UV-induced mutations accumulate during early clonal expansion in aneuploid subtypes of pediatric B-cell precursor acute lymphoblastic leukemia

Lianne C. Suurenbroek,[1*] Cédric G. van der Ham,[1*] Judith M. Boer,[1] Lucca L.M. Derks,[1,2] Rico Hagelaar,[1,2] Markus J. van Roosmalen,[1,2] Puk Klamer,[1] Edwin Sonneveld,[1,3] Monique L. den Boer,[1] Ruben van Boxtel,[1,2] Frank N. van Leeuwen,[1] Jan L.C. Loeffen[1] and Roland P. Kuiper[1,4]

[1]Princess Máxima Center for Pediatric Oncology; [2]Oncode Institute; [3]Dutch Childhood Oncology Group and [4]Department of Genetics, Utrecht University Medical Center, Utrecht University, Utrecht, the Netherlands

**LCS and CGvdH contributed equally as first authors.*

Correspondence: R.P. Kuiper
r.kuiper@prinsesmaximacentrum.nl

Received: October 29, 2025.
Accepted: March 30, 2026.
Early view: April 9, 2026.

https://doi.org/10.3324/haematol.2025.300118

©2026 Ferrata Storti Foundation
Published under a CC BY license

UV-induced mutations accumulate during early clonal expansion in aneuploid subtypes of pediatric B-cell precursor acute lymphoblastic leukemia

Table of contents

Data	Page
Supplementary methods	3-7
Supplementary figure S1. Genes in the NER pathway show no differential expression between SBS7a-positive and SBS7a-negative BCP-ALL	8
Supplementary figure S2. Differential expression analysis for samples with and without SBS7a	10
Supplementary figure S3. Gene set enrichment analysis comparing aneuploid BCP-ALL subtypes to *ETV6*::*RUNX1* BCP-ALL	12
Supplementary figure S4. Over-representation analysis of genes which are differentially expressed in both iAMP21 and high hyperdiploid BCP-ALL compared to *ETV6*::*RUNX1* BCP-ALL	13
Supplementary figure S5. Somatic SBS7a-associated mutations are accumulated after chromosome 21 amplification in high hyperdiploid and iAMP21 BCP-ALL	14
Supplementary figure S6. The allele frequency of ten out of 25 SBS7a-positive tumors shows a bimodal distribution, of which six present with at least 200 subclonal mutations	15
Supplementary figure S7. Sample P0429Dx and P0633Dx contain subclonal mutations without SBS7a, whereas sample P0557Dx contains subclonal SBS7a-associated mutations	16
Supplementary figure S8. 10 mutations in driver genes have a high probability of being caused by SBS7a	17
Supplementary figure S9. Branch-specific quality metrics for single-cell WGS data	18
Supplementary figure S10. Lineage tree based on the single cell WGS data of patient P0624	19
Supplementary Table S1: Info on whole genome sequencing	
Supplementary Table S2: Mutational signatures	
Supplementary Table S3: Signatures pan-cancer	
Supplementary Table S4: Localization of ALCL lesions	
Supplementary Table S5: Nucleotide excision repair pathway genes	
Supplementary Table S6: Nucleotide excision repair pathway mutations in SBS7a-positive BCP-ALL	
Supplementary Table S7: Over-representation analysis of genes which are differentially expressed in aneuploid BCP-ALL	
Supplementary Table S8: Percentage of chromosome gains and other common alterations in SBS7a-positive and SBS7a-negative hyperdiploid B-ALL	

Supplementary Table S9: Mean number of gained chromosomes in SBS7a-positive and SBS7a-negative hyperdiploid B-ALL

Supplementary tables can be found in the Supplementary Excel Table.

Supplementary methods

Reference and validation cohorts

We received a pan-cancer pediatric WGS cohort and a BCP-ALL RNAseq cohort from the Princess Máxima Center biobank (PMCLAB2021.279 & PMCLAB2022.343). RNA and DNA were isolated automatically using the Allprep DNA/RNA/miRNA Universal Kit (Qiagen, 80224) and the QIAcube Connect (Qiagen, 9002864). Library prep for WGS was performed automatically using 150 ng DNA, the KAPA HyperPlus Kit (Roche, 07962428001) and the epMotion 5075 (Eppendorf, 12018865). Library prep for RNAseq was performed manually using 300 ng RNA and the KAPA RNA HyperPrep kit with RiboErase (Roche, 08098140702). Subsequent sequencing was performed on Illumina Novaseq 6000 (Illumina) to generate 150 base-pair paired-end reads.

As an additional validation cohort we used RNAseq and WGS data of high hyperdiploid and iAMP21 BCP-ALL samples from St. Jude Children's Research Hospital in Memphis, Philadelphia, which we requested through the St. Jude Cloud.[1] The relative contributions of mutational signatures were either used as reported in literature,[2] or extracted with the same methods used for in-house data. A small subset of the RNAseq data was generated using unstranded instead of stranded RNAseq. Calling of overlapping genes is greatly improved in stranded RNAseq compared to unstranded RNAseq, causing significant bias.[3] As such we excluded the two unstranded RNAseq high hyperdiploid BCP-ALL samples and analyzed the unstranded (n=15) and stranded (n=12) RNAseq iAMP21 BCP-ALL samples separately.

WGS data analysis

Reads generated by WGS were mapped to the GRCh38 human reference genome using the Burrows-Wheeler Aligner (BWA) v0.7.13.[4] Picard v2.20.1[5] was used to mark duplicate reads, followed by base quality score recalibration using Genome Analysis ToolKit (GATK) v4.0.1.2.[6] Somatic mutations were called using Mutect2 from GATK v4.1.1.0 and annotated using Variant Effect Predictor v105.[7] R v4.1.2 was used to filter out mutations overlapping with centromeric regions or with a population frequency above 1% according to either gnomAD v3.0[8] or GoNL.[9] Mutations were kept if they were supported by 5 alternative reads at an allele frequency of at least 0.15 and a minimum coverage of 20.

