7/15/2022

Signature SBS7a in pediatric BCP-ALL

Lianne Suurenbroek
PRINCESS MÁXIMA CENTER FOR PEDIATRIC ONCOLOGY

Internship report for Cancer, Stem Cells and Developmental Biology
Supervisors: Freerk van Dijk, Cédric van der Ham and Dr. Roland Kuiper
Examiners: Dr. Roland Kuiper and Dr. Ruben van Boxtel

Table of Contents

Abstract

Acute lymphoblastic leukemia (ALL) is the most common childhood cancer in the Netherlands, with an incidence of around 115 new cases per year. Although the survival rates have reached 90%, the underlying mutational mechanisms are still not fully understood. Multiple genetic subtypes have been identified, with large differences in prognosis. While patients with ALL typically have a very low tumor mutation burden (TMB), some patients show the presence of an active mutational process which can increase the TMB[1–3]. To gain insight into which mutational processes play a role in the carcinogenesis and drug resistance of ALL, we extracted mutational signatures from whole genome sequencing data. Strikingly, we have identified 14 patients with ALL across multiple cohorts who show a mutational profile similar to single base substitution signature 7a (SBS7a), which is typically seen in melanomas and has been associated with damage by ultraviolet (UV) light.

Here, we aim to study the etiology of mutational signature SBS7a in ALL by analyzing common features between SBS7a-positive ALL samples and comparing them with those found in melanomas. Major copy number alterations (CNAs) were often found in chromosome 21. Additionally, copy number analysis showed that the presence of signature SBS7a is enriched within the group of ALL patients with intrachromosomal amplifications of chromosome 21 (iAMP21). While amplifications of chromosome 21 were found to be neither sufficient nor necessary for the development of signature SBS7a, CNAs were always present prior to SBS7a mutations.

Next, we compared SBS7a-postive ALL patients to patients with melanomas, in which SBS7a is known to be caused by UV light, to investigate a possible role for UV light in the mutagenesis of ALL. While some characteristics of SBS7a in melanomas could also be found in ALL, such as an overrepresentation of SNVs on the untranscribed strand of the deoxyribonucleic acid (DNA), no replication bias was found. All patients with signature SBS7a also showed the presence of doublet base substitution signature 1 (DBS1), which can be caused by UV light. The presence of insertion and deletion signature 13 (indel/ID13), however, could not be identified in patients with ALL, as opposed to most melanoma cases. Patients who had developed relapses provided information about the timing of SBS7a, showing that no new SBS7a mutations were gained after the initial diagnosis in 3 out of 5 patients. In the other 2 patients, however, new SBS7a mutations were present in the first relapse.

Lastly, our data shows that SBS7a can influence disease progression through the inactivation of glucocorticoid-related genes, potentially leading to prednisolone resistance. Although UV light is unlikely to play a role in the development of ALL, our results show the presence of a similar type of damage in ALL for which the underlying mutational process has not yet been identified.

Layman’s summary

Acute lymfatische leukemie (ALL) is de meest voorkomende vorm van kanker bij kinderen: elk jaar komen er in Nederland ongeveer 115 nieuwe patiënten bij. Bij acute lymfatische leukemie blijven de B- of T-cellen delen in het beenmerg, waardoor zich uiteindelijk te veel van deze cellen in het bloed bevinden en deze daarmee de werking en productie verhinderen. Dit leidt vervolgens tot ziekte. Alhoewel kinderen met ALL door de toediening van chemotherapie een goede overlevingskans hebben (ca. 90%), komt de kanker bij een deel van de patiënten terug, ook wel recidieven genoemd. Om meer inzicht te krijgen in welke processen een rol spelen bij het ontstaan van de kanker en recidieven, kijken we in dit onderzoek naar mutatiepatronen.

Bepaalde stoffen, maar ook intrinsieke processen van de cel, kunnen het DNA veranderen, en laten hierbij hun eigen herkenbare spoor achter. Deze sporen noemen we mutatiepatronen. Ultraviolet (UV) licht laat bijvoorbeeld typische schade achter, te herkennen aan C naar T mutaties op specifieke posities. Dit mutatiepatroon van UV-schade, genaamd SBS7a, hebben wij gevonden bij 14 patiënten met ALL, terwijl UV-straling niet in het beenmerg door kan dringen. In deze studie richten wij ons op de vraag hoe het kan dat we dan toch het mutatiepatroon van UV-straling zien bij patiënten met ALL.

Ten eerste kijken we naar de karakteristieken van ALL-patiënten met SBS7a. 4 van deze patiënten hebben het zeldzame iAMP21 subtype, waarbij delen van chromosoom 21 meer dan 2 keer aanwezig zijn in de cel. In de algemene ALL-populatie vormt dit maar 1-2% van alle gevallen. Veel van de andere patiënten in ons cohort hebben ook meerdere kopieën van chromosoom 21; in totaal hebben 13 patiënten meer kopieën van chromosoom 21 dan van de andere chromosomen. We hebben nog geen regio kunnen aanwijzen die specifiek gekopieerd is bij patiënten met SBS7a. We zien wel dat de extra kopieën van chromosoom 21 eerder aanwezig waren dan de SBS7a mutaties.

Vervolgens hebben we de patiënten met ALL vergeleken met patiënten met huidkanker, omdat dat vaak door UV-straling wordt veroorzaakt. We identificeren in beide groepen patiënten een ander mutatiepatroon van UV-straling, namelijk DBS1. Mutatiepatroon ID13, wat ook veel in huidkanker voorkomt en geassocieerd is met UV-straling, vinden we niet terug in ALL.

Doordat UV-schade wordt gerepareerd door een proces dat gekoppeld is aan transcriptie, is SBS7a vooral aanwezig op de streng die niet getranscribeerd wordt. Dit is zowel in huidkanker als in ALL het geval. In huidkanker is SBS7a ook vooral aanwezig in regio’s van het genoom die laat gerepliceerd worden, omdat de replicatie dan minder betrouwbaar is. In ALL is dit niet het geval.

Tot slot hebben we naar de timing van SBS7a in ALL gekeken, door het DNA van meerdere tumoren van 5 patiënten met recidieven uit te lezen. Bij 3 van deze patiënten zien we geen nieuwe SBS7a-mutaties opkomen na de initiële diagnose. Bij de andere 2 patiënten daarentegen zien we bij het eerste recidief nog nieuwe mutaties opkomen. We zien in geen van de patiënten nieuwe mutaties opkomen na het eerste recidief, dus het lijkt erop de het proces niet altijd actief blijft. Alhoewel het niet waarschijnlijk is dat UV-straling een rol speelt bij ALL, laat onze data zien dat er wel vergelijkbare schade aanwezig is waarvan de oorzaak nog onbekend is.

List of abbreviations

ALL	Acute Lymphoblastic Leukemia
BCP-ALL	B-Cell Precursor Acute Lymphoblastic Leukemia
BWA	Burrows-Wheeler Aligner
CNA	Copy Number Alteration
DBS	Doublet Base Substitution
DNA	Deoxyribonucleic Acid
GATK	Genome Analysis ToolKit
GoNL	Genome of the Netherlands
iAMP21	Intrachromosomal Amplification of chromosome 21
ID/Indel	Insertion/Deletion
MNV	Multi-Nucleotide Variant
NER	Nucleotide Excision Repair
NMF	Non-negative matrix factorization
SBS	Single Base Substitution
SNV	Single-Nucleotide Variant
TMB	Tumor Mutation Burden
UV	UltraViolet
VAF	Variant Allele Frequency
VEP	Variant Effect Predictor
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Introduction

ALL is the most common type of pediatric cancer, with an incidence of around 115 new cases per year in the Netherlands[4]. ALL shows a peak incidence in children between 3 and 5 years old, and 55-60% of these patients are boys[4,5]. Although survival rates have reached 90%, the underlying mutational mechanisms are not yet fully understood[4,5]. Multiple different genetic subtypes have been identified, and prognosis largely differs between these subgroups[6]. Whereas some subtypes are relatively well understood, the mechanisms of others remains to be elucidated[6]. Additionally, many patients are still classified as having B-other ALL, meaning that no clear subtype could be assigned. Improving our knowledge of the identified subtypes and understanding the processes underlying B-other ALLs might help us generate more accurate prognoses. Furthermore, understanding the mechanisms of ALL could eventually help the development of targeted therapies.

ALL is characterized by neoplasms of progenitors of either T- or B-cells, with 86% of pediatric cases having B-cell precursor (BCP)-ALL. This percentage correlates with the age of the patients: 94% of the cases younger than 5 years have BCP-ALL, while 73% of patients aged 15-17 years have BCP-ALL[4].

