Molecular Characterization of Pediatric Acute Lymphoblastic Leukemia via Integrative Transcriptomics: A Multicenter Study in Argentina

Authors: María Sol Ruiz[1,2], María Mercedes Abbate[1,2], Ezequiel Sosa[1,2], Daniel Avendaño[1,2], Ignacio Gomez Mercado[1,2], María Laura Lacreu[1,2], María Cecilia Riccheri[3], Virginia Schuttenberg[4], Luis Aversa[5,] Elba Vazquez[1,2], Geraldine Gueron[1,2], Javier Cotignola[1,2]

1 Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Química Biológica, Laboratorio de Inflamación y Cáncer, Buenos Aires, Argentina
2 CONICET - Universidad de Buenos Aires, Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), Buenos Aires, Argentina.
3 Hospital Nacional Posadas, Servicio de Pediatría, El Palomar, Buenos Aires, Argentina.
4 Hospital de Niños Sor María Ludovica, La Plata, Buenos Aires, Argentina.
5 Hospital de Niños Ricardo Gutiérrez, Ciudad Autónoma de Buenos Aires, Argentina.

Correspondence
Javier Cotignola (jcotignola@qb.fcen.uba.ar), María Sol Ruiz (ma.sol.ruiz@gmail.com)

Data availability statement
Transcriptome sequencing data is available at the European Nucleotide Archive repository under the Accession PRJEB80172. Variants, fusion transcripts, molecular subtypes, gene lists and primers are available upon reasonable request to the authors.

Funding statement
This work was supported by grants from ANPCyT-PICT 2020-00326 (Argentina), ANPCyT-PICT 2017-1186 (Argentina), Fundación para el Progreso de la Medicina 2017-2019 and Instituto Nacional del Cáncer 2016-2018 (Argentina). MSR, EV, GG, JC are researchers from the Consejo Nacional de Investigaciones Científicas y Tecnológicas of Argentina (CONICET). ES is a member of the professional support staff from CONICET. A received a PhD fellowship from CONICET. DA received an undergraduate fellowship from Universidad de Buenos Aires. IGM received an undergraduate fellowship from Consejo Interuniversitario Nacional. MLL received a postgraduate fellowship from ANPCyT.

Conflict of interest disclosure
All the authors declare no conflicts of interest to disclose.

Ethics approval statement
All procedures involving human participants complied with the ethical guidelines outlined by the Institutional Review Boards (IRB) of the participating centers and with the Declaration of Helsinki. Additionally, the study protocol received approval from Argentina's National Drug, Food and Technology Administration (ANMAT).

Patient consent statement
The inclusion criteria of the study were: 1) IRB protocol approval, 2) *de novo* primary ALL diagnosis, 3) patient age between 1 and 18 years old, 4) written informed consent from parents or legal guardians, with written informed assent from patients when applicable.

Acknowledgements
We thank members of the GATLA (Grupo Argentino de Tratamiento de Leucemias Agudas) working group, and all physicians, biochemists and patients involved in this study. We thank Santiago Di Lella, PhD., for his contributions in the structural analysis of novel variants and visualization of protein structures.

Abstract

Pediatric acute lymphoblastic leukemia (ALL) is a highly heterogeneous disease at the molecular level, with over 20 subtypes of B-cell ALL (B-ALL) identified to date. These subtypes are critical for guiding risk-adapted therapy and precision medicine, as they are defined by specific chromosomal rearrangements, gene expression profiles, aneuploidies and point mutations. However, the ability to accurately classify these subtypes is limited by the technical and economic challenges faced by many centers, particularly in low- and middle-income countries. In this study, we employed transcriptome sequencing to perform molecular classification of B-ALL at diagnosis in pediatric patients enrolled in the multicentric ALLIC-GATLA-2010 clinical protocol in Argentina. Using a combination of bioinformatic tools, we successfully identified single nucleotide variants, fusion transcripts and gene expression profiles, achieving molecular classification in over 90% of patients. Our analysis also revealed high-risk molecular features and novel genetic alterations. These findings hold potential clinical value for improving risk stratification and identifying therapeutic targets, particularly for patients who remain unclassified by conventional diagnostic methods.

Introduction

Acute lymphoblastic leukemia (ALL) is the most common childhood cancer globally. Advances in chemotherapy regimens and the stratification of patients into risk groups have led to steady improvements in overall and relapse-free survival. However, despite these advancements, 15-30% of patients still experience relapse, often accompanied by severe treatment complications and poor overall survival rates[1]. The further intensification of traditional chemotherapy is unlikely to yield better patient outcomes due to the associated acute toxicity and an unfavorable toxicity/benefit ratio[2,3]. Therefore, understanding the intrinsic and extrinsic factors that contribute to disease development and treatment failure is crucial.

Childhood ALL outcomes vary significantly worldwide, with survival rates exceeding 80% in most high-income countries, but as low as 20% in low- and middle-income countries[4,5,6]. In Argentina, ALL accounted for 29.3% of all pediatric cancers diagnosed between 2000 and 2019, with an average of 395 new cases annually and an age-standardized incidence rate of 4.93 per 100,000 individuals[7]. According to the Argentinian Oncopediatric Hospital Registry (ROHA) the five-year survival rate was 72.2% between 2005 and 2014[7]. Poorer outcomes in low- and middle-income countries are generally attributed to multiple factors, such as limited access to healthcare facilities, lack of standard-of-care diagnostics and restricted access to effective treatments. Moreover, recent evidence also suggests that ALL molecular subtypes and patient prognosis may be influenced by genetic ancestry[8]. However, how regional genetic and environmental factors affect treatment response and disease development remains poorly understood.

B-cell ALL (B-ALL) is the most frequent type of ALL, with over 20 subtypes identified to date. These subtypes guide risk-adapted therapies and lay the bases for precision medicine. They are distinguished by specific chromosomal alterations, gene expression profiles, aneuploidies and point mutations[9]. However, the ability to accurately identify each subtype is dependent on the availability of molecular diagnostic techniques, which vary across medical centers. In this context, the increasing complexity of molecular classification presents significant technical, analytical, and economic challenges, particularly in low- and middle-income countries, which can hinder the widespread implementation of these advanced diagnostic approaches[10].

The general aim of this study was to perform an integrated molecular characterization of childhood ALL in Argentine patients enrolled in the ALLIC-GATLA-2010 protocol. We used transcriptome sequencing (RNA-seq) on bone marrow aspirates to identify chromosomal rearrangements, small insertions/deletions (InDels) and single nucleotide variants (SNVs), and evaluated their association with clinical outcome. Additionally, an *in silico* analysis was conducted to further explore the novel genetic variants identified. The findings of this research offer valuable insights into the molecular heterogeneity of B-ALL and provide a robust framework for incorporating transcriptome sequencing into clinical diagnostics.

Methods

Patients and samples

Bone marrow aspirates (1-5mL) were obtained from newly diagnosed, untreated pediatric ALL patients enrolled in the national multicenter clinical protocol ALLIC-GATLA 2010 across three hospitals in Argentina: Hospital Nacional Posadas (n=8), Hospital de Niños Dr. Ricardo Gutierrez (n=5), and Hospital Sor María Ludovica (n=19). Clinico-pathological characteristics and disease outcome were evaluated by specialized oncohematologists. All procedures involving human participants complied with the ethical guidelines outlined by the Institutional Review Board (IRB) and with the Declaration of Helsinki. Additionally, the study protocol received approval from Argentina's National Drug, Food and Technology Administration (ANMAT).

The inclusion criteria were: 1) IRB protocol approval, 2) *de novo* primary ALL diagnosis, 3) patient age between 1 and 18 years old, 4) written informed consent from parents or legal guardians, with written informed assent from patients when applicable.

Transcriptome sequencing

Frozen cell pellets in Trizol or purified total RNA were shipped to the Inflammation and Cancer Laboratory at Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina. Total RNA was isolated using either Trizol (Thermo Fisher Scientific, USA) or Quick-Zol (Kallium technologies, Argentina) following the manufacturer's protocols at either the extraction site or in our laboratory. The RNA was then resuspended in RNAse/DNase-free distilled water and stored at -80°C. RNA quantification and integrity were evaluated using a Nanodrop (Thermo Fisher Scientific, USA) and Bioanalyzer (Agilent Technologies).

Transcriptome sequencing (Macrogen, Korea) was performed on 32 bone marrow aspirate samples collected at the time of diagnosis. RNA-seq libraries were prepared following ribosomal RNA depletion using Ribo-Zero Plus rRNA Depletion Kit (Illumina, USA). Paired-end RNA sequencing was conducted on HiSeq2000 or HiSeq2500 Systems (Illumina, USA), in five separate batches over the course of three years, with low quality samples excluded from the final analysis.

