Real time-PCR a diagnostic tool for reporting copy number variation and relative gene-expression changes in pediatric B-cell acute lymphoblastic leukemia—a pilot study

Zoha Sadaqat[1,2], Smitha Joseph (ID)[1,3], Chandrika Verma[2], Jyothi Muni Reddy[4], Anand Prakash[4], Tinku Thomas (ID)[5], Vandana Bharadwaj[4,*], Neha Vyas (ID)[2]

[1]Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India
[2]Division of Molecular Medicine, St John's Research Institute, St John's National Academy of Health Sciences (a Unit of CBCI Society for Medical Education), Bangalore 560034, Karnataka, India
[3]Division of Epidemiology and Biostatistics, St John's Research Institute, St John's National Academy of Health Sciences (a Unit of CBCI Society for Medical Education), Bangalore 560034, Karnataka, India
[4]Department of Pediatric Hematology Oncology and Bone Marrow Transplantation, St John's Medical College and Hospital, St John's National Academy of Health Sciences (a Unit of CBCI Society for Medical Education), Bangalore 560034, Karnataka, India
[5]Department of Biostatistics, St John's Medical College and Hospital, St John's National Academy of Health Sciences (a Unit of CBCI Society for Medical Education), Bangalore 560034, Karnataka, India

*Correspondence: Department of Pediatric Hematology Oncology, St John's Medical College and Hospital, SJNAHS, Bangalore, Karnataka, India. Tel: +91 080 22065000 / 5001; Fax: +91 80 25501088; E-mail: vandana.b@stjohns.in (V.B.)

Abstract

Real time-polymerase chain reaction (RT-PCR) is used routinely in clinical practice as a cost-effective method for molecular diagnostics. Research in pediatric B-cell Acute Lymphoblastic Leukemia (ped B-ALL) suggests that apart from cytogenetics and clinical features, there is a need to include Copy number variation (CNV) in select genes at diagnosis, for upfront stratification of patients. Using ped B-ALL as a model, we have developed a RT-PCR-based iterative probability scoring method for reporting CNVs, and relative gene-expression changes. Our work highlights that once genes of interest and hotspots of CNVs are identified in discovery phase, our proposed method can be used as a cost-effective and user-friendly diagnostic tool for the identification of changes at genomic or transcriptomic level. It has the potential to be incorporated in routine diagnostics in resource constrained settings and be tailored for different diseases as per need.

Keywords: CNV; BCP-ALL; Shh; CGH; RT-PCR; risk-stratification

Introduction

Treatment of pediatric acute lymphoblastic leukemia (ped ALL) has evolved considerably over the last six to seven decades becoming increasingly risk stratified and Minimal Residual Disease (MRD)-driven. Clinical features including age, white blood cell count at diagnosis, hepatomegaly/splenomegaly, along with cytogenetics, and MRD are used to classify patients to standard-risk (SR), intermediate-risk (IR), or high-risk (HR) categories [1, 2]. Over time, the cure rates of pediatric B-cell ALLs (ped B-ALLs) have improved dramatically to over 90% in high-income countries [3–5]. But despite the advancements in therapy and supportive care, ped B-ALL continues to pose a challenge to clinicians in lower middle-income countries, including India, with a relapse rate (RR) ranging from 18% to 41% [6–9]. While improved supportive care contributes to the increased overall survival (OS), refining risk stratification remains the need of the hour to reduce relapses in the IR group, to which ∼40% to 50% of patients in India belong [10, 11].

Recent genomic findings have recognized somatic copy number variations (CNVs) in ped B-ALL as an additional factor for risk-stratification [10, 12–15]. Efforts are ongoing to conjoin the present risk stratification method (based on clinical features and conventional cytogenetics/FISH and molecular panels) with the prognostic bearing of CNVs in genes of interest [12]. Molecular techniques such as multiplex ligation-dependent probe amplification (MLPA), cytogenetics-based array, comparative genomic hybridization (CGH), and fluorescence in-situ hybridization (FISH) have served as a benchmark for CNV detection. The prevalence of CNVs in genes involved in cell cycle, apoptosis, differentiation, and lymphogenesis have been well acknowledged by clinical studies on ped B-ALL [12]. Deletion of the transcription factor, Ikaros family zinc finger 1 (*IKZF1*) is predominantly found

in the HR-group with low OS and high RR [15]. *IKZF1* deletion in combination with deletions in select few other genes namely—paired box 5 *(PAX5)*, cyclin-dependent kinase inhibitor 2A/2B *(CDKN2A/2B)*, or pseudoautosomal region 1 *(PAR1)*, are also associated with higher MRD and RR [16].

Additionally, drug response profiling (DRP) is also being tried to tailor the therapy for ped ALL patients with very high risk, relapsed or refractory disease [17]. Synergistic effect is observed when Asparaginase (ASNase) is combined with exportin-1 inhibitors *(XPO1;* selinexor, eltanexor), *BCL2* inhibitor (venetoclax), or proteasomal inhibitors (bortezomib, carfilzomib). This suggests that relative differences in cellular processes are also important and need to be accounted for to further improve outcomes, especially for IR or HR-categories of patients, upon relapse or treatment failure.

We have developed an iterative scoring method to evaluate the usage of real time-polymerase chain reaction (RT-PCR), for relative gene-expression and CNV determination, a cost-effective and user-friendly screening tool. In correlation with clinical data, our proposed method here could potentially help identify the genetic profiles of ped B-ALL patients that fail therapy on contemporary treatment protocols [2], and for further molecular stratification to enable additional suitable therapeutic considerations.

Methodology

Patient sample collection protocol

Patients diagnosed with pre-B-ALL (aged 1–18 years) based on standard morphologic, immunophenotypic, and genetic features (presented at Department of Pediatric Haematology Oncology and BMT, SJMCH), during the study period were sampled. Patients at the time of diagnosis were classified as SR/IR/HR based on ICiCLe-2014 protocol [2]. MRD status determined at day 29 post-induction led to the stratification of patients into two groups: MRD positive patients (blast cells >0.001%) and MRD negative patients (blast cells <0.001%). Patient with immune thrombocytopenia was selected as the non-malignant age matched control (NAC). Other healthy controls were sampled randomly from the staff in the department. Supplementary Fig. 1 gives a breakup of the samples that underwent CNV and/or relative gene expression analysis. Ethical approval was taken for the study from the Institutional Ethical Committee (IEC, Study reference number IEC/132/2022, dated 5th August 2022), SJMCH. Written consent was obtained from all parents and patients.

BM and peripheral blood processing

BM aspirates from patients and peripheral blood from healthy adults were collected (1.0–1.5 ml) in EDTA tubes. Plasma (upper layer) was separated from BM aspirates by centrifuging for 5 min at 1000 rpm. It was further centrifuged at 2500 rpm for 25 min, aliquoted, and stored at −80°C. The remaining sample was then treated with 1× red blood cell (RBC) lysis buffer, and kept for 15 min on Rocker. Cells were then centrifuged at 1000 rpm for 15 min. Another RBC lysis buffer wash was given. The cells were further washed with 1× phosphate buffer solution. Cells were aliquoted in different tubes (including an aliquot for RNA extraction using Trizol Reagent) and stored at −80°C.

Cell culture

K562 cells were grown in RPMI-1640 Medium (Gibco, Life Technologies) supplemented with 10% (v/v) Fetal Bovine Serum (Gibco, Life Technologies), 1% (v/v) Pen-Strep (Gibco, Life Technologies), and 1% (v/v) GlutaMAX (Gibco, Life-Technologies). The cell line was maintained in a 5% CO_2 humidified Galaxy 170R incubator (Eppendorf) at 37°C.

DNA extraction

Genomic DNA from ped B-ALL patients, healthy adults, and cell line was extracted using a commercially available kit (QIAmp DNA Mini Kit, Qiagen) following the manufacturer's instructions.

RNA extraction

RNA was isolated using TRIzol reagent (Invitrogen). Samples were treated with Chloroform (0.2 ml per 1 ml of TRIzol) used and incubated at room temperature for 15 min. Samples were then centrifuged at 13 000 rpm for 15 min at 4°C. The aqueous phase (top layer) was carefully transferred to a new tube and isopropanol (0.5 ml per 1 ml of TRIzol used) along with GlycoBlue (∼15–20 μg/ml) (Invitrogen) was added and incubated at room temperature for 15 min. After centrifuging at 13 000 rpm for 15 min at 4°C, the supernatant was discarded and pellets were washed with 70% and 100% ethanol subsequently. Pellets were finally resuspended in double-distilled water and stored at −80°C.

Primer design

Primers (Pr) were designed in-house for gene expression *(SHH, BCL2, GLI1)* and CNV *(RB1, EBF1, ETV6, IKZF1, CDKN2A, BTG1, CDKN2A/2B)* analysis. CNV primers were designed using the available protocol [18], specific to the different exonic regions of the genes based on the human genome—GRCh38/hg38. The primers used for gene expression and CNV analysis are mentioned in Supplementary Table 1.

cDNA conversion

Approximately 2 μg of RNA was synthesized into cDNA using the Applied Biosystems' High-Capacity cDNA Reverse Transcription Kit (Ref: 4368813) using the manufacturer's protocol.

RT-PCR

For RT-PCR, ∼15 ng of genomic DNA or ∼100 ng of cDNA template was used for CNV or gene expression analysis, respectively.