Comparison with SBS7a-positive skin cancer

Somatic data for a cohort of adult metastatic skin cancers were received from the Hartwig Medical Foundation,[10] and further filtered with the same methods used for in-house data. To determine the mutational load, the total number of SBSs was calculated. The number of CC>TT mutations was determined using MutationalPatterns v3.4.1.[11] Transcriptional strand bias was calculated using MutationalPatterns v3.4.1 and the known genes table from UCSC for hg38.[11,12]

RNA expression analysis

Reads generated by RNA sequencing were mapped to GRCh38 (gencode version 31) using STAR v2.7.2d,[13] after which count matrices were generated with Rsubread v1.6.4.[14] Further analysis was performed using DESeq2 v.1.34.0.[15] For each dataset, genes without any counts were removed. For NER pathway analysis, counts were transformed using Variance Stabilizing Transformation, genes in the NER pathway in the Kyoto Encyclopedia of Genes and Genomes were selected and the deviation from the average per gene was calculated. Heatmaps were generated using pheatmap v1.0.12.[16] Differential expression analysis was

performed with DESeq2, comparing SBS7a-positive to SBS7a-negative samples and correcting for sex as a confounder. An adjusted *P*-value of 0.05 was used as a threshold to identify differentially expressed genes. Volcano plots were made using EnhancedVolcano v1.12.0.[17]
For subtype comparisons, 10 samples of iAMP21 BCP-ALL and 72 samples of high hyperdiploid BCP-ALL were separately compared to 58 cases of *ETV6*::*RUNX1*-positive or *ETV6*::*RUNX1*-like BCP-ALL. For gene set enrichment analysis, genes were ordered on log2 fold change, and gene set enrichment analysis was performed on GO terms using clusterProfiler v4.2.2. Differentially expressed genes in both subtypes were identified using DESeq2 v1.34.0, using a *P*-value cutoff of 0.05 and a log2 fold change cutoff of 1 for upregulated and -1 for downregulated genes. Over-representation analysis of GO terms on the overlapping differentially expressed genes was performed using clusterProfiler v4.2.2.

Copy number analysis
Copy number alterations were called according to the GATK best practices, using GATK v4.1.7.0 and an in-house panel of normals. To study the timing of SBS7a, we selected somatic mutations on chromosome 21 with an allele frequency of at least 0.1. As a control, we called germline mutations on chromosome 21 using HaplotypeCaller and selected mutations with an allele frequency between 0.25 and 0.75 in the germline and an allele frequency of at least 0.1 in the tumor. Somatic and germline mutations overlapping with gained regions were selected with R v4.1.2. Karyotypes for patients with high hyperdiploid BCP-ALL were determined manually based on the CNA plots. Karyotype plots were generated using pheatmap v1.0.12[16] and samples were clustered using hierarchical clustering with complete linkage and euclidean distance. Prognostic risk group of high hyperdiploid patients was determined according to criteria identified by Enshaei et al[18]. For patients with iAMP21 BCP-ALL, segment files were used to visualize copy numbers with karyoploteR v1.20.3.[19]

Subclonal mutations
Mutations on amplified chromosomes were filtered out for analysis of subclones. Peaks in the allele frequency (AF) were identified using LaplacesDemon v16.1.6.[20] The minimum between these peaks was found using base R v4.1.2, and used as threshold between clonal and subclonal mutations. Tumors with a bimodal distribution and at least 200 subclonal mutations were selected for further inspection. For two patients, where the allele frequencies of the clonal and subclonal mutations showed great overlap, manual thresholds were applied to identify clonal and subclonal mutations.

Analysis of acquired mutations
SBSs were clustered based on their allele frequency dynamics over time,[21] and a refit with the previously identified mutational signatures for BCP-ALL was performed on the clusters. Based on these clusters, fishplots were generated using fishplot v0.5.1.[22]

Amplicon sequencing
We selected relapse-specific mutations, based on WGS data, that were likely caused by SBS7a using R v4.1.2 and the method described by Morganella et al, 2016 and Brady et al, 2019.[23,24] We then excluded mutations in repeat-rich regions and aimed to include at least 15 mutations per patient, preferably in coding regions. We designed primers for 300-400 bp

fragments using primer3[25] and included 15 mutations for P0557, 19 mutations for P0608 and 16 mutations for P0611. Each primer pair was first validated separately and then multiplexed in sets of three and one set of two, for a total of 17 pools, and validated again. Multiplex PCR was then performed using Taq polymerase (Roche, 11647679001) with an initial 2 minutes at 94°C, followed by 35 cycles of 30s at 94°C, 45s at 60°C and 60s at 72°C, then finalized with 7 minutes at 72°C. 150 ng DNA was used for initial diagnosis samples and 50 ng DNA was used for relapse samples. Successful fragment amplification was confirmed using gel electrophoresis. Samples were purified using AMPure XPbeads (Beckmann Coulter, B37419AB), and DNA concentration was measured using the Qubit dsDNA Quantitation High Sensitivity kit (Thermo Fisher, Q32854) and Qubit™ 4 Fluorometer (Thermo Fisher, Q33238).

Illumina DNA Prep (Illumina, 20015826) and Nextera DNA CD Indexes (24 indexes, 24 samples) (Illumina, 20015881) were used for library prep. PCR amplicons of each sample were pooled in one library, i.e. the 17 multiplex PCR amplicons generated for the initial diagnosis of P0557 were pooled in one library. The six resulting libraries were then pooled, with initial diagnosis samples and relapse samples in a 3:1 ratio. sequencing was performed on the Illumina iSeq 100 (Illumina, 20021532) with the iSeq 100 i1 Reagent v2 (300-cycle) (Illumina, 20031371) for 2 x 150 bp paired-end reads.
BWA v0.7.13 was used to map the resulting reads to the GRCh38 human reference genome. The number of reads supporting each allele for the selected positions was extracted using bam-readcount v0.8.[26] As all 50 mutations were sequenced in all samples, the samples that did not carry a mutation could serve as a negative control for the sample that did carry said mutation.