Several subtypes of BCP-ALL are characterized by large CNAs. The most common subtype in BCP-ALL is hyperdiploidy, which is seen in roughly 25% of pediatric cases, is characterized by a chromosome number of at least 51 and is associated with a good prognosis[6]. In contrast, patients with the hypodiploid subtype generally have a poor prognosis[6]. These patients have a chromosome number of less than 39[6]. The last subtype characterized by large CNAs is a group of patients with iAMP21[6]. This subtype is quite rare as it is seen in around 1% of pediatric BCP-ALL patients, and is associated with a poor prognosis[6]. iAMP21 is thought to be caused by breakage-fusion-bridge cycles[7,8]. Additionally, the presence of a Robertsonian translocation of chromosome 15 and 21 can contribute to iAMP21[7,8]. Other subtypes are typically characterized by the presence of a fusion gene[6]. One of these subtypes is *ETV6::RUNX1*, which accounts for 20-25% of cases and is associated with a very good prognosis[6]. *BCR::ABL1*, which is also known as the Philadelphia chromosome and has been linked to chronic myeloid leukemia, is often seen in BCP-ALL as well[6]. Some subtypes have been identified which are similar to subtypes with fusion genes on a transcriptomic level but in which no fusion gene could be found, including *ETV6::RUNX1*-like ALL and Philadelphia-like ALL[6]. Tumors for which the subtype cannot be determined are called B-other ALL. The size of this group depends on which technique is used for the classification, since some subtypes can only be detected by certain methods. The percentage can therefore be as low as 5% or above 50% depending on the sensitivity of the techniques[9].

Although some subtypes are associated with a good prognosis, patients with other subtypes are still very prone to relapses[6]. To identify subtype-specific mutational processes, the extraction of mutational signatures is a very strong tool. The identified signatures in BCP-ALL include signature SBS1, which is associated with aging, SBS2 and SBS13, which are associated with APOBEC activity, and SBS87, which has been linked to thiopurine therapy[3,10]. Strikingly, signature SBS7a has also been found in ALL, characterized by C>T mutations at C$\underline{\text{C}}$N and T$\underline{\text{C}}$N trinucleotides[1,3,10–12]. This signature, which is shown in Figure 1, is typically found in melanomas and other types of skin cancer[12–15]. Signature SBS7a has therefore been associated with exposure to UV light, which has also been proven *in vitro*[16,17].

UV light can form pyrimidine dimers, which are usually repaired by nucleotide excision repair (NER)[18,19]. When NER fails to repair these pyrimidine dimers, cytosines can be deaminated and turn into uracil. This leads to the introduction of C>T mutations with the characteristic pattern of signature SBS7a. Because UV damage is partly repaired by transcription-coupled NER, signature SBS7a has a bias towards the untranscribed strand[13,20]. For the same reasons, SBS7a mutations are more prevalent in lowly transcribed genes than in genes with a high expression[20]. Tumors with SBS7a typically have a very high TMB, with cutaneous melanomas having an average TMB of 49.17 single-nucleotide variants (SNVs) per megabase[15]. ALL, on the other hand, typically presents with a very low TMB of on average 0.34 SNVs and indels per megabase[1,2]. This number differs a lot between subtypes, with iAMP21 tumors showing the highest TMB and KMT2A-rearranged tumors showing a low TMB[1]. Additionally, studies have shown an association between the presence of SBS7a and TMB in ALL[1].

Apart from SBS7a, UV damage can also result in the presence of signature DBS1, which is characterized by CC>TT mutations[12,21,22]. This signature is therefore often seen in melanomas and has also been validated *in vitro*[12,16,22]. In addition, UV light is thought to cause thymine deletions at thymine-thymine dinucleotides, causing signature ID13[12]. Although this signature has not yet been experimentally validated, there is a strong statistical association between the presence of SBS7a, DBS1 and ID13[12].

Although UV light can reach the dermis, it cannot penetrate the skin. UV light should therefore not be able to reach the bone marrow and damage lymphocytic precursors. This study focuses on identifying the underlying mutational process of signature SBS7a in ALL. First, we have identified similarities shared between ALL patients with SBS7a.

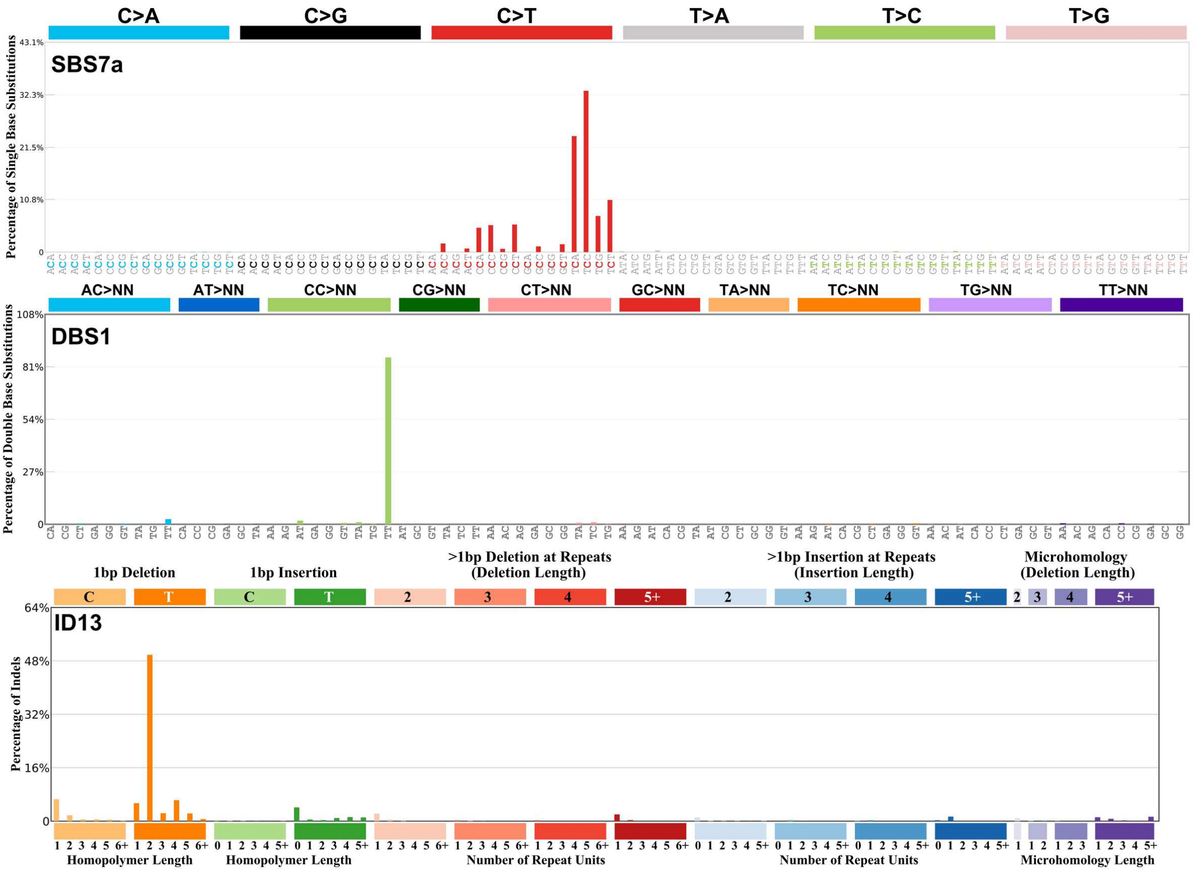


Figure 1: Signatures which have been associated with damage by UV light. *SBS7a is a single base substitution signature and consists of C>T mutations. DBS1 is a doublet base substitution signature and consists of CC>TT mutations. ID13 is an indel signature and mostly shows 1 base-pair deletions of thymines with a homopolymer length of 2. While SBS7a and DBS1 have been validated in vitro, the association between ID13 and UV radiation is based on statistics.*

We have also looked at similarities and differences between the presentation of SBS7a in melanomas and in ALL. Furthermore, we have studied the timing of the underlying mutational process of SBS7a in ALL. Lastly, we have determined the effects of SBS7a on further disease progression.

Materials and Methods

Patient samples

Patients with a mutational pattern resembling SBS7a were selected as follows: patients P0608 and P0609 were selected for whole genome sequencing (WGS) for a study on patients who had developed multiple relapses. They were then included in this study due to the high cosine similarity of their mutational profile to signature SBS7a. Patients P0610, P0611, P0557 and P0621 were selected for whole exome sequencing (WES) as part of a screening of relapsed ALL, and then selected for WGS based on high cosine similarity to SBS7a based on the WES results. Patients P0612, P0622, P0623, P0624, P0625, P0626, P0627 and P0628 were selected for WGS as part of routine diagnostics in the Princess Máxima Center for Pediatric Oncology and used for this study due to the high cosine similarity of their mutational profile to signature SBS7a.

In addition, we used a cohort of BCP-ALL patients with CNAs affecting chromosome 21. This includes BCP-ALL patients with the hyperdiploid and iAMP21 subtype and 1 patient with Down syndrome. These samples were whole exome sequenced and data were kindly provided by the den Boer group (Princess Máxima Center for Pediatric Oncology). The generation, mapping and somatic variant calling of these data was performed by the den Boer group according to Genome Analysis Toolkit (GATK) best practices[23].