The reaction was done using SYBR green Master Mix (MM) (PowerUp™ SYBR™ Green Master Mix—Applied Biosystems by Life Technologies, Cat# A25742) on Applied Biosystems™ StepOnePlus™ Real-Time PCR System. About 0.1 μM concentration of forward and reverse primers each, were used per reaction. Pre incubation and initial denaturation of the DNA template were performed at 95°C for 2 min, followed by amplification for 40 cycles at 95°C for 15 s and 60°C for 30 s. Cycle threshold (CT) values were normalized to a reference/internal control gene for each sample. Relative expression was calculated using the known $2^{-\Delta\Delta Ct}$ method where $\Delta CT = CT_{Target} - CT_{Reference}$ for test and healthy samples, represented as Log_2 fold change. Log_2-fold change ≥ 1 or ≤ -1 was assigned as amplification or deletion, respectively.

For gene expression analysis *GAPDH* was used as the reference gene; while for CNV analysis *RPP30* was used as the reference gene.

Iteration-based probability scoring

Using the delta-CT values for healthy adults ($n = 6$), median delta-CT and sigma delta-CT were determined based on an iteration method (described in the results section). Median was then compared to the delta-CT for individual samples and probability of deletion/downregulation (P) was acquired.

Next-generation sequencing

The Next-generation Sequencing (NGS) analysis was performed by a commercial service platform (Genotypic Technology Pvt. Ltd). Whole exome sequencing was carried out by them for K562 cell line, on the Illumina platform. The raw reads were trimmed, clipped, and mapped against the reference human genome—GRCh38/hg38 using BWA-v0.7.5 [2] aligner, and showed 99.9% reads mapping to the reference genome. Using Samtools-v1.9. [3], aligned data were processed. SureSelectXT HS Human All Exon V8 was used for target enrichment of the exonic regions, while variant calling in the same regions was done using GATK-v4.2.6.15 HaplotypeCaller. For copy number variant analysis, CONTRA-v 2.0.8 [7] tool was used to determine CNVs between the control sample (NAC) and the test samples (K562-cell line). By removing GC-content bias using base-level log-ratios, correction for an imbalanced library size effect on log-ratios, and the estimation of log-ratio variations via binning and interpolation, the tool was able to call for gains and losses for focus regions.

Comparative genomic hybridization

The CGH analysis was outsourced to a commercial service platform (Genotypic Technology Pvt. Ltd). CGH analysis was performed by them using Agilent Human Genome CGH oligonucleotide array, 180k, following the manufacturer's protocol. The microarray comprised of 180 000, 60-mer oligonucleotide probes. Agilent references (Male/Female) and test samples' DNA were fluorescently labeled with Cy3 and Cy5, respectively, after which they were hybridized and microarray slides were scanned. Log_2-intensity ratios (Cy5-test/Cy3-control) were calculated and CNVs were called for either deletion (log_2-intensity ratio ≤ -1), or amplification (log_2-intensity ratio ≥ 2) if a minimum number of three adjacent probes were impacted. On the other hand, losses (log_2-intensity ratio -0.25 to -1) or gains (log_2-intensity ratio 0.25–2) could be identified with a single probe as well, but they were not deemed as CNVs per se. For genes with a single probe, only deletion or amplification cut-off value was taken into consideration (not the minimum number of probes) to call for CNV. Genomic coordinates were based on the human genome (GRCh37/hg38). Genomic DNA used for analysis includes: ped B-ALL patients' bone-marrow aspirates at diagnosis, K562 cell line, healthy adults' (controls) peripheral blood, and a NAC's bone-marrow aspirate. For accurate comparison with the RT-PCR primers, these probes were also mapped on GRCh38/hg38 genome by our laboratory.

Results

Iteration-based probability scoring method for reporting genomics and transcriptomics changes in ped B-ALL

To determine relative differences in the CNV or transcript levels in samples collected from ped B-ALL bone-marrow aspirates (Table 1), peripheral blood from six healthy adults was collected as controls for necessary correlations.

Iteration-based probability scoring method was devised to determine and compute the spread in gene-expression or CNV levels in healthy control samples. A conventional RT-PCR based approach was used using cellular RNA with *GAPDH* as internal control for gene expression analysis; while DNA was used for the CNV analysis with *RPP30* gene as an internal reference gene (Fig. 1A). In this method, median delta-CT values were derived from six healthy controls for each gene-of-interest (represented as *X*) and used to derive delta delta-CT per ped ALL sample

Table 1. Clinical characteristic of Ped-B-ALL patients.

Clinical details	N (%)
Total number of patients	14
Gender	
Female	5 (35.7)
Male	9 (64.3)
Age at diagnosis	
1–9	9 (64.3)
10–14	4 (28.6)
≥15	1 (7.1)
Cytogenetics	
Hyperdiploidy	4 (28.6)
ETV6::RUNX1	1 (7.1)
BCR::ABL1	1 (7.1)
Normal karyotype	8 (57.2)
Risk status at diagnosis	
Standard	1 (7.1)
Intermediate	12 (85.7)
High	1 (7.1)
MRD status[a]	
MRD positive	0 (0)
MRD negative	13 (100)

[a] One patient lost to follow up. Percentages may not add up to 100 due to rounding-off.

(represented as *x*). The variability (represented as sigma, σ) between the delta-CT values for the healthy controls was also estimated. The Median delta-CT from healthy controls for each gene-of-interest was then used to derive delta delta-CT per ped B-ALL sample (represented as *x*) and z-scores were derived to identify probabilities of upregulation in gene-expression or deletion for CNV analysis (Fig. 1A). Due to the low sample-size, it was not possible to identify the distribution of the data to calculate z-scores. Hence the z scores are calculated from the following formula:

$$z = \frac{x - X}{\sigma}, \tag{1}$$

where *x* is the delta-CT of ped B-ALL (test)-sample individually, *X* is the location estimate (derived Median delta-CT from healthy control samples, $n = 6$) and σ is the scale estimate (variability in delta-CT of control samples, $n = 6$) [19, 20].

For the calculation of z score, estimates of X and σ were determined from delta CT values for each gene-of-interest using healthy controls through robust analysis for small samples [21]. X and σ are estimated iteratively using Newton–Raphson method as T_n and S_n, respectively where T_n is the M-estimator of location as described in the equation given below:

$$\sum_{i=1}^{k} \psi\left(\frac{(x_i - T_n)}{S_n}\right) = 0, where \ \psi(x) = \frac{e^x - 1}{e^x + 1} \tag{2}$$

where T_n is the location estimator and S_n is the scale estimator. The initial value used for T_n was $T_1 = median(x_i)$ and the scale estimator, S_n (median absolute deviation) was calculated as [19]

$$S_n = c.1.483.median(|x_i - T_n|) \tag{3}$$

where, *c* is the small sample correction used to ensure that S_n is unbiased. Then the location estimator was obtained as $T_{n+1} = T_n + \frac{S_n * f_n}{0.4132}$, where $f_n = \frac{1}{k}\sum_{i=1}^{k}\psi\left(\frac{(x_i - T_n)}{S_n}\right)$.

This procedure was repeated until there was no change in estimated values from one iteration to the next or until abs(f_n) < 0.000001. T_n and S_n were fixed as the values from that iteration.
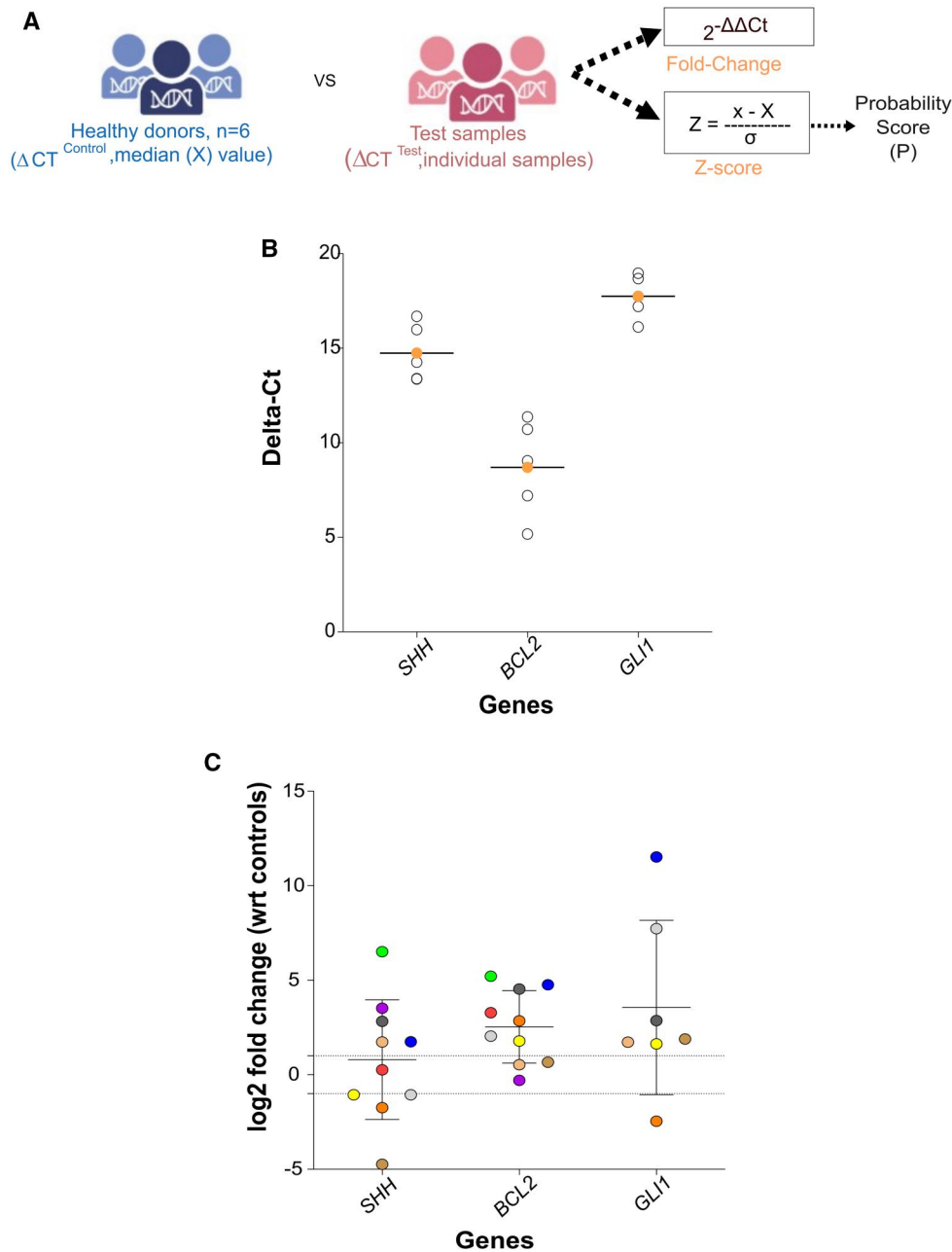
Figure 1. RT-PCR-based relative gene-expression analysis using pediatric B-ALL bone-marrow aspirates. (**A**) Cartoon representing RT-PCR based method of relative gene expression using RNA or Copy number variation determination using genomic DNA for Test samples with reference to healthy adults (median derived using $n = 6$ healthy adults for RNA or DNA, using peripheral blood samples). Fold change and Z-score based probability scores are derived for necessary comparison. Created using Biorender.com. (**B**) Graph representing spread in Delta-CT value for each gene of interest as specified, solid coloured data-point represents the median value, derived using iterative scoring method for each primer. (**C**) Dot-plot representing relative expression of *SHH*-signalling genes for Ped B-ALL patients. Each colour represents a patient sample and relative expression of each gene with reference to healthy control. Conventionally used cut-offs for fold change, 2 to (−2) fold change, are represented as dotted lines. Y-axis represents Log2 scale. Error bars represent ± standard deviations (SD).

