npj | genomic medicine

Article

Published in partnership with CEGMR, King Abdulaziz University

https://doi.org/10.1038/s41525-026-00560-5

Proof-of-concept study for the detection of somatic structural variant driver alterations using HiFi long-read sequencing in a pediatric leukemia cohort

Lisa A. Lansdon[1,2], Byunggil Yoo[1,3], Ayse Keskus[4], Irina Pushel[3], Chengpeng Bi[2,3], Tanveer Ahmad[4], Asher Bryant[4], Adam Walter[3], Margaret Gibson[1,3], Mary Rindler[1], Weijie Li[1,2], Sultan M. Habeebu[1,2], Linda D. Cooley[1,2], John Herriges[1,2], Elena Repnikova[1,2], Lei Zhang[1,2], Keith J. August[2,5], Terrie G. Flatt[2,5], Alan S. Gamis[2,5], Erin M. Guest[2,5], J. Allyson Hays[2,5], Maxine Hetherington[2,5], Karen Lewing[2,5], Tomi Pastinen[2,3], Mikhail Kolmogorov[4] & Midhat S. Farooqi[1,2,3] ✉

Gene fusions are common primary drivers of pediatric leukemias and are the result of underlying structural variants (SVs). Current clinical workflows to detect such alterations rely on a multimodal approach, which often increases analysis time and overall cost of testing. In this study, we used long-read sequencing (lrSeq) as a proof-of-concept to determine whether clinically relevant (cr) SVs could be detected within a small (n = 17) pediatric leukemia cohort. We show that this methodology successfully determined all known crSVs (n = 5/5) detected through routine clinical testing. This approach also identified crSVs that resulted in the classification of a leukemia genetic subtype for four additional patients (n = 4/12), such as an ins(11;10)(q23.3;p12p12) forming a *KMT2A*::*MLLT10* fusion, that were missed by routine clinical approaches. This study demonstrates the diagnostic potential of lrSeq as an assay for SV detection in pediatric leukemia and supports lrSeq as a valuable tool for the accurate detection of crSVs.

Pediatric leukemias are the most common cancers of childhood and are largely defined by genetically distinct subgroups that confer prognostic significance[1–5]. Gene fusions are the genetic drivers of many of these subgroups and often result from underlying chromosomal translocations and/or other structural variants (SVs)[1,3,6,7]. Current routine clinical diagnostic testing algorithms for pediatric leukemia rely on a series of complementary yet distinct approaches to identify SVs, including karyotyping, fluorescence in situ hybridization (FISH), chromosomal microarray, DNA sequencing, RNA sequencing (followed by expression profiling and/or fusion panel analysis), and optical genome mapping: an emerging technology in clinical laboratory practice[6,8–11]. Although generally effective, this approach has several drawbacks: it necessitates that sufficient neoplastic sample be available to perform a battery of tests until a genetic driver alteration is identified; it is often limited to neoplastic-only (also known as "tumor-only") analyses, which cannot distinguish between true somatic events versus those that are in the patient's germline; and, finally, it results in an overall increased cost of testing and time to diagnosis due to the use of multiple different methods.

Long-read sequencing (lrSeq) allows for the improved resolution of historically difficult-to-sequence regions of the genome, as well as superior detection and delineation of SVs compared with other existing technologies[12,13]. This approach has been used for variant detection in rare disease[14,15], and other studies have demonstrated the utility of lrSeq in adult cancer testing, which successfully identified known SVs, and revealed novel SVs, in colorectal cancer[16], lung cancer[17], mantle cell lymphoma[18], acute myeloid leukemia (AML)[19], and hereditary cancer[20,21], among others[22,23]. Some lrSeq studies have been performed in a pediatric cancer setting, notably for pediatric medulloblastoma[24] and pediatric leukemia[25,26]; however, the diagnostic utility of whole genome HiFi lrSeq

[1]Department of Pathology and Laboratory Medicine, Children's Mercy Kansas City, Kansas City, MO, USA. [2]Departments of Pathology and Pediatrics, University of Missouri Kansas City – School of Medicine, Kansas City, MO, USA. [3]Genomic Medicine Center, Children's Mercy Kansas City, Kansas City, MO, USA. [4]Cancer Data Science Laboratory, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [5]Division of Hematology/Oncology, Children's Mercy Kansas City, Kansas City, MO, USA. ✉e-mail: midhatsf@gmail.com

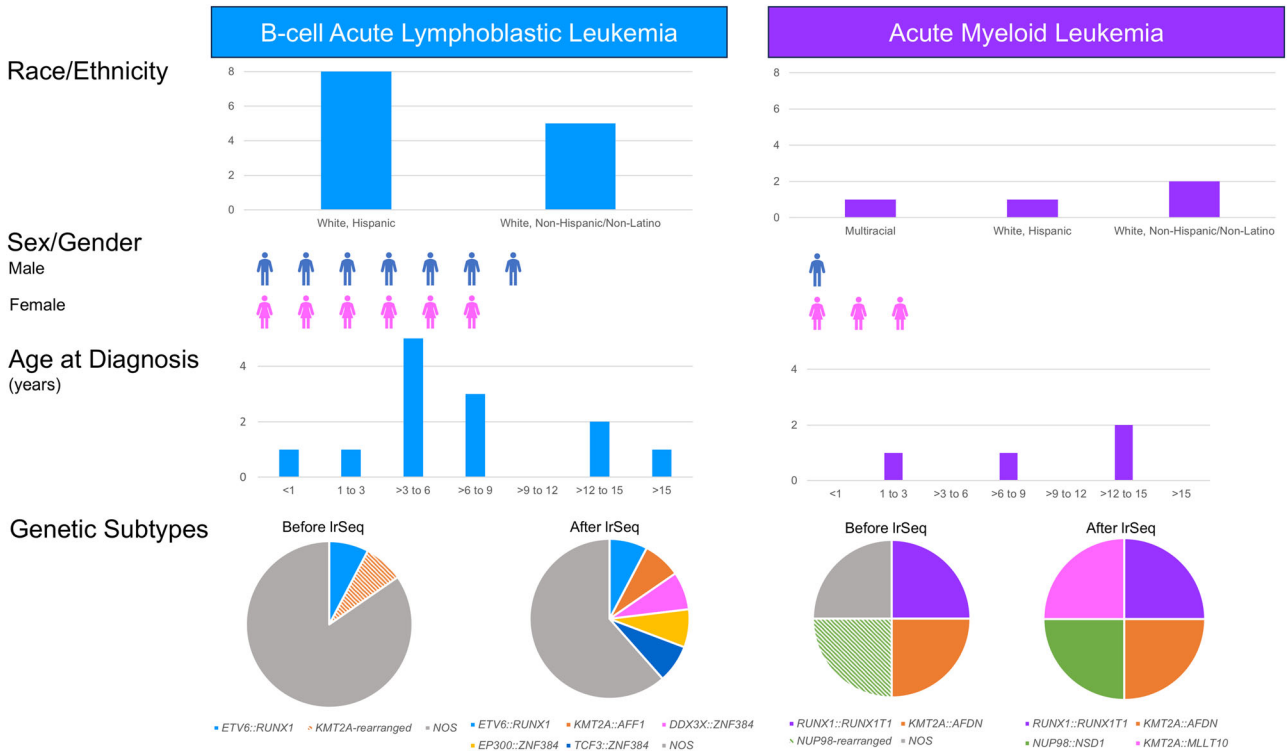

Fig. 1 | Pediatric leukemia cohort demographics. Seventeen pediatric patients were included in the present study, comprising of 13 individuals with B-cell acute lymphoblastic leukemia and 4 individuals with acute myeloid leukemia. The race/ethnicity, sex/gender, and age at diagnosis for each patient are specified. In addition, genetic subtypes of the 5 "known" and 12 "unknown" genetic subtypes (including the 4 "unknown" cases that were resolved in the present study) are depicted. Striped pie chart segments indicate cases for which one of the two partner genes involved in the structural variant was not known prior to lrSeq.

for SV calling within pediatric cancer (and pediatric leukemia, specifically) has not been explored to date.