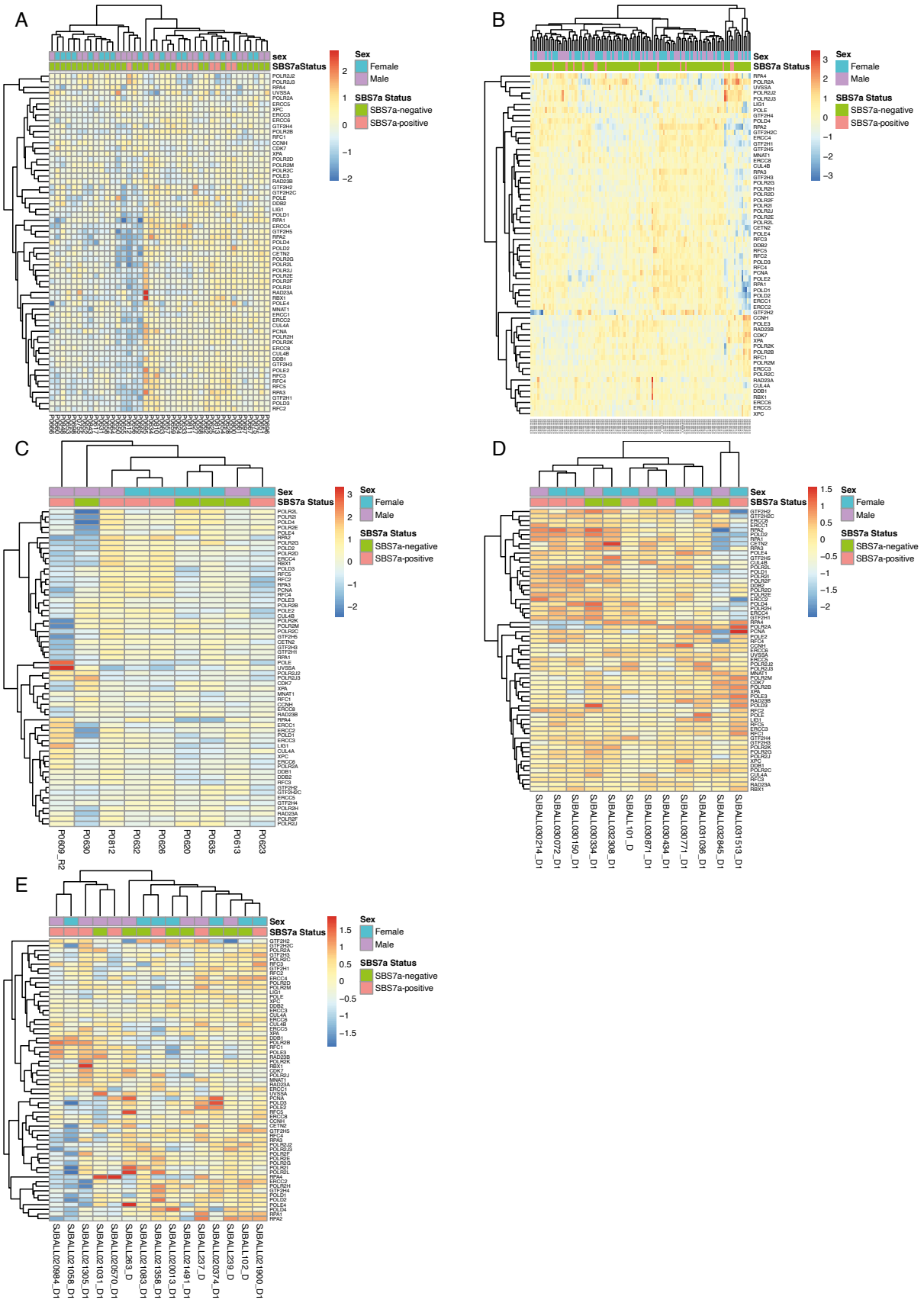
Mutational driver analysis
Clustered mutations which are either exonic or affect splicing were filtered for a CADD score of at least 20, and mutations were selected based on a list of known ALL driver genes.[21] Using the relative contributions per cluster, the probability of each mutation being caused by SBS7a was calculated based on the mutation type.[2,21,23] Mutations are shown which have a trinucleotide context which is also present in SBS7a and have a predicted probability to be caused by SBS7a of at least 50%.

Single-cell WGS data analysis
Reads were aligned to the GRCh38 human reference genome using BWA-MEM v2.2.1,[27] and duplicates were marked using GATK v4.4.0.0.[6] Mutations were called using GATK's HaplotypeCaller v4.4.0.0 and filtered using PTATO v1.3.3.[28] Mutations that, in one or more cells, had not been flagged as artifact by PTATO, had at least 5 supporting reads, had an allele frequency of at least 0.25 and had a mapping quality of at least 59, were used as positions for variant calling in all cells, the MSCs and the tumor bulk sample to ensure a minimal retention of artifacts. Variant calling was performed using Haplotypecaller and GenotypeGVCFs from GATK 4.2.0.0,[6] and SBSs were selected using bcftools v1.17.[29] An in-house blacklist of PTA artifacts, defined by a presence in multiple unrelated patients, was used to further filter the mutations. Lineage trees were constructed from the resulting mutations using CellPhyWrapper and visualized using CellPhyWrapperPlotting.[30,31] A refit using the previously identified signatures for BCP-ALL and the known PTA artifact signature was performed to obtain signature contributions for each branch that contained at least 100 mutations.