The estimators of location and scale parameters were calculated from the reference population using the above-mentioned method and z score for the cases (or patient samples) was calculated using Equation (1) (Fig. 1A). Using the z-scores calculated from the equation mentioned above, the probability score (P) was determined.

Fold change and the z score-derived probabilities were used together to evaluate the reliability of the data. Standard cut-offs for fold change were used to call for upregulation or downregulation in gene expression; and amplification or deletion for CNV calling (Fold-change $\leq -2 =$ downregulation or deletion; fold-change $\geq 2 =$ upregulation or amplification) in correlation with probability scores.

Gene-expression analysis in ped B-ALL

To develop a quantitative method for reporting relative differences in gene expression levels, select genes from the Sonic Hedgehog *(SHH)* signaling pathway were used as a model given its role in B-cell maintenance [22] and our lab's interest too [23].

Here, ped B-ALL patients' bone-marrow aspirates were used to evaluate relative expression of select genes of Shh-signaling pathway, namely—SHH; Glioma-associated oncogene family zinc finger 1 *(GLI1)*; and B-cell leukemia/lymphoma 2 protein *(BCL2)*. Delta-CT values *SHH*, *BCL2*, and *GLI1* demonstrate significant variability between the peripheral blood samples derived from healthy controls, using GAPDH as an internal control (Fig. 1B, $n = 6$). Fold change in expression of *SHH, GLI1* and *BCL2* in ped B-ALL patients, at diagnosis, was determined using the median delta-CT values from healthy controls (Fig. 1C). We find patients with ∼91 fold up-regulation in *SHH* (Fig. 1C, green data-point), ∼37 fold up-regulation in *BCL2* expression (Fig. 1C, green data-point), or up to ∼2930 fold up-regulation in *GLI1* expression (Fig. 1C, blue data-point). Upregulation in expression of these genes is observed in several ped B-ALL patients (Fig. 1C; Table 2, >2 fold). The probability scores for upregulation of these genes-of-interest ranged from ∼0.6 to 1.0 (Table 2), which corroborated with the significant upregulation in their gene expression levels. However, all subjects were MRD negative and clinically stable. This suggests that the current intensive, combinatorial chemotherapeutic regime [3] is effective, despite the elevated gene expression levels in the *SHH* pathway. It must be noted that, our method has thus, not scored for any HR patients here, leaving the clinically relevant cut-offs (if any) of the gene expression levels undetermined.

Importantly, this method can be easily tested for other diseases or can be extended to include other genes-of-interest in ped B-ALL such as the ones identified using DRP-based approach for targeted therapy, like *XPO1* or proteasomal genes, etc. [17].

CNV determination using RT-PCR-based approach

Considered a gold standard for CNV determination, a commercially available CGH array was used to analyse the 5 ped B-ALL samples, K562 cell-line, and a NAC.

The CGH array used here includes 180 000 probes, 60mer each, spanning the whole human genome. However, in view of the objective here, we have confined our analysis to the eight genes of interest, namely—*EBF1* (early B-cell factor 1), *RB1* (retinoblastoma 1), *BTG1* (B cell translocation 1 gene 1), *ETV6* (ETS variant transcription factor 6), *PAX5*, *CDKN2A*, *CDKN2B*, and *IKZF1*, already implicated in ped B-ALL therapy response [12, 14, 24]. Although *IKZF1* deletion is known to have the most prognostic impact in B-ALL patients [15, 16, 25], *CDKN2A* stands out as the most frequently deleted gene [15]. Consequently, our discussion here primarily revolves around these two genes, along with *ETV6*, which is associated with favorable outcomes upon deletion [12, 26].

K562 cells are known to be deleted for *CDKN2A* and *CDKN2B* genes [27]. We hence performed NGS-based exome sequencing and CGH array to confirm these deletions. No reads were obtained for *CDKN2A/2B* using the NGS based approach (Supplementary Fig. 2); while the CGH array demonstrated major deletions in and around *CDKN2A/2B* genes, unlike NAC (Fig. 2A(i and ii). It must be noted that the number of CGH probes for some of our genes of interest is not optimal here (Fig. 2B; Supplementary Fig. 3A), as this is not a customized array. Overall, K562 cell line demonstrated deletion in *CDKN2A* (Fig. 2C (i)), and region-specific losses in *ETV6* (Fig. 2C(ii), black data-points). According to the threshold (log2-intensity ratio ≤ -1), *CDKN2B* (Supplementary Fig. 3B(i), black data-points) also exhibited deletions. In contrast *PAX5* (Supplementary Fig. 3B(ii), black data-points), and *RB1* (Supplementary Fig. 3B(iii), black data-points) displayed contiguous losses within the lower threshold range (log2-intensity ratio −0.25 to −1).

Based on the threshold (log$_2$-intensity ratio ≤−1), only NAC exhibited some deletions (2 of 41) in the *RB1* gene (Supplementary Fig. 3B(iii), blue data-points). However, the deletions observed were not consecutive and hence, were not called out as CNVs based on the current CGH analysis rules, as technical limitations cannot be ruled out unless these are verified using alternate method. NAC also demonstrated a few losses (15 of 41) interspersed with probes showing gains and no-changes. No deletions were observed in the *EBF1* gene for either NAC or K562 (Supplementary Fig. 3B(iv), blue and black data-points, respectively).

Table 2. Log2 fold change (Log2 FC) and Iteration-based probability scores of gene-expression analysis of ped-B-ALL patients bone-marrow aspirates using select genes as model.

Sample code	Analysis type	Gene expression analysis (using cellular RNA)		
		SHH	BCL2	GLI1
B-ALL 1	Log2 FC	−1.06	2.04	7.73
	Probability	0.29	0.66	1.00
B-ALL 2	Log2 FC	1.74	4.76	11.52
	Probability	0.86	1.00	1.00
B-ALL 3	Log2 FC	1.73	0.53	1.72
	Probability	0.86	0.38	0.96
B-ALL 4	Log2 FC	0.26	3.28	NA
	Probability	0.44	0.97	NA
B-ALL 5	Log2 FC	−1.06	1.78	1.63
	Probability	0.29	0.59	0.95
B-ALL 10	Log2 FC	2.82	4.53	2.86
	Probability	1.00	1.00	1.00
B-ALL 11	Log2 FC	6.51	5.21	NA
	Probability	1.00	1.00	NA
B-ALL 12	Log2 FC	−4.74	0.66	1.89
	Probability	0.21	0.39	0.98
B-ALL 13	Log2 FC	3.52	−0.30	NA
	Probability	1.00	0.32	NA
B-ALL 14	Log2 FC	−1.74	2.85	−2.46
	Probability	0.26	0.89	0.18