In accordance with the Declaration of Helsinki, informed written consent was obtained from all patients and/or their legal guardians before enrolment in the study and the DCOG institutional review board approved the use of excess diagnostic material for this study (PMCLAB2019.054, PMCLAB2021.279).

Library preparation

Mononuclear cells were obtained from either bone marrow or peripheral blood using Ficoll-Paque (Cytiva, Marlborough, United States). DNA was then isolated using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). Libraries were prepared using the Illumina TruSeq Nano DNA Library Prep kit (Illumina, San Diego, United States).

Whole genome sequencing

WGS for patients P0608, P0609 and P0611 was performed at the Hartwig Medical foundation (Amsterdam, The Netherlands). For patients P0557 and P0621, WGS was performed at USEQ (Utrecht, The Netherlands). For patient P0610, the sample from the initial diagnosis was sequenced at USEQ, while the remission and relapse sample were sequenced at the Hartwig Medical Foundation. Except for patient P0608, all samples were sequenced on an Illumina NovaSeq 6000 platform using 150 base-pair paired-end reads, at a target depth of 30 for tumor samples and 15 for remission samples. For patient P0608, all samples were sequenced on an Illumina NovaSeq 6000 platform using 150 base-pair paired-end reads, and the reached depth is shown in Supplementary Table 1[3]. The samples of patients P0612, P0622, P0623, P0624, P0625, P0626, P0627 and P0628 were sequenced at the Princess Máxima Center for Pediatric Oncology (Utrecht, The Netherlands), with a target depth of 68 for tumor samples and 30 for remission samples.

Data analysis and quality control

For each sample, the reads were mapped to the GRCh38 human reference genome using the Burrows-Wheeler aligner (BWA)[24]. Afterwards, duplicate reads were marked with

Picard[25], after which GATK[23] was used to perform base quality score and variant quality recalibration. Finally, germline variants were called using GATK HaplotypeCaller followed by GenotypeGVCF. All this was done according to GATK best-practices guidelines[23]. Afterwards, all samples were checked to see if they passed quality control.

Somatic single-nucleotide variant calling and filtering

Somatic SNVs were called using Mutect2 of GATK version 4.1.1.0[23]. Filters were applied using FilterMutectCalls, and only variants with a PASS filter were used[23]. SNVs were selected with SelectVariants[23]. The SNVs were annotated using variant effect predictor (VEP) version 92[26]. Additionally, population frequencies from gnomAD version 3.0 and population frequencies from Genome of the Netherlands (GoNL) [26–28] were added. Further filtering of the variants was performed with R version 4.1.2[29,30]. Variants within centromeric regions as defined by the UCSC genome browser, variants with reads in the remission sample and variants with a population frequency of at least 0.01 in either gnomAD or GoNL were filtered out. Furthermore, only variants with a coverage of at least 20X, 5 or more supporting reads of the alternative allele and a minimal variant allele frequency (VAF) of 0.25 were selected for further analyses. For the clustering of variants based on VAF at multiple timepoints, variants with a coverage of at least 20X in all samples, 5 or more supporting reads of the alternative allele in at least one sample and a minimal VAF of 0.25 in at least one sample were used.

Somatic multi-nucleotide variant calling and filtering

Somatic multi-nucleotide variants (MNVs) were called using Mutect2 of GATK version 4.1.1.0[23]. Filters were applied using FilterMutectCalls, and only variants with a PASS filter were used[23]. MNVs were selected with SelectVariants[23]. The MNVs were annotated using VEP version 92, population frequencies from gnomAD version 3.0 and population frequencies from GoNL[26–28]. Germline variants were extracted from the remission samples using HaplotypeCaller from GATK version 4.0.1.2[23]. These variants were filtered according to GATK best practices[23]. Only germline variants with a PASS filter were selected. Next, MNVs at the same position of germline variants and MNVs which are directly adjacent to germline variants were filtered out using tabix from samtools version 1.3[31]. Further filtering of the MNVs was performed with R version 4.1.2, using the same selection criteria as were described for the SNVs.

Somatic indel calling and filtering

Indels were called using Mutect2 of GATK version 4.1.1.0, excluding soft clipped bases from calling[23]. Indels were selected with SelectVariants[23]. The indels were annotated using VEP version 92, population frequencies from gnomAD version 3.0 and population frequencies from GoNL[26–28]. Variants were then filtered using the Encode DAC Exclusion List using vcftools version 0.1.14[32,33]. Further filtering of the indels was performed with R version 4.1.2, in the same way as was done for the SNVs. Additionally, indels with a length of at least 10 base pairs and indels within 20 base pairs of each other were filtered out. Lastly, only indels for which the mean mapping quality of both alleles was 60 were selected.

Mutational profile analysis

Count matrices and mutation profiles for SNVs, DBSs and indels were made for each timepoint using the R package MutationalPatterns version 3.4.1[34]. MutationalPatterns was also used for calculating transcriptional strand bias and replication bias. For transcriptional strand bias, UCSC known genes for reference genome GRCh38 were used, as extracted with R package TxDb.Hsapiens.UCSC.hg38.knownGene version 3.14.0[35,36]. Repli-seq data from several cell lines of the ENCODE project (Gm06990, Gm12801, Gm12812, Gm12813, Gm12878, K562, Bg02es, Bj and MCf7) were used for replication bias[2,37]. The extracted signatures were compared to known signatures from COSMIC 3.2 by calculating the cosine similarity.

Mutational signature extraction

R packages MutationalPatterns version 3.4.1 and non-negative matrix factorization (NMF) version 0.24.0 were used for de novo mutational signature extraction[34,38]. As input, we used the filtered SNVs from the first timepoint of each patient. Additionally, data from 214 external pediatric ALL patients were added as input to gain power and prevent overfitting[11]. The cophenetic correlation coefficient was used to determine the optimal rank. This resulted in the extraction of 6 mutational signatures. The extracted signatures were compared to known signatures from COSMIC 3.2 by calculating the cosine similarity.

Copy number alterations

For patients P0612, P0622, P0623, P0624, P0625, P0626, P0627 and P0628, copy numbers were calculated by routine diagnostics, according to GATK best practices[23]. For the other patients, copy numbers were calculated using GATK version 4.1.7.0, according to GATK best practices[23], which was also used for the routine diagnostics samples. For denoising the read counts, an internal panel of normals from the Princess Máxima Center for Pediatric Oncology was used, which is based on WGS data from either healthy blood or skin samples of patients with varying diagnoses. The resulting segment files were annotated in R version 4.1.2, where segments with a copy number above 1.2 were called as a gain and segments with a copy number below 0.6 were called as a loss. Segments with a copy number between 0.6 and 1.2 were called as neutral and filtered out. The annotated files were further processed in R version 4.1.2.

Mutation timing

For the timing of SNVs compared to the timing of CNAs, we determined whether the position of SNVs overlapped with the annotated CNAs.

For patients P0557, P0608, P0609, P0610 and P0621, WGS data at multiple timepoints was available. To look at mutation timing, the filtered somatic SNVs were clustered based on their VAF at different timepoints. This was done using k-means clustering with the R package stats version 4.1.2[3,29]. A k of 10 was used, and after clustering the clusters were manually merged, split and cleaned to obtain biologically relevant clusters[3]. Signatures were extracted from the separate clusters using de novo extraction as described above. As input, we used the separate clusters, the SNVs from the remaining 9 patients with SBS7a and data from 214 external pediatric ALL patients. This resulted in the extraction of 9 signatures.

Targeted deep sequencing

Custom probes for all SNVs of patient P0608 were designed using Roche HyperDesignTool (Roche, Basel, Switzerland), along with variants found in other patients from the multiple relapses cohort. Using these probes, DNA from patient P0608 and 5 other patients with multiple relapses were selectively amplified. Sequencing was performed at the Hartwig Medical foundation (Amsterdam, The Netherlands) on an Illumina NovaSeq 6000 platform using 150 base-pair paired-end reads, at a target depth of 1000. Data was analyzed using the same pipeline as described above (Data analysis and quality control). The data were then further processed using R version 4.1.2. Variants with less than 200 reads across all samples were filtered out. Around 14% of variants were filtered out due to a low number of reads, meaning that the efficiency was close to the expected 85.2%.

Pathogenic variants

SNVs were filtered as described above, and variants in protein coding regions were selected based on the consequence as assigned by VEP. Variants with the following consequences were selected: synonymous_variant, inframe_insertion, inframe_deletion, missense_variant, protein_altering_variant, transcript_ablation, splice_acceptor_variant, splice_donor_variant, stop_gained, stop_lost, start_lost, frameshift_variant, transcript_amplification. A list of ALL driver genes was made, based on the Cosmic Cancer Gene Census and the top 250 mutated genes in B-ALL based on 1588 samples from the St. Jude Cloud[39,40]. Additionally, genes that have been identified as ALL driver genes or associated with therapy resistance in several studies were added to the list (**Error! Reference source not found.**) [2,10,41]. All identified SNVs that occurred in the list were manually checked. The probability of a mutation being caused by signature SBS7a was calculated by multiplying the contribution of SBS7a to the sample with the contribution of the mutation type to signature SBS7a. All probabilities were scaled to add up to one.