Here, we investigate the feasibility of SV detection by HiFi lrSeq in a small pediatric leukemia cohort ($n = 17$ cases) selected from patients with known ($n = 5$) and unknown ($n = 12$) genetic subtypes. The demographics of the cohort assessed in this study are summarized in Fig. 1 (see also Supplementary Data 1). The lrSeq data generated from archived DNA for each case were assessed using a recently developed somatic SV detection tool, Severus[27], which uses a phased breakpoint graph approach for the resolution of complex rearrangements[28–31]. Severus can utilize paired tumor/normal data for improved, true somatic SV detection, or assess neoplastic data in a "tumor-only" fashion (i.e., without a paired normal sample). It also showed the best performance compared to currently available lrSeq SV callers when assessing a panel of tumor cell lines[27]. We first aimed to determine whether lrSeq, paired with SV calling by Severus in both tumor/normal and tumor-only modes, could result in the accurate detection of SVs in cases with known alterations. We then explored whether additional cases with unknown genetic subtypes after routine clinical testing might be resolved by this approach.

Results

Quality control metrics of lrSeq data

For the lrDNA-seq data, an average of 77.7 Gb of data were generated per SMRT Cell (with 1 SMRT Cell sequenced per sample; range: 18.28–107.6 Gb), with an average read length of 14.7 kb (range: 10.86–17.43 kb), average N50 of 15.3 kb (range: 11.2–18.1 kb), and average read quality score of 33.6 (range: 29–39). For the lrRNA-seq data, an average of 29.8 Gb of data were generated per SMRT Cell (13.9 Gb and 45.7 Gb for Case 3 and Case 16, respectively), with an average read length of 4.3 kb (3.94 kb and 4.65 kb, respectively), average N50 of 4.1 kb (3.77 and 4.44, respectively), and average read quality score of 39.5 (38 and 41, respectively). Please see Supplementary Data 3 for sample-specific QC metrics.

Tumor/normal analysis of leukemia samples with known structural variants

Five tumor/normal pairs had known, diagnostically significant, SVs previously detected by routine clinical testing. All five of these cases were successfully detected by lrSeq and Severus. This included two B-ALLs, one with a t(12;21)(p13;q22) resulting in *ETV6::RUNX1* fusion and the other with a t(4;11)(q21;q23) expected to result in *KMT2A::AFF1* fusion (formerly known as *MLL::AF4*), and three AMLs, one with t(8;21)(q22;q22) resulting in *RUNX1::RUNX1T1* fusion, one with t(6;11)(q27;q23) resulting in *KMT2A::AFDN* fusion, and one with a *NUP98* rearrangement identified via FISH but without additional testing performed to identify the partner gene. Breakend analysis of the lrSeq data using Severus detected each of these rearrangements. In addition, Severus provided the exact genomic breakpoints within each gene, confirmed that *AFF1* was rearranged with *KMT2A* in the t(4;11)(q21;q23) B-ALL, and also identified the unknown fusion partner in the AML case with the *NUP98* rearrangement, calling a t(5;11)(q35;p15.5) resulting in *NUP98::NSD1* fusion. The specific details of these rearrangements, including the breakpoints within each gene, are listed in Table 1 (see also Supplementary Data 1).

Tumor/normal analysis of genetically undefined leukemia samples

We next examined four tumor/normal pairs with a heretofore genetically undefined leukemia subtype. These four cases had previously undergone routine clinical genetic characterization but did not have a WHO diagnostic genetic subtype identified. All four had lrSeq DNA data analyzed for clinically relevant SVs. On average, Severus detected 14 somatic breakend calls per sample (range 0–60; see Supplementary Data 4), and two of these four genetically unknown cases were successfully defined by this approach (see Table 1 and Supplementary Data 1). One was an AML with an ins(11;10)(q23.3;p12p12) resulting in *KMT2A::MLLT10* fusion (Fig. 2), and

https://doi.org/10.1038/s41525-026-00560-5

Table 1 | Summary of crSV findings for the study cohort

Case number	Study case type	Leukemia genetic subtype post-clinical testing	FISH panel SVs	Clinically detected diagnostic SV	Array	RNA Fusion Testing	Expected SV for lrSeq (if applicable)	lrSeq Diagnostic Finding (Severus Run Mode)	Genes involved in translocation called by lrSeq	Transcript	Translocation breakpoints (GRCh38)	Breakpoint location	Leukemia genetic subtype post-lrSeq
Case 1	Known Diagnostic SV	B-lymphoblastic leukemia/ lymphoma with t(12;21) (p13.2;q22.1); *ETV6::RUNX1*	*ETV6::RUNX1*	t(12;21) (p13;q22)	No patterns suggestive of diagnostic SVs	Not Performed	*ETV6::RUNX1*	*ETV6::RUNX1*	*ETV6*	NM_001987.5	chr12:11880011	Intron 5	B-lymphoblastic leukemia/ lymphoma with t(12;21) (p13.2;q22.1); *ETV6::RUNX1*
								(TN & TO)	*RUNX1*	NM_001754.5	chr21:34906687	Intron 2	
Case 2	Known Diagnostic SV	Acute myeloid leukemia, t(8;21) (q22;q22.1); *RUNX1::RUNX1T1*	*RUNX1::RUNX1T1*	t(8;21) (q22;q22)	No patterns suggestive of diagnostic SVs	Not Performed	*RUNX1::RUNX1T1*	*RUNX1::RUNX1T1*	*RUNX1*	NM_001754.5	chr21:34858376	Intron 6	Acute myeloid leukemia, t(8;21) (q22;q22.1); *RUNX1::RUNX1T1*
								(TN & TO)	*RUNX1T1*	NM_175634.3	chr8:92049540	Intron 2	
Case 3	Known Diagnostic SV	Acute myeloid leukemia, t(6;11) (q27;q23.3); *KMT2A::AFDN*	*KMT2A* rearrangement	t(6;11) (q27;q23.3)	No patterns suggestive of diagnostic SVs	Not Performed	*KMT2A::AFDN*	*KMT2A::AFDN*	*KMT2A*	NM_001197104.2	chr11:118486554	Intron 10	Acute myeloid leukemia, t(6;11) (q27;q23.3); *KMT2A::AFDN*
								(TN & TO)	*AFDN*	NM_001386888.1	chr6:167846482	Intron 1	
Case 4	Known Diagnostic SV (Fusion Partner Resolved by Study)	B-lymphoblastic leukemia/ lymphoma with t(v;11q23.3); *KMT2A*-rearranged	*KMT2A* rearrangement	t(4;11) (q21.3;q23.3)	No patterns suggestive of diagnostic SVs	Not Performed	*KMT2A::?*	*KMT2A::AFF1*	*KMT2A*	NM_001197104.2	chr11:118488445	Intron 10	B-lymphoblastic leukemia/ lymphoma with t(4;11) (q21.3;q23.3); *KMT2A::AFF1*
								(TN & TO)	*AFF1*	NM_001166693.3	chr4:87078149	Intron 4	
Case 5	Known Diagnostic SV (Fusion Partner Resolved by Study)	Acute myeloid leukemia, *NUP98*-rearranged	*NUP98* rearrangement	None	No patterns suggestive of diagnostic SVs	Not Performed	*NUP98::?*	*NUP98::NSD1*	*NUP98*	NM_016320.5	chr11:3743500	Intron 12	Acute myeloid leukemia, t(5;11) (q35.3;p15.4); *NUP98::NSD1*
								(TN & TO)	*NSD1*	NM_022455.5	chr5:177226488	Intron 5	
Case 7	Unknown; Later Resolved by Study	Acute myeloid leukemia, NOS	Negative AML Panel	None	No patterns suggestive of diagnostic SVs	Not Performed	N/A	*KMT2A::MLLT10*	*KMT2A*	NM_001197104.2	chr11:118483699	Intron 8	Acute myeloid leukemia, ins(11;10) (q23.3;p12p12) *KMT2A::MLLT10*
								(TN & TO)	*MLLT10*	NM_001195626.3	chr10:21615044	Intron 7	
Case 17	Unknown; Later Resolved by Study	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel*	None	No patterns suggestive of diagnostic SVs	Not Performed	N/A	*TCF3::ZNF384*	*TCF3*	NM_003200.5	chr19:1621339	Intron 11	B-lymphoblastic leukemia/ lymphoma with t(12;19) (p13.31;p13.3) *TCF3::ZNF384*
								(TN & TO)	*ZNF384*	NM_001135734.1	chr12:6681474	Intron 2	
Case 11	Unknown; Later Resolved by Study	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel	None	No patterns suggestive of diagnostic SVs	Not Performed	N/A	*DDX3X::ZNF384*	*DDX3X*	NM_001356.3	chrX:41343609	Intron 7	B-lymphoblastic leukemia/ lymphoma with t(X;12) (p11.4;p13.31) *DDX3X::ZNF384*
								(TO)	*ZNF384*	NM_001385745.1	chr12:6687022	Intron 2	
Case 16	Unknown; Later Resolved by Study	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel	None	No patterns suggestive of diagnostic SVs	Negative Fusion Multiplex Assay	N/A	*EP300::ZNF384*	*EP300*	NM_001429.4	chr22:41133508	Intron 6	B-lymphoblastic leukemia/ lymphoma with t(12;22) (p13.31;q13.2) *EP300::ZNF384*
								(TO)	*ZNF384*	NM_001385745.1	chr12:6689740	Upstream (5'UTR)	
Case 6	Unknown; Remained Unknown	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel	None	No patterns suggestive of diagnostic SVs	Negative Fusion Multiplex Assay	N/A	No clinically significant SVs	N/A	N/A	N/A	N/A	B-lymphoblastic leukemia/ lymphoma, NOS
								(TN & TO)					