Probability of deletion represented as values from 0 to 1; NA = No amplification in RT-PCR.
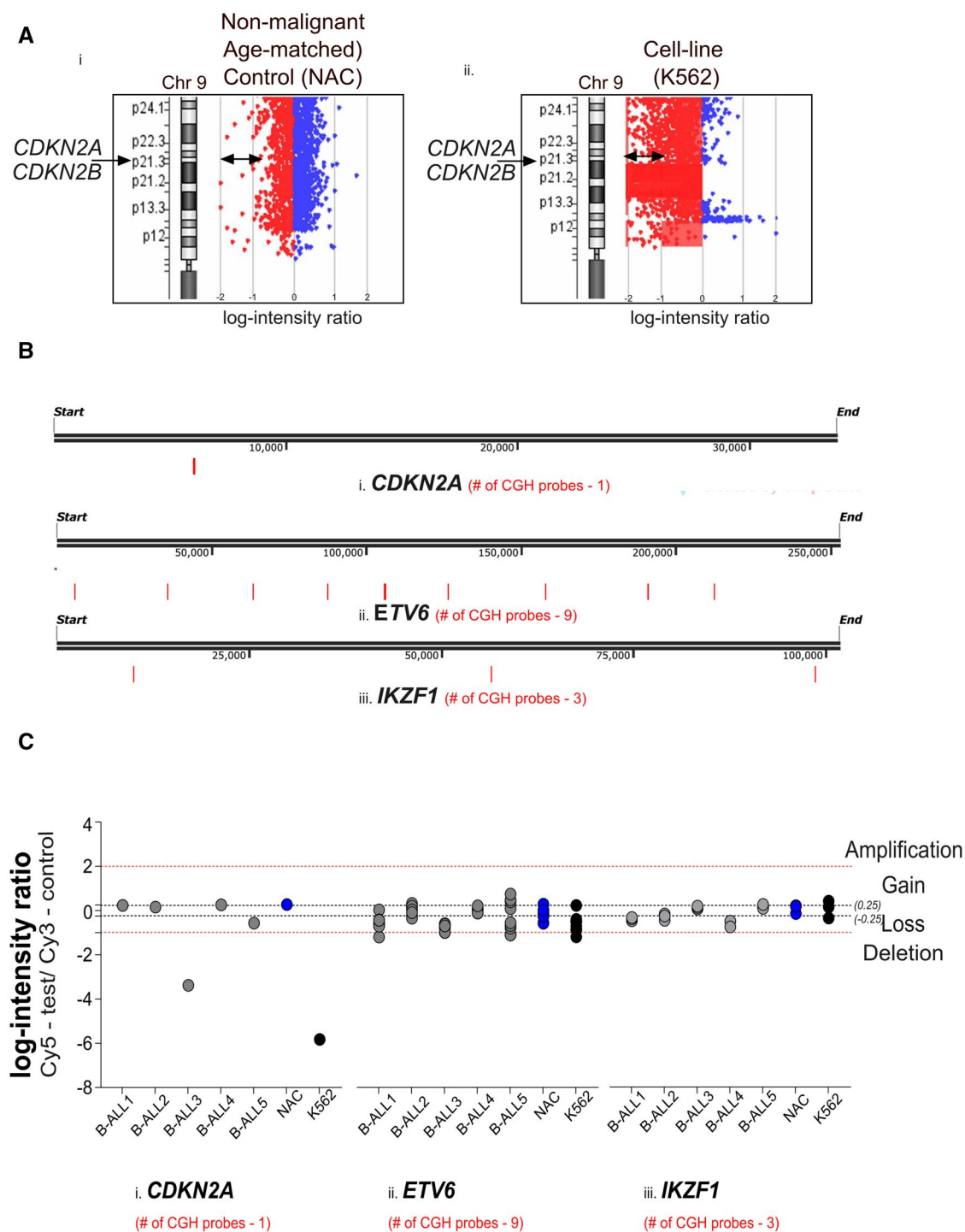
Figure 2. Comparative Genome Hybridisation (CGH) array identifies CNVs in K562 cell-line and Ped B-ALL derived bone-marrow aspirates. (**A**) Scatter-plot of region flanking CDKN2A/2B probes (p21.3 region—chromosome 9) in (**i**) Non-malignant Age-matched Control (NAC) and (**ii**) K562. Each data point here represents a single CGH probe. The x-axis represents the log-intensity ratio. Double head arrow highlights the log-intensity ratio for calling deletions. (**B**) Cartoon representing CGH probes positions (red-marks) mapped onto GRCh38/hg38 genome, using Snapgene, for (**i**) CDKN2A, (**ii**) ETV6, and (**iii**) IKZF1. (**C**) Represents dot-plot of CNV status using CGH array. Log-intensity ratios for CGH probes of (**i**) CDKN2A, (**ii**) ETV6, and (**iii**) IKZF1 for ped B-ALL bone-marrow aspirates (B-ALL 1–5), NAC, and K562 are presented. For all images, Gain = ≥0.25 to < 2; Amplification = ≥ 2; Loss = −0.25 to <−1; Deletion = ≤ −1. Each dot represents 1 probe for the given gene as mentioned. Black-dotted line represents cut offs for loss or gain and red-dotted lines represents cut-offs for calling deletions or amplifications.

Further, CNVs from 5 Ped B-ALL bone-marrow aspirates were also estimated using the CGH array (Fig. 2C; Supplementary Fig. 3B, grey data-points; Table 3). Apart from K562, CDKN2A was deleted in B-ALL3 too (Fig. 2C(i), grey data-point; Table 3). Most samples, except B-ALL2 & B-ALL4 showed region-specific losses, or deletions for ETV6 gene (Fig. 2C(ii), grey data-points; Table 3); While only in B-ALL1 and B-ALL4, all three probes for IKZF1 suggest a clear loss (Fig. 2C(iii), grey data-points; Table 3).

Table 3. CGH array based CNV callings using reference DNA provided by manufacturer as control.

Genes	Sample codes (cellular DNA)						
	B-ALL1	B-ALL2	B-ALL3	B-ALL4	B-ALL5	NAC	K562
RB1	–	–	–	–	–	–	Lc (21 of 41)
BTG1[a]	–	–	–	–	–	–	–
ETV6	Lc (7 of 9)	–	Lc (9 of 9)	–	Lc (3 of 9)	–	Lc (7 of 9)
IKZF1	Lc (3 of 3)	Lc (2 of 3)	–	Lc (3 of 3)	–	–	–
EBF1	–	–	–	–	–	–	–
PAX5	–	–	–	–	Lc (3 of 6)	–	Lc (4 of 6)
CDKN2A	–	–	Dc (1 of 1)	–	Lc (1 of 1)	–	Dc (1 of 1)
CDKN2B	–	–	–	–	Lc (2 of 4)	–	Dc (4 of 4)

[a] All 3 *BTG1* probes did not map on GRCh38/hg38, but did on GRCh37/hg19.
Lc = Loss by CGH (-0.25 to < -1), Dc = Deletion by CGH (≤ -1), Dash = No changes (-0.25 to 0.25) + Gain ($0.25 \leq 2$) + Amplification (>2).

For the CNV calling using RT-PCR, DNA from the same six healthy controls was used to derive the control delta-CT value (median and sigma) for each primer (Fig. 3A). Two primer sets were designed, in-house, per gene for this analysis. Only *RB1* gene has only one primer set (Supplementary Table 1).

While DNA from NAC or K562 demonstrate RT-PCR based amplification of the internal control gene—*RPP30* (Fig. 3B(i-ii), yellow arrows); Only NAC shows amplification for *CDKN2A* (Fig. 3B(i and ii), green arrows). DNA from K562 show lack of amplification or CT-values for *CDKN2A* using either primer, Pr 1 (Fig. 3(i), white arrow) or Pr 2 (Fig. 3B(ii), white arrow), and for *CDKN2B* (data not shown), confirming the major deletion status of the loci (p21.3 region in chromosome 9) in K562 cell line. This is in corroboration with the CGH (Fig. 2A and C) and Exome sequencing data (Supplementary Fig. 2). As observed in CGH analysis, one of the two primers also showed deletion in *CDKN2A* for B-ALL3 (Fig. 3C (i), grey data-point below dotted-line), unlike NAC (Fig. 3C(i), blue data-points). Additionally, we find that B-ALL1, B-ALL4, and B-ALL5 also show deletion in *CDKN2A* using one of the primers' set, but were not picked by the CGH array (compare Fig. 3C(i) and Fig. 2C(i)). The probability of deletion for *CDKN2A* using RT-PCR method was also in the higher range for B-ALL1 ($P = .92$), B-ALL3 ($P = .74$), B-ALL4 ($P = 1.00$), and B-ALL5 ($P = 1.00$) (Table 4).

Deletion in *ETV6* was observed using both primer sets, in most of the patients tested, B-ALL1 ($P = .99$, 1.00), B-ALL2 ($P = 1.00$, 1.00), B-ALL3 ($P = 1.00$, 1.00), B-ALL8 ($P = .95$, .96), B-ALL9 ($P = .93$, 1.00), K562 ($P = .85$, 1.00) (Table 4), unlike NAC. The probabilities of deletions ranged from $P = .5$–1 for either *ETV6* primer sets below the reference line (Fig. 3C(ii); Table 4). Only B-ALL4 showed deletion in *IKZF1* using Pr1 primer sets (Fig. 3C(iii); Table 4, $P = .83$). While the Pr2 primer do not show value below the reference line, the probability of deletion is still high here (Fig. 3C(iii); Table 4, $P = 0.77$). Largely however, comparable trends were observed between the two RT-PCR primer sets for *ETV6* and *IKZF1*, unlike *CDKN2A* primer sets (Fig. 3C, compare the data-points per sample per gene).

It must be noted that the RT-PCR primers' and CGH probes' locations are completely different for each gene here (Fig. 3D; Supplementary Fig. 3A). We hence conclude, that for comparing the accuracy of microdeletion callings (unlike major deletions), the CGH probes and RT-PCR primers need to be tested together and should be designed for overlapping positions.