Bioinformatic methods

Sequencing quality was assessed using *FastQC* v0.11.5. We used *Kallisto*[11] to pseudoalign the reads to the GRCh38 reference genome and obtain transcript counts. Transcripts were annotated via the EnsDb.Hsapiens.v86 package (Bioconductor). SNVs and small insertions/deletions (InDels) were evaluated using the *RNAmut* software[12] based on a custom index (https://www.sanger.ac.uk/science/tools/rnamut/rnamut/#/) of 114 genes and 79 fusion genes frequently altered in ALL patients (Supplementary Material 1). SNVs/InDels were annotated and filtered using an in-house pipeline, applying thresholds for the number of reads (MutReads>=15), Variant Allele Fraction in RNA (VAF>=0.2) and frequency in population databases (Freq_1000genomes<1% or not available). Additional filtering was

performed based on clinical significance in ClinVar, association with Leukemia phenotype and pathogenicity predictions from Polyphen. RT-PCR and Sanger sequencing were used to confirm identified variants (see Supplementary Materials 1 and 2). The discovery of novel fusion genes was conducted using *STAR-Fusion (v1.11.0)* software[13], with further refinement of candidate fusions using the *FusionInspector* tool[14]. Both, *STAR-Fusion* and *FusionInspector* were implemented via *singularity* at the CCAD-UNC (Centro de Cómputo de Alto Desempeño, Universidad Nacional de Córdoba). Molecular subtypes were predicted from transcriptome sequencing data using the ALLsorts[15] classifier, which was deployed on the high-performance cluster at the Biological Chemistry department (School of Sciences, University of Buenos Aires). Following classification, subtypes "high sig", "high hyperdiploid" and "low hyperdiploid" were combined into a single hyperdiploid subtype.

Statistical analysis, data filtering, graphics

All statistical analysis, tables and graphics were generated using Rstudio (packages *tidyverse*, *gtsummary*, *plotly, survival, survminer*). The heatmap of variants was created using a script available in: https://github.com/sarahet/The_Distinct_DNA_Methylome_ALL/blob/main/Figure5_Extended_Data_Figure5.R. Circos plots were generated using the *chimeraviz* package in R.

Evaluation of variants and fusion transcripts by RT-PCR and Sanger sequencing

Total RNA was first treated with RNase-Free DNase (Promega) according to the manufacturer's protocol. Complementary DNAs (cDNAs) were synthesized using RevertAidTM First Strand cDNA Synthesis Kit (Thermo Scientific) following the manufacturer's guidelines with random primers. The cDNA was then used for PCR amplification, targeting specific regions of interest, including each variant or fusion transcript, with primers detailed in Supplementary Material 1. PCR amplification was conducted on a Veriti™ 96-well Thermal Cycler (Applied Biosystems); with cycling conditions optimized for each amplicon. PCR products were purified using a 2% agarose gel electrophoresis and extracted with the Monarch® DNA Gel Extraction Kit (New England Biolabs). Sequencing was performed with the forward and reverse primers using capillary electrophoresis on an ABI 3730xl system, provided by Macrogen.

Results

Patient cohort and transcriptome sequencing

The Acute Lymphoblastic Leukemia Intercontinental Consortium (ALLIC) congregates countries from Europe and Latin America and, originally, from Asia (Israel and Hong-Kong) with restricted resources, and has held two large clinical trials in childhood ALL[16]. Patients included in the current study were part of ALLIC BFM-2009 (Berlin-Frankfurt-Münster) through the multicentric protocol ALLIC-GATLA (Grupo Argentino de Tratamiento de la Leucemia Aguda) 2010-2020 that was carried out in Argentina. This protocol incorporated the minimal residual disease determined by multiparametric flow cytometry and evaluated the impact of early intensification and methotrexate dose on survival rates[17]. We recruited 32 patients newly diagnosed with ALL. Integrity of RNA was highly variable among hospitals, therefore this factor, together with sequencing batch, was taken into account in statistical models used for comparisons among samples. Patients were stratified according to age, white blood cell number at diagnosis, number of blasts at day 8, MRD at day 15 and 33 by flow cytometry, hypodiploidy, t(9;22) and t(4;11) (TABLE 1). Median age at diagnosis was 7 years (range 2-18 years); sex distribution was 47% boys and 53% girls. The cohort comprised 3 standard-risk (SR, 9%), 22 intermediate-risk (IR, 69%) and 7 high-risk (HR,

22%) patients. Median follow-up time was 32 months (TABLE 1). Overall survival (OS) and Event-Free Survival (EFS) at 36 months were 84.2% [Standard Error=6.5%] and 75.6% [Standard Error=8.3%], respectively (FIG. 1).

Molecular Subtyping of B-ALL via an RNA-Seq-based Classifier

To identify molecular subtypes of B-ALL, we used ALLSorts classifier, a machine learning algorithm designed to categorize samples into 18 predefined molecular subtypes using whole transcriptome sequencing data[15]. The ALLSorts algorithm assigned a molecular subtype to 30 out of 32 samples (93.75%) (FIG. 2), and demonstrated a high concordance rate with conventional subtyping methods such as karyotyping and RT-PCR (17 out of 18 samples with available genetic data; 94%). It is important to note that current molecular guidelines for ALL subtyping mandate testing for 5 of the 18 subgroups characterized by ALLSorts; therefore, we were unable to confirm ALLSorts classification for all subtypes. We found a discordant result in one sample classified as DUX4 by ALLSorts and hyperdiploid by traditional karyotyping. The misclassified sample was predicted as DUX4 but had a hyperdiploid karyotype, so we classified it as DUX4/Hyperdiploid. Taking these caveats into account, we could identify 9 different subtypes, being hyperdiploid, DUX4 and *ETV6::RUNX1* the most frequent in this cohort (FIG. 2), providing a molecular classification to over 90% of samples. In contrast, traditional methods used in this protocol evaluated a limited number of molecular features, therefore classification only a 35% of samples (TABLE 2). Frequencies and age distribution were compared to those reported by Brady *et al* for the St. Jude cohort[9] , obtaining similar results; though a proper statistical analysis could not be applied due to the low number of patients in each category (TABLE 2). Finally, we evaluated an alternative molecular subtype classifier, ALLCatchR[18], that requires fewer computing resources. Considering “high confidence” and “candidate” categories from AllCatchR (which excludes “unclassified” samples), we found 82.1% agreement with ALLsorts (23 out of 28 samples), whereas evaluation of only “high confidence” predictions resulted in 100% agreement (19 out of 19 samples) (Supplementary Material 1). Four out of 32 samples (12.5%) remained unclassified by ALLCatchR.

Identification of fusion transcripts

Bioinformatic methods used to discover fusion genes through RNA-seq data are usually based on identifying reads aligned to two different genes (spanning chimeric junctions) or detecting split reads into two genes in paired-end sequencing data. Traditional methods like *STAR-Fusion*[13], Arriba[19], Fusion Catcher[20], or CICERO[21], initially perform a whole-genome alignment, requiring significantly large computational resources and processing time. In contrast, methods such as *RNAmut*[12], which rely on a predefined list of candidate genes as input, align reads to the reference sequence of the candidate genes, significantly reducing processing time and computational demands, but limiting the identification of novel fusion transcripts. With the aim of comparing the capabilities, sensitivity and speed of different bioinformatic approaches, we searched for fusion genes in transcriptomic data using *STAR-Fusion*, reported to be one of the methods with highest accuracy for this task[22], and *RNAmut* through a custom gene index built based on bibliography. *STAR-Fusion* identified a total of 82 chimeric transcripts: 19 (23%) interchromosomal and 63 (77%) intrachromosomal (Supplementary Material 1). The intergenic distance for intrachromosomal fusions varied between <100 kb and >1 Mb and their presence was suggestive of intrachromosomal deletions. Eight out of 82 fusion transcripts were previously reported in ALL patients and associated with molecular subtypes of prognostic significance[23] (*EP300::ZNF384, ETV6::RUNX1, PAX5::ETV6, TCF3::FLI1, TCF3::PBX1, P2RY8::CRLF2, DUX4::IGH,*

MEF2D::BCL9) (FIG. 2). Additionally, given the association between *CRLF2* rearrangements and Hispanic ancestry[24], we evaluated the incidence of *P2RY8::CRLF2* in additional samples by RT- PCR. We detected the fusion transcript in 4 out of 56 patients (7%), being the group of patients harboring *P2RY8::CRLF2* enriched in relapsed patients (50% vs 13%, Supplementary Material 2). There was no association with other clinical variables (p>0.05); however, the small number of patients may have hindered small size effects.

