1 non-deleted

Figure 6.

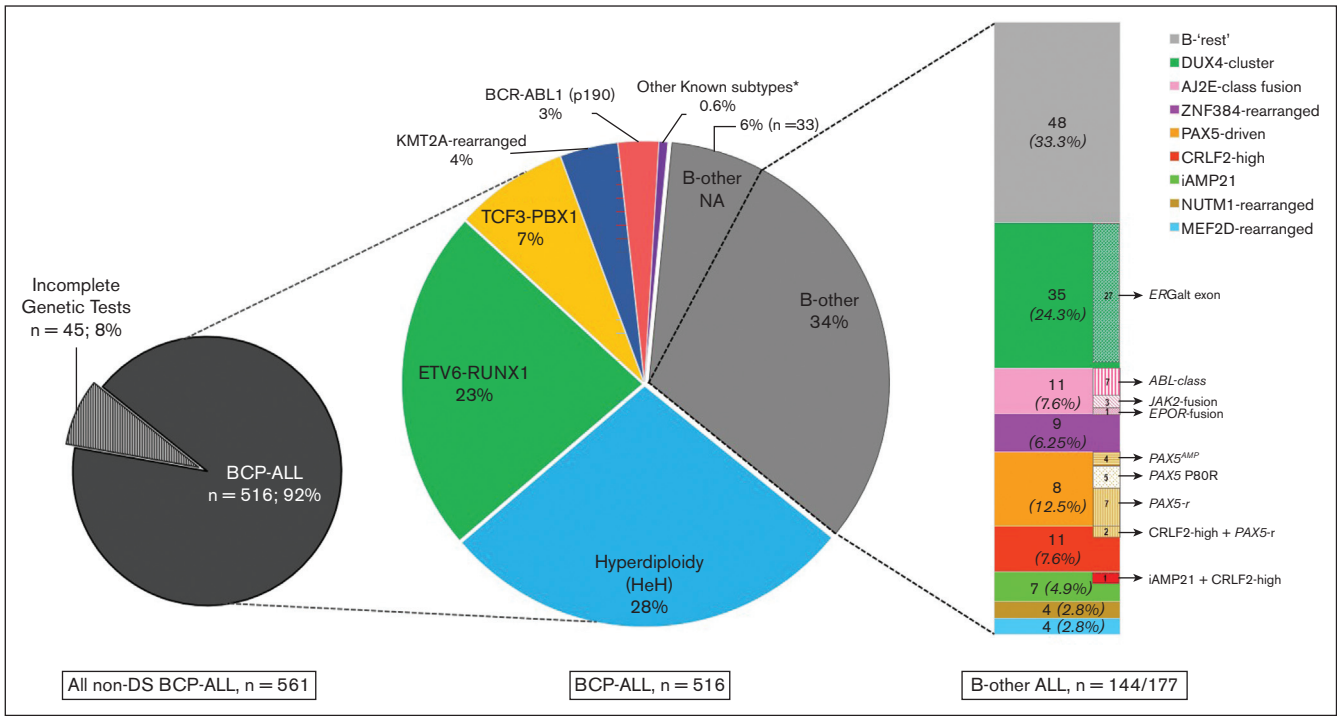

Figure 7. Frequency and distribution of BCP-ALL subtypes among all 516 patients with non-DS BCP-ALL registered in the GBTLI LLA-1999 and LLA-2009 in the Boldrini Children's Center. The pie chart depicts the percentages of the classical subtypes. Patients with incomplete genetic tests (n = 45) could not be classified. Of these, 177 (34%) cases were classified as B-other ALL, because they lacked t(12;21)(*ETV6-RUNX1*), t(1;19)(*TCF3-PBX1*), t(9:22)(*BCR-ABL1*), and *KMT2A*-rearrangements, and had ≥44 and <51 chromosomes (or a DNA index of <1.16). Not all B-other cases had samples available for the analyses (n = 33). RNA targeted sequencing was performed on 144 B-other and 40 classical ALL samples. The number of cases belonging to the novel molecular subgroups of B-other ALL are shown in the right bar graph. A comprehensive table with the genetic and demographic information on all the 184 cases is provided in the supplemental material (supplemental Table 8). AJ2E, group of *ABL*-, *JAK2*-, and *EPOR*-fusion cases.

necessary to analyze IKZF1del and other copy number alterations of prognostic significance.

Although B-other ALL is more frequent in males,[13] we found, to the best of our knowledge, for the first time, that PAX5-driven B-other ALL and B-rest ALL groups had a significantly higher proportion of males and females, respectively. We have no explanation for these gender biases. As expected, *ZNF384*-rearrangements were associated with lack of CD10[53] and *CRLF2*-high with an initial WBC count of ≥50 × 10^3/μL and *GATA3* risk alleles (rs3781093 and rs3824662).[22] The group of *ABL/JAK2/EPOR*-fusions was associated with high initial WBC counts, high risk per the NCI criteria, and IKZF1del, all of which are markers of poor prognosis.