Table 1 (continued) | Summary of crSV findings for the study cohort

Case number	Study case type	Leukemia genetic subtype post-clinical testing	FISH panel SVs	Clinically detected diagnostic SV	Array	RNA Fusion Testing	Expected SV for lrSeq (if applicable)	lrSeq Diagnostic Finding (Severus Run Mode)	Genes involved in translocation called by lrSeq	Transcript	Translocation breakpoints (GRCh38)	Breakpoint location	Leukemia genetic subtype post-lrSeq
Case 8	Unknown; Remained Unknown	B-lymphoblastic leukemia/ lymphoma, BCR-ABL1-like	Negative ALL Panel	None	No patterns suggestive of diagnostic SVs	Not Performed	N/A	No clinically significant SVs (TO)	N/A	N/A	N/A	N/A	B-lymphoblastic leukemia/ lymphoma, BCR-ABL1-like
Case 9	Unknown; Remained Unknown	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel	None	Not Performed	Not Performed	N/A	No clinically significant SVs (TN & TO)	N/A	N/A	N/A	N/A	B-lymphoblastic leukemia/ lymphoma, NOS
Case 10	Unknown; Remained Unknown	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel	None	Not Performed	Not Performed	N/A	No clinically significant SVs (TO)	N/A	N/A	N/A	N/A	B-lymphoblastic leukemia/ lymphoma, NOS
Case 13	Unknown; Remained Unknown	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel	None	Not Performed	Not Performed	N/A	No clinically significant SVs (TO)	N/A	N/A	N/A	N/A	B-lymphoblastic leukemia/ lymphoma, NOS
Case 14	Unknown; Remained Unknown	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel	None	No patterns suggestive of diagnostic SVs	Not Performed	N/A	No clinically significant SVs (TO)	N/A	N/A	N/A	N/A	B-lymphoblastic leukemia/ lymphoma, NOS
Case 15	Unknown; Remained Unknown	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel	None	No patterns suggestive of diagnostic SVs	Negative Fusion Multiplex Assay	N/A	No clinically significant SVs (TO)	N/A	N/A	N/A	N/A	B-lymphoblastic leukemia/ lymphoma, NOS
Case 12	Unknown; Remained Unknown	B-lymphoblastic leukemia/ lymphoma, NOS	Negative ALL Panel	None	No patterns suggestive of diagnostic SVs	Not Performed	N/A	No clinically significant SVs (TO)	N/A	N/A	N/A	N/A	B-lymphoblastic leukemia/ lymphoma, NOS

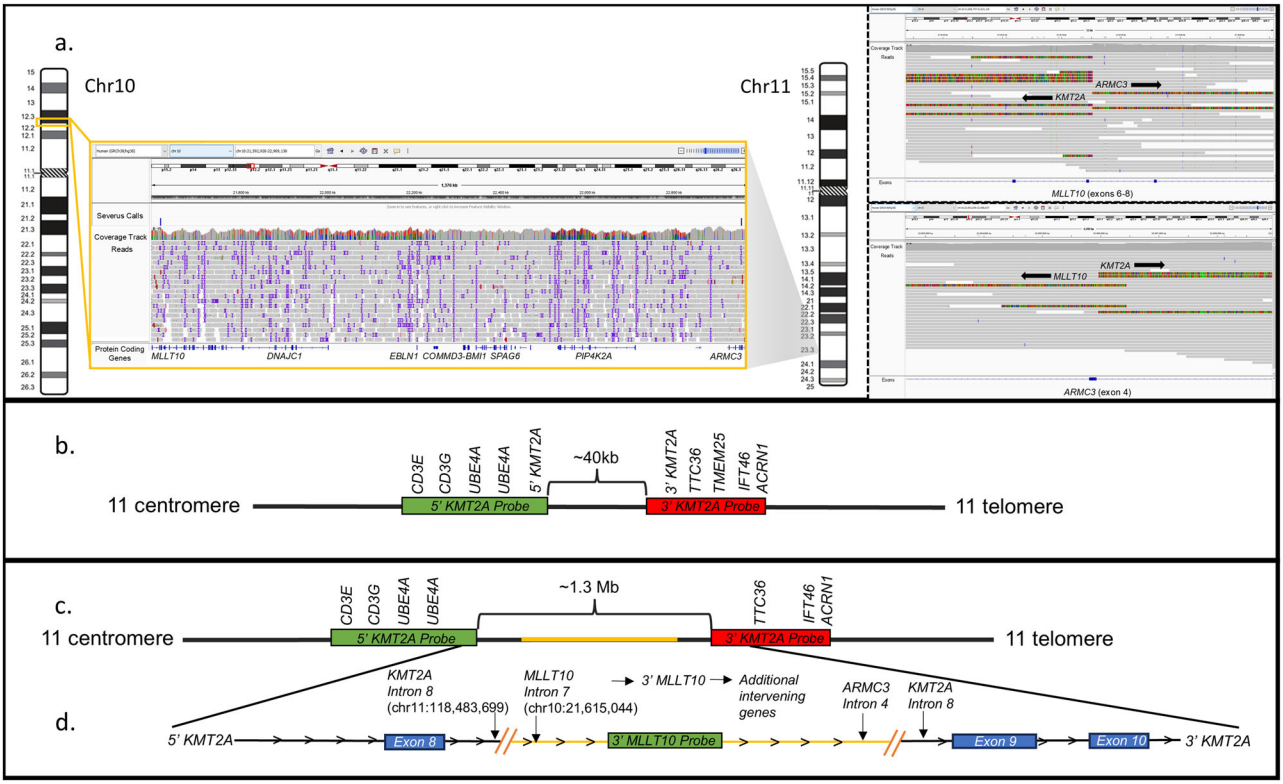

Fig. 2 | *KMT2A::MLLT10* fusion detected by Severus in a previously genetically undefined acute myeloid leukemia case. a Detection of an ~1.3 Mb segment of 10p12 that is called by Severus (indicated by the blue bars in the "Severus Calls" track of the IGV screen capture) and corresponding long-read sequencing (lrSeq) reads. This segment is inserted into the *KMT2A* gene on chromosome 11q23, as demonstrated by the zoomed-in version of the reads within the dashed box. **b** Typical spacing of the *KMT2A* break-apart FISH probe utilized for the in-house acute myeloid leukemia (AML) panel. The probes are represented by their respective colors. **c** Illustration of the increased spacing of the probes that is expected after the insertion of the ~1.3 Mb segment of 10p12 (indicated by the yellow bar) into *KMT2A*. **d** Schematic of the orientation of the *KMT2A* (NM_001197104.2) and *MLLT10* (NM_001195626.3) genes including the intronic location of each breakpoint; a functional fusion product is predicted.

the other was a B-ALL with a t(12;19)(p13;p13) resulting in *TCF3::ZNF384* fusion (Fig. 3).

The AML with the ins(11;10)(q23.3;p12p12) was further investigated to compare lrSeq findings with those of prior clinical genetic testing, especially since a FISH AML panel (which included a *KMT2A* breakapart probe) had been performed clinically but was negative. Manual review of the lrSeq data demonstrated that an ~1.3 Mb segment of chromosome 10p within 10p12.33p12.2 (extending from intron 7 of *MLLT10* (NM_001195626.3) to intron 4 of *ARMC3* (NM_173081.5)) was inserted into intron 8 of the *KMT2A* gene (NM_001197104.2) on 11q23.3 (Fig. 2a). This insertion would cause the 5′ and 3′ probes of *KMT2A* to further separate from their baseline of ~40 kb (i.e., no rearrangement present; Fig. 2b), but only to a total distance of ~1.4 Mb; a difference too small to visualize in FISH or chromosome analyses (i.e., cytogenetically cryptic; Fig. 2c). Follow-up testing using a breakapart *MLLT10* FISH probe confirmed the presence of a rearrangement involving *MLLT10*. Furthermore, the use of lrSeq data allowed for the exact breakpoints of this balanced insertion to be assessed, which demonstrated that a functional fusion product would be expected due to the 5′ to 3′ orientation of the inserted 10p sequence (Fig. 2d). Of note, RNA was not available for this case for additional confirmatory testing.