Supplementary References

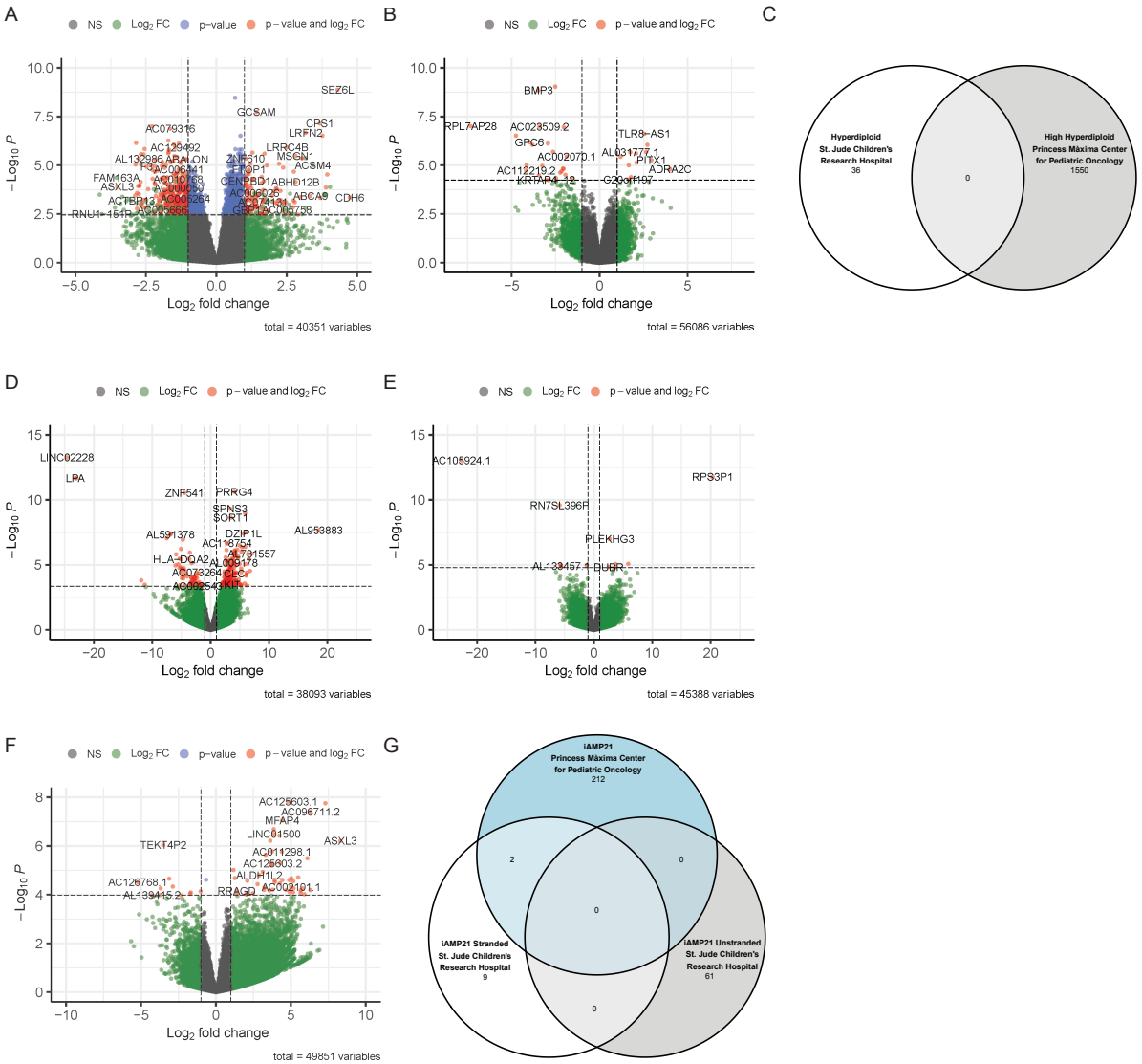
1. McLeod C, Gout AM, Zhou X, et al. St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. Cancer Discov 2021;11(5):1082–1099.
2. Brady SW, Roberts KG, Gu Z, et al. The genomic landscape of pediatric acute lymphoblastic leukemia. Nat Genet 2022;54(9):1376–1389.
3. Zhao S, Zhang Y, Gordon W, et al. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. BMC Genomics 2015;16(1):675.
4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–1760.
5. Picard. http://broadinstitute.github.io/picard/ (accessed June 12, 2025).
6. Van der Auwera GA, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. "O'Reilly Media, Inc."; 2020. 496 p.
7. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17(1):122.
8. Chen S, Francioli LC, Goodrich JK, et al. A genomic mutational constraint map using variation in 76,156 human genomes. Nature 2024;625(7993):92–100.
9. Boomsma DI, Wijmenga C, Slagboom EP, et al. The Genome of the Netherlands: design, and project goals. Eur J Hum Genet 2014;22(2):221–227.
10. Priestley P, Baber J, Lolkema MP, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature 2019;575(7781):210–216.
11. Manders F, Brandsma AM, de Kanter J, et al. MutationalPatterns: the one stop shop for the analysis of mutational processes. BMC Genomics 2022;23(1):134.
12. Perez G, Barber GP, Benet-Pages A, et al. The UCSC Genome Browser database: 2025 update. Nucleic Acids Res 2025;53(D1):D1243–D1249.
13. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29(1):15–21.
14. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 2019;47(8):e47.
15. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *bioRxiv*.
16. Kolde R. pheatmap: Pretty Heatmaps. *R package version 1.0.12*. https://github.com/raivokolde/pheatmap (2025).
17. Blighe K, Rana S, Lewis M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. *R package version 1.12.0*. https://bioconductor.org/packages/EnhancedVolcano (2025).
18. Enshaei A, Vora A, Harrison CJ, Moppett J, Moorman AV. Defining low-risk high hyperdiploidy in patients with paediatric acute lymphoblastic leukaemia: a retrospective analysis of data from the UKALL97/99 and UKALL2003 clinical trials. Lancet Haematol 2021;8(11):e828–e839.
19. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics 2017;33(19):3088–3090.
20. Statisticat, LLC. LaplacesDemon: Complete Environment for Bayesian Inference. *R package version 16.1.6*. https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software (2021).

21. van der Ham CG, Suurenbroek LC, Kleisman MM, et al. Mutational mechanisms in multiply relapsed pediatric acute lymphoblastic leukemia. Leukemia 2024;38(11):2366–2375.
22. Miller CA, McMichael J, Dang HX, et al. Visualizing tumor evolution with the fishplot package for R. BMC Genomics 2016;17(1):880.
23. Morganella S, Alexandrov LB, Glodzik D, et al. The topography of mutational processes in breast cancer genomes. Nat Commun 2016;7(1):11383.
24. Brady SW, Ma X, Bahrami A, et al. The clonal evolution of metastatic osteosarcoma as shaped by cisplatin treatment. Mol Cancer Res 2019;17(4):895–906.
25. Kõressaar T, Lepamets M, Kaplinski L, Raime K, Andreson R, Remm M. Primer3_masker: integrating masking of template sequence with primer design software. Bioinformatics 2018;34(11):1937–1938.
26. Khanna A, Larson DE, Srivatsan SN, et al. Bam-readcount - rapid generation of basepair-resolution sequence metrics. ArXiv 2021;7(69):3722.
27. Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). IEEE; p.
28. Middelkamp S, Manders F, Peci F, et al. Comprehensive single-cell genome analysis at nucleotide resolution using the PTA Analysis Toolbox. Cell Genom 2023;3(9):100389.
29. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. Gigascience;10(2):.
30. Bertrums EJM, de Kanter JK, Derks LLM, et al. Selective pressures of platinum compounds shape the evolution of therapy-related myeloid neoplasms. Nat Commun 2024;15(1):6025.
31. Kozlov A, Alves JM, Stamatakis A, Posada D. CellPhy: accurate and fast probabilistic inference of single-cell phylogenies from scDNA-seq data. Genome Biol 2022;23(1):37.