For RNAmut we built an index of fusion transcripts based on literature review and STAR-Fusion results. It is important to note that this index cannot include non-coding transcripts, transcripts without Consensus CDS, and variable regions such as *IGH* genes, resulting in a final list of 101 candidate fusion genes. Agreement with STAR-Fusion was low (8/37, 21.6%), but for gene fusions frequently found in ALL patients, agreement increased to 85.7%. Differences between methods resulted in RNAmut unable to detect, in our study, *MEF2D::BCL9,* and unable to evaluate *DUX4::IGH* because it could not be included in the index. On the other hand, STAR-Fusion was not able to detect *ETV6::RUNX1* in 2 out of 3 samples. To explore whether this was a limitation of our approach, we analyzed results reported in a recent study of childhood cancer that included 908 B-ALL patients[25], in which we estimated a false negative rate for *ETV6::RUNX1* detection of 9.6% (STAR-Fusion), 6.4% (CICERO), 5.6% (Arriba) and 0% (Fusion Catcher). Finally, two novel fusion transcripts were detected by both methods: *ABHD17B::CEMIP2* and *DBF4B::EFTUD2.*

Thirteen fusion transcripts involving long non-coding RNAs were detected by STAR-Fusion. Of these, only *ERG::LINC01423* was highly expressed (65 junction reads) and breakpoint sites predicted a 112,000 bp deletion in chromosome 21 leading to partial loss of *ERG* gene (exons 2 to 10, NM_182918.4) (Supplementary Material 2). In this patient, ALLsorts predicted a DUX4 subtype, which has been associated with *ERG* deletions (HLP016, FIG. 2)[26]. Accordingly, *ERG::LINC01423* has been previously reported in DUX4 subtype B-ALL patients[27]. Finally, fusions not detected by RNAmut were manually reviewed. Apart from *MEF2D::BCL9*, *SCAF8::FER1L4* was the only inter-chromosomal fusion transcript with more than 10 junction reads included in this group. This is a novel, highly expressed (16 junction reads), interchromosomal fusion transcript involving *SCAF8* and pseudogene *FER1L4*; *SCAF8::FER1L4* expression was confirmed by RT-PCR followed by Sanger sequencing (FIG. 3A). This patient (HNG008) of 14 years of age was classified as intermediate risk group, ALLsorts predicted a ETV6::RUNX1-like subtype, harbored a *TCF3::FLI1* fusion transcript and overexpressed *FER1L4* compared to samples lacking *SCAF8::FER1L4*. On the other hand, *SCAF8* was highly expressed in all the samples included in this study (FIG.3B). Regarding novel fusions, *DBF4B::EFTUD2* would arise from an intrachromosomal fusion at a distance of ~110,000bp in chromosome 17, joining *DBF4B* exon 3 to *EFTUD2* exon 10 (FIG. 3C), as an in-frame fusion. This fusion transcript was identified in a 2-year-old patient (HNG007) with a hyperdiploid molecular subtype classified as intermediate risk. This patient also harbored a C>G substitution at the RNA level in *CSF3R* gene (NP_000751.1:p.Gly147Arg), described in the following section. mRNA levels of *DBF4B* suggest absence or loss of expression of DBF4B non-translocated allele (FIG. 3D). *ABHD17B::CEMIP2* would arise from an intrachromosomal fusion at a distance of ~150,000bp in chromosome 9. This fusion transcript was identified in 5 patients but only by both methods in one patient (POS032), and with low abundance (6 junction reads).

Identification of SNVs and small InDels by transcriptome analysis.

In order to identify clinically relevant somatic variants through transcriptomic data, we applied *RNAmut* software which is of interest by its feasibility of implementation in a clinical center, due to its low computational requirements and high processing speed. For this, we

designed a custom index of 114 genes commonly mutated in pediatric ALL (Supplementary Material 1). We found a total of 9,594 variants among all samples. After annotation, filtering and exclusion of patients with incomplete follow-up or low-quality samples, we identified 20 different SNVs/InDels in 16 genes in 14/32 patients (42%) (FIG. 2). According to OncoKB classification, we identified 4 oncogenic (*NSD2* E1099K, *PTPN11* D61Y, *FLT3* N676K, and *FLT3* D835E), 7 likely oncogenic (*ATM* P604S in two patients, *NRAS* G12S, *TP53* P152L, *KRAS* Q61P, *TP53* R267P, *NF1* R1306*, *TCF3* S338fs*10), and 8 unclassified variants (*PAX5* P34L, *ZEB2* H1038R, *PAX5* R38C, *IKZF1* D186Y, *CREBBP* G1542V, *DUX4* I65N, *ETV6* E12_DEL:Q, *CSF3R* G147R). 19 out of 20 variants were confirmed by an independent bioinformatic analysis of whole-transcriptome SNV detection (CTAT-Mutations[28], Supplementary Material 1). Four variants were not reported in ALL patients or other pathologies and were further characterized by *in silico* tools. Three of these variants (DUX4 I65N, CREBBP G1542V, CSF3R G147R) were predicted to affect protein function and were located at highly conserved residues (TABLE 3, FIG. 4). The novel variant found in DUX4 is located in homeodomain 1, and the residue change is compatible with alterations in the interactions between DUX4 and target DNA (FIG. 4A). Given that DUX4 is a transcription factor, minor changes in the interaction between DUX4 and its target DNA sequence could have consequences in DUX4 regulatory functions. The novel variant found in CREBBP protein is located in the histone acetyltransferase domain, in which a great number of variants have been found in relapsed ALL patients[29]. The location of this residue in a disordered region might compromise its stability or interactions with other molecules (FIG. 4B). Regarding the novel variant found in CSF3R, despite being located in the extracellular region of the receptor, it is not predicted to directly affect the interaction with the ligand CSF3. However, the mutation observed may be responsible for destabilization of the loop containing the residue, or the interaction with other molecules, since a positive-charged residue in the loop provides new possible intermolecular interactions; or a combination of both (FIG. 4C).

Integration of multiple data types for characterization of molecular subtypes and novel variants.

ALLIC-GATLA-2010 protocol included the mandatory evaluation of the fusion transcripts *ETV6::RUNX1*, *BCR::ABL1*, *KMT2A::AFF1,* and karyotyping by conventional cytogenetics. Therefore, it was not possible to confirm by an independent study the accuracy of the bioinformatic predictions in the remaining molecular subtypes identified by ALLsorts. To provide additional support to the predicted subtypes, we integrated fusion transcripts, SNVs/InDels and gene expression analysis from transcriptomic sequencing. Regarding **ZNF384-alt** subtype**,** the patient identified by ALLsorts (POS007) harbored a *EP300::ZNF384* fusion transcript. Gene expression analysis was concordant with the expected immunophenotype for this subtype: low CD10 and aberrant CD33[30] (FIG. 5A).

Regarding **MEF2D subtype**, the patient harbored a *MEF2D9::BCL9* fusion transcript and increased levels of *HDAC9,* as expected in this subtype[31] (FIG. 5B). Under the criteria of the ALLIC-BFM protocol, this patient was classified as intermediate risk. Despite a good clinical response at day 8 and 15, the patient relapsed at 11 months and died at 16 months, which is concordant with the poor prognosis of the MEF2D subtype[32].

Regarding the DUX4 subtype, we could identify the canonical fusion transcript *DUX4::IGH* in only 1 out of 5 patients (POS003). In accordance, this patient showed overexpression of *DUX4* mRNA and the highest expression among all samples of *DUX4::IGH* direct transcriptional targets[33], *CCNJ* and *ITGA6* (FIG. 5C), and mutations in *ZEB2*, *TP53* and

PTPN11 (FIG. 2). Surprisingly, despite the absence of reports of point mutations in *DUX4* in ALL, we identified 1 sample (HLP011) harboring a *DUX4* I65N variant (TABLE 3) at the RNA level, located in a homeodomain coding region (FIG. 4), overexpression of *DUX4* mRNA, and normal levels of *CCNJ* and *ITGA6* (FIG. 5C), suggesting a different expression profile compared to *DUX4::IGH* samples. A third patient (HLP016) harbored an intrachromosomal fusion transcript involving *ERG*, suggesting a 112,000 bp deletion in chromosome 21 leading to partial *ERG* loss, frequently found in this subtype[26]. In the remaining samples, no canonical *DUX4* alterations were found by the methods applied in this study. However, expression of CD371 marker, previously reported as enriched in the DUX4 subtype and associated to lineage switch[34] was increased in all DUX4 samples compared to most non-DUX4 samples (FIG. 5C).

Association of molecular features with clinical variables

We found an association between the molecular subtype, relapse ($p<0.05$, TABLE 4), and minimal residual disease at day 15 ($p<0.05$, TABLE 5). Interestingly, 3 out of 5 patients that relapsed were classified as intermediate risk, had <10% Minimal Residual Disease at day 15 and were classified as B-other subtype according to the clinical protocol guidelines. However, this study identified high risk molecular features in 2 out of 3 patients: NSD2_E1099K and *MEF2D::BCL9* fusion transcript[32,35]. It is of interest to assess in future studies if the evaluation of these alterations at diagnosis may be of clinical utility for early detection of relapses or identification of candidates for clinical trials[36]. No significant association was detected between molecular subtype and death, response to prednisone at day 8, severe acute toxicity related to treatment, sex or age ($p>0.05$), though the small number of patients in each group may have hindered minor size effects.