The SV calls from the B-ALL with the t(12;19)(p13;p13) were similarly compared with prior clinical testing results to assess the discrepancy in crSV detection. Analysis of the SV breakpoints demonstrated that this was a balanced reciprocal translocation between chromosomes 12p13.31 and 19p13.3 (Fig. 3a, b), with a presumed functional product resulting from the fusion of the 5′ end of *TCF3* (NM_003200.5) through intron 11 with intron 2 of *ZNF384* (NM_001385745.1) through the 3′ end of the gene on the derivative chromosome 19 (Fig. 3b). Given its cytogenetically cryptic and balanced nature, this SV was not detected by clinical chromosome and array studies; RNA-based fusion testing had not been performed. A clinical FISH ALL panel had been performed, but neither of the genes involved in the SV (*ZNF384* or *TCF3*) was tested. These genes are commonly absent from standard clinical laboratory ALL FISH panels, since translocations involving these genes are infrequently observed in pediatric B-ALL[32]. Although *TCF3* is now a part of our clinical ALL FISH panel, a *ZNF384* FISH probe has not yet been validated.

In silico assessment of limit of detection

We next explored the ability of lrSeq and Severus to detect SVs across a range of tumor amounts. Notably, the blast percentages of the neoplastic samples utilized for this study ranged from 20% to 94% blasts at diagnosis (per Pathology review; the blast percentages of each sample can be found in Supplementary Data 1). In silico modeling was performed to determine the limit of detection (LOD) for this approach at an average of 30× depth of coverage (which corresponds to using 1 HiFi sequencing SMRT cell per sample). Reads from the seven tumor samples with paired normal data and a detected clinically relevant SV were 'diluted' using sequence data that had been generated from the normal sample, then run through the Severus TN and TO pipelines (see Supplementary Data 5 and 6). This analysis demonstrated that the lower limit of detection was ~3–4 reads, corresponding with a 1:5 dilution (i.e., LOD of ~15% VAF in a sample with ~30× coverage) for both the TN and TO analyses and suggests a minimum required blast or tumor percentage of ~20% per neoplastic sample for clonal SV detection. However, variability in the coverage of genes (as well as variability in intragenic coverage) could also play a role in SV detection (see Supplementary Figs. 1–37 for gene coverage plots). Our coverage analysis

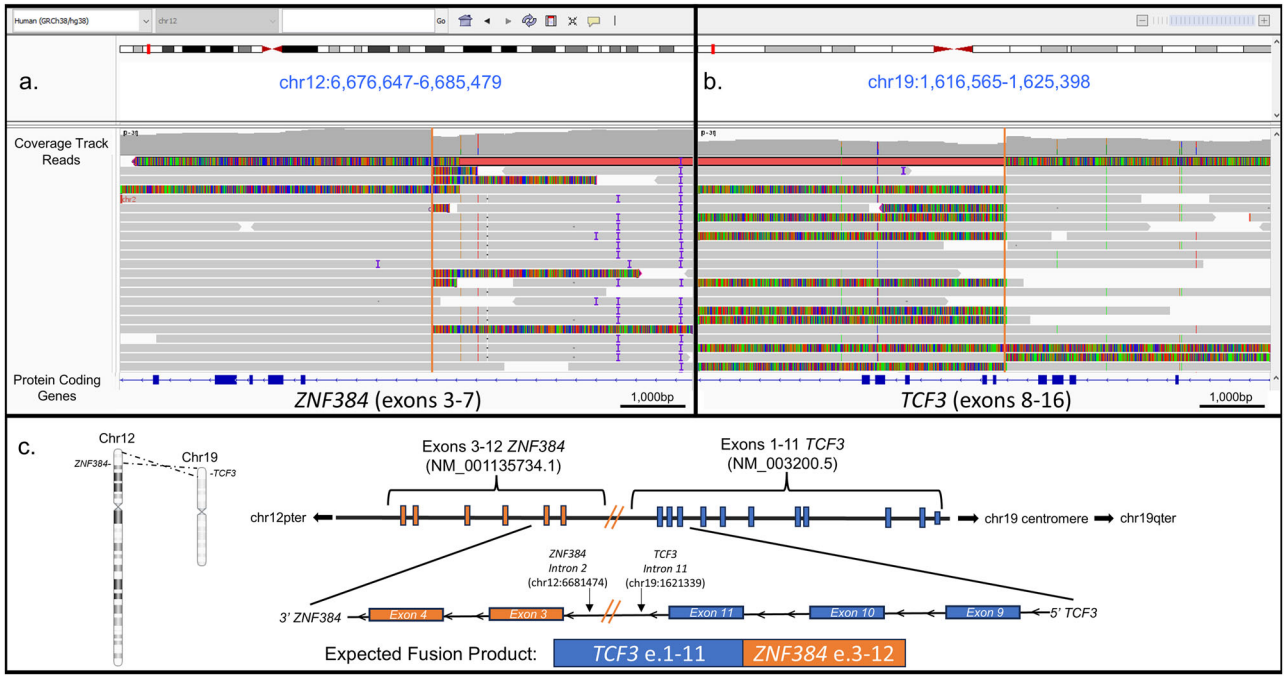

Fig. 3 | t(12;19)(p13;p13) detected by Severus in a previously genetically undefined precursor B-cell acute lymphoblastic leukemia case. a IGV demonstrating the presence of chimeric reads, indicating a translocation between *ZNF384* and (**b**) *TCF3*. **c** Schematic of the orientation of *TCF3* (NM_003200.5) and *ZNF384* (NM_001385745.1) genes including the intronic location of each breakpoint. Note that a functional fusion product is predicted.

demonstrated random areas of reduced coverage for various genes in our sample cohort, and this variability was not systematic.

Assessment of tumor-only structural variant calling

Although paired tumor/normal analysis enables true somatic SV calling from lrSeq data, obtaining a germline sample within a clinically relevant timeframe is not always feasible in pediatric leukemia testing, nor are paired 'normal' samples always available for research cohorts. Thus, we ran Severus in "tumor-only mode" (i.e., without the paired normal sample) for the previously analyzed tumor/normal pairs above ($n = 9$). There was an expected increase in the average number of cluster IDs (i.e., breakend calls that comprise unique SVs) in the unmatched samples versus the matched dataset (Supplementary Fig. 38; see Supplementary Data 7 and 8 for the resulting breakend call files from the matched tumor/normal versus tumor-only analyses, respectively). Severus called an average of 285 break-end calls per case (range 124–440). However, when restricting to breakends occurring within a gene, the number of calls was reduced by approximately half (see Supplementary Data 4 for counts per case by calling mode). To further screen these calls, we employed a list of genes known to be mutated in pediatric leukemia per the National Cancer Institute's Childhood Cancer Genomics summary and/or the World Health Organization's diagnostic genetic subtypes of B-ALL and AML[33,34]. On average, these genes were covered at a depth of at least 30×, with the exception of cases 6 and 9 which were sequenced at a lower depth of coverage (see Supplementary Figs. 1–37). This approach, combined with an additional filter requiring that a breakend occur between two different genes (supporting a rearrangement that would more probably result in a fusion product), readily identified the crSVs in all 7 of 7 neoplastic samples (100%) from the tumor/normal analyses above (see Supplementary Data 2 for the gene lists utilized), resulting in an average of 1 breakend call per case (range 0–4; see Supplementary Data 4). This suggests that the use of appropriate filters in a tumor-only dataset is equally effective as paired tumor/normal testing despite the larger number of SVs initially detected. Of note, a few ongoing studies are aiming to generate population-scale SV databases using lrSeq data, which could be used as a panel of normals in tumor-only sample analysis in the future[35,36], and the Severus package includes a 'panel of normals' that can be utilized as a filter SV calling[27].