Code availability

All code used for the analyses in this study are available on https://bitbucket.org/lsuurenbroek/sbs7ainall/src/master/.

Results

ALL patients with SBS7a share CNAs on chromosome 21

Using WGS, we have identified 14 patients with BCP-ALL who carry signature SBS7a. The tumors of these patients showed a mutational profile with a high cosine similarity to SBS7a (Figure 2a, Supplementary Figure 1) and we were able to extract an SBS7a-like signature using de novo signature extraction (Figure 2b, Supplementary Figure 2). De novo signature extraction also showed that all 14 samples had a contribution of at least 20% of the SBS7a-like signature (Figure 2c). In some patients, SBS1 had a high contribution as well. Of the 14 patients, 5 have high hyperdiploid BCP-ALL (P0621, P0624, P0625, P0627 and P0628), 4 have iAMP21 BCP-ALL (P0609, P0611, P0623 and P0626), 1 has near haploid BCP-ALL (P0622) and 1 has Down syndrome/high hyperdiploid BCP-ALL (P0557) (Supplementary Figure 3, Supplementary Figure 4). The remaining patients (P0608, P0610 and P0612) have B-other ALL. These numbers deviate from the distribution of subtypes in the general BCP-ALL population. Since only 1% of BCP-ALL cases are classified as iAMP21[6], our cohort contains a relatively high number of iAMP21 patients (4 out of 14 patients, 29%). Previous research has also shown an overrepresentation of SBS7a in patients with iAMP21 ALL[1].

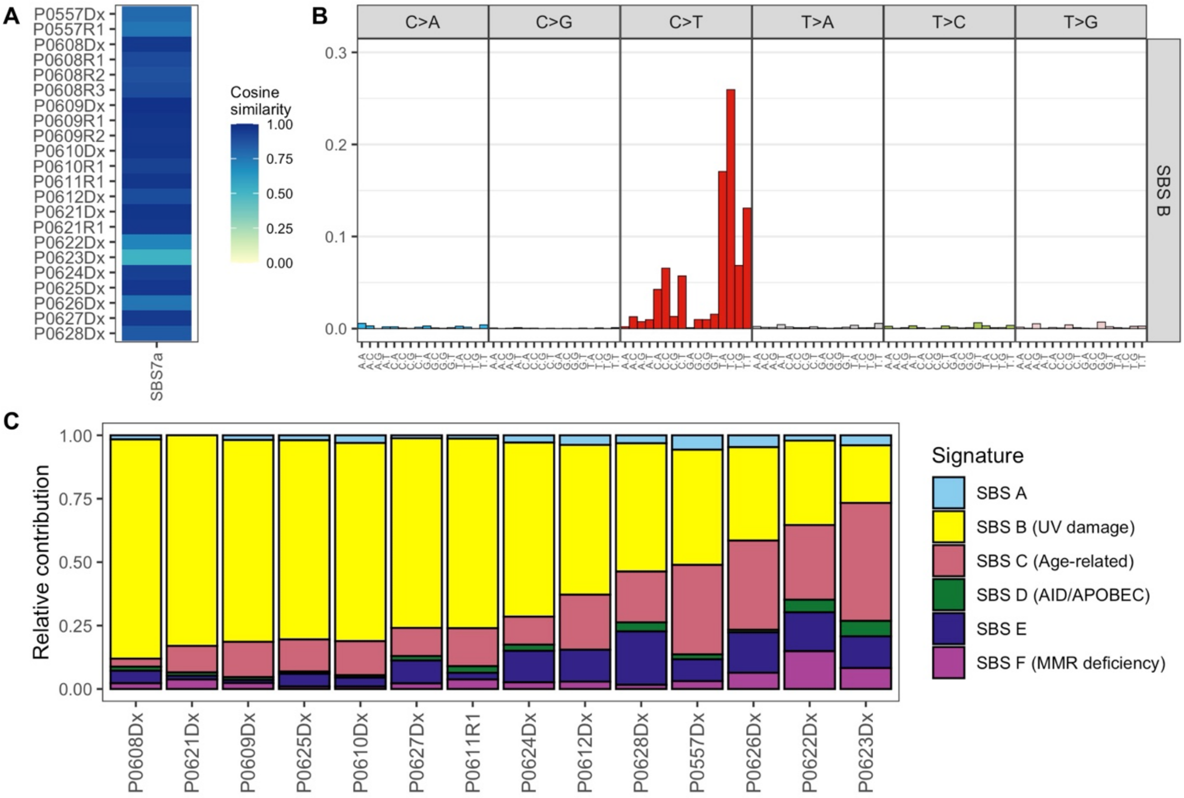


__Figure 2: 14 patients with ALL show the presence of signatures SBS7a.__ __a__ Heatmap showing cosine similarity of mutational profiles with COSMIC signature SBS7a. 14 patients with BCP-ALL were identified whose mutational profiles show a high similarity to mutational signature SBS7a. All patients show a cosine similarity of at least 0.5. __b__ De novo extraction using WGS data from ALL patients resulted in the extraction of signature B, which shows a high similarity to COSMIC signature SBS7a (cosine similarity = 0.986). __c__ Relative contribution of de novo extracted signatures in 14 patients with SBS7a. De novo signature extraction shows a contribution of SBS7a of at least 0.2 in 14 patients with BCP-ALL. Dx: initial diagnosis, R: relapse

All patients with high hyperdiploid BCP-ALL in this cohort have a gain of chromosome 21, as does patient P0612. Additionally, patient P0610 has a CNA plot similar to that of patients with the iAMP21 subtype, with many rearrangements of chromosome 21 (Supplementary Figure 3). In total, 13 out of 14 SBS7a-positive BCP-ALL have at least partial amplifications of chromosome 21 compared to the other chromosomes, suggesting that major CNAs on chromosome 21 are associated with the presence of SBS7a.

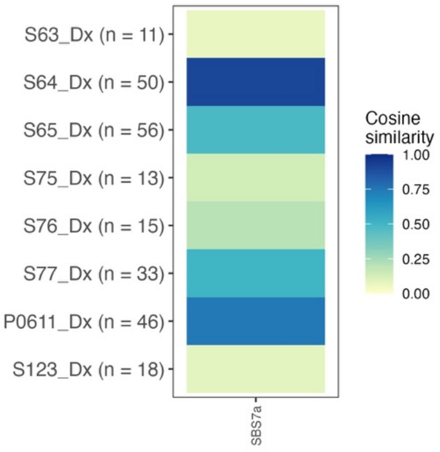


Figure 3: Heatmap of cosine similarity between mutational profiles and SBS7a. *Based on WES data, the mutational profile of 4 out of 8 patients with iAMP21 shows high cosine similarity with COSMIC signatures SBS7a.*

To further investigate the overrepresentation of chromosome 21 amplifications in SBS7a-positive ALL, we have analyzed a WES cohort of BCP-ALL patients with major CNAs of chromosome 21. This cohort includes patients with the high hyperdiploid and iAMP21 subtype and 1 patient with Down syndrome. Due to the low number of SNVs, it is challenging to construct a mutational profile based on WES data. 2 out of 8 patients with the iAMP21 subtype showed a high cosine similarity to signature SBS7a (Figure 3, Supplementary Figure 5). 2 other patients showed some, albeit low, cosine similarity to signature SBS7a. For the other 4 patients less than 20 mutations could be identified, meaning that no reliable mutational profile could be constructed. In conclusion,