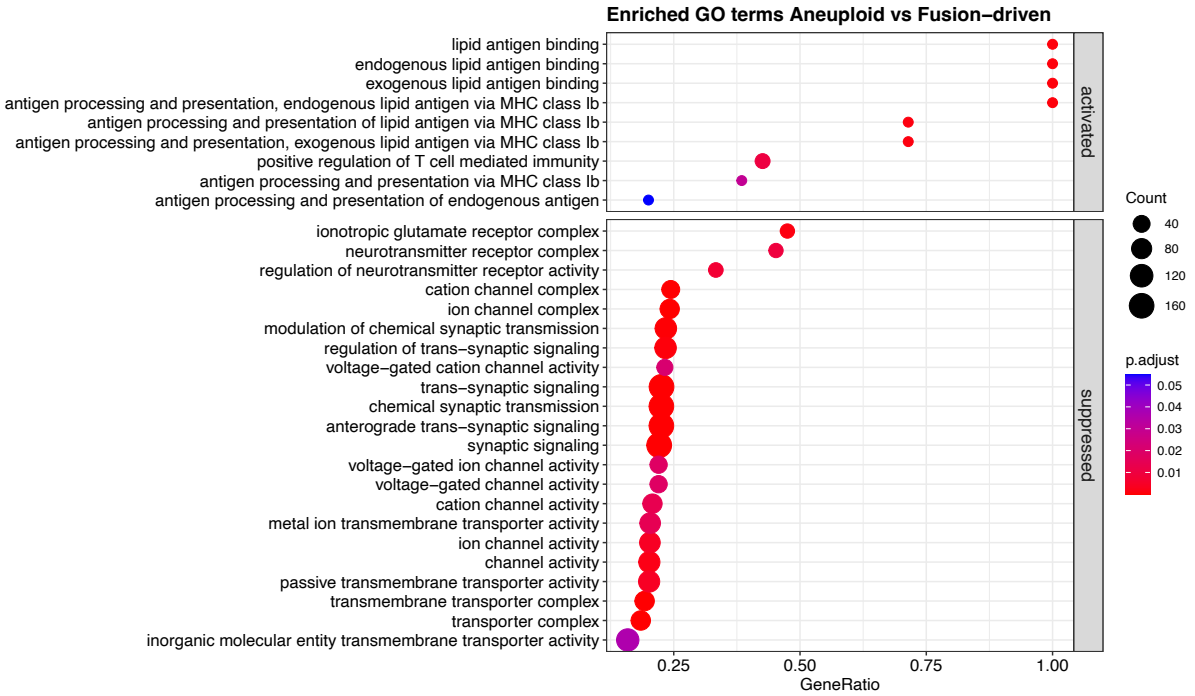
Supplementary figure S1. Genes in the NER pathway show no differential expression between SBS7a-positive and SBS7a-negative BCP-ALL. Whereas the NER pathway is upregulated in high hyperdiploid BCP-ALL in the cohort from the Princess Máxima Center (A), this is not seen in the high hyperdiploid BCP-ALL dataset from St. Jude (B), or in iAMP21

BCP-ALL, as seen in the cohort from the Princess Máxima Center (C) and the stranded (D) and unstranded (E) cohort from St. Jude. A-E) Heatmaps of differentially expressed genes between SBS7a-positive and SBS7a-negative samples in different cohorts. Bars above the plot show sex and whether a sample is SBS7a-positive (red) or SBS7a-negative (green). All genes from the NER pathway according to the Kyoto Encyclopedia of Genes and Genomes (Table S5) with expression in at least one sample are shown on the y-axis, sample names are shown on the x-axis.



Supplementary figure S2. Differential expression analysis for samples with and without SBS7a. A-B) Volcano plots of differentially expressed genes in hyperdiploid samples with SBS7a compared to samples without SBS7a. The y-axis represents the -log10 adjusted *P*-value and the x-axis represents the log2 fold change. Each gene is plotted as a circle and highlighted in green if there is a significant log2 fold change, highlighted in blue if there is a significant adjusted *P*-value and highlighted in red if there is both a significant log2 fold change and a significant adjusted *P*-value. A) Volcano plot of differential expression analysis between SBS7a-positive and SBS7a-negative high hyperdiploid BCP-ALL samples collected at the Princess Maxima center for Pediatric Oncology. B) Volcano plot of differential expression analysis between SBS7a-positive and SBS7a-negative high hyperdiploid BCP-ALL of the validation cohort collected by St. Jude children's hospital. C) Venn diagram showing the overlap in differentially expressed genes detected in our high hyperdiploid BCP-ALL samples and the St. Jude children's hospital hyperdiploid BCP-ALL samples. D) Volcano plot of differential expression analysis between SBS7a-positive and SBS7a-negative iAMP21 BCP-ALL collected at the Princess Maxima center for Pediatric Oncology. E) Volcano plot of differential expression analysis between SBS7a-positive and SBS7a-negative iAMP21 BCP-ALL samples of the validation cohort collected by St. Jude children's hospital on which stranded RNAseq was performed. F) Volcano plot of differential expression analysis between SBS7a-positive and

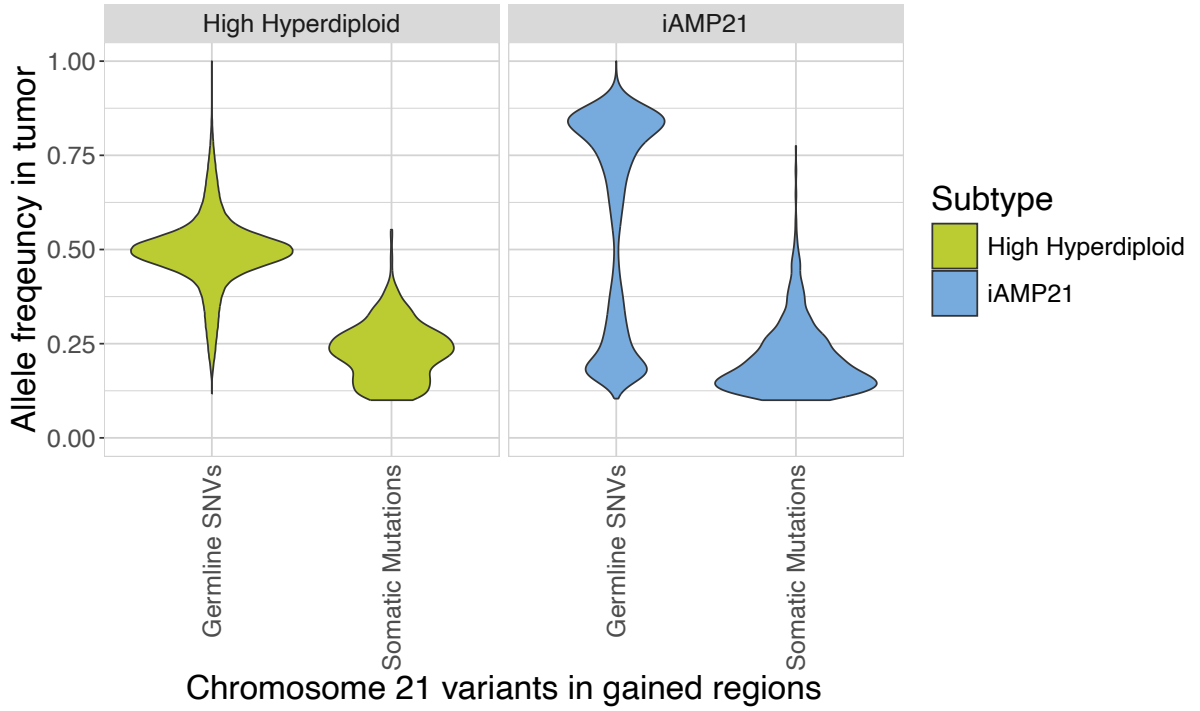
SBS7a-negative iAMP21 BCP-ALL samples of the validation cohort collected by St. Jude children's hospital on which unstranded RNAseq was performed. G) Venn diagram showing the overlap in differentially expressed genes detected in our iAMP21 BCP-ALL samples, the St. Jude children's hospital iAMP21 BCP-ALL samples with stranded RNAseq data and the St. Jude children's hospital iAMP21 BCP-ALL samples with unstranded RNAseq data. The two upregulated genes detected in both iAMP21 cohorts are *TMEM40* and *ZSCAN23*, which encode a transmembrane protein and a transcription factor, respectively.