To further investigate the utility of tumor-only lrSeq and analysis using Severus, eight additional B-ALL cases without an available germline sample (and also without a known genetic subtype after routine clinical testing) were analyzed. By using the filtering approach outlined above, a crSV was identified in two of these cases: one t(X;12)(p11.4;p13.31) resulting in *DDX3X::ZNF384* fusion; and one t(12;22)(p13.31;q13.2) resulting in *EP300::ZNF384* fusion. A summary of all fusions identified by Severus in both the 'known' and 'unknown' cases analyzed can be found in Fig. 1 (with breakpoints of each SV listed in Table 1 and Supplementary Data 1). As before, the existing clinical data for these 'tumor-only, unknown' cases were reviewed to determine why these rearrangements had not been previously detected.

Both cases had received karyotyping, FISH B-ALL panel, and chromosomal microarray testing. However, *ZNF384* rearrangements are commonly cytogenetically cryptic[37,38] and thus were not detected by chromosome analysis. Because *ZNF384* probes are not part of standard clinical B-ALL FISH panels, *ZNF384* was not investigated by FISH for either case. Finally, the *EP300::ZNF384* rearrangement was genomically balanced and thus not detected by microarray. For the other case (*DDX3X::ZNF384*), the microarray analysis showed significant genomic complexity in the 12p region, where *ZNF384* resides, including several copy number alterations, which made the genetic interpretation difficult. No additional testing was performed for the case with the *DDX3X::ZNF384* rearrangement, and RNA was not available for follow-up testing. Interestingly, the case with the *EP300::ZNF384* fusion did receive RNA-based fusion analysis testing as well as next-generation sequencing (NGS). However, *ZNF384* was not included on the RNA fusion panel utilized for the analysis, nor was the NGS assay validated for SV detection. Thus, neither rearrangement was detected via the clinical test algorithm selected for each case.

lrRNA-seq fusion detection supports the crSVs identified by lDNA-seq

In addition to the lrSeq of DNA for SV detection, two cases underwent lrRNA sequencing (lrRNA-seq) to assess the feasibility of this approach for fusion transcript detection as well as to determine whether the crSVs detected by Severus were supported by the presence of RNA fusion

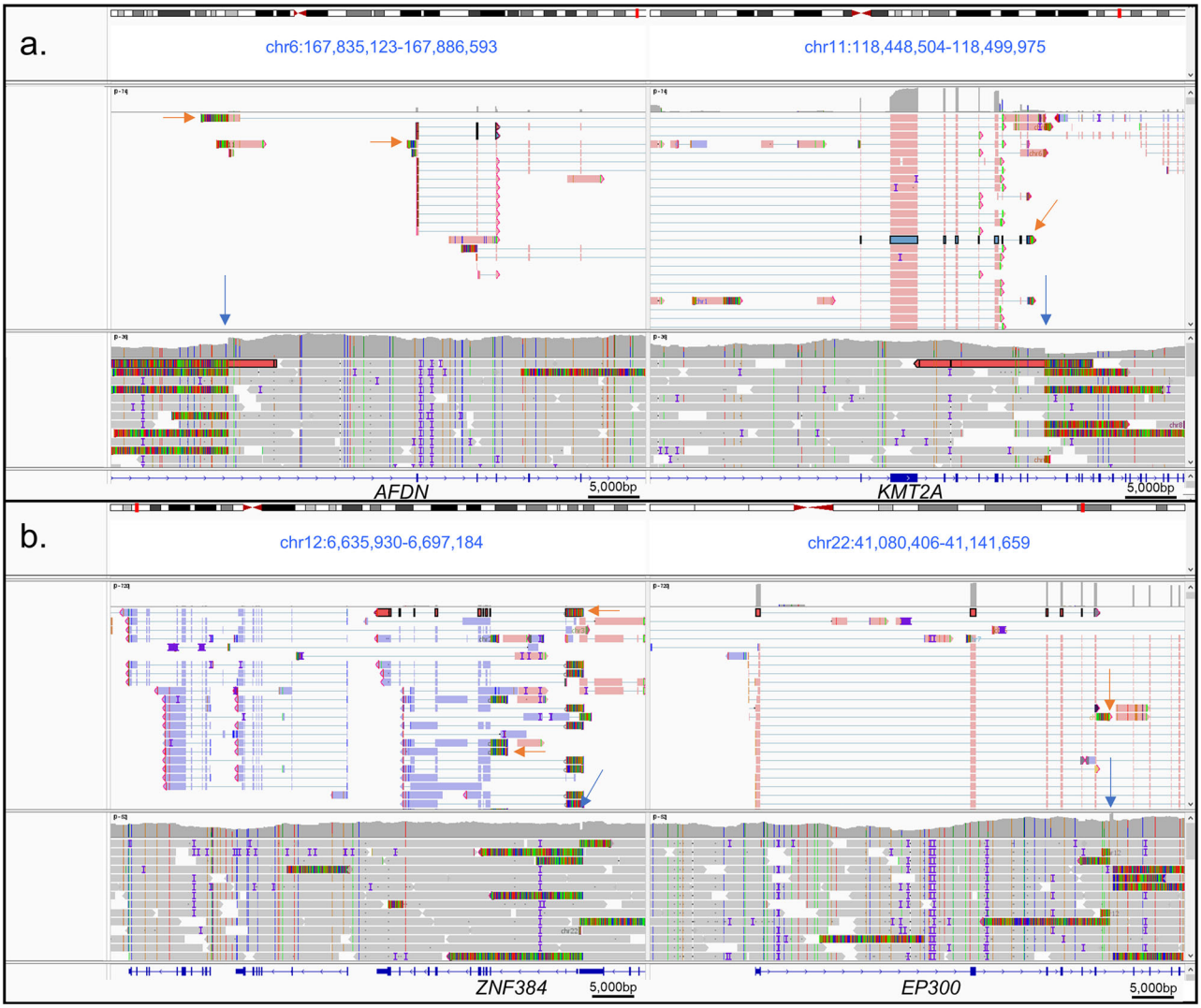

Fig. 4 | lrRNA-seq successfully detects fusion transcripts within pediatric leukemia cases. a IGV demonstrating the presence of a *KMT2A::AFDN* fusion in the lrRNA-seq data (top panel), which corresponds with the lrSeq DNA breakpoints (bottom panel). Fusion transcripts corresponding with different isoforms are highlighted by the orange arrows, while the blue arrows indicate the location of the breakpoints within the DNA. **b** IGV demonstrating the presence of an *EP300::ZNF384* in the lrRNA-seq data (top panel), which corresponds with the lrSeq DNA breakpoints (bottom panel). Orange arrows indicate a representative fusion transcript, whereas the blue arrows indicate the SV breakpoints as called by Severus in the lrSeq DNA data.

transcripts. Diagnostically relevant fusion transcripts were successfully detected in 2/2 cases, which demonstrated a *KMT2A::AFDN* and an *EP300::ZNF384* fusion product; these transcripts resulted in the fusion of exon 11 of *KMT2A* (NM_001197104.2) with exon 2 of *AFDN* (NM_001386888.1), and exon 7 of *EP300* (NM_001429.4) with exon 1 of *ZNF384* (NM_001385745.1), respectively (Fig. 4; the mean depth of exonic coverage for each gene and the respective exons involved in each fusion product can be found in Supplementary Data 9). Overall, the successful identification of fusion transcripts in two cases supporting the crSVs and their corresponding DNA breakpoints highlights the clinical utility of lrSeq in pediatric leukemia crSV detection.

Discussion

Long-read sequencing (lrSeq) is an effective technology for structural variant (SV) detection in inherited and neoplastic settings, especially for SVs in repetitive regions of the genome or those that are of high genetic complexity[13,19,39–41]. Although several studies have assessed SV calling using lrSeq in various types of adult cancer[16–23], the potential to use it as a clinical diagnostic tool for patients with cancer has not been evaluated systematically. In addition, extremely limited focus has been given to pediatric cancer, with only one published study to date investigating the utility of SV detection using lrSeq in pediatric medulloblastoma[24], and two studies investigating lrSeq in pediatric leukemia[25,26]. Furthermore, these studies were performed using existing bioinformatics SV callers[24,25], which had primarily been built for germline analyses, or in-house bioinformatics solutions[26]. Severus (a bioinformatics tool specifically optimized to detect somatic SVs using lrSeq data) has only recently been developed[27]. Since Severus outperformed other long-read somatic SV detection tools[27], this study was performed to assess the feasibility of detecting diagnostically relevant SVs using break-end calling from lrSeq data in a small clinical pediatric leukemia cohort.