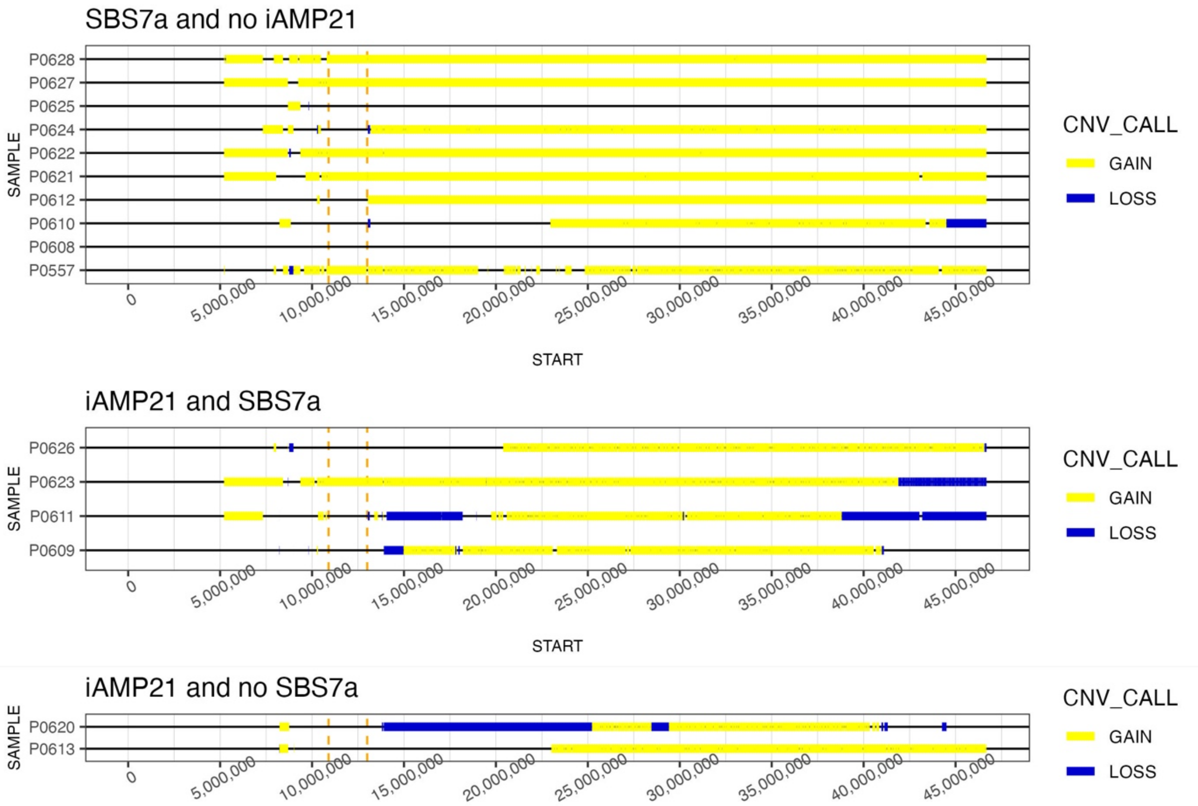


Figure 4: Overview of CNAs of chromosome 21 in 14 BCP-ALL patients with and without signature SBS7a and iAMP21. *Although many patients with signature SBS7a have gains of large regions on chromosome 21, some patients with signature SBS7a do not show any CNAs of chromosome 21. Gains are shown in yellow; losses are shown in blue. Orange dashed lines mark the position of the centromere.*

at least 25% of iAMP21 patients in this cohort show the presence of signature SBS7a, while around 10% of all BCP-ALL patients show some contribution of SBS7a[1].
The overrepresentation of CNAs on chromosome 21 in the WGS cohort and the high number of iAMP21 patients with SBS7a in the WES cohort might indicate a causal relationship between amplifications of chromosome 21 and signature SBS7a. To determine whether a specific region of chromosome 21 was amplified in SBS7a-positive patients, we compared the locations of CNAs. No regions were found to be exclusively altered in all patients with SBS7a (Figure 4). Additionally, not all BCP-ALL cases with the iAMP21 or hyperdiploid subtype show the presence of signature SBS7a[1]. Thus, while CNAs of chromosome 21 might play a role in the development of signature SBS7a, it is neither necessary nor sufficient for the activation of the mutational process behind SBS7a.

Amplifications of chromosome 21 are present prior to SBS7a mutations

To further investigate the probability of a causal relationship between amplifications of chromosome 21 and signature SBS7a, we have compared the relative timing of SNVs within gained regions of chromosome 21 with SNVs outside of CNAs. This was done by comparing the VAF of both groups. While clonal SNVs typically have a VAF of 0.5, SNVs in amplified

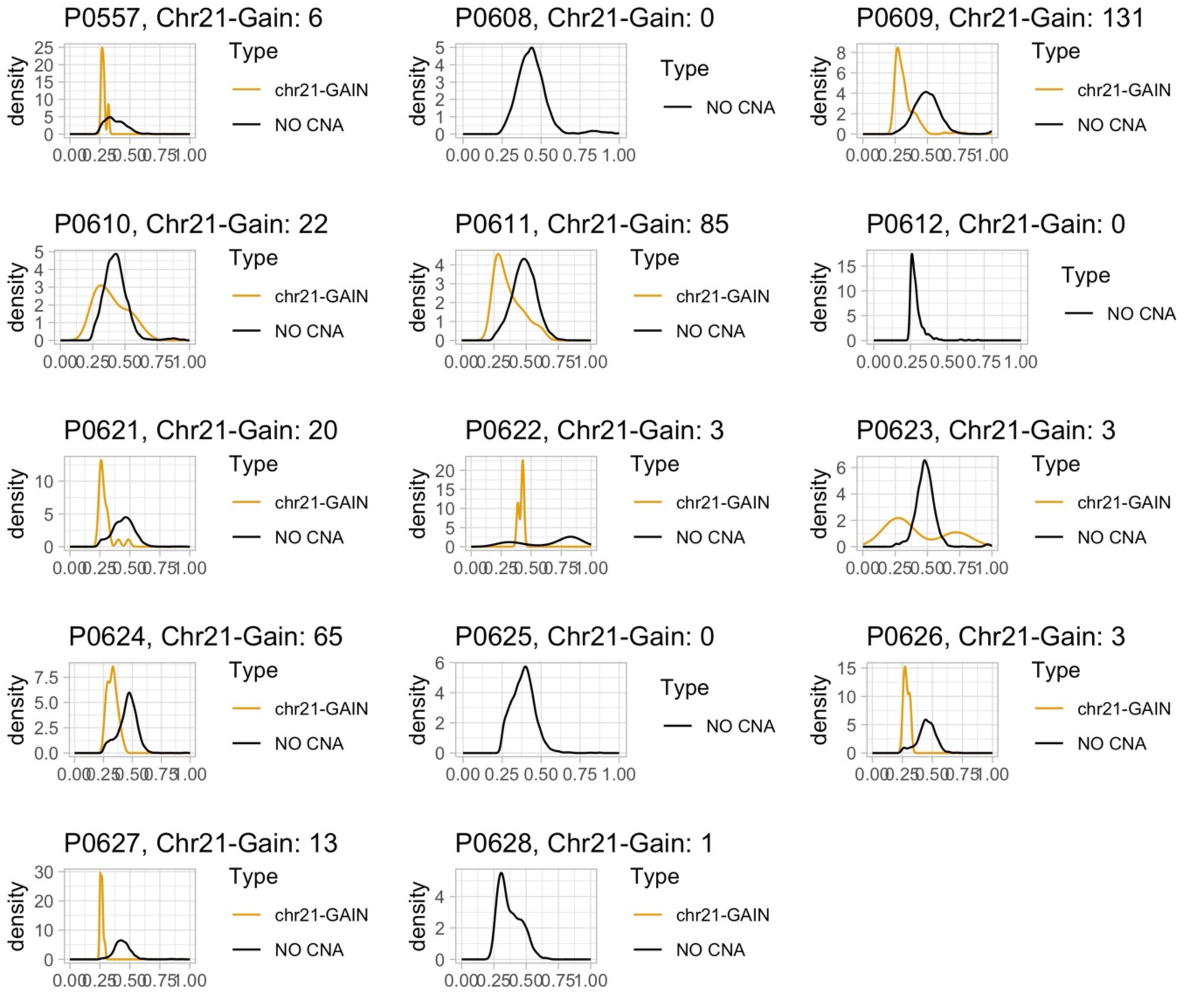


Figure 5: CNAs on chromosome 21 were present before the formation of SNVs. While SNVs outside of CNAs (black) have a VAF which usually peaks at 0.5, as expected, SNVs within amplified regions of chromosome 21 (orange) generally show a lower peak. For patients P0608, P0612, P0625 and P0628 no density could be plotted due to the low number of SNVs within gains of chromosome 21. The number of SNVs within gains of chromosome 21 is shown for each sample.

regions are expected to have a VAF of 0.33 if they were either formed after the region was amplified or if the other copy of the chromosome was amplified. SNVs which were on the amplified copy before the alteration, however, are expected to have a VAF of 0.67. Our results show a peak at 0.5 for SNVs outside of CNAs, as was expected (Figure 5). The SNVs within amplified regions of chromosome 21, however, show a peak around 0.33, while no peak at 0.67 is visible (Figure 5). These results show that CNAs on chromosome 21 were present before the activation of the mutational process behind signature SBS7a. While this could point to a role for chromosome 21 amplifications in the formation of SBS7a, it might also be explained by a process which causes both structural rearrangements of chromosome 21 and the formation of signature SBS7a.

DBS1 and transcriptional strand bias confirmed in ALL

Apart from C>T mutations, UV light can inflict other types of damage. In melanomas, signatures DBS1 and ID13 have been linked to UV damage and often co-occur with SBS7a. WGS results of 14 ALL patients with SBS7a show a high number of CC>TT mutations, resulting in a very high cosine similarity to signature DBS1 (Figure 6a, Supplementary Figure 6). We also see a high similarity to DBS11, which is similar to DBS1 and probably caused by APOBEC activity[12]. Since we do not find SBS2 and SBS13 in these samples, APOBEC is unlikely to play a role in these tumors. These results suggest a mutagenesis similar to damage by UV light in BCP-ALL, through the formation of pyrimidine dimers.