Supplementary figure S3. Gene set enrichment analysis comparing aneuploid BCP-ALL subtypes to *ETV6*::*RUNX1* BCP-ALL. Dotplot of gene sets which are enriched in both iAMP21 and high hyperdiploid BCP-ALL compared to *ETV6*::*RUNX1* BCP-ALL. None of the gene sets identified in this analysis are indicative of a higher tendency of aneuploid BCP-ALL subtypes to migrate.



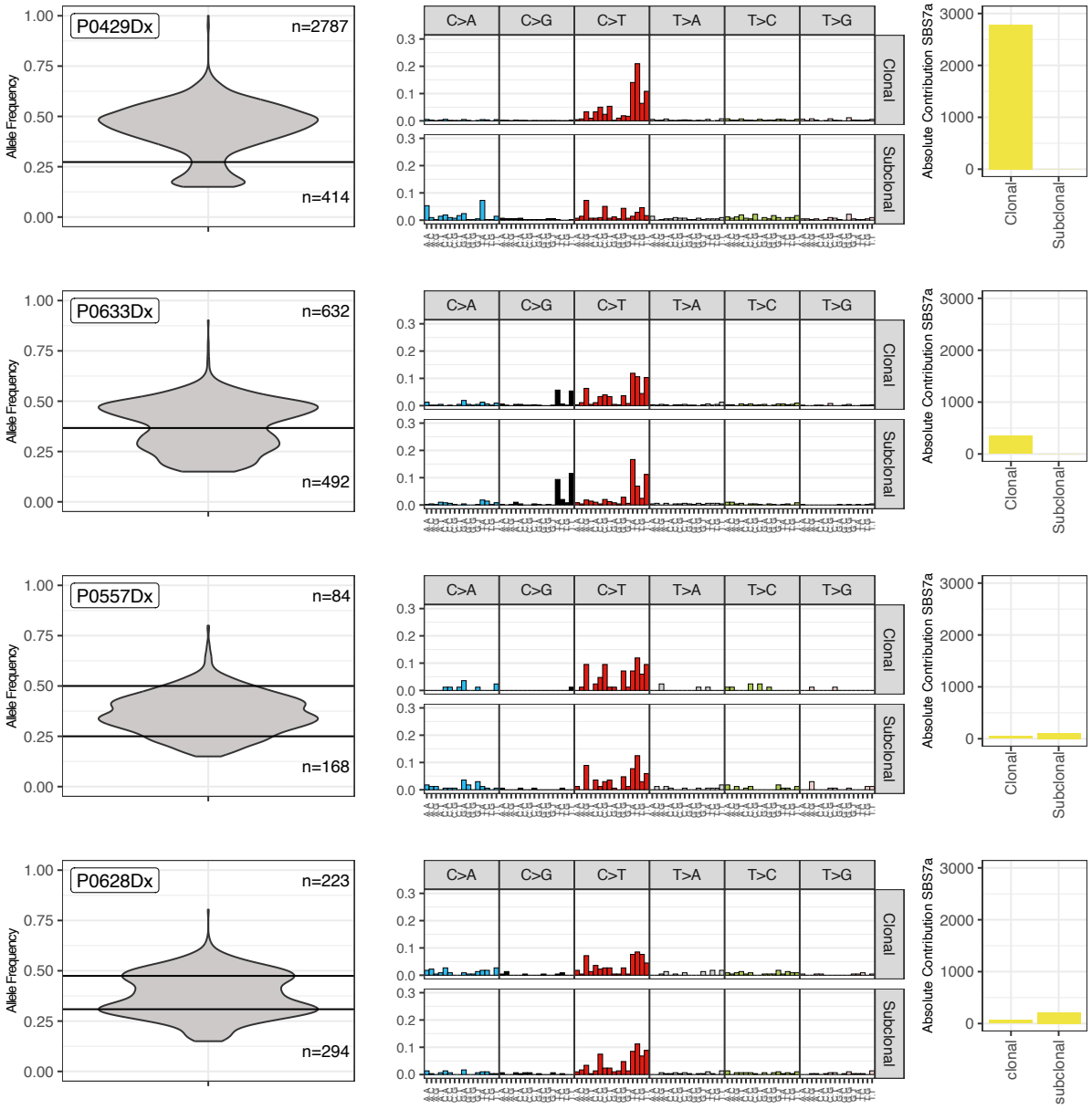
Supplementary figure S4. Over-representation analysis of genes which are differentially expressed in both iAMP21 and high hyperdiploid BCP-ALL compared to *ETV6*::*RUNX1* BCP-ALL. Dotplots of the top-50 enriched GO terms in upregulated (A) and downregulated (B) genes in aneuploid (iAMP21 and high hyperdiploid) BCP-ALL compared to *ETV6*::*RUNX1* BCP-ALL are shown. GO terms associated with migration are indicated with red arrows and are all enriched in upregulated genes.