Overall, we find that the RT-PCR-based method can identify the deletions in ped B-ALL patients unlike in the NAC. We have compared the overall callings between the two methods (Table 5; Supplementary Table 2). The samples with no CNV callings in genes of interest by the CGH array, also failed to demonstrate any significant fold change via the RT-PCR-based approach (with a low probability of deletion, between 0 and 0.5) (Table 4; Table 5; Supplementary Table 2). While the three major deletions (in K562—*CDKN2A* and *CDKN2B*; and B-ALL3—*CDKN2A*) were detected by both CGH and RT-PCR-based approach with the probability scores ranging from 0.73–1 (Table 4; Table 5; Supplementary Table 2). RT-PCR successfully detected 10 of the 12 losses, with a probability score of 0.5–1 (Tables 4 and 5; Supplementary Table 2). In summary, while the RT-PCR-based method effectively identified major deletions or lack of deletions; microdeletion callings require more rigorous corroboration with overlapping primer-probe locations.

Discussion
Risk stratification in Ped B-ALL
Multiple efforts are being made to improve the risk stratification in Ped B-ALL by developing better risk algorithms based on cytogenetics and clinical signs [28, 29]. It is now consistently observed that CNVs in select genes also carry prognostic value [10, 30–32]. Additionally, preliminary evidence suggests that upon relapse or refractory disease, patients might need a more tailored therapeutic approach based on their relative cellular activities [17]. In light of this, we devised a cost-effective RT-PCR-based method to aid molecularly stratification of ped B-ALL patients. Currently, our RT-PCR-based CNV calling or gene expression analysis uses conventional cut-offs for fold change or probability scores. It is thus, necessary to identify clinically relevant cut-off values using a prospective clinical study. Also, the accuracy of RT-PCR-based CNV-callings needs to be compared with the current methods using overlapping primer-probe locations.

Measures to improve the RT-PCR based callings
In this study, we identified a total of 56 callings by analyzing eight genes across seven samples (five Ped B-ALLs, NAC, and K562), using CGH and RT-PCR techniques. It should be emphasized that CGH necessitates at the least three adjacent probes to report deletion or amplification as per current commercially available kits (see section Methodology), whereas for RT-PCR, a single primer demonstrating $\log_2$-fold change ≤ -1, was classified as a deletion by us.

We observe that most of the CNVs deemed as losses (12 of 56) and deletions (3 of 56) by CGH, were captured by at least one of the primer sets using the RT-PCR method -10 of 12 losses, (~83.3%; Table 5 and Supplementary Table 2) and all three deletions (100%; Table 5 and Supplementary Table 2). This was observed despite non-overlapping primer-probe positions. This suggests presence of either macro-deletions or frequent micro-deletions in these genes. ~73% (41 of 56) callings had no deletions
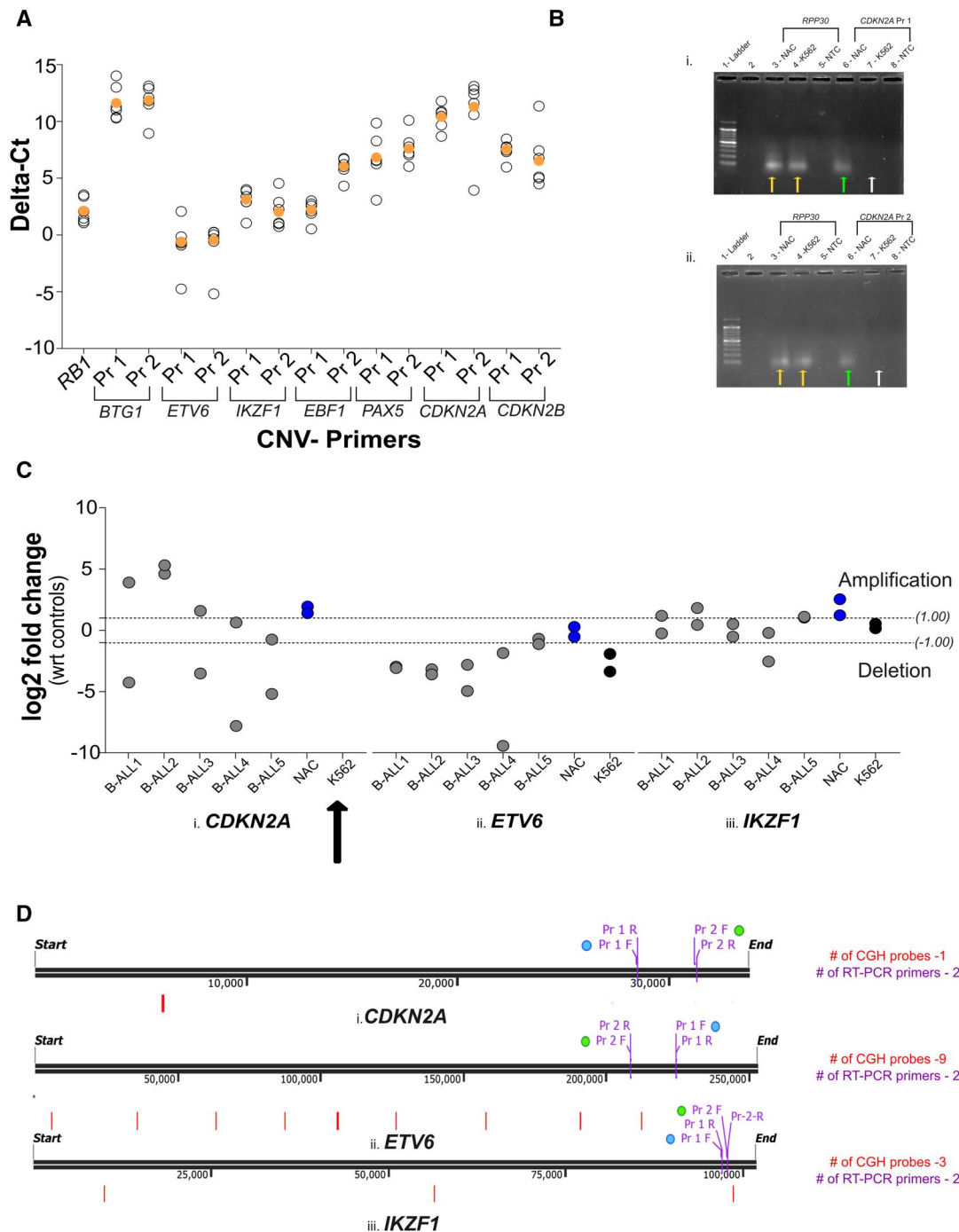
Figure 3 RT-PCR-based CNV determination of K562 cell-line and Ped B-ALL derived bone-marrow aspirates. (**A**) Graph represents delta-CT value (using genomic DNA) for each primer designed (two per gene, except RB1) using *RPP30* as internal control. Coloured, solid data-point represents the median value, derived using iterative scoring method for each primer. (**B**) Agarose gel electrophoresis of RT-PCR amplification products of (**i**) *CDKN2A* Pr 1, (**ii**) *CDKN2A* Pr 2, using RPP30 as the internal control, for NAC, K562 and no-template control (NTC). For both K562 and NAC, arrow points towards the amplified products of *RPP30* gene, or *CDKN2A* (Pr 1 or Pr 2) gene as mentioned. 100 bp DNA ladder is used as the size reference. NTC = No template control. (**C**) Graph representing RT-PCR determined CNV status of ped B-ALL patients, NAC and K562 cell-line for (**i**) *CDKN2A*, (**ii**) *ETV6*, and (**iii**) *IKZF1*, represented as log2-fold change, normalized to healthy adults. Deletion = ≤ -1; Amplification = ≥ 1. Each data-point represents one RT-PCR primer-set. Two primer sets are used for each gene as mentioned. No amplification was obtained for *CDKN2A* primers (Pr 1 and Pr 2) in K562 as represented by the black arrow. (**D**) Image represents location of the RT-PCR primers and CGH probes mapped using GRCh38/hg38 genome, using Snapgene, for (**i**) *CDKN2A*, (**ii**) *ETV6*, and (**iii**) *IKZF1* to compare their relative positions.

or losses as per CGH, out of which ∼65% (27 of 41) were reckoned by RT-PCR as well (Table 5 and Supplementary Table 2).

While MLPA provides better throughput and can cover the genome more widely, it involves higher skills. The costs vary depending on the lab's settings, the number of probes utilized, and the quantity of reagents used for the analysis. MLPA currently stands as the method of choice for the detection of CNVs, single nucleotide polymorphisms, and other aberrations in

Table 4. Log2-fold change (Log2 FC) and iteration-based probability scores of RT-PCR primers for CNV analysis of ped-B-ALL patients.