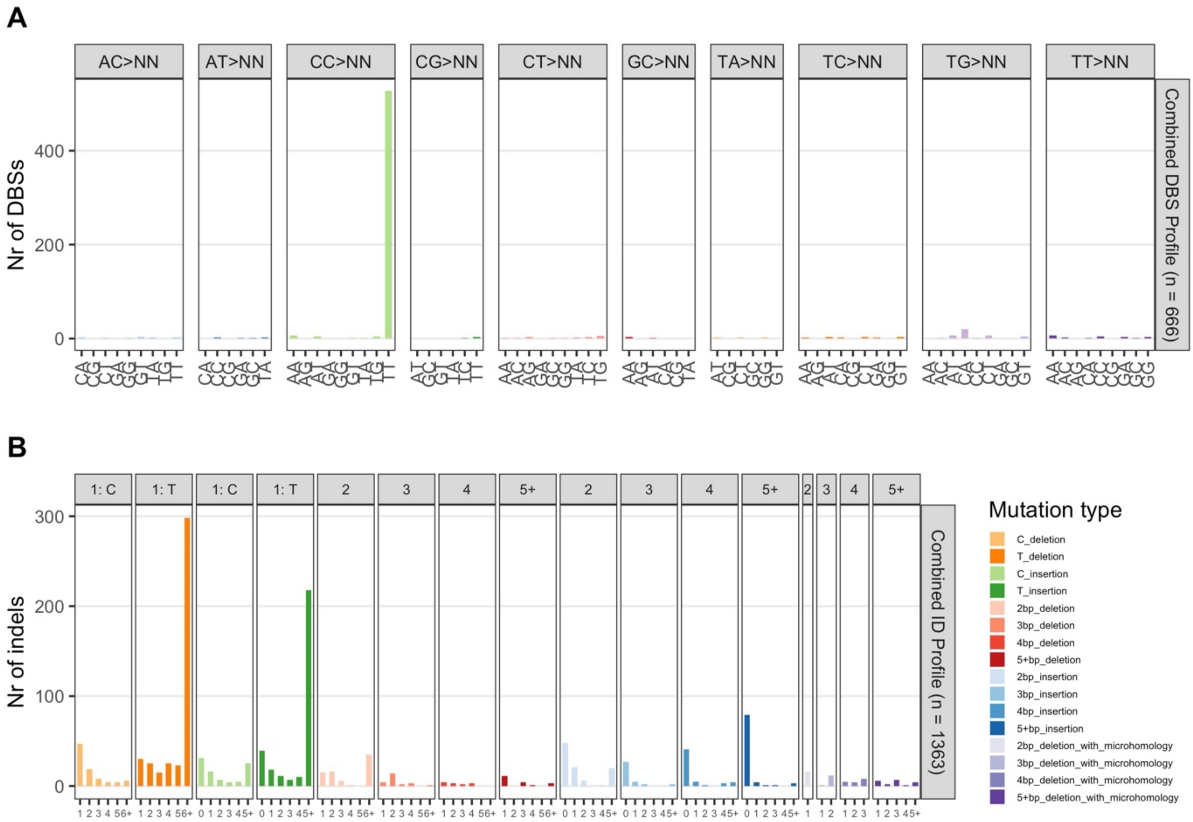


***Figure 6: 14 patients with ALL show the presence of signatures SBS7a and DBS1 but not ID13. a** The combined DBS profile of 14 SBS7a-positive ALL cases. The combined profile shows a high similarity to signature DBS1 (cosine similarity = 0.998), which has also been linked to damage by UV light. **b** The combined ID profile of 14 ALL patients with SBS7a. This combined profile does not show any similarity to UV-associated signature ID13 (cosine similarity = 0.137).*

Additionally, we have analyzed the presence of ID signatures. Signatures ID1 and ID2, which are caused by slippage during DNA replication and are commonly found in cancer[12], were identified in our samples (Figure 6b, Supplementary Figure 7). The presence of ID13 could not be confirmed in our samples, and no peak of 1 base pair thymine deletions with a homopolymer length of 2 was visible. This is a clear difference between the presentation of SBS7a in BCP-ALL and melanomas, suggesting the presence of a differences in the mutational process in BCP-ALL.

Furthermore, we have determined the presence of transcriptional strand bias of SBS7a in BCP-ALL. Since pyrimidine dimers are repaired by transcription-coupled NER, the resulting mutations are typically enriched on the untranscribed strand. Our results show that this is also the case in pediatric BCP-ALL, with 8 patients showing a significant bias towards the untranscribed strand (Figure 7, Supplementary Figure 8). This suggests that NER plays a role in BCP-ALL as well, which would be the case if SBS7a in BCP-ALL is indeed caused by the formation of pyrimidine dimers.

Apart from transcription, DNA repair mechanisms can be linked to other cellular processes such as replication. DNA synthesis tends to be more reliable during early replication, while error-prone translesion synthesis is more common during late replication[42]. This can lead to the introduction of more mutations in late replicating regions, which is also the case for SBS7a in melanomas. In BCP-ALL, however, no replication bias was found (Figure 7, Supplementary Figure 9). A difference of the underlying mutational process between melanomas and BCP-ALL would provide an explanation for the absence of replication bias. Alternatively, repair mechanisms might have different activities in BCP-ALL compared to melanomas. Lastly, we might not be able to fully separate SBS7a and SBS7b due to small sample sizes. Since SBS7b has a bias toward early replicating regions[43], the combination of SBS7a and SBS7b would not necessarily show replication bias. Therefore, the absence of replication bias does not necessarily exclude a role for UV in the mutagenesis of BCP-ALL.

SBS7a is not caused by a continuous process

To identify the underlying cause of SBS7a in ALL, we have determined the timing of mutational processes. 5 out of the 14 patients in our cohort have developed relapses, so for these patients WGS could be performed at multiple timepoints. Using the data from multiple timepoints, we could cluster the variants based on their VAF throughout time and construct the mutational profiles of separate clusters[3]. We combined the separate clusters with a WGS dataset of 214 ALL patients and were able to extract 9 signatures, including an SBS7a-like signature (Supplementary Figure 10). Out of 5 patients for whom multiple timepoints were available, 3 only showed SBS7a in clusters which were already present at the timepoint of the initial diagnosis (Figure 8A, Supplementary Figure 11). For the other 2 patients, however, new mutations showing the SBS7a signature arose at the first relapse (Figure 8B, Supplementary Figure 12). These mutations where therefore either gained after the initial diagnosis, or they were already present at the timepoint of initial diagnosis at a VAF below the detection limit. Targeted deep sequencing results could not confirm the presence of rising SBS7a mutations in the initial diagnosis of patient P0608 (Supplementary Figure 13). No new SBS7a mutations were identified after the first relapse in any of the patients. In conclusion, although new SBS7a mutations can arise before the first relapse, the underlying mechanism does not seem to be a continuous process.