In conclusion, Pan-Cancer RNA sequencing identified all the different subgroups of BCP-ALL, except for high-hyperdiploid, hypodiploid, iAMP21, and whole gene (*IKZF1, ERG,* and *PAX5*) deletions. Previous associations between clinico-biological features and the different subgroups of ALL were mostly confirmed, and a gender bias toward males in PAX5-driven ALL and females in B-rest ALL was discovered.

Acknowledgments

The authors thank Maria Alessandra Salgado and Eduardo William Rodrigues Amâncio for clinical data management. The authors thank Lilian Girotto Zambaldi and Daniele Ribeiro Lucon for the

Figure 6. Frequencies and associations of the molecular subgroups of B-other ALL. (A) Graphical overview of the molecular subgroups, demography, and clinical characteristics of the 144 B-other ALL cases. Subgroups were clustered based on the recurrent genetic abnormalities defined by fusion transcripts, gene-expression profile, and point mutations, as identified by using the TruSight Pan-Cancer targeted RNA sequencing. Each column represents a single patient. A single *PAX5-JAK2* was grouped with *JAK2*-rearranged cases. Three *CRLF2*-high cases also grouped with the *PAX5*-driven (2 cases) or iAMP21 (1 case) groups. *GATA3* analysis was performed in only 142 cases and MLPA in 136 of 144 B-other cases. (B) Gender biases observed in B-rest and the *PAX5*-driven subgroups of B-other ALL. (C and D) Associations between the molecular subgroups of B-other ALL and the different prognostic factors at diagnosis; (C) the *GATA3* risk alleles are C for rs3781093 and A for rs3824662. *GATA3* allele frequencies were compared with the frequencies of 1171 admixed individuals from São Paulo (Brazil) by the Fisher test (risk allele frequency for rs3781093 was 0.221; and for rs3824662 was 0.218), *$P < .05$; **$P < .005$; ****$P < .0001$; (D) stacked bar chart showing the frequency of each patient's demographic information by subgroup. Comparisons were made by the Fisher test followed by Monte Carlo correction, *$P < .05$ (supplemental Table 7). Numbers inside bars represent absolute numbers of cases. WBC, white blood cell; AJ2E, group of *ABL*-, *JAK2*-, and *EPOR*-fusion cases.

excellent technical help with cytogenetics analysis, and Guilherme Giusti Navarro for helping with bioinformatics data management.

This work was supported by grants to J.A.Y. from Institute Ronald McDonald (project IRM 2016109) and CNPq (process 429811/2016-0). J.A.Y. received a Productivity fellowship from the Brazilian National Counsel of Technological and Scientific Development (CNPq, process 301596/2017-4 and 308399/2021-8). K.B.M. received financial support through FAPESP (grants 14/21704-9 and 14/50897-0). N.A.M., V.S.V., and G.L.C. received scholarships from Brazilian funding agencies (CAPES or CNPq).

Authorship

Contribution: N.A.M. and J.A.Y. conceived and designed the study, performed the data interpretation, and wrote the manuscript; N.A.M. and P.Y.J. performed all targeted RNA-sequencing experiments; N.A.M. performed the validation assays by RT-PCR; P.Y.J., N.P.d.N., and G.L.C. contributed with *GATA3* polymorphisms analysis by Sanger sequencing; N.A.M., V.S.V., and K.B.M. contributed with bioinformatics analysis; A.C.d.A. and S.R.B. were responsible for patients' treatment and provided their demographic and clinical data; N.A.M. and J.A.Y. performed statistical analysis; and all authors revised the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: N.A.M., 0000-0002-1326-3841; P.Y.J., 0000-0001-9820-8210; N.P.d.N., 0009-0007-0511-2131; V.S.V., 0000-0002-6134-067X; G.L.C., 0000-0001-8892-2680; K.B.M., 0000-0001-6390-2560; A.C.d.A., 0000-0002-8293-6448; S.R.B., 0000-0003-3696-9852; J.A.Y., 0000-0002-1316-3525.

Correspondence: José Andrés Yunes, Centro Infantil Boldrini, Rua Gabriel Porto 1270, Campinas, SP 13083-210, Brazil; email: andres@boldrini.org.br.