Sample code	Analysis type	RB1	BTG1		ETV6		IKZF1		EBF1		PAX5		CDKN2A		CDKN2B	
			Pr 1	Pr 2	Pr 1	Pr 2	Pr 1	Pr 2	Pr 1	Pr 2	Pr 1	Pr 2	Pr 1	Pr 2	Pr 1	Pr 2
B-ALL 1	Log2 FC (wrt controls)	0.61	1.25	7.63	−2.56	−2.56	1.43	1.63	0.66	2.75	2.6	−1.51	−1.36	5.72	0.36	4.1
	Probability of Deletion	0.35	0.27	0	1	0.99	0.03	0.16	0.15	0	0.05	0.97	0.92	0	0.1	0
B-ALL 2	Log2 FC (wrt controls)	0.12	−0.62	6.32	−2.74	−3.06	2.14	2.26	0.37	0.36	4.77	−0.76	7.51	7.14	6.36	4.8
	Probability of Deletion	0.47	0.63	0	1	1	0	0.08	0.26	0.27	0	0.85	0	0	0	0
B-ALL 3	Log2 FC (wrt controls)	0.63	0	3.28	0.66	−4.32	2.2	1.1	0.44	0.16	0.96	0.61	−0.62	3.42	0.86	0
	Probability of Deletion	0.34	0.51	0.03	1	1	0	0.53	0.23	0	0.26	0	0.74	0.03	0	0.5
B-ALL 4	Log2 FC (wrt controls)	−0.54	−3.84	4.22	NA	−1.32	−0.86	0.24	NA	4.02	−1.51	0.82	−5.06	2.45	−1.79	3.9
	Probability of Deletion	0.64	0.17	0	–	0	0.83	0.77	–	0	0.82	0	1	0.04	1	0.5
B-ALL 5	Log2 FC (wrt controls)	1.08	1.61	0.83	−0.3	−0.58	2.73	1.56	0.14	3.56	−0.34	−1.03	−2.32	1.08	−2.84	3.2
	Probability of Deletion	0.24	0.22	0	0.54	0	0	0.17	0.38	0	0.57	0	1	0.22	1	0.8
B-ALL6	Log2 FC (wrt controls)	1.8	1.83	5.69	−0.81	0.56	5.26	2.34	2.5	−1.09	4.86	2.87	4.21	5.81	2.12	2
	Probability of Deletion	0.12	0.19	0	0.83	0.12	0	0.07	0	0	0	0	0	0	0	0.2
B-ALL 7	Log2 FC (wrt controls)	−6.64	3.53	NA	4.22	4.1	2.63	2.92	7	2.24	−0.64	6.15	4.3	2.01	7.7	7.4
	Probability of Deletion	1	0.04	-	0	0	0	0.03	0	0	0.65	0	0	0.11	0	0
B-ALL 8	Log2 FC (wrt controls)	2.42	NA	NA	−1.29	−1.94	2.42	1.79	3.38	3.36	3.04	5.64	7.28	8.35	5.99	5.3
	Probability of Deletion	0.06	–	–	0.96	0.95	0	0.14	0	0	0.02	0	0	0	0	0
B-ALL 9	Log2 FC (wrt controls)	−0.56	0.23	NA	−1.94	−1.79	NA	3.1	2.3	3.8	5.85	−0.45	4.28	7.5	2.47	2.5
	Probability of Deletion	0.65	0.46	-	1	0.93	-	0.03	0	0	0	0.75	0	0	0	0.2
NAC	Log2 FC (wrt controls)	3.05	−2.74	8.39	0.29	−0.54	1.25	2.57	−1.64	0.03	−0.45	−5.64	1.95	1.4	2.63	2.3
	Probability of Deletion	0	0.65	0	0.06	0.29	0	0.03	0.49	0	0.22	1	0	0.03	0	0
K562	Log2 FC (wrt controls)	−1.47	−0.6	0.36	−3.32	−1.94	0.56	0.18	−0.67	0.49	1.17	−3.84	NA	NA	NA	NA
	Probability of Deletion	0.5	0.25	0.02	1	0.85	0	0.36	0.07	0	0.04	0.96	-	-	-	-

Probability of deletion represented as values from 0 to 1; Dash = Probability callings not possible, NA = No amplification in RT-PCR. Pr1 = CNV Primer set 1, Pr2 = CNV Primer set 2, Total samples used= 9 Ped B-ALL samples, 1 NAC, and K562 cell line.

Table 5. Overlap of CGH callings found in RT-PCR.

Category	CGH callings count (out of 56)[a]	Overlap between CGH & RT-PCR (%)
Deletions	3	100 (3 of 3)
Losses	12	83.4 (10 of 12)
Not called for deletions or losses	41	65.8 (27 of 41)
deletions unique to RT-PCR	–	25 (14 of 56)

RT-PCR base deletion called based probability scores using conventional cut-off >0.5 (not using clinical data).
[a] Total samples for both CGH array and RT-PCR based analysis = 5 Ped-B-ALL, 1 NAC, & K562 cell line. Total 56 CNV callings = 8 genes × 7 samples.

genetic disorders such as Duchenne Muscular Dystrophy (DMD) and Spinal Muscular Atrophy (SMA), where probing one or few select genes is necessary [33, 34]. The kits employed in diagnostics to detect CNVs in DMD and SMA facilitate the utilization of a minimum of 1-4 probes per exon, and over 40-60 target regions at a time [35, 36]. By analyzing the ratios (above and below the threshold) of neighboring probes, deletions and/or duplications in specific regions are identified and assigned [37, 38].

Few MLPA kits have also been created and utilized in research only (Probemix iAMP21 ERG [39], Probemix D007 [40], P335 ALL-IKZF1 [41]) to determine CNV profiles in B-ALL [10, 12, 16, 42]. Probemix P335 ALL-IKZF1 has been CE-marked and is employed in diagnostics in selected countries. It covers genes relevant to B-ALL with limited number of probes (PAX5 –7; EBF1 –4; ETV6 –6; RB1–5; IKZF1–8; the CDKN2A/CDKN2B region –3; additionally BTG1 and downstream region –4; and Xp22.33/Yp11.32 region including PAR1 –5) [41]. Additionally, for CNV detection in B-ALL, a robust technique combining MLPA-based next-generation sequencing (NGS), also known as digital-MLPA or MLPA-NGS, has emerged [43]. This advanced method has proven to enhance accuracy by analyzing a larger array of target regions, assessing sub-clonal alterations, allowing smaller initial DNA amounts, and eliminating the need to discriminate fragments based on length.

It must be noted that even for the MLPA-based approach, CNV calling using less than two probes are considered as a loss and recommended to be tested by alternative techniques [42, 45]. Losses identified by CGH or MLPA if not validated by alternative

methods, are currently not considered for clinical decisions. This reflects on the technical limitation of these approaches, rather than the clinical consequences of such micro-deletions.

Continuing in the same manner with CGH, regions that lie outside the target sequence of MLPA probes will not be accounted for and there is a possibility of changes in those regions that remain undetected, rather than absent. Single nucleotide variations and point mutations in the target sequence may also lead to false positive results in MLPA [41]. Copy number changes detected by more than one consecutive probe in MLPA, are also recommended for further validation via other techniques such as array CGH, sequencing, long PCR as well as RT-PCR [41].

Overall, these limitations are shared across techniques (CGH, MLPA), and we propose that a similar criterion can be formulated for RT-PCR-based approach too, which would be a more cost-effective alternative in resource-constrained settings.

Studies have also brought to light how CNVs differ across nation, with geography and regional assessment playing a critical role in establishing their prevalence and frequency. In UKALL2003 [12] study, demonstrated ∼15% cases with IKZF1 deletion versus 40% in another group, whereas, the AIEOP-BFM [16] study conducted in Italy, and Austria, demonstrate only ∼8.3% (83/991) cases with IKZF1 deletions. They go on to define another class of IKZF1 deletion as well—IKZF1^{plus}, where IKZF1 deletion was accompanied by other deletions (CDKN2A/2B, PAR1, PAX5; in the absence of ERG1 deletion), which they observe ∼6.3% (63/991) patients. On the other hand, ICiCLe-2014 [15] conducted in India,

reflected *IKZF1* deletion alone in ~19.5% cases whereas, ~13.4% cases demonstrated *IKZF1*^{plus} deletions. As *IKZF1* is known to impact prognosis and is considered a HR CNV, this data also reflects on the importance of understanding CNVs in control population based on the geography.

To derive sensitivity, specificity, and reproducibility of the proposed RT-PCR based method for clinical setting and application, a rigorous comparison with well-established methods like MLPA or CGH array is inevitable. Similar efforts have been made using a customized NGS panel for CNV detection in Ped B-ALL patients [46].

To determine specificity and sensitivity, each region (especially the deletion hotspots) for CNV callings needs to be compared with the established technique. While we have tried making this comparison, we are currently not able to accurately determine this due to the lack of overlapping primer probes' loci.

Given the limited number and non-overlapping positions of the primers and probes, refining these metrics is a requisite to further develop this method. Additionally, cloning the gene-of-interest(s), and generating knock-out lines for the same are likely to enable in defining optimal PCR conditions for determining specificity and sensitivity of this method. The limit of detection (LoD), reflecting the lowest number of clones with the said CNV that would be detectable via RT-PCR needs to be determined too.

To determine reproducibility, it would be ideal to analyze the same samples repeatedly to compare the callings. Samples must be run in replicates for the same assay. Same sample should be analysed across laboratories, to compare the callings. For this however, overall stability of the nucleic acids must be ensured. RT-PCR relies on an internal normalizing control-based approach. Here we have used *RPP30* as an internal control to determine the reliability of CNV callings. The CT values for *RPP30* expressed as median (interquartile range): 24.545 (23.90–25.31), was consistent across the samples. However, choosing an appropriate internal control is necessary too. Efforts should be made to ensure that the internal control(s) amplify at specific range of CT value. This can ensure that the quantity of the nucleic acids or quality is not limiting.