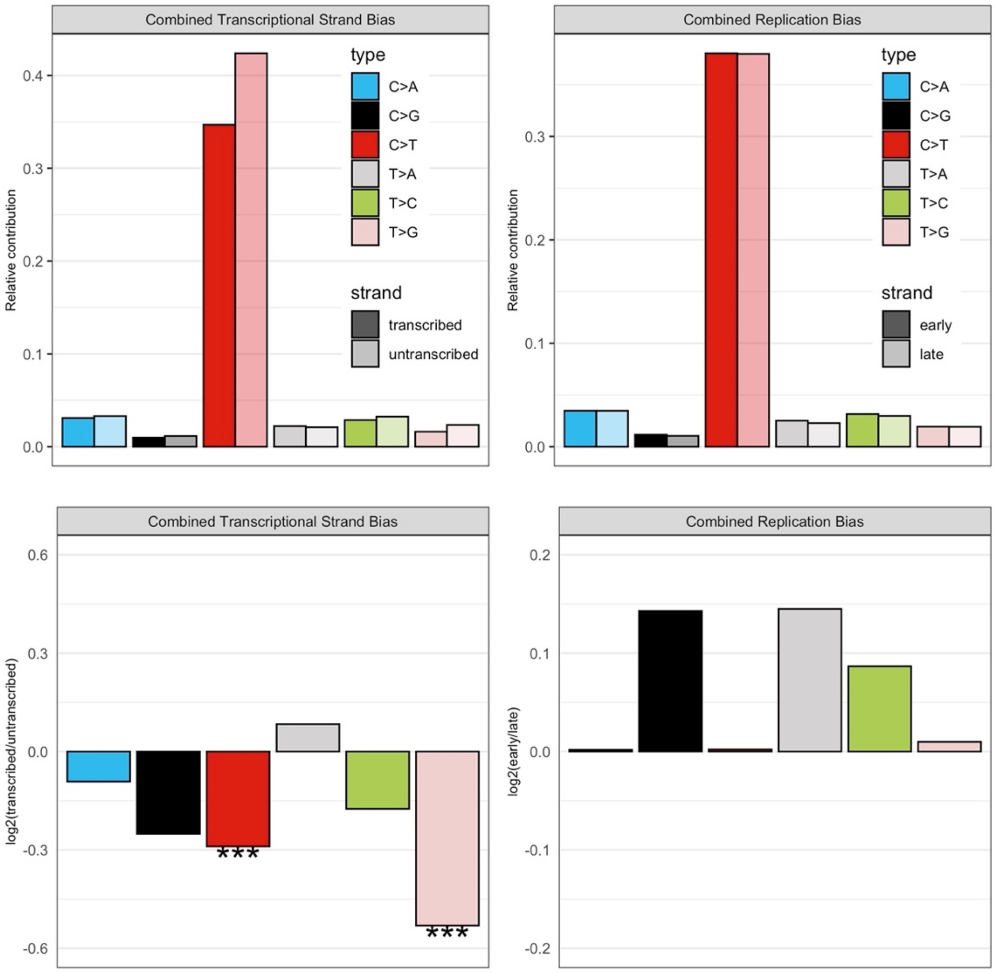


Figure 7: SBS7a in ALL shows transcriptional strand bias but no replication bias. *Left panels show the combined transcriptional strand bias of 14 BCP-ALL patients with SBS7a (top) and the ratio of the number of variants on the transcribed strand compared to variants on the untranscribed strand (bottom). Most patients show a significant bias of C>T mutations towards the untranscribed strand. Right panels show the combined replication bias of 14 BCP-ALL patients with SBS7a (top) and the ratio of the number of variants in early replicating regions compared to variants in late replicating regions (bottom). C>T mutations show no replication bias in BCP-ALL. *: FDR<0.05, **: FDR<0.01, ***: FDR<0.005*

Driver mutations caused by the SBS7a-associated mutational mechanism

SBS7a was initially identified in patients with relapses and is overrepresented in the iAMP21 subtype, which is correlated with a poor prognosis. To investigate whether SBS7a also directly influences disease outcome, we have developed a method to calculate the probability of single mutations being caused by the underlying process of SBS7a. With this method, variants with a mutation type which has a high contribution to SBS7a in samples with a high contribution of SBS7a will be assigned a high probability of being caused by SBS7a. Two mutations with a predicted effect on disease progression were found. Patient P0608 has a C>T missense mutation in the NR3C1 gene which has very likely been caused by SBS7a (probability of 0.88, Supplementary Table 3). NR3C1 is a corticosteroid receptor and therefore plays a role in the glucocorticoid response. Mutations in this gene can alter the prednisolone sensitivity of the tumor[10]. In patient P0624, a mutation caused by SBS7a was found in the CREBBP gene (probability of 0.95, Supplementary Table 3). This gene encodes for an epigenetic regulator which plays a role in the glucocorticoid response. The inactivity in this gene can thus also lead to reduced prednisolone sensitivity[10]. In conclusion, SBS7a has the potential to alter genes which play a role in drug sensitivity, and might therefore influence disease progression.

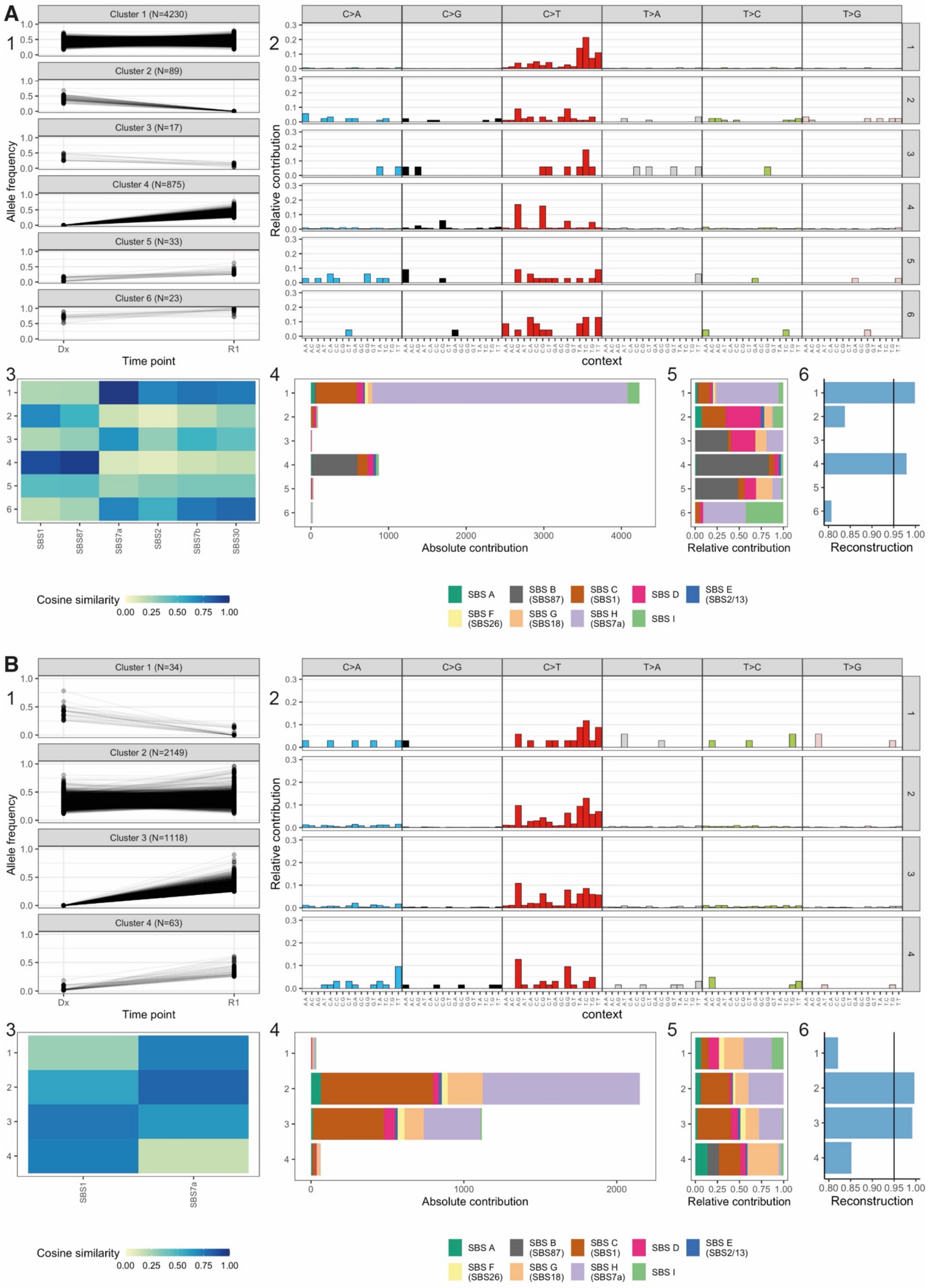


Figure 8: While patient P0610 has not gained signature SBS7a mutations after initial diagnosis, patient P0557 shows a high contribution of signature SBS7a in the rising cluster. a *The variants of patient P0610 can be grouped into 5 clusters based on the development of VAF over time (a1). While the founding clone (cluster 1) is similar to SBS7a (a2,3) and shows a high contribution of signature SBS H (a4,5), which is very similar to signature SBS7a, the rising clone (cluster 4) seems to consist mainly of SBS87 mutations(a2,3,4,5). Both clones show a high cosine similarity between the original and recontructed profile (a6), meaning signature extraction was reliable for these clusters.* ***b*** *The variants of patient P0557 show 4 clusters (b1). Both the founding clone (cluster 2) and the rising clone (cluster 3) show a profile which is similar to SBS1 and SBS7a and a combination of de novo extracted signature SBS C, which is similar to SBS1, and SBS H (SBS7a-like) (b2,3,4,5). Both clusters show a high cosine similarity between the original and recontructed profile (b6). Dx: initial diagnosis, R: relapse*

Discussion

Although signature SBS7a is commonly seen in skin cancer, we have identified this signature in 14 Dutch BCP-ALL patients across multiple cohorts. In our unselected WGS cohort which was generated by routine diagnostics, we have identified signature SBS7a in 8 out of 111 patients (7.2%), and literature shows a prevalence of around 10%[1]. In patients with CNAs of chromosome 21, however, this percentage is larger. Especially patients with the iAMP21 subtype often show a high contribution of SBS7a[1]. In this study, we have analyzed the characteristics of the group of BCP-ALL patients with signature SBS7a. These characteristics were then compared to melanomas, in which SBS7a is caused by UV light, to investigate a possible role for UV light in the mutagenesis of ALL.