References

1. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(2):83-103.
2. Schrappe M, Reiter A, Ludwig WD, et al. Improved outcome in childhood acute lymphoblastic leukemia despite reduced use of anthracyclines and cranial radiotherapy: results of trial ALL-BFM 90. German-Austrian-Swiss ALL-BFM Study Group. *Blood*. 2000;95(11):3310-3322.
3. Moorman AV, Ensor HM, Richards SM, et al. Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK medical research council ALL97/99 randomised trial [published correction appears in Lancet Oncol. 2010 Jun;11(6):516]. *Lancet Oncol*. 2010;11(5):429-438.
4. Hunger SP, Lu X, Devidas M, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the children's oncology group. *J Clin Oncol*. 2012;30(14):1663-1669.
5. Roberts KG, Morin RD, Zhang J, et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer Cell*. 2012;22(2):153-166.
6. Inaba H, Mullighan CG. Pediatric acute lymphoblastic leukemia. *Haematologica*. 2020;105(11):2524-2539.
7. Borowitz MJ, Devidas M, Hunger SP, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood*. 2008;111(12):5477-5485.
8. Schrappe M, Bleckmann K, Zimmermann M, et al. Reduced-intensity delayed intensification in standard-risk pediatric acute lymphoblastic leukemia defined by undetectable minimal residual disease: results of an international randomized trial (AIEOP-BFM ALL 2000). *J Clin Oncol*. 2018;36(3):244-253.
9. Hunger SP, Mullighan CG. Acute lymphoblastic leukemia in children. *N Engl J Med*. 2015;373(16):1541-1552.
10. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.
11. Liu YF, Wang BY, Zhang WN, et al. Genomic profiling of adult and pediatric b-cell acute lymphoblastic leukemia. *EBioMedicine*. 2016;8:173-183.
12. Gu Z, Churchman ML, Roberts KG, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet*. 2019;51(2):296-307.
13. Zaliova M, Stuchly J, Winkowska L, et al. Genomic landscape of pediatric B-other acute lymphoblastic leukemia in a consecutive European cohort. *Haematologica*. 2019;104(7):1396-1406.
14. Roberts KG, Li Y, Payne-Turner D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N Engl J Med*. 2014;371(11):1005-1015.
15. Tran TH, Loh ML. Ph-like acute lymphoblastic leukemia. *Hematology Am Soc Hematol Educ Program*. 2016;2016(1):561-566.
16. Inaba H, Azzato EM, Mullighan CG. Integration of next-generation sequencing to treat acute lymphoblastic leukemia with targetable lesions: the St. Jude Children's Research Hospital approach. *Front Pediatr*. 2017;5:258.
17. Brandalise SR, Pinheiro VR, Aguiar SS, et al. Benefits of the intermittent use of 6-mercaptopurine and methotrexate in maintenance treatment for low-risk acute lymphoblastic leukemia in children: randomized trial from the Brazilian Childhood Cooperative Group–protocol ALL-99. *J Clin Oncol*. 2010;28(11):1911-1918.
18. Silva KAS, Rechenmacher C, de Morais RV, Michalowski MB, Daudt LE. Are there regional variations in the presentation of childhood leukemia? *Clin Biomed Res*. 2021;41(3):192-198.