PCR based CNV detection is currently not a method of choice for diagnostics, however it holds significant potential. On the contrary, the utilization of RT-PCR to serially monitor the progression of a disease via assessing the number of transcripts has been an established practice for Chronic Myeloid Leukemia (CML). Concentric efforts are made to use *BCR::ABL/ABL* expression to report the disease at the time of diagnosis and for regular monitoring [47–49]. Similar meticulous efforts are necessary to further the use of this method for diagnostic purpose.

Limitations of the study and the way ahead

While our proposed method might hold diagnostic potential, there are several aspects that need to be worked upon for developing this method further.

i) This study is performed using limited samples for the method development. To assess the clinical relevance of the method in future, sample size needs to be calculated so that the z score is calculated considering the actual distribution of the data which helps to improve the reliability of the relative callings.

ii) Hotspots in CNVs are identified and may not be a rare feature. DMD [50–53] and SMA [54] genes have been extensively analyzed for characteristic patterns of deletions and/or duplications, and a similar approach needs to be applied

for a detailed analysis of our genes of interest. Currently we do not have any information about the indels or CNVs in healthy population from the geography. Given that sub-microscopic CNVs or indels are common, it is important to identify these for each gene-of-interest. Identifying frequently deleted regions per gene of interest in controls and patients is necessary to identify the appropriate locations for RT-PCR primers.

Similarly, care must be taken to account for regions containing low-frequency, rare variants, and/or geography specific polymorphisms. The need to include more than one primer sets to account for such polymorphisms must be reviewed using healthy populations as reference for each geography.

Accuracy of CNV callings in samples with hyperdiploidy and low blast percentage (minimum 20% blasts) must be evaluated critically to understand if CGH/MLPA versus RT-PCR approach has common or specific limitations.

iii) The non-overlapping position between CGH probes and RT-PCR primers has proven to be a major limitation of the study. It is imperative for the probes and primers to be at the same loci to identify relative specificity and sensitivity of the RT-PCR based CNV detection method. A customized CGH array or MLPA based analysis in comparison to RT-PCR is a must to evaluate the true positives and true negatives or vice versa.

iv) Here, we have used only one internal control gene for CNV or gene expression analysis. Identifying more appropriate internal controls using DNA or cDNA respectively, can provide better handle on the reliability of the CNV callings or relative gene-expression levels.

v) The necessary technical considerations (CT-value cut-offs for control genes, LoD for DNA or cDNA, acceptable variation between duplicates/triplicates of individual samples, technical considerations for reproducibility between laboratories and human resource, quality of reagents, RT-PCR machine, etc.) to define PCR conditions that can ensure reproducibility or reliability of the method have not been achieved.

In conclusion, applying our devised iterative-scoring method in genetic profiling and testing could advance patient risk-stratification. Due to the affordability and minimal technical demands, RT-PCR tests could serve as a primary screening tool during diagnosis for upfront stratification and classification of ped B-ALL patients based on CNV analysis. This would be particularly beneficial in clinical and hospital settings with limited resources. In contrast, methods such as MLPA, digital-MLPA, or CGH for CNV reporting can be used if our RT-PCR based approach fails to identify any CNVs for upfront stratification and/or in patients with relapsed or refractory disease. While gene-expression analysis using RT-PCR based approach can be initially compared with DRP analysis to develop tailored therapeutic approach, it can be restricted to patients with high risk, relapsed or refractory disease, for now. Such measures are likely to contribute significantly towards improving molecular risk-stratification, at different stages, to improve clinical management and outcomes of ped B-ALL patients.

Author contributions

Zoha Sadaqat (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Validation [lead],

Visualization [lead], Writing—original draft [equal], Writing—review & editing [equal]), Smitha Joseph (Data curation [equal], Formal analysis [equal], Methodology [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [supporting]), Chandrika Verma (Validation [supporting], Visualization [supporting], Writing—original draft [supporting], Writing—review & editing [supporting]), Jyothi Muni Reddy (Resources [equal], Writing—review & editing [equal]), Anand Prakash (Project administration [equal], Resources [equal], Writing—review & editing [equal]), Tinku Thomas (Investigation [equal], Methodology [equal], Supervision [equal], Validation [equal], Visualization [equal], Writing—review & editing [equal]), Vandana Bharadwaj (Conceptualization [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Project administration [equal], Resources [equal], Supervision [equal], Writing—original draft [equal], Writing—review & editing [equal]), and Neha Vyas (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [lead], Methodology [equal], Project administration [lead], Resources [lead], Supervision [lead], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal])

Supplementary data

Supplementary data is available at *Biology Methods and Protocols* online.

Conflict of interest statement. None declared.

Funding

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

1. Carobolante F, Chiaretti S, Skert C, Bassan R. Practical guidance for the management of acute lymphoblastic leukemia in the adolescent and young adult population. *Ther Adv Hematol* 2020;**11**: 2040620720903531. https://doi.org/10.1177/2040620720903531
2. Das N, Banavali S, Bakhshi S *et al*. Protocol for ICiCLe-ALL-14 (InPOG-ALL-15-01): a prospective, risk stratified, randomised, multicentre, open label, controlled therapeutic trial for newly diagnosed childhood acute lymphoblastic leukaemia in India. *Trials* 2022;**23**:102. https://doi.org/10.1186/S13063-022-06033-1
3. Pui CH, Evans WE. A 50-year journey to cure childhood acute lymphoblastic leukemia. *Semin Hematol* 2013;**50**:185–96. https://doi.org/10.1053/J.SEMINHEMATOL.2013.06.007
4. *Surveillance, Epidemiology, and End Result Programs (SEER)*. Bethesda, MD: National Cancer Institute (NCI). https://seer.cancer.gov/statistics-network/explorer/application.html?site=92&data_type=4&graph_type=6&compareBy=age_range&chk_age_range_1=1&chk_age_range_9=9&chk_age_range_141=141&chk_age_range_157=157&chk_age_range_62=62&sex=1&race=1&hdn_stage=101&advopt (18 May 2024, date last accessed).
5. Moorman AV, Antony G, Wade R *et al*. Time to cure for childhood and young adult acute lymphoblastic leukemia is independent of early risk factors: long-term follow-up of the UKALL2003 trial. *J Clin Oncol* 2022;**40**:4228–39. https://doi.org/10.1200/JCO.22.00245/ASSET/IMAGES/LARGE/JCO.22.00245APP2.JPEG
6. Arya LS, Kotikanyadanam SP, Bhargava M *et al*. Pattern of relapse in childhood ALL: challenges and lessons from a uniform treatment protocol. *J Pediatr Hematol Oncol* 2010;**32**:370–5. https://doi.org/10.1097/MPH.0B013E3181D7AE0D
7. Magrath I, Shanta V, Advani S *et al*. Treatment of acute lymphoblastic leukaemia in countries with limited resources; lessons from use of a single protocol in India over a twenty year period. *Eur J Cancer* 2005;**41**:1570–83. https://doi.org/10.1016/J.EJCA.2004.11.004
8. Kulkarni KP, Marwaha RK, Trehan A, Bansal D. Survival outcome in childhood ALL: experience from a tertiary care centre in North India. *Pediatr Blood Cancer* 2009;**53**:168–73. https://doi.org/10.1002/PBC.21897
9. Arora RS, Arora B. Acute leukemia in children: a review of the current Indian data. *South Asian J Cancer* 2016;**5**:155–60. https://doi.org/10.4103/2278-330X.187591
10. Thakral D, Kaur G, Gupta R *et al*. Rapid identification of key copy number alterations in B- and T-cell acute lymphoblastic leukemia by digital multiplex ligation-dependent probe amplification. *Front Oncol* 2019;**9**:871. https://doi.org/10.3389/FONC.2019.00871/BIBTEX
11. Patkar N, Subramanian P, Tembhare P *et al*. An integrated genomic profile that includes copy number alterations is highly predictive of minimal residual disease status in childhood precursor B-lineage acute lymphoblastic leukemia. *Indian J Pathol Microbiol* 2017;**60**:209–13. https://doi.org/10.4103/IJPM.IJPM_466_16
12. Moorman AV, Enshaei A, Schwab C *et al*. A novel integrated cytogenetic and genomic classification refines risk stratification in pediatric acute lymphoblastic leukemia. *Blood* 2014;**124**:1434–44. https://doi.org/10.1182/BLOOD-2014-03-562918
13. Stanulla M, Cavé H, Moorman AV. IKZF1 deletions in pediatric acute lymphoblastic leukemia: still a poor prognostic marker? *Blood* 2020;**135**:252–60. https://doi.org/10.1182/BLOOD.2019000813
14. Ostergaard A, Enshaei A, Pieters R *et al*. The prognostic effect of IKZF1 deletions in ETV6:: RUNX1 and high hyperdiploid childhood acute lymphoblastic leukemia. *HemaSphere* 2023;**7**:E875. https://doi.org/10.1097/HS9.0000000000000875
15. Gupta SK, Singh M, Chandrashekar PH *et al*. Clinical and prognostic impact of copy number alterations and associated risk profiles in a cohort of pediatric B-cell precursor acute lymphoblastic leukemia cases treated under ICiCLe protocol. *HemaSphere* 2022;**6**:E782. https://doi.org/10.1097/HS9.0000000000000782
16. Stanulla M, Dagdan E, Zaliova M, International BFM Study Group *et al*. IKZF1plus defines a new minimal residual disease-dependent very-poor prognostic profile in pediatric B-cell precursor acute lymphoblastic leukemia. *J Clin Oncol* 2018;**36**:1240–9. https://doi.org/10.1200/JCO.2017.74.3617
17. Sidhu J, Steffen FD, Jimenez IA *et al*. Phenotypic drug response profiling identifies asparaginase-based synergistic combinations for very high risk acute lymphoblastic leukaemia. *Blood* 2023;**142**:417–417. https://doi.org/10.1182/BLOOD-2023-172885
18. Bell AD, Usher CL, McCarroll SA. Analyzing copy number variation with droplet digital PCR. *Methods Mol Biol* 2018;**1768**:143–60. https://doi.org/10.1007/978-1-4939-7778-9_9
19. Szewczak E, Bondarzewski A. Is the assessment of interlaboratory comparison results for a small number of tests and limited number

of participants reliable and rational? *Accred Qual Assur* 2016;**21**: 91–100. https://doi.org/10.1007/S00769-016-1195-Y/TABLES/1