Using this approach, all diagnostically relevant somatic SVs were identified with the exact breakpoints demonstrated in the five patients who had a genetic classification of their leukemia known through prior clinical diagnostic testing. Of note, for two of these five cases, lrSeq resolved the partner gene, which enhanced the prior clinical results. In addition, lrSeq and Severus analysis resolved the genetic driver alterations for four cases (4/12 total unknowns) that had not been genetically classified by prior clinical testing. For the eight remaining unresolved cases, there were no notable technical limitations for crSV detection (e.g., inadequate blast percentage

and gaps in coverage of key genes) with the exception of case 6 which had a lower than expected depth of coverage and may have impacted SV calling; an additional review of the Severus data was performed for these cases without utilizing the gene panel to determine whether a rare but genetic subtype defining SV, CNV, or inversion was present. No such findings were identified; thus, in this small exploratory cohort, the present approach resulted in a 33% increase in diagnostic yield over current clinical diagnostic methodologies employed by our laboratory. Of note, a limitation of the current gene-focused filtering approach is that it does not detect regulatory rearrangements (e.g., enhancer hijacking or promoter swapping), which are recognized genetic driver events in pediatric malignancies. A filtering approach that annotates regulatory elements combined with Severus calls could potentially detect these other driver events. Though a comprehensive assessment of other genetic events called by Severus was outside the scope of the present work, it is important to note that such findings are identified via the Severus workflow and could also be incorporated into future analyses.

The clinical utility of RNA-based profiling with fusion transcript detection and optical genome mapping (OGM)[10,42,43] is of known benefit for genetically subtyping pediatric leukemia[8,9]; however, these assays have historically not been a routine part of our clinical testing workflow. Had they been, it may have aided in the detection of underlying structural variants that were cryptic by other methodologies. However, both approaches present their unique challenges. For one, RNA is not always available, especially for historical cases, and requires specific isolation and storage conditions. It is also a challenge to maintain optimal conditions during transport for such specimens if they are being shipped out to external reference laboratories for testing. Additionally, for clinical laboratories that employ RNA-based fusion panels, if one of the genes involved in the fusion is not specifically included on the panel, then that fusion transcript will be missed. Such is the case for *ZNF384*, which is not part of the gene list for multiple clinically available RNA-based hematologic cancer fusion panels, even at Children's Oncology Group (COG) approved laboratories or those at large pediatric academic medical centers. For OGM, although it is a DNA-based technology, it presently requires a specific isolation procedure for ultra-long DNA molecules, which is not always compatible with banked samples. In contrast, the lrSeq data analyzed in our study were generated using DNA that was isolated using routine clinical protocols (see Methods). Additionally, the breakpoint resolution of SVs detected via OGM can be pinpointed only to a genomic interval that is a few kilobases in length, as it is dependent upon the location of the breakend relative to the repeat sequences that OGM examines. For comparison, lrSeq allows for the unbiased and successful detection of the exact nucleotide-level breakpoints of each rearrangement. Thus, although the diagnostic yield of our small cohort would likely be impacted if RNA profiling, fusion analysis, and/or OGM were incorporated into our clinical workflow, this study demonstrates a proof-of-principle for DNA-based lrSeq to be successfully, and perhaps more feasibly, implemented clinically for SV detection in pediatric leukemia.

Importantly, the successful detection of these SVs provides highly relevant information that has a significant and direct impact on patient management. This is indeed the case for the four genetically 'unknown' cases above that were solved by lrSeq and Severus. For example, the AML patient with the ins(11;10)(q23.3;p12p12) identified in the present study would have been managed as high-risk, whereas patients without a diagnostic genetic alteration are managed as intermediate risk. Patients who are high-risk undergo more aggressive treatment, such as hematopoietic stem cell transplantation[44], and *KMT2A*-rearranged AML has been shown to respond more favorably to gemtuzumab ozogamicin[45], which would have been an additional therapy that could have been pursued for this patient.

Similarly, *ZNF384* rearrangements, like the three identified in the previously 'unknown' cohort, are known to be karyotypically cryptic within the context of pediatric B-ALL, requiring detection via *ZNF384* FISH testing and/or next-generation DNA or RNA sequencing. *ZNF384* is also known to rearrange with one of several partner genes, most commonly *EP300*, *TCF3*, *TAF15*, and *CREBBP*[46]. Identifying the fusion partner is important in *ZNF384*-rearranged B-ALL, as *EP300* confers an inferior prognosis and high-risk disease[46]. In contrast, the partner genes *TCF3* or *DDX3X* are not expected to be of prognostic significance, but the identification of a *TCF3::ZNF384* or *DDX3X::ZNF384* translocation helps rule out the presence of a different, previously undetected, high-risk genetic alteration.

The utility of lrRNA-seq for isoform detection and inferring structural variants has been described previously in pediatric oncology settings[47–49]; however, it has not yet been investigated using HiFi data within individuals with pediatric leukemia. Therefore, we were interested in determining whether diagnostically relevant fusion transcripts supporting the structural variants detected via DNA-based lrSeq and Severus could be captured by this approach. Both fusion isoforms (i.e., *KMT2A::AFDN* and *EP300::ZNF384*) were successfully detected, suggesting that this test could be feasibly implemented as an orthogonal method to further support findings from DNA-based SV detection and/or as a replacement for short-read RNA-seq. Furthermore, lrRNA-seq may additionally complement DNA-based SV detection by allowing for the identification of genomic rearrangement-independent fusions[50]. Finally, lrRNA-seq offers increased resolution of fusion transcripts in that it accurately captures specific isoforms; this study demonstrated that these isoforms matched the DNA-based lrSeq breakpoints with remarkably high precision (as shown in Fig. 4). Future work will be needed to determine the most clinically relevant isoform(s) for each patient as lrRNA-seq becomes more widely implemented, as well as to determine whether lrRNA-seq, including RNA expression profiling from lrRNA-seq data, could further assist in defining pediatric leukemia cases with unknown genetic drivers.

In summary, this study establishes the successful detection of crSVs by lrSeq in a small pediatric leukemia cohort, including an increased diagnostic yield over commonly used clinical workflows, though we note the potential for ascertainment bias due to the size and nature of this study. Given that lrSeq can generate complete genomic data for a sample in a few days and its cost has decreased to less than $1000 USD per genome, we may be reaching a point where its clinical implementation in cancer, especially in a pediatric setting, is now feasible. In addition, although lrSeq is more expensive than other assays that are routinely used in clinical settings for crSV detection, it should be noted that the data can be easily reanalyzed; this allows for seamless reclassification of previously undefined cases as novel genetic subtypes are discovered without necessitating that new data be generated. Finally, since lrSeq provides additional genetic information not investigated in the present study, as mentioned above, it is likely that the utility of lrSeq as a comprehensive cancer profiling test will continue to grow.

Methods

Study participant details

Nine pediatric leukemia cases with corresponding normal (remission) samples were selected for HiFi sequencing from the Children's Mercy Research Institute Biorepository (CRIB) following informed consent. These cases were enrolled in the CRIB between 2017 and 2023 and had sufficient material available for HiFi sequencing. The sample type(s) utilized for each case can be found in Supplementary Data 1. Samples were stored in accordance with the CRIB protocol (i.e., a peripheral blood or bone marrow aspirate white blood cell pellet was snap frozen after red blood cell lysis), or lrSeq was performed on archival clinically isolated DNA. The pathology diagnosis for five of these cases was precursor B-cell acute lymphoblastic leukemia (B-ALL), and for the remaining four was acute myeloid leukemia (AML). Five of the cases (2 B-ALL; 3 AML) had diagnostic translocations that had been identified clinically using routine cytogenetic testing (FISH, chromosomes, and microarray). The remaining four cases (3 B-ALL; 1 AML) had undergone routine clinical testing (FISH, chromosomes, microarray, and/or next-generation sequencing), and were selected because no diagnostic genetic driver alteration (i.e., genetic driver "not otherwise specified"; NOS) had been identified. In addition, eight B-ALL NOS cases for which a paired normal sample was not available were selected. The age, sex, race, and ethnicity of each patient are listed in Supplementary Data 1, when provided by self or physician report.