19. Tian L, Li Y, Edmonson MN, et al. CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biol*. 2020;21(1):126.
20. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26.
21. John CR, Watson D, Russ D, et al. M3C: Monte Carlo reference-based consensus clustering. *Sci Rep*. 2020;10(1):1816.
22. Perez-Andreu V, Roberts KG, Harvey RC, et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat Genet*. 2013;45(12):1494-1498.
23. Naslavsky MS, Scliar MO, Yamamoto GL, et al. author correction: whole-genome sequencing of 1171 elderly admixed individuals from Brazil. *Nat Commun*. 2022;13(1):1831.
24. Katz Y, Wang ET, Silterra J, et al. Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics*. 2015;31(14):2400-2402.
25. Harvey RC, Mullighan CG, Chen IM, et al. Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. *Blood*. 2010;115(26):5312-5321.
26. Russell LJ, Capasso M, Vater I, et al. Deregulated expression of cytokine receptor gene, CRLF2, is involved in lymphoid transformation in B-cell precursor acute lymphoblastic leukemia. *Blood*. 2009;114(13):2688-2698.
27. Yoda A, Yoda Y, Chiaretti S, et al. Functional screening identifies CRLF2 in precursor B-cell acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A*. 2010;107(1):252-257.
28. Roll JD, Reuther GW. CRLF2 and JAK2 in B-progenitor acute lymphoblastic leukemia: a novel association in oncogenesis. *Cancer Res*. 2010;70(19):7347-7352.
29. Lilljebjörn H, Fioretos T. New oncogenic subtypes in pediatric B-cell precursor acute lymphoblastic leukemia. *Blood*. 2017;130(12):1395-1401.
30. Zhang J, McCastlain K, Yoshihara H, et al. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat Genet*. 2016;48(12):1481-1489.
31. Hormann FM, Hoogkamer AQ, Beverloo HB, et al. NUTM1 is a recurrent fusion gene partner in B-cell precursor acute lymphoblastic leukemia associated with increased expression of genes on chromosome band 10p12.31-12.2. *Haematologica*. 2019;104(10):e455-e459.
32. Boer JM, Valsecchi MG, Hormann FM, et al. Favorable outcome of NUTM1-rearranged infant and pediatric B cell precursor acute lymphoblastic leukemia in a collaborative international study. *Leukemia*. 2021;35(10):2978-2982.
33. Treviño LR, Yang W, French D, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet*. 2009;41(9):1001-1005.
34. Raca G, Abdel-Azim H, Yue F, et al. Increased Incidence of IKZF1 deletions and IGH-CRLF2 translocations in B-ALL of Hispanic/Latino children–a novel health disparity. *Leukemia*. 2021;35(8):2399-2402.
35. Yang H, Zhang H, Luan Y, et al. Noncoding genetic variation in GATA3 increases acute lymphoblastic leukemia risk through local and global changes in chromatin conformation. *Nat Genet*. 2022;54(2):170-179.
36. Lilljebjörn H, Henningsson R, Hyrenius-Wittsten A, et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. *Nat Commun*. 2016;7:11790.
37. Yap KL, Furtado LV, Kiyotani K, et al. Diagnostic evaluation of RNA sequencing for the detection of genetic abnormalities associated with Ph-like acute lymphoblastic leukemia (ALL). *Leuk Lymphoma*. 2017;58(4):950-958.
38. Li JF, Dai YT, Lilljebjörn H, et al. Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases. *Proc Natl Acad Sci U S A*. 2018;115(50):E11711-E11720.
39. Grioni A, Fazio G, Rigamonti S, et al. A simple RNA target capture NGS strategy for fusion genes assessment in the diagnostics of pediatric b-cell acute lymphoblastic leukemia. *Hemasphere*. 2019;3(3):e250.
40. Ueno H, Yoshida K, Shiozawa Y, et al. Landscape of driver mutations and their clinical impacts in pediatric B-cell precursor acute lymphoblastic leukemia. *Blood Adv*. 2020;4(20):5165-5173.
41. Brown LM, Lonsdale A, Zhu A, et al. The application of RNA sequencing for the diagnosis and genomic classification of pediatric acute lymphoblastic leukemia. *Blood Adv*. 2020;4(5):930-942. Blood Adv. 2020;4(7):1217.
42. Hayette S, Grange B, Vallee M, et al. Performances of targeted RNA sequencing for the analysis of fusion transcripts, gene mutation, and expression in hematological malignancies. *Hemasphere*. 2021;5(2):e522.
43. Schieck M, Lentes J, Thomay K, et al. Implementation of RNA sequencing and array CGH in the diagnostic workflow of the AIEOP-BFM ALL 2017 trial on acute lymphoblastic leukemia. *Ann Hematol*. 2020;99(4):809-818.
44. Chouvarine P, Antić Ž, Lentes J, et al. Transcriptional and mutational profiling of b-other acute lymphoblastic leukemia for improved diagnostics. *Cancers*. 2021;13(22):5653.
45. Tran TH, Langlois S, Meloche C, et al. Whole-transcriptome analysis in acute lymphoblastic leukemia: a report from the DFCI ALL Consortium Protocol 16-001. *Blood Adv*. 2022;6(4):1329-1341.
46. Brady SW, Roberts KG, Gu Z, et al. The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat Genet*. 2022;54(9):1376-1389.
47. Kim B, Lee H, Shin S, Lee ST, Choi JR. Clinical evaluation of massively parallel RNA sequencing for detecting recurrent gene fusions in hematologic malignancies. *J Mol Diagn*. 2019;21(1):163-170.
48. Lee SHR, Antillon-Klussmann F, Pei D, et al. Association of genetic ancestry with the molecular subtypes and prognosis of childhood acute lymphoblastic leukemia. *JAMA Oncol*. 2022;8(3):354-363.

49. Boddicker RL, Razidlo GL, Dasari S, et al. Integrated mate-pair and RNA sequencing identifies novel, targetable gene fusions in peripheral T-cell lymphoma. *Blood*. 2016;128(9):1234-1245.

50. Papp E, Hallberg D, Konecny GE, et al. Integrated genomic, epigenomic, and expression analyses of ovarian cancer cell lines. *Cell Rep*. 2018;25(9):2617-2633.

51. Mullighan CG, Su X, Zhang J, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med*. 2009;360(5):470-480.

52. Stanulla M, Cavé H, Moorman AV. IKZF1 deletions in pediatric acute lymphoblastic leukemia: still a poor prognostic marker? *Blood*. 2020;135(4):252-260.

53. Hirabayashi S, Ohki K, Nakabayashi K, et al. ZNF384-related fusion genes define a subgroup of childhood B-cell precursor acute lymphoblastic leukemia with a characteristic immunotype. *Haematologica*. 2017;102(1):118-129.