20. ISO-10534-1-1996. International standard iTeh standard iTeh standard preview. *Int Organ Stand* 2015;**20**:3–6.

21. Rousseeuw PJ, Verboven S. Robust estimation in very small samples. *Comput Stat Data Anal* 2002;**40**:741–58. https://doi.org/10.1016/S0167-9473(02)00078-6

22. Sacedón R, Díez B, Nuñez V *et al.* Sonic hedgehog is produced by follicular dendritic cells and protects germinal center B cells from apoptosis. *J Immunol* 2005;**174**:1456–61. https://doi.org/10.4049/JIMMUNOL.174.3.1456

23. Dalal H, Subramanian S, V P S *et al.* Exovesicular-Shh confers Imatinib resistance by upregulating Bcl2 expression in chronic myeloid leukemia with variant chromosomes. *Cell Death Dis* 2021;**12**:259–17. https://doi.org/10.1038/s41419-021-03542-w

24. Schwab CJ, Chilton L, Morrison H *et al.* Genes commonly deleted in childhood B-cell precursor acute lymphoblastic leukemia: association with cytogenetics and clinical features. *Haematologica* 2013;**98**:1081–8. https://doi.org/10.3324/HAEMATOL.2013.085175

25. Boer JM, Van Der Veer A, Rizopoulos D *et al.* Prognostic value of rare IKZF1 deletion in childhood B-cell precursor acute lymphoblastic leukemia: an international collaborative study. *Leukemia* 2016;**30**:32–8. https://doi.org/10.1038/LEU.2015.199

26. Fang Q, Song Y, Gong X *et al.* Gene deletions and prognostic values in B-linage acute lymphoblastic leukemia. *Front Oncol* 2021; **11**:677034.https://doi.org/10.3389/FONC.2021.677034/BIBTEX

27. Iolascon A, Giordani L, Moretti A *et al.* Analysis of CDKN2A, CDKN2B, CDKN2C, and cyclin Ds gene status in hepatoblastoma. *Hepatology* 1998;**27**:989–95. https://doi.org/10.1002/HEP.510270414

28. Enshaei A, O'Connor D, Bartram J *et al.* A validated novel continuous prognostic index to deliver stratified medicine in pediatric acute lymphoblastic leukemia. *Blood* 2020;**135**:1438–46. https://doi.org/10.1182/BLOOD.2019003191

29. DelRocco NJ, Loh ML, Borowitz MJ *et al.* Enhanced risk stratification for children and young adults with B-cell acute lymphoblastic leukemia: a children's oncology group report. *Leuk* 2024; **38**:720–8. https://doi.org/10.1038/s41375-024-02166-1

30. Lejman M, Chałupnik A, Chilimoniuk Z, Dobosz M. Genetic biomarkers and their clinical implications in B-cell acute lymphoblastic leukemia in children. *Int J Mol Sci* 2022;**23**:2755.https://doi.org/10.3390/ijms23052755

31. Forero-Castro M, Robledo C, Benito R *et al.* Genome-wide DNA copy number analysis of acute lymphoblastic leukemia identifies new genetic markers associated with clinical outcome. *PLoS One* 2016;**11**:e0148972. https://doi.org/10.1371/journal.pone.0148972

32. Moorman AV, Enshaei A, Schwab C *et al.* A novel integrated cytogenetic and genomic classification refines risk stratification in pediatric acute lymphoblastic leukemia. *Blood* 2014;**124**: 1434–44. https://doi.org/10.1182/blood-2014-03

33. Schouten JP, McElgunn CJ, Waaijer R *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 2002;**30**: e57-e57. https://doi.org/10.1093/NAR/GNF056

34. Stuppia L, Antonucci I, Palka G, Gatta V. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci* 2012;**13**:3245–76. https://doi.org/10.3390/IJMS13033245

35. MRC Holland. *SALSA Probemix for SMA.* https://www.mrcholland.com/product/P021/634 (16 May 2024, date last accessed).

36. MRC Holland. *SALSA Probemix for DMD, BMD.* https://www.mrcholland.com/product/P034/744 (16 May 2024, date last accessed).

37. Jeuken J, Cornelissen S, Boots-Sprenger S *et al.* Multiplex ligation-dependent probe amplification: a diagnostic tool for simultaneous identification of different genetic markers in glial tumors. *J Mol Diagn* 2006;**8**:433–43. https://doi.org/10.2353/JMOLDX.2006.060012

38. Lalic T, Vossen RHAM, Coffa J *et al.* Deletion and duplication screening in the DMD gene using MLPA. *Eur J Hum Genet* 2005;**13**: 1231–4. https://doi.org/10.1038/sj.ejhg.5201465

39. *SALSA MLPA Probemix P327 iAMP21-ERG.* https://www.mrcholland.com/product/P327 (12 December 2024, date last accessed).

40. *SALSA MLPA Probemix D007 Acute Lymphoblastic Leukemia.* https://www.mrcholland.com/product/D007 (7 December 2024, date last accessed).

41. *SALSA MLPA Probemix P335 C2 ALL-IKZF1.* MRC Holland. https://www.mrcholland.com/product/P335 (5 December 2024, date last accessed).

42. Chang TC, Chen W, Qu C *et al.* Genomic determinants of outcome in acute lymphoblastic leukemia. *J Clin Oncol* 2024;**42**:3491–503. https://doi.org/10.1200/JCO.23.02238/SUPPL_FILE/DS_JCO.23.02238.XLSX

43. Benard-Slagter A, Zondervan I, de Groot K *et al.* Digital multiplex ligation-dependent probe amplification for detection of key copy number alterations in T- and B-cell lymphoblastic leukemia. *J Mol Diagn* 2017;**19**:659–72. https://doi.org/10.1016/J.JMOLDX.2017.05.004

44. *SMA MLPA kit.* MRC Holland. https://www.mrcholland.com/product/P021 (18 July 2024, date last accessed).

45. *DMD MLPA kit.* MRC Holland. https://www.mrcholland.com/product/P034 (18 July 2024, date last accessed).

46. Montaño A, Hernández-Sánchez J, Forero-Castro M *et al.* Comprehensive custom NGS panel validation for the improvement of the stratification of B-acute lymphoblastic leukemia patients. *J Pers Med* 2020;**10**:1–21. https://doi.org/10.3390/JPM10030137

47. Wang L, Pearson K, Pillitteri L *et al.* Serial monitoring of BCR-ABL by peripheral blood real-time polymerase chain reaction predicts the marrow cytogenetic response to imatinib mesylate in chronic myeloid leukaemia. *Br J Haematol* 2002;**118**:771–7. https://doi.org/10.1046/J.1365-2141.2002.03705.X

48. Branford S, Hughes TP, Rudzki Z. Monitoring chronic myeloid leukaemia therapy by real-time quantitative PCR in blood is a reliable alternative to bone marrow cytogenetics. *Br J Haematol* 1999;**107**: 587–99. https://doi.org/10.1046/J.1365-2141.1999.01749.X

49. Hochhaus A, Lin F, Reiter A *et al.* Quantification of residual disease in chronic myelogenous leukemia patients on interferon-α therapy by competitive polymerase chain reaction. *Blood* 1996;**87**:1549–55. https://doi.org/10.1182/BLOOD.V87.4.1549.BLOODJOURNAL8741549

50. White SJ, Aartsma-Rus A, Flanigan KM *et al.* Duplications in the DMD gene. *Hum Mutat* 2006;**27**:938–45. https://doi.org/10.1002/HUMU.20367

51. Zimowski JG, Pawelec M, Purzycka JK *et al.* Deletions, not duplications or small mutations, are the predominante new mutations in the dystrophin gene. *J Hum Genet* 2017;**62**:885–8. doi: 10.1038/jhg.2017.70

52. Thakur N, Abeysekera G, Wanigasinghe J, Dissanayake VHW. The spectrum of deletions and duplications in the dystrophin (DMD) gene in a cohort of patients with Duchenne muscular dystrophy in Sri Lanka. *Neurol India* 2019;**67**:714–5. https://doi.org/10.4103/0028-3886.263235

53. Shariati G, Shakerian S, Mohammadi Anaie M *et al.* Deletion and duplication mutations spectrum in Duchenne muscular dystrophy in the southwest of Iran. *Meta Gene* 2020;**23**:100641. https://doi.org/10.1016/J.MGENE.2019.100641

54. Vijzelaar R, Snetselaar R, Clausen M *et al.* The frequency of SMN gene variants lacking exon 7 and 8 is highly population dependent. *PLoS One* 2019;**14**:e0220211. https://doi.org/10.1371/JOURNAL.PONE.0220211