Discussion

Therapeutic response rates are highly variable and tend to be lower in patients from low- and middle-income countries[4,5]. According to the US SEER Program, the age-adjusted incidence rate of ALL patients in 2014 was higher for “Latinos” compared to Non-Hispanic whites[37]. However, studying the genetic components of Latin American populations presents challenges due to the high heterogeneity in ancestry across different regions. For instance, in Argentina, a study revealed a predominantly Caucasian ancestry with some admixture of Amerindian and African ancestry[38].
Regarding subtype distribution, a recent research involving over 2,000 patients from different genetic ancestry reported that Native American and African ancestries were associated with poorer prognosis, even after accounting for molecular subtype and clinical features[8]. In comparison to a cohort from StJude, the patients in our study exhibited a similar distribution of subtypes. However, we observed a marginally lower proportion of the favorable prognosis subtype ETV6::RUNX1^{+} and an increased occurrence of the variable prognosis subtype DUX4[9]. Expanding the characterization of patients in this region is contingent on the technical feasibility and financial resources available at each center. In this work we evaluated the performance of various bioinformatic tools that facilitate molecular classification using raw RNA-seq data. Notably, both RNAmut and ALLCatchR can be applied without requiring high computing resources.

For the first time in pediatric ALL we utilized the *RNAmut* tool, which was originally designed to detect mutations in Acute Myeloid Leukemia. To support this, we developed a custom gene index aimed at identifying clinically relevant SNVs/InDels and gene fusions. By

combining RNAmut with *STAR-Fusion* and CTAT-mutations, we were able to validate the results and efficiently identify most known clinically significant fusion transcripts and SNVs, without requiring advanced programming knowledge or access to supercomputers. The custom index is now available for testing in other patient cohorts. Gene expression analysis further provided supporting evidence for the molecular subtypes identified in this study, confirming *CRLF2* overexpression in Ph-like P2RY8-CRLF2, *HDAC* overexpression in MEF2D, *CD33/CD10* deregulation in ZNF384, and *DUX4/CD371/ERG* deregulation in DUX4. However, neither of the fusion gene detection tools used in this study achieved a 100% detection rate, suggesting that additional analyses would be needed for samples that test negative. In a recent work by Vicente Garces et al, authors compared five fusion calling pipelines, including STAR Fusion, and found that none achieved perfect sensitivity and precision[22], implying that a combination of multiple tools is necessary, and indicating that there is still room for improvement in the individual fusion callers.

This study identified three novel SNV variants in the *CREBBP*, *DUX4* and *CSF3R* genes. *CREBBP* mutations have been previously associated with a highe incidence of relapse in ALL patients[29]. The patient harboring this variant also carried a truncating oncogenic mutation in *NF1*, was classified as intermediate risk and relapsed at 33 months. Truncating, activating mutations in *CSF3R* have been recognized as drivers in chronic neutrophilic leukemia and atypical chronic myeloid leukemia[39], although reports of *CSF3R* mutations in B-ALL are rare. This gene holds clinical relevance, as mutations in its intracellular domain have been linked to specific therapeutic vulnerabilities including sensitivity to dasatinib and resistance to ruxolitinib[40]. As for DUX4, the overexpression of full length or C-truncated protein can be toxic to cells[41]. The variant identified in this study is predicted to alter DNA binding, which could explain the patient's tolerance to the high levels of *DUX4* mRNA observed. However, alternative interactions or reduced affinity interactions may also be occurring, as evidenced by the deregulation of *CD371* and *ERG* in this patient.

This is the first comprehensive analysis of the molecular landscape and its association with clinical parameters in pediatric ALL patients from Argentina, treated under the ALLIC-GATLA 2010 protocol. Although the study was limited by the inability to detect copy-number alterations, transcriptomic sequencing enabled the analysis and integration of multiple molecular features. These insights have significant clinical utility, offering potential improvements in risk stratification and the identification of therapeutic targets. Furthermore, these approaches are practical for complementing routine diagnostics, especially in cases where the molecular subtype cannot be determined using traditional methods.

Acknowledgements

We thank members of the GATLA (Grupo Argentino de Tratamiento de Leucemias Agudas) working group, and all physicians, biochemists and patients involved in this study. We thank Santiago Di Lella, PhD., for his contributions in the structural analysis of novel variants and visualization of protein structures.

References

1. Oskarsson, T. *et al.* Treatment-related mortality in relapsed childhood acute lymphoblastic leukemia. *Pediatr. Blood Cancer* **65**, (2018).
2. McNeer, J. L., Rau, R. E., Gupta, S., Maude, S. L. & O'Brien, M. M. Cutting to the Front

of the Line: Immunotherapy for Childhood Acute Lymphoblastic Leukemia. *Am. Soc. Clin. Oncol. Educ. Book* e132–e143 (2020) doi:10.1200/EDBK_278171.
3. Tallen, G. *et al.* Long-Term Outcome in Children With Relapsed Acute Lymphoblastic Leukemia After Time-Point and Site-of-Relapse Stratification and Intensified Short-Course Multidrug Chemotherapy: Results of Trial ALL-REZ BFM 90. *J. Clin. Oncol.* **28**, 2339–2347 (2010).
4. Allemani, C. *et al.* Global surveillance of trends in cancer survival: analysis of individual records for 37,513,025 patients diagnosed with one of 18 cancers during 2000–2014 from 322 population-based registries in 71 countries (CONCORD-3). *Lancet Lond. Engl.* **391**, 1023–1075 (2018).
5. Chatenoud, L. *et al.* Childhood cancer mortality in America, Asia, and Oceania, 1970 through 2007. *Cancer* **116**, 5063–5074 (2010).
6. Pui, C.-H., Yang, J. J., Bhakta, N. & Rodriguez-Galindo, C. Global efforts toward the cure of childhood acute lymphoblastic leukemia. *Lancet Child Adolesc. Health* **2**, 440–454 (2018).
7. Moreno, F. & Chaplin, M. A. Registro Oncopediátrico Argentino -7a ed. - Ciudad autónoma de Buenos Aires: Instituto Nacional del Cáncer, 2021.
8. Lee, S. H. R. *et al.* Association of Genetic Ancestry With the Molecular Subtypes and Prognosis of Childhood Acute Lymphoblastic Leukemia. *JAMA Oncol.* **8**, 354–363 (2022).
9. Brady, S. W. *et al.* The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat. Genet.* **54**, 1376–1389 (2022).
10. Jerez, J. *et al.* Epidemiological and clinical characteristics of adult acute lymphoblastic leukemia patients in Chile: A single-center analysis. *Leuk. Res. Rep.* **21**, 100405 (2024).
11. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
12. Gu, M. *et al.* RNAmut: robust identification of somatic mutations in acute myeloid leukemia using RNA-sequencing. *Haematologica* **105**, e290–e293 (2020).
13. Haas, B. J. *et al.* Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* **20**, 213 (2019).
14. Haas, B. J. *et al.* Targeted in silico characterization of fusion transcripts in tumor and normal tissues via FusionInspector. *Cell Rep. Methods* **3**, 100467 (2023).
15. Schmidt, B. *et al.* ALLSorts: an RNA-Seq subtype classifier for B-cell acute lymphoblastic leukemia. *Blood Adv.* **6**, 4093–4097 (2022).
16. Ribeiro, R. C. & Conter, V. Optimizing Pediatric Leukemia Care in Countries With Limited Resources. *J. Clin. Oncol.* **41**, 3482–3485 (2023).
17. Campbell, M. *et al.* Childhood Acute Lymphoblastic Leukemia: Results of the Randomized Acute Lymphoblastic Leukemia Intercontinental-Berlin-Frankfurt-Münster 2009 Trial. *J. Clin. Oncol.* **41**, 3499–3511 (2023).
18. Beder, T. *et al.* The Gene Expression Classifier ALLCatchR Identifies B-cell Precursor ALL Subtypes and Underlying Developmental Trajectories Across Age. *HemaSphere* **7**, e939 (2023).
19. Uhrig, S. *et al.* Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res.* **31**, 448–460 (2021).
20. Nicorici, D. *et al.* FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. 011650 Preprint at https://doi.org/10.1101/011650 (2014).
21. Tian, L. *et al.* CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biol.* **21**, 126 (2020).
22. Vicente-Garcés, C. *et al.* Fusion InPipe, an integrative pipeline for gene fusion detection from RNA-seq data in acute pediatric leukemia. *Front. Mol. Biosci.* **10**, 1141310 (2023).
23. Gu, Z. *et al.* PAX5-driven Subtypes of B-progenitor Acute Lymphoblastic Leukemia. *Nat. Genet.* **51**, 296–307 (2019).
24. Harvey, R. C. *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* **115**, 5312–5321 (2010).