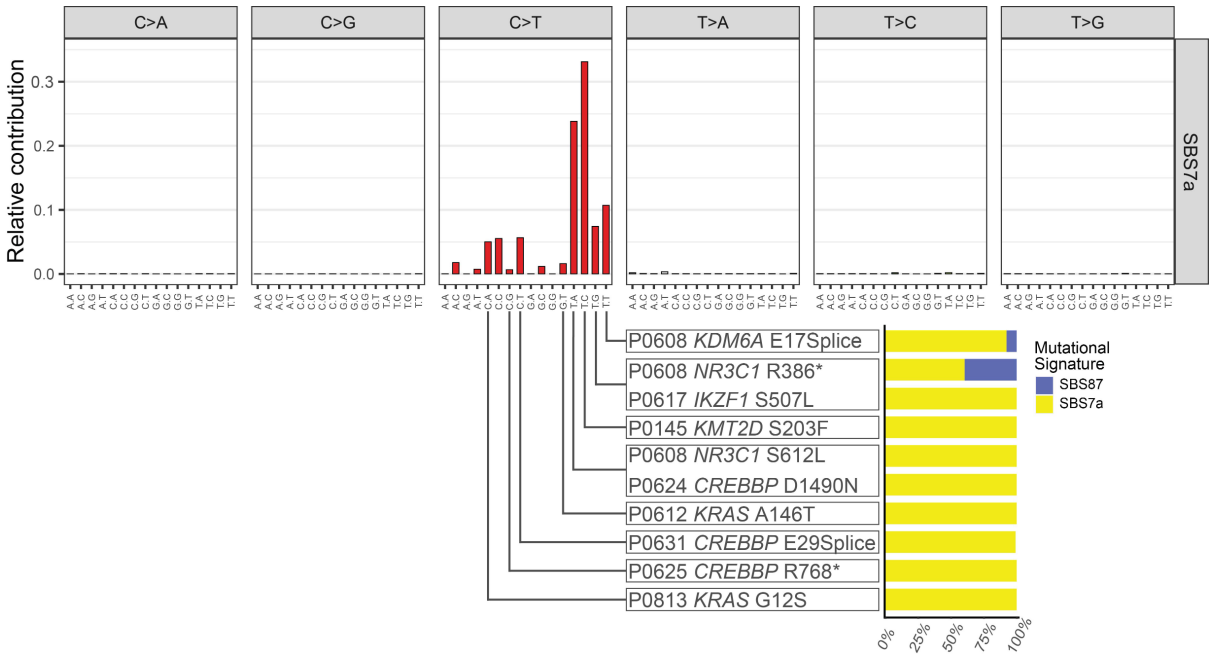
Supplementary figure S5. Somatic SBS7a-associated mutations are accumulated after chromosome 21 amplification in high hyperdiploid and iAMP21 BCP-ALL. Violin plots depicting the allele frequency of germline and somatic single base substitutions on chromosome 21 in high hyperdiploid (n=16) and iAMP21 BCP-ALL (n=9). In high hyperdiploid BCP-ALL, chromosome 21 typically has 4 copies, and germline SNVs of chromosome 21 therefore retain their allele frequency of 0.5. Somatic chromosome 21 mutations, however, present with an allele frequency of approximately 0.25, suggesting that they accumulated after amplification. Chromosome 21 in iAMP21 BCP-ALL has more complex CNAs, resulting in germline SNVs of chromosome 21 having either a high (~0.85) or a low (~0.15) allele frequency. Somatic mutations typically present with a low allele frequency, again confirming that they accumulated after the primary driver event.



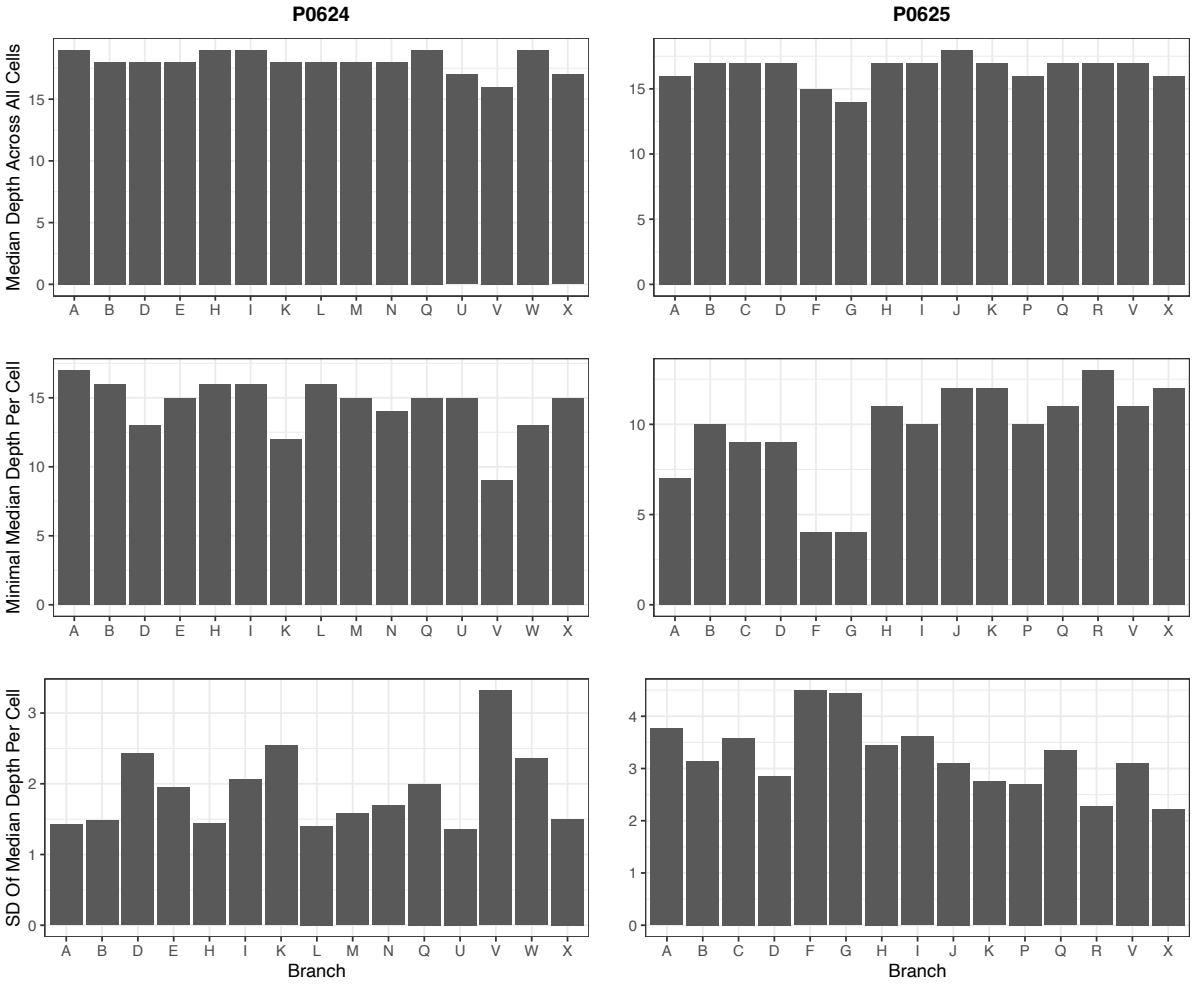
Supplementary figure S6. The allele frequency of ten out of 25 SBS7a-positive tumors shows a bimodal distribution, of which six present with at least 200 subclonal mutations. Violin plots of the allele frequency of all somatic single base substitutions in the initial diagnosis sample of 25 SBS7a-positive BCP-ALL. Horizontal lines show the split between subclone and major clone. Tumors P0557Dx, P0628Dx, P0621Dx, P0633Dx, P0629Dx and P0429Dx show a bimodal distribution with a subclone containing at least 200 mutations, and were therefore included in further analyses (red box). Four other tumors show a bimodal distribution with a subclone of less than 200 mutations, and one tumor (P0626Dx) shows a trimodal distribution. The remaining 14 tumors show a unimodal distribution. P0632 and P0612 were excluded due to insufficient tumor purity.