DNA sample Prep and lrSeq

Genomic DNA from each sample was extracted using a Chemagen instrument (PerkinElmer, Baesweiler, Germany) or via manual TKM isolation. DNA samples were mechanically fragmented on the Diagenode Megaruptor 3, and library preparation was performed with the SMRTbell Prep Kit 3.0 (Pacific Biosciences ("PacBio"), Menlo Park, CA, USA) using 2.5 µg of DNA per sample. Size selection for intact library fragments greater than 10 kb was performed using the PippinHT (Sage Science, Beverly, MA, USA). The concentration of each library was determined by Qubit, and libraries were sequenced on a PacBio Revio instrument using supported reagents (Revio Polymerase Kit, Revio Sequencing Plate, and Revio SMRT Cell Tray) to a target depth of 30×.

lrDNA-seq data processing and analysis

The PacBio Human Whole Genome Sequencing workflow was used to process HiFi reads for haplotagged alignment and phasing to GRCh38, followed by structural variant (SV) calling from each tumor/normal (TN) pair using Severus[27]. Specific code utilized for data processing can be found in the Supplementary Information. All prior clinical genetic test data were collated and used for comparisons with the breakend analysis findings (see Supplementary Data 1). The TN Severus output was filtered for break-end calls that overlapped genes known to be relevant to pediatric leukemia (see Supplementary Data 2). All data filtering was performed in Microsoft Excel, and SVs were visualized using Integrative Genomics Viewer (IGV)[51]. Each neoplastic sample was additionally processed by Severus without the paired normal sample to assess the feasibility of diagnostic SV detection in a tumor-only (TO) setting. The TO Severus output was similarly filtered for break-end calls overlapping genes known to be relevant to pediatric leukemia (see Supplementary Data 2), with an additional filter to prioritize breakends occurring between two different genes. The average read depth after base-calling for the genes assessed was determined; representative coverage plots can be found in Supplementary Figs. 1–37. To investigate the theoretical limit of detection of this assay, reads from the paired normal sample were used to "dilute" the tumor sample at varying concentrations, and the bioinformatically "diluted" BAM files were run through Severus in TN and TO modes.

RNA sample Prep and lrSeq

For the lrRNA-seq, RNA was isolated from cell pellets in TRIzol (ThermoFisher, Cat. No. 15596026) using a TRIzol/chloroform precipitation and cleaned with a RNeasy Mini Kit (Qiagen, Cat. No. 74104). Sample concentrations were quantified with a Qubit RNA BR Assay Kit (ThermoFisher, Cat. No. Q10210). Samples with a high concentration (>800 ng/µL) were diluted 1:10 with RNase-free water and were quantified again. The 1:10 diluted samples were checked on an RNA ScreenTape (Agilent, Cat. No. 5067-5577 and 5067-5576) on the TapeStation platform to assess whether RIN scores were greater than or equal to 7.0. Long-transcript cDNA libraries were prepared as described in the PacBio Iso-Seq™ Express Template Preparation for Revio Systems protocol. The maximum recommended input (300 ng) of RNA was aliquoted from the 1:10 diluted RNA samples and was used as the sample input. Samples were enriched for long transcripts greater than 3 kb in length. Quantification with a Qubit dsDNA HS Assay Kit (ThermoFisher, Cat. No. Q32854) determined that the cDNA yield of the samples was less than the minimum required yield (160 ng) for the Revio system and that the samples would need additional PCR cycling (as described in Appendix 1 of the Iso-Seq procedure[52]). Three additional PCR cycles were performed on the samples, as recommended, with NEBNext High-Fidelity 2× PCR Master Mix (NEB, Cat. No. M0541S) used in the PCR reamplification master mix. The samples were then bead-cleaned following the procedure for low-yielded samples and were quantified again with a Qubit dsDNA HS Assay Kit to ensure there was enough cDNA for SMRTbell library preparation.

Up to 500 ng of cDNA from each sample was used as an input into SMRTbell library preparation without pooling using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Cat. No. 100-938-900). After the SMRTbell library cleanup, the libraries were quantified with a Qubit dsDNA HS Assay Kit to determine their final concentration. The library size was determined with a High Sensitivity D5000 ScreenTape (Agilent, Cat. No. 5067-5592 and 5067-5593) on the TapeStation platform. Resulting libraries were sequenced with one Revio SMRT Cell (Pacific Biosciences, Cat. No. 102-202-200) each on the Revio System using the Revio Polymerase Kit (Pacific Biosciences, Cat. No. 102-739-100) according to standard manufacturer's directions with loading at 90pM per library.

lrRNA-seq data processing and analysis

Starting from HiFi reads, data were processed using the standard isoseq v4.0.0 pipeline as described at isoseq.how, using the cluster2 algorithm for clustering. FLNC reads were aligned to GRCh38 using minimap2 v2.28-r1209[53]. Average coverage for each gene was calculated over exons only using the samtools v1.16.1 coverage command. Data analysis for expected crSV products was performed in IGV.

Data availability

HiFi DNA-seq data are publicly available from dbGaP as of the date of publication (study ID phs002529) and the dbGaP identifiers corresponding to each case alias utilized in this manuscript can be found in Supplementary Table 1. Any additional information required to reanalyze the data reported in this paper is available from the corresponding author upon reasonable request.

Received: 29 September 2025; Accepted: 6 March 2026;
Published online: 09 April 2026

References

1. Bolouri, H. et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* **24**, 103–112 (2018).
2. Brady, S. W. et al. The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat. Genet.* **54**, 1376–1389 (2022).
3. Iacobucci, I., Kimura, S. & Mullighan, C. G. Biologic and therapeutic implications of genomic alterations in acute lymphoblastic leukemia. *J. Clin. Med.* **10**, 3792 (2021).
4. Roberts, K. G. & Mullighan, C. G. The biology of B-progenitor acute lymphoblastic leukemia. *Cold Spring Harb. Perspect. Med.* https://doi.org/10.1101/cshperspect.a034835 (2020).
5. Tarlock, K. & Meshinchi, S. Pediatric acute myeloid leukemia: biology and therapeutic implications of genomic variants. *Pediatr. Clin. North Am.* **62**, 75–93 (2015).
6. Akkari, Y. M. N. et al. Evidence-based review of genomic aberrations in B-lymphoblastic leukemia/lymphoma: Report from the cancer genomics consortium working group for lymphoblastic leukemia. *Cancer Genet.* **243**, 52–72 (2020).
7. Abla, O. et al. Structural variants involving MLLT10 fusion are associated with adverse outcomes in pediatric acute myeloid leukemia. *Blood Adv.* **8**, 2005–2017 (2024).
8. Hiemenz, M. C. et al. A multimodal genomics approach to diagnostic evaluation of pediatric hematologic malignancies. *Cancer Genet.* **254-255**, 25–33 (2021).
9. Hu, Z. et al. Transcriptome sequencing allows comprehensive genomic characterization of pediatric B-acute lymphoblastic leukemia in an Academic Clinical Laboratory. *J. Mol. Diagn.* **26**, 49–60 (2024).
10. Rack, K. et al. Optimizing the diagnostic workflow for acute lymphoblastic leukemia by optical genome mapping. *Am. J. Hematol.* **97**, 548–561 (2022).
11. Vicente-Garces, C. et al. Technical validation and clinical utility of an NGS targeted panel to improve molecular characterization of pediatric acute leukemia. *Front. Mol. Biosci.* **9**, 854098 (2022).
12. Wenger, A. M. et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162 (2019).