25. Liu, Y. *et al.* Etiology of oncogenic fusions in 5,190 childhood cancers and its clinical and therapeutic implication. *Nat. Commun.* **14**, 1739 (2023).
26. Zhang, J. *et al.* Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat. Genet.* **48**, 1481–1489 (2016).
27. Migita, N. A. *et al.* Classification and genetics of pediatric B-other acute lymphoblastic leukemia by targeted RNA sequencing. *Blood Adv.* **7**, 2957–2971 (2023).
28. https://github.com/NCIP/ctat-mutations.
29. Mullighan, C. G. *et al.* CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature* **471**, 235–239 (2011).
30. Wang, Y. *et al.* Immunophenotypic characteristics of ZNF384 rearrangement compared with BCR-ABL1, KMT2A rearrangement, and other adult B-cell precursor acute lymphoblastic leukemia. *Cytometry B Clin. Cytom.* **102**, 360–369 (2022).
31. Ohki, K. *et al.* Clinical and molecular characteristics of MEF2D fusion-positive B-cell precursor acute lymphoblastic leukemia in childhood, including a novel translocation resulting in MEF2D-HNRNPH1 gene fusion. *Haematologica* **104**, 128–137 (2019).
32. Gu, Z. *et al.* Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. *Nat. Commun.* **7**, 13331 (2016).
33. Tanaka, Y. *et al.* Transcriptional activities of DUX4 fusions in B-cell acute lymphoblastic leukemia. *Haematologica* **103**, e522–e526 (2018).
34. Buldini, B. *et al.* CD371+ pediatric B-cell acute lymphoblastic leukemia: propensity to lineage switch and slow early response to treatment. *Blood* blood.2023021952 (2024) doi:10.1182/blood.2023021952.
35. Pierro, J. *et al.* The NSD2 p.E1099K Mutation Is Enriched at Relapse and Confers Drug Resistance in a Cell Context–Dependent Manner in Pediatric Acute Lymphoblastic Leukemia. *Mol. Cancer Res.* **18**, 1153–1165 (2020).
36. Li, J. *et al.* PRC2 Inhibitors Overcome Glucocorticoid Resistance Driven by NSD2 Mutation in Pediatric Acute Lymphoblastic Leukemia. *Cancer Discov.* **12**, 186–203 (2022).
37. Quiroz, E. *et al.* The emerging story of acute lymphoblastic leukemia among the Latin American population – biological and clinical implications. *Blood Rev.* **33**, 98–105 (2019).
38. Avena, S. *et al.* Heterogeneity in Genetic Admixture across Different Regions of Argentina. *PLOS ONE* **7**, e34695 (2012).
39. Maxson, J. E. *et al.* Oncogenic CSF3R mutations in chronic neutrophilic leukemia and atypical CML. *N. Engl. J. Med.* **368**, 1781–1790 (2013).
40. Schwartz, M. S. & Wieduwilt, M. J. CSF3R truncation mutations in a patient with B-cell acute lymphoblastic leukemia and a favorable response to chemotherapy plus dasatinib. *Leuk. Res. Rep.* **14**, 100208 (2020).
41. Tian, L. *et al.* Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia. *Nat. Commun.* **10**, 2789 (2019).

FIG. 1

Survival analysis of patients included in this study, as part of the ALLIC-GATLA-2010 protocol. Event was defined as relapse or death.

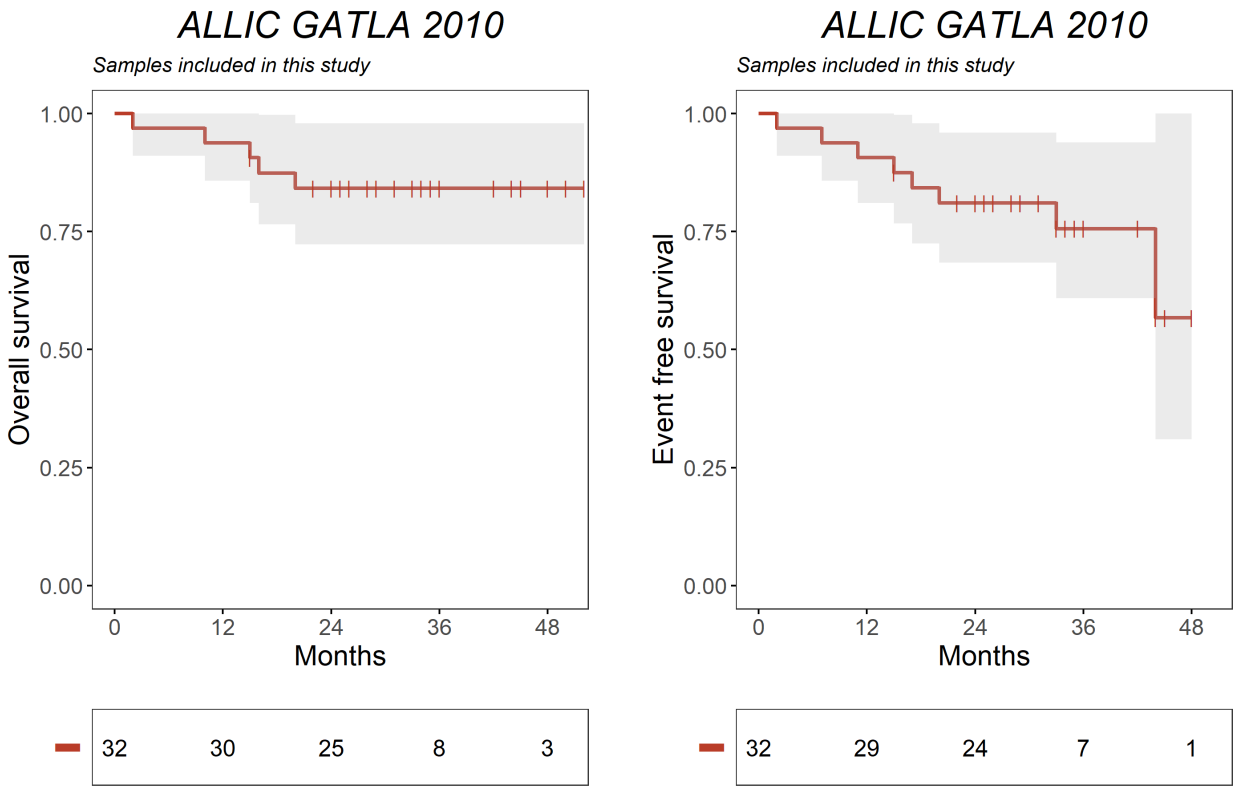


FIG. 2

Molecular alterations identified in this study. For each patient (column), SNVs/Indels (in orange, blue, green) and fusion transcripts (in yellow) are shown, ordered by frequency in the cohort. Molecular subtype and clinical features are shown in the lower rows. MRDCat=minimal residual disease measured at day 15 by flow cytometry.