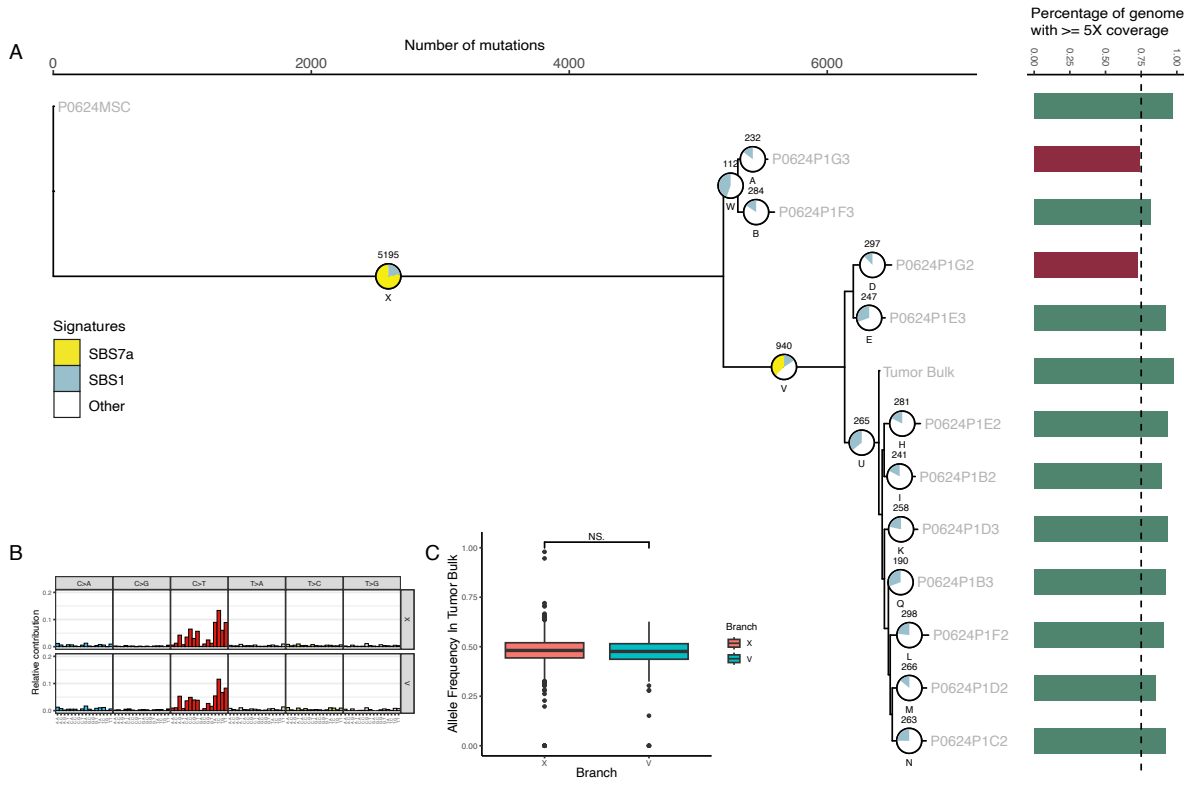
Supplementary figure S7. Sample P0429Dx and P0633Dx contain subclonal mutations without SBS7a, whereas samples P0557Dx and P0628Dx contain subclonal SBS7a-associated mutations. The first column depicts a violin plot of the allele frequency of somatic single base substitutions. Lines were added to indicate the split between mutations belonging to the major clone and subclonal mutations. The second column shows the mutational profile of the clonal and subclonal mutations as identified in the first column. The third column shows the relative contribution of SBS7a-associated mutations to the clonal and subclonal mutations identified in the first column.



Supplementary figure S8. 10 mutations in driver genes have a high probability of being caused by SBS7a. COSMIC signature SBS7a is shown on top, with on the bottom potentially pathogenic mutations in known BCP-ALL driver or relapse-associated genes. Bar plots on the right show the probability of being caused by SBS7a (yellow) or SBS87 (purple) as calculated by *a posteriori* probability analysis.



Supplementary figure S9. Branch-specific quality metrics for single-cell WGS data. Quality metrics for the branches of the lineage trees depicted in supplementary figure S10 (P0624, left column) and figure 6 (P0625, right column). The letter corresponding to each branch is displayed on the x-axis. The first row shows the median depth in each branch. The second row shows the minimal median depth in each branch. The third row shows the standard deviation of median depth in each branch.



Supplementary figure S10. Lineage tree based on the single cell WGS data of patient P0624. A) Lineage tree. The trunk of the lineage tree shows the number of shared mutations between all following clones. If a clone branches off, it means that it shares no more mutations with other clones. The number of mutations is depicted above each branch along with a pie chart that shows the percentage of mutations associated with each mutational signature. SBS7a is shown in yellow, SBS1 is shown in blue and other mutational signatures are shown in white. To the right of the lineage tree is a bar plot corresponding to each clone in the lineage tree. This bar plot shows whether over 75% of the genome is covered with a depth of at least 5 (green bars). B) Mutational profiles of each group of shared mutations that had a contribution of SBS7a. C) Boxplot that shows the allele frequency for each group of shared mutations that contained SBS7a-associated mutations.