13. Kosugi, S. & Terao, C. Comparative evaluation of SNVs, indels, and structural variations detected with short- and long-read sequencing data. *Hum. Genome Var.* **11**, 18 (2024).
14. Cohen, A. S. A. et al. Genomic answers for children: dynamic analyses of >1000 pediatric rare disease genomes. *Genet. Med.* **24**, 1336–1348 (2022).
15. Negi, S. et al. Advancing long-read nanopore genome assembly and accurate variant calling for rare disease detection. *medRxiv* https://doi.org/10.1101/2024.08.22.24312327 (2024).
16. Sakamoto, Y., Sereewattanawoot, S. & Suzuki, A. A new era of long-read sequencing for cancer genomics. *J. Hum. Genet.* **65**, 3–10 (2020).
17. Haga, Y., Sakamoto, Y., Arai, M., Suzuki, Y. & Suzuki, A. Long-read whole-genome sequencing using a nanopore sequencer and detection of structural variants in cancer genomes. *Methods Mol. Biol.* **2632**, 177–189 (2023).
18. Hansen, M. H. et al. Toward cytogenomics: technical assessment of long-read nanopore whole-genome sequencing for detecting large chromosomal alterations in mantle cell lymphoma. *J. Mol. Diagn.* **25**, 796–805 (2023).
19. Klever, M. K. et al. AML with complex karyotype: extreme genomic complexity revealed by combined long-read sequencing and Hi-C technology. *Blood Adv.* **7**, 6520–6531 (2023).
20. Thibodeau, M. L. et al. Improved structural variant interpretation for hereditary cancer susceptibility using long-read sequencing. *Genet. Med.* **22**, 1892–1897 (2020).
21. Lee, O. W. et al. Targeted long-read sequencing of the Ewing sarcoma 6p25.1 susceptibility locus identifies germline-somatic interactions with EWSR1-FLI1 binding. *Am. J. Hum. Genet.* **110**, 427–441 (2023).
22. Fujimoto, A. et al. Whole-genome sequencing with long reads reveals complex structure and origin of structural variation in human genetic variations and somatic mutations in cancer. *Genome Med.* **13**, 65 (2021).
23. Zhou, L. et al. Long-read sequencing unveils high-resolution HPV integration and its oncogenic progression in cervical cancer. *Nat. Commun.* **13**, 2563 (2022).
24. Rausch, T. et al. Long-read sequencing of diagnosis and post-therapy medulloblastoma reveals complex rearrangement patterns and epigenetic signatures. *Cell Genom.* **3**, 100281 (2023).
25. Kato, S. et al. Genome profiling with targeted adaptive sampling long-read sequencing for pediatric leukemia. *Blood Cancer J.* **14**, 145 (2024).
26. Geyer, J. et al. Real-time genomic characterization of pediatric acute leukemia using adaptive sampling. *Leukemia* **39**, 1069–1077 (2025).
27. Keskus, A. G. et al. Severus detects somatic structural variation and complex rearrangements in cancer genomes using long-read sequencing. *Nat. Biotechnol.* **44**, 247–257 (2025).
28. Hadi, K. et al. Distinct classes of complex structural variation uncovered across thousands of Cancer Genome Graphs. *Cell* **183**, 197–210 e132 (2020).
29. Choo, Z. N. et al. Most large structural variants in cancer genomes can be detected without long reads. *Nat. Genet.* **55**, 2139–2148 (2023).
30. Aganezov, S. & Raphael, B. J. Reconstruction of clone- and haplotype-specific cancer genome karyotypes from bulk tumor samples. *Genome Res.* **30**, 1274–1290 (2020).
31. Shale, C. et al. Unscrambling cancer genomes via integrated analysis of structural variation and copy number. *Cell Genom.* **2**, 100112 (2022).
32. Choi, J. K. F., K. E., Hodge, J. C., Yasuda, T., Greipp, P. T. & Mejstrikova, E., Hunger, S. in *WHO Classification of Tumours Editorial Board. Haematolymphoid tumours [Internet]*. Vol. WHO classification of tumours series, 5th ed.; vol. 11 (0 Rita Alaggio, Brent Lee Wood, Sumeet Gujral, Yassmine Akkari) (International Agency for Research on Cancer, Lyon (France), 2024).
33. *Childhood Cancer Genomics (PDQ(R)): Health Professional Version*. https://www.ncbi.nlm.nih.gov/pubmed/27466641 (2002).
34. Haematolymphoid tumours. https://tumourclassification.iarc.who.int/chapters/63 (2024).
35. Gustafson, J. A. et al. High-coverage nanopore sequencing of samples from the 1000 Genomes Project to build a comprehensive catalog of human genetic variation. *Genome Res.* https://doi.org/10.1101/gr.279273.124 (2024).
36. Schloissnig, S. et al. Structural variation in 1,019 diverse humans based on long-read sequencing. *Nature* **644**, 442–452 (2025).
37. Berg, H. E. et al. Detection of a Cryptic EP300/ZNF384 gene fusion by chromosomal microarray and next-generation sequencing studies in a pediatric patient with B-lymphoblastic leukemia. *Lab. Med.* **52**, 297–302 (2021).
38. Janet, N. B. et al. Systematic application of fluorescence in situ hybridization and immunophenotype profile for the identification of ZNF384 gene rearrangements in B cell acute lymphoblastic leukemia. *Int. J. Lab. Hematol.* **43**, 658–663 (2021).
39. Romagnoli, S., Bartalucci, N. & Vannucchi, A. M. Resolving complex structural variants via nanopore sequencing. *Front Genet* **14**, 1213917 (2023).
40. Ahsan, M. U., Liu, Q., Perdomo, J. E., Fang, L. & Wang, K. A survey of algorithms for the detection of genomic structural variants from long-read sequencing data. *Nat. Methods* **20**, 1143–1158 (2023).
41. Mastrorosa, F. K., Miller, D. E. & Eichler, E. E. Applications of long-read sequencing to Mendelian genetics. *Genome Med.* **15**, 42 (2023).
42. Suttorp, J. et al. Optical genome mapping as a diagnostic tool in pediatric acute myeloid leukemia. *Cancers (Basel)* https://doi.org/10.3390/cancers14092058 (2022).
43. Luhmann, J. L. et al. The clinical utility of optical genome mapping for the assessment of genomic aberrations in acute lymphoblastic leukemia. *Cancers (Basel)* https://doi.org/10.3390/cancers13174388 (2021).
44. Guest, E. M. et al. Prognostic significance of 11q23/MLL fusion partners in children with acute myeloid leukemia (AML)—results from the Children's Oncology Group (COG) Trial AAML0531. *Blood* **128**, 1211–1211 (2016).
45. Pollard, J. A. et al. Gemtuzumab ozogamicin improves event-free survival and reduces relapse in pediatric KMT2A-rearranged AML: results from the phase III Children's Oncology Group Trial AAML0531. *J. Clin. Oncol.* **39**, 3149–3160 (2021).
46. Hirabayashi, S. et al. Clinical characteristics and outcomes of B-ALL with ZNF384 rearrangements: a retrospective analysis by the Ponte di Legno Childhood ALL Working Group. *Leukemia* **35**, 3272–3277 (2021).
47. Miller, A. R. et al. Pacific biosciences fusion and long isoform pipeline for cancer transcriptome-based resolution of isoform complexity. *J. Mol. Diagn.* **24**, 1292–1306 (2022).
48. Lysenkova Wiklander, M. et al. A multiomic characterization of the leukemia cell line REH using short- and long-read sequencing. *Life Sci. Alliance* **7**, e202302481 (2024).
49. Rybacki, K. et al. Combining panel-based and whole-transcriptome-based gene fusion detection by long-read sequencing. *Cell Rep. Methods* **5**, 101111 (2025).
50. Dorney, R., Dhungel, B. P., Rasko, J. E. J., Hebbard, L. & Schmitz, U. Recent advances in cancer fusion transcript detection. *Brief Bioinform.* **24**, bbac519 (2023).
51. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
52. PacBio. *Preparing Iso-Seq v2 libraries using SMRTbell prep kit 3.0 Procedure & checklist*. https://www.pacb.com/wp-content/uploads/Procedure-checklist-Preparing-Iso-Seq-libraries-using-SMRTbell-prep-kit-3.0.pdf (2025).
53. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).

Acknowledgements

This study was funded by Braden's Hope for Childhood Cancer, Elizabeth and Monte McDowell, the Black & Veatch Foundation, and Big Slick. A.K., T.A., A.B., and M.K. were supported in part by the Intramural Research Program of the NIH. PacBio provided partial reagent support for this study. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this paper. M.S.F. and E.G. thank these funders for their generous support and thank Nhu Bui and Mindy Spano for their assistance with financial administration. M.S.F., E.M.G., and L.A.L. also thank Children's Mercy Oncology Biorepository study personnel: Judy Vun, Amie Hatfield, and Robin Ryan; as well as Jason Seymour and Keiondra Sanders in the Children's Mercy Research Institute (CMRI) Biorepository, for their assistance with sample collection and processing. M.S.F. and L.A.L. are grateful to Jonas Korlach for his expertise and thoughtful feedback on the paper, and to Jonathan Bibliowicz, Gary Gelland, Khi Pin Chua, Aaron Wenger, and Alex Sockell for their collaborative support.

Author contributions

M.S.F. and L.A.L. conceptualized this analysis. L.A.L., B.Y., A.K., M.K., and M.S.F. analyzed the data. L.A.L., B.Y., I.P., M.K., and M.S.F. drafted the paper. All authors, including those aforementioned as well as C.B., T.A., A.B., A.W., M.G., M.R., W.L., S.M.H., L.D.C., J.H., E.R., L.Z., K.J.A., T.G.F., A.S.G., E.M.G., J.A.H., M.H., K.L., and T.P. provided edits and insights for the final submission. All authors read and approved the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-026-00560-5.

Correspondence and requests for materials should be addressed to Midhat S. Farooqi.