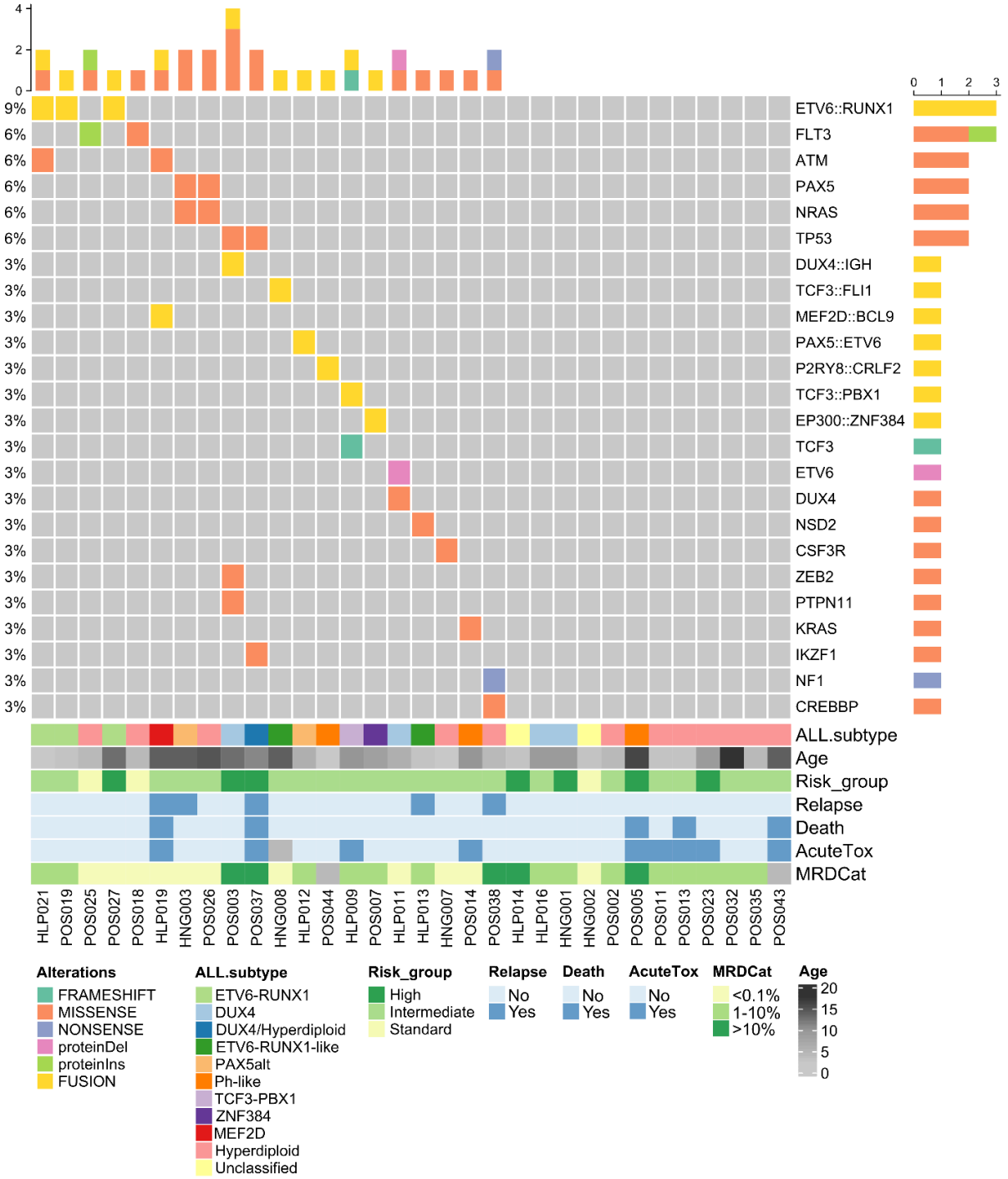
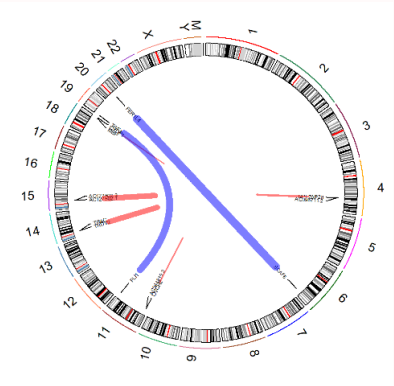


FIG. 3

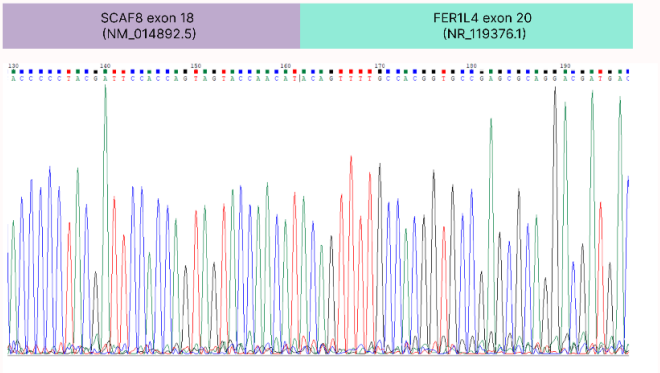
Novel fusion transcripts identified by RNA-seq and confirmed by Sanger sequencing. *SCAF8::FER1L4* was detected (16 junction reads and 1 spanning read) in a patient with subtype *ETV6::RUNX1*-like, harboring a *TCF3::FLI* fusion transcript (A). *FER1L4* mRNA levels were increased whereas *SCAF8* mRNA levels were not altered in this patient compared to those not harboring *SCAF8::FER1L4* (B). *DBF4B::EFTUD2* (7 junction reads and 1 spanning read) was detected in a patient with subtype hyperdiploid (C). *DBF4B* mRNA levels suggest a loss of the expression of the non-translocated DBF4B gene, whereas *EFTUD2* mRNA was highly abundant in all samples (D).

A

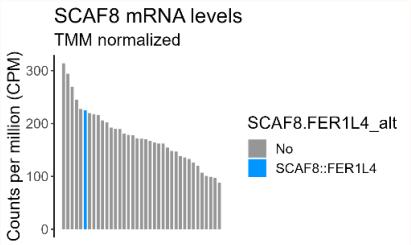
HNG008, *ETV6::RUNX1 like*

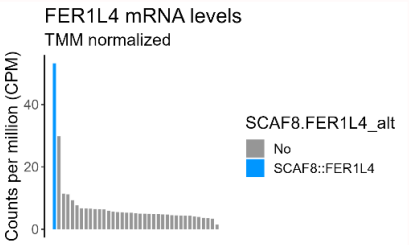


SCAF8::FER1L4



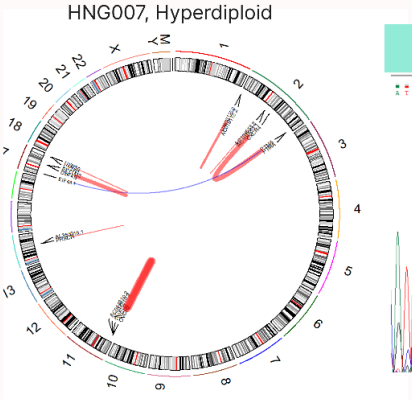
B



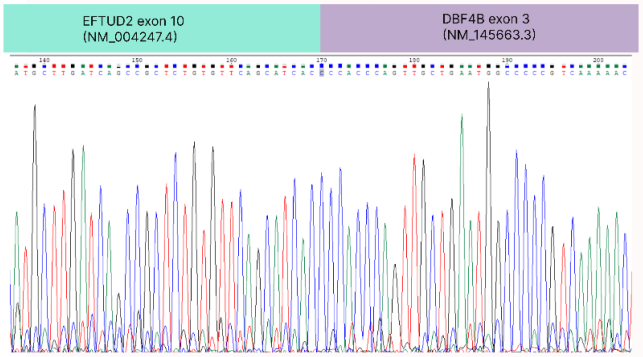


C

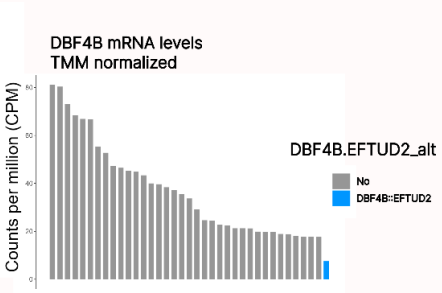
HNG007, Hyperdiploid



DBF4B::EFTUD2



D



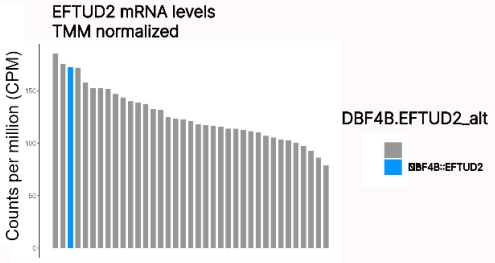


FIG. 4

Novel single nucleotide variants identified at diagnosis in bone marrow samples from LLA patients. Superimposition of wildtype structure with mutated residues in the described variants are shown. A) DUX4 a3 helix of HD1 domain inserted in the DNA major groove, with the side chain of residue 65 making base-specific contacts (sticks representation). Hydrogen bonds are depicted by a dotted line. Note that, while Ile65 in the wild-type (ghost licorice side-chain representation) DUX4 version does not directly interact with any base, Ile65Asn variant (solid licorice side-chain representation) offers new interactions with the base, affecting the original interactions, since the N in the base is now able to form hydrogen bond interactions with the Asn side chain groups, as depicted. PDBid: 5Z6Z. B) Cryo-EM structure (PDBid: 8HAL) of the CBP catalytic core of CREB-binding protein (green cartoon representation) bound to histone H2B and acetylated H4 (ghost gray representation) and DNA (transparent blue). As noted, Gly1542Val variant occurs in a disordered region (purple) connecting two α-helices. While not directly located in the protein-histone complex interface, the variant might compromise the stability of this region, as also reported for structurally near variants identified by Mullighan *et al*[29]. C) CSF3R (cyan cartoon representation) interacting with CSF3 ligand (blue cartoon representation). As noted, while Gly147Arg variant (Arg side chain overlaps with the short Gly side chain, which is not visible) does not directly affect the intermolecular interaction with CSF3, we can hypothesize that it may be responsible of i) a destabilization of the loop containing the residue; ii) the interaction with other molecules, since a positive-charged residue in the loop provides new possible intermolecular interactions; or a combination of both. PDBid: 2D9Q

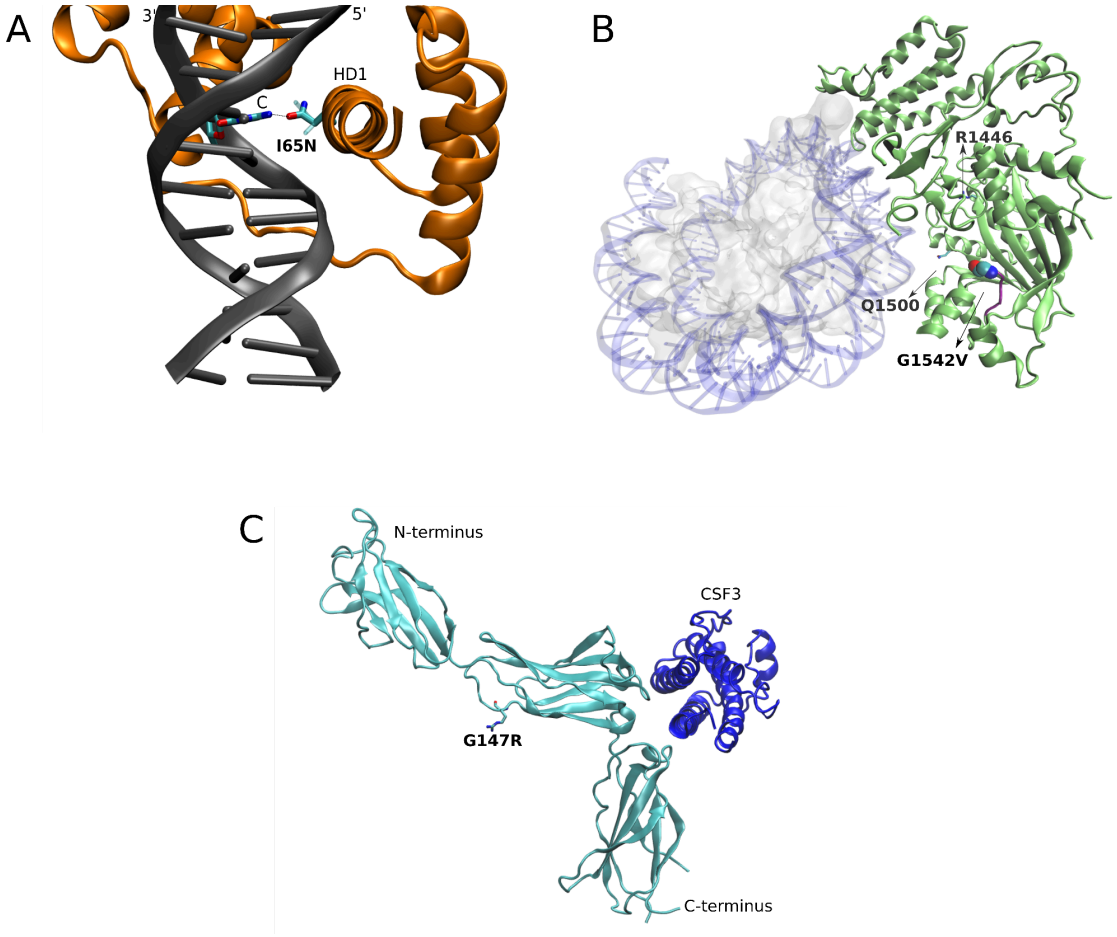


FIG. 5

Integration of multiple data types for characterization of molecular subtypes and novel variants. ZNF384 subtype was predicted in a patient harboring *EP300::ZNF384* fusion transcript, reduced abundance of *CD10* and increased *CD33* (A). MEF2D was predicted in a patient harboring *MEF2D::BCL9* fusion transcript and increased levels of *HDAC9* mRNA (B). DUX4 subtype was predicted in 5 patients with different molecular features. In some but not all patients there was an association with upregulation of *DUX4::IGH* target genes (*ITGA6*, *CCNJ*), increased levels of *DUX4* and *CD371*, and reduced levels of *ERG* (C).

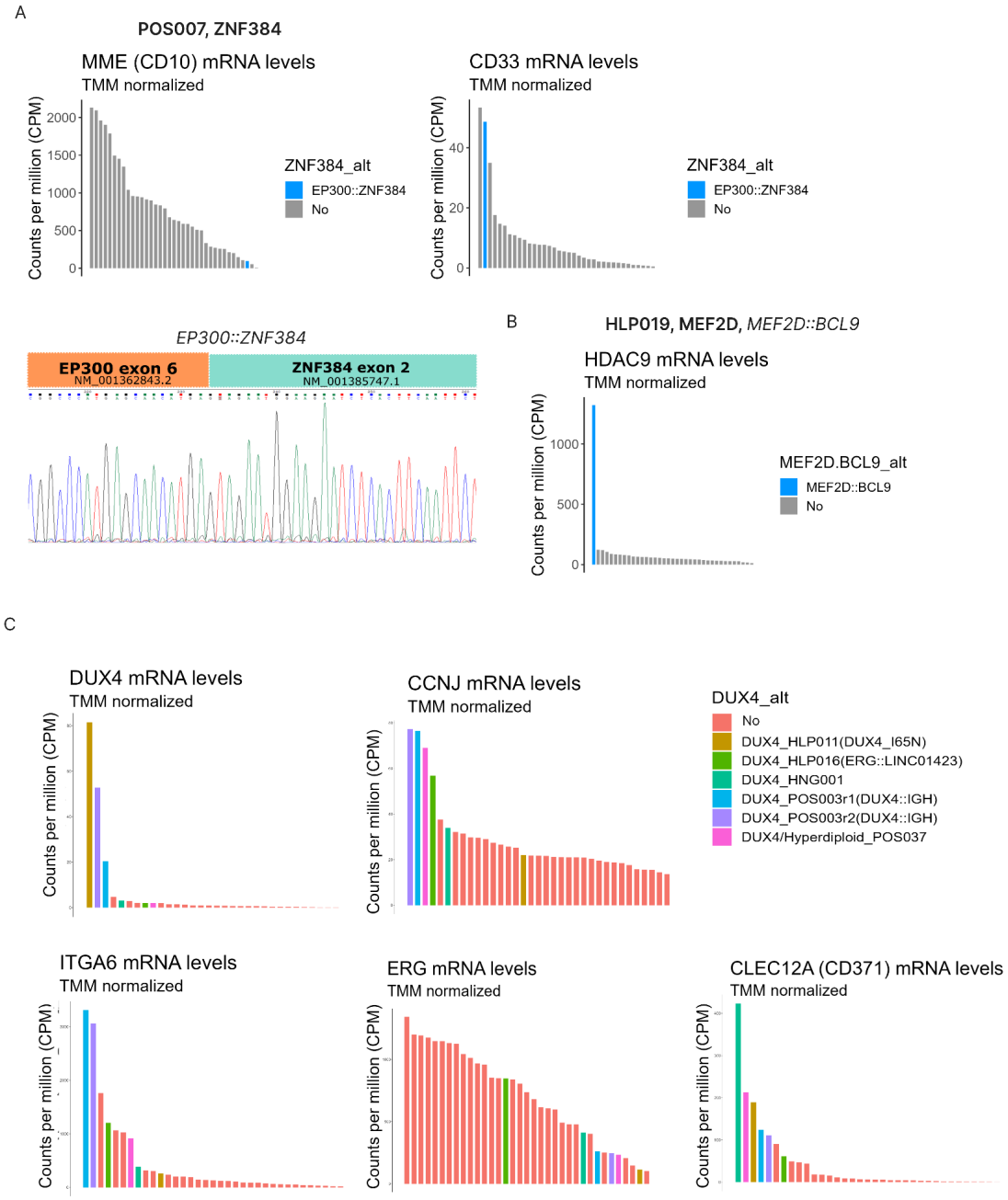


TABLE 1

Variable	N = 32[1]
Relapse	5 (16%)
Age	7.0 (3.0, 13.0)
Sex	
F	17 (53%)
M	15 (47%)
ALL_subtype	
common	26 (81%)
pre-B	6 (19%)
Risk_group*	
high	7 (22%)
intermediate	22 (69%)
standard	3 (9.4%)
Response_to_Prednisone	
good	28 (88%)
poor	4 (13%)
MRD_15	
<0,1%	11 (37%)
>10%	5 (17%)
1-10%	14 (47%)
Unknown	2
Acute_toxicity	9 (29%)
Unknown	1
Death	5 (16%)
Follow_up_time	32 (25, 35)

[1] n (%); Median (IQR)

*SR (age:1-5yrs, WBC<20x10^9/L, blasts D8<1000), IR (age<1yr or >=6yrs, WBC>=20x10^9/L, blasts D8<1000), HR (t(9;22) or t(4;11), or blasts D8 >=1000, or hypodiploidy). Patients were re-classified to higher risk groups according to their MRD at D15 and D33 assessed by flow cytometry.

TABLE 2 (StJude=Childhood Standard Risk and High Risk)

Variable	Cohort			
	Freq_GATLA, N = 32[1]			**Freq_StJude**, N = 2,059[1]
	Protocol	**ALLSorts**	**Final classification**	
B-ALL_subtype				
B-other/Unclassified	10 (31.2%)	2 (6.3%)	2 (6.3%)	208 (10%)
BCL2/MYC	N/A	0	0 (0%)	2 (<0.1%)
DUX4	N/A	5 (15.6%)	4 (13%)	74 (3.6%)
DUX4/Hyperdiploid	N/A	0	1 (3.1%)	0 (0%)
ETV6-RUNX1	2 (6.2%)	2 (6.3%)	3 (9.4%)	398 (19%)
ETV6-RUNX1-like	N/A	2 (6.3%)	2 (6.3%)	35 (1.7%)
HLF	N/A	0	0 (0%)	9 (0.4%)
Hyperdiploid	8 (25%)	13 (40.6%)	12 (38%)	510 (25%)
iAMP21	N/A	0	0 (0%)	129 (6.3%)
IKZF1 N159Y	N/A	0	0 (0%)	5 (0.2%)
KMT2A	0	0	0 (0%)	74 (3.6%)
KMT2A-like	N/A	0	0 (0%)	2 (<0.1%)
Low hypodiploid	0	0	0 (0%)	26 (1.3%)
MEF2D	N/A	1 (3.1%)	1 (3.1%)	30 (1.5%)
Near haploid	0	0	0 (0%)	41 (2.0%)
NUTM1	N/A	0	0 (0%)	12 (0.6%)
PAX5 P80R	N/A	0	0 (0%)	19 (0.9%)
PAX5alt	N/A	2 (6.3%)	2 (6.3%)	133 (6.5%)
Ph	0	0	0 (0%)	68 (3.3%)
Ph-like(CRLF2+other)	N/A	3 (9.3%)	3 (9.4%)	149 (7.2%)
TCF3-PBX1	1 (3.1%)	1 (3.1%)	1 (3.1%)	105 (5.1%)
ZEB2/CEBP	N/A	N/A	N/A	2 (<0.1%)
ZNF384	N/A	1 (3.1%)	1 (3.1%)	28 (1.4%)

Cytogenetic failed or unavailable	11 (34.3%)	N/A	N/A	N/A
[1] n (%)				
Age (yrs) median (IQR)		7 (3;13)		6 (3;10)

TABLE 3

Gene	Gene description	Localization	Variant	Reads variant /wt	VAF*	Type of variant	FRANKLIN**	Varsome*	OncoKB	ClinVar	COSMIC	Cancer Genome Interpreter
CREBBP	Tumor suppressor, transcriptional coactivator	HAT (histone acetylase protein) domain; close to ligand binding residues	NP_004371.2:p.Gly1542Val	39/0	1	missense	Likely pathogenic	Likely pathogenic	No info	No info	No info. G1542S, G1542D	Driver
CSF3R	Known cancer gene, cytokine receptor; germline heterozygous variants as risk factors for myeloid and lymphoid malignancies.	extracellular domain	NP_000751.1:p.Gly147Arg	92/82	0.52	missense	VUS	Uncertain significance	No info	No info	1 report in basal cell carcinoma	Driver
DUX4	Known cancer gene, transcription factor	homeodomain, DNA interaction site	NP_001280727.1:p.Ile65Asn	657/0	1	missense	VUS	Likely benign	No info	No info	No substitutions in DUX4.	Not available

*in RNA-seq, *germline classification

TABLE 4

	Relapse		
	No, N = 27[1]	**Yes, N = 5[1]**	**p-value[2]**
Follow_up_time	31 (25, 35)	34 (16, 50)	0.7
Gene_Alteration--subtype			0.025
ABHD17B-CEMIP2_FUSION--Hyperdiploid	1 (3.7%)	0 (0%)	
ATM_P604S--ETV6-RUNX1_FUSION–ETV6-RUNX1	1 (3.7%)	0 (0%)	
ATM_P604S--MEF2D-BCL9_FUSION–MEF2D	0 (0%)	1 (20%)	
CSF3R_G147R--DBF4B-EFTUD2_FUSION--Hyperdiploid	1 (3.7%)	0 (0%)	
DUX4	3 (11%)	0 (0%)	
EP300-ZNF384_FUSION--ZNF384	1 (3.7%)	0 (0%)	
ETV6-RUNX1_FUSION--ETV6-RUNX1	2 (7.4%)	0 (0%)	
FLT3mut--Hyperdiploid	2 (7.4%)	0 (0%)	
Hyperdiploid	6 (22%)	0 (0%)	
KRAS_Q61P-Ph-like	1 (3.7%)	0 (0%)	
NF1_R1306*--CREBBP_G1542V--Hyperdiploid	0 (0%)	1 (20%)	
NRAS_G12D--PAX5_R38C--Hyperdiploid	1 (3.7%)	0 (0%)	
NRAS_G12S--PAX5_P34L--PAX5alt	0 (0%)	1 (20%)	
NSD2_E1099K--ETV6-RUNX1-like	0 (0%)	1 (20%)	
PAX5-ETV6_FUSION--PAX5alt	1 (3.7%)	0 (0%)	
Ph-like (CRLF2 and non-CRLF2 rearranged)	2 (7.4%)	0 (0%)	
PTPN11_D61Y--TP53_P152L--ZEB2_H1038R--DUX4-IGH_FUSION--DUX4	1 (3.7%)	0 (0%)	
TCF3-FLI1_FUSION--ETV6-RUNX1-like	1 (3.7%)	0 (0%)	
TCF3_S338fs*10--TCF3-PBX1_FUSION--TCF3-PBX1	1 (3.7%)	0 (0%)	
TP53_R267P--IKZF1_D186Y--DUX4-Hyperdiploid	0 (0%)	1 (20%)	
Unclassified	2 (7.4%)	0 (0%)	

1 Median (IQR); n (%)
2 Wilcoxon rank sum test; Fisher's exact test

TABLE 5

	MRD_15			
	1-10%, N = 14[1]	**<0,1%, N = 11**[1]	**>10%, N = 5**[1]	p-value[2]
Follow_up_time	35 (32, 36)	28 (25, 33)	31 (15, 48)	0.2
Gene_Alteration--subtype				0.018
ABHD17B-CEMIP2_FUSION--Hyperdiploid	1 (7.1%)	0 (0%)	0 (0%)	
ATM_P604S--ETV6-RUNX1_FUSION--ETV6-RUNX1	1 (7.1%)	0 (0%)	0 (0%)	
ATM_P604S--MEF2D-BCL9_FUSION--MEF2D	0 (0%)	1 (9.1%)	0 (0%)	
CSF3R_G147R--DBF4B-EFTUD2_FUSION--Hyperdiploid	0 (0%)	1 (9.1%)	0 (0%)	
DUX4	2 (14%)	1 (9.1%)	0 (0%)	
EP300-ZNF384_FUSION--ZNF384	1 (7.1%)	0 (0%)	0 (0%)	
ETV6-RUNX1_FUSION--ETV6-RUNX1	1 (7.1%)	1 (9.1%)	0 (0%)	
FLT3mut--Hyperdiploid	0 (0%)	2 (18%)	0 (0%)	
Hyperdiploid	5 (36%)	0 (0%)	0 (0%)	
KRAS_Q61P-Ph-like	0 (0%)	1 (9.1%)	0 (0%)	
NF1_R1306*--CREBBP_G1542V--Hyperdiploid	0 (0%)	0 (0%)	1 (20%)	
NRASmut--PAX5mut--Hyperdiploid	0 (0%)	1 (9.1%)	0 (0%)	
NRASmut--PAX5mut--PAX5alt	0 (0%)	1 (9.1%)	0 (0%)	
NSD2_E1099K--ETV6-RUNX1-like	1 (7.1%)	0 (0%)	0 (0%)	
PAX5-ETV6_FUSION--PAX5alt	1 (7.1%)	0 (0%)	0 (0%)	
Ph-like	0 (0%)	0 (0%)	1 (20%)	
PTPN11_D61Y--TP53_P152L--ZEB2_H1038R--DUX4-IGH_FUSION--DUX4	0 (0%)	0 (0%)	1 (20%)	
TCF3-FLI1_FUSION--ETV6-RUNX1-like	0 (0%)	1 (9.1%)	0 (0%)	
TCF3_S338fs*10--TCF3-PBX1_FUSION--TCF3-PBX1	1 (7.1%)	0 (0%)	0 (0%)	
TP53_R267P--IKZF1_D186Y--DUX4-Hyperdiploid	0 (0%)	0 (0%)	1 (20%)	
Unclassified	0 (0%)	1 (9.1%)	1 (20%)	

1 Median (IQR); n (%)
2 Kruskal-Wallis rank sum test; Fisher's exact test