Although the underlying mutational process of SBS7a in ALL is still unknown, our results suggest a similar mutagenesis to damage by UV light through the formation of pyrimidine dimers. This is shown by the presence of both signature DBS1 and transcriptional strand bias in 14 BCP-ALL patients. While many other mutational processes lead to C>T mutations, these are not expected to result in DBS1. Some mutagens, such as N-acetoxy-2-acetylaminofluorene, 4-nitroquinoline-1-oxide, cisplatin, and psoralen, can form adducts with DNA, triggering a response by NER[44]. Although this would result in transcriptional strand bias, these mutagens are not expected to form pyrimidine dimers and leave SBS7a and DBS1. Cisplatin, for example, has its own distinct signatures: SBS31, SBS35 and DBS5[16,45]. Mutagenesis through the formation of pyrimidine dimers does therefore not directly lead to the identification of the underlying mutational process.

Based on the signatures and transcriptional strand bias, we cannot fully exclude a role for UV light in BCP-ALL. Although we did not find signature ID13, the association between this signature and damage by UV light has not been validated *in vitro*. The absence of ID13 can therefore not fully eliminate the possibility of UV light playing a role in BCP-ALL. We could not confirm the presence of replication bias either, but this can also be explained by variations in timing of repair mechanisms between different cell types. Alternatively, we might have extracted a mixture of SBS7a and SBS7b, which does not have a bias towards late replicating regions[43,46]. Since SBS7a and SBS7b are very similar and are both caused by UV damage, separating these signatures can be challenging. However, we only extracted one UV-like signature, which was very similar to SBS7a (cosine similarity with SBS7a = 0.99; cosine similarity with SBS7b = 0.78). In conclusion, the absence of ID13 and replication bias does not necessarily exclude a role for UV light in the mutagenesis of BCP-ALL.

When taking the timing into account, damage by UV light does not provide the most likely explanation. For cells to accumulate UV damage, they would have to travel from the bone marrow to the skin. After gaining a great number of mutations, in some cases around 4000, the preleukemic cell should return to the bone marrow. Additionally, the accumulation of a new wave of SBS7a mutations was seen in 2 patients. If SBS7a was indeed caused by UV light, this would either mean that preleukemic cells remain in the skin where they accumulate mutations until they return to the bone marrow to form a relapse, or that leukemic cells in the bone marrow travel to the skin during treatment, accumulate mutations and return to the bone marrow. Both scenarios are highly unlikely. Additionally, our results show the presence of CNAs of chromosome 21 prior to the accumulation of SBS7a mutations. Therefore, if B-cell precursors did travel to the skin they would already have to possess preleukemic features at that stage. While both findings cannot fully eliminate the possibility that SBS7a in BCP-ALL is caused by UV light, this explanation does lead to very unlikely scenarios based on the mutational timing.

Based on timing, we can also exclude most intrinsic mutational processes. If SBS7a in BCP-ALL was caused by, for example, a defective repair mechanism, the process would be expected to have a continuous activity. In 4 out of 5 cases, however, we see clusters of rising mutations which cannot be attributed to SBS7a. The underlying process must thus have a fluctuating activity. Therefore, an external process seems more likely to be causative of SBS7a in BCP-ALL.

Although external processes provide a more probable explanation for SBS7a in BCP-ALL, a clear overrepresentation of CNAs of chromosome 21 and the iAMP21 subtype were present in our cohort. Since these amplifications were present before the accumulation of SBS7a mutations, the CNAs might play a role in the formation of SBS7a. Possibly, SBS7a is caused by an external mutagen in cells which have been sensitized by amplifications of chromosome 21.

To further study internal processes, we will have to analyze pathogenic variants in our cohort more systematically. So far, no commonly mutated genes were found, but germline mutations have not been studied yet. Combining different types of somatic and germline variants, such as SNVs, indels, CNAs and structural variants, might give new insights into the possibility of an intrinsic process. Additionally, differential expression analysis based on RNAseq data could be useful for identifying up- or downregulated pathways. Further analyzing internal processes could explain the overrepresentation of CNAs of chromosome 21 in our cohort.

More insight into the possibility of UV light being the causative process could be gained from further studying comparisons between our cohort and melanoma samples. While we could compare our data to information about UV damage found in literature, we were not able to obtain a cohort of melanoma patients. This comparison could give more insight into the indel signature, the ratio of C>T and CC>TT mutations and the topography of the UV-related signatures[46].

Since the development of relapses were a selection criterium for some of the cohorts from which we got our data, our final cohort is not informative in terms of clinical outcome. We could, however, show that at least two SBS7a mutations may have led to therapy resistance. An unselected BCP-ALL cohort would be very useful for determining both the proportion of patients with SBS7a for each subtype and the clinical outcome of patients with SBS7a.

In conclusion, while our results could not fully exclude a role for UV light in BCP-ALL, another external process leading to pyrimidine dimers in combination with an internal sensitivity seems most probable based on our data. Further research will have to show which mutational process is responsible for SBS7a in BCP-ALL.

References

1. Studd, J. B. *et al.* Cancer drivers and clonal dynamics in acute lymphoblastic leukaemia subtypes. *Blood Cancer J* **11**, 177 (2021).

2. Waanders, E. *et al.* Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. *Blood Cancer Discov* **1**, 96–111 (2020).

3. Antić, Ž. *et al.* Unravelling the Sequential Interplay of Mutational Mechanisms during Clonal Evolution in Relapsed Pediatric Acute Lymphoblastic Leukemia. *Genes (Basel)* **12**, 214 (2021).

4. Reedijk, A. M. J. *et al.* Progress against childhood and adolescent acute lymphoblastic leukaemia in the Netherlands, 1990-2015. *Leukemia* **35**, 1001–1011 (2021).

5. Hunger, S. P. & Mullighan, C. G. Acute Lymphoblastic Leukemia in Children. *N Engl J Med* **373**, 1541–1552 (2015).

6. Iacobucci, I., Kimura, S. & Mullighan, C. G. Biologic and Therapeutic Implications of Genomic Alterations in Acute Lymphoblastic Leukemia. *J Clin Med* **10**, 3792 (2021).

7. Harrison, C. J. Blood Spotlight on iAMP21 acute lymphoblastic leukemia (ALL), a high-risk pediatric disease. *Blood* **125**, 1383–1386 (2015).

8. Li, Y. *et al.* Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* **508**, 98–102 (2014).

9. Fioretos, T. Why B(-)other? About the gap of unknowns in ALL. *Blood* **139**, 3455–3457 (2022).

10. Li, B. *et al.* Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. *Blood* **135**, 41–55 (2020).

11. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).

12. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

13. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

14. Degasperi, A. *et al.* Substitution mutational signatures in whole-genome–sequenced cancers in the UK population. *Science* **376**, abl9283 (2022).

15. Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180 (2017).

16. Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**, 821-836.e16 (2019).

17. Nik-Zainal, S. *et al.* The genome as a record of environmental exposure. *Mutagenesis* **30**, 763–770 (2015).

18. Mullenders, L. H. F. Solar UV damage to cellular DNA: from mechanisms to biological effects. *Photochem Photobiol Sci* **17**, 1842–1852 (2018).

19. Spivak, G. Nucleotide excision repair in humans. *DNA Repair (Amst)* **36**, 13–18 (2015).

20. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).

21. Brash, D. E. *et al.* A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proc Natl Acad Sci U S A* **88**, 10124–10128 (1991).

22. Chen, J.-M., Férec, C. & Cooper, D. N. Patterns and Mutational Signatures of Tandem Base Substitutions Causing Human Inherited Disease. *Human Mutation* **34**, 1119–1130 (2013).

23. Auwera, G. A. V. der & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. (O'Reilly Media, Inc., 2020).

24. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

25. Picard Tools - By Broad Institute. http://broadinstitute.github.io/picard/.

26. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).

27. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

28. Boomsma, D. I. *et al.* The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* **22**, 221–227 (2014).

29. *R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. (2021).

30. *RStudio Team. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA*. (2021).

31. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

32. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019).

33. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

34. Manders, F. *et al.* MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genomics* **23**, 134 (2022).

35. *Bioconductor Core Team and Bioconductor Package Maintainer. TxDb.Hsapiens.UCSC.hg38.knownGene: Annotation package for TxDb object(s). R package version 3.14.0.* (2021).

36. Hsu, F. *et al.* The UCSC Known Genes. *Bioinformatics* **22**, 1036–1046 (2006).

37. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

38. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).

39. Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* **18**, 696–705 (2018).

40. McLeod, C. *et al.* St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. *Cancer Discov* **11**, 1082–1099 (2021).

41. Ueno, H. *et al.* Landscape of driver mutations and their clinical impacts in pediatric B-cell precursor acute lymphoblastic leukemia. *Blood Adv* **4**, 5165–5173 (2020).

42. Tomkova, M., Tomek, J., Kriaucionis, S. & Schuster-Böckler, B. Mutational signature distribution varies with DNA replication timing and strand asymmetry. *Genome Biol* **19**, 129 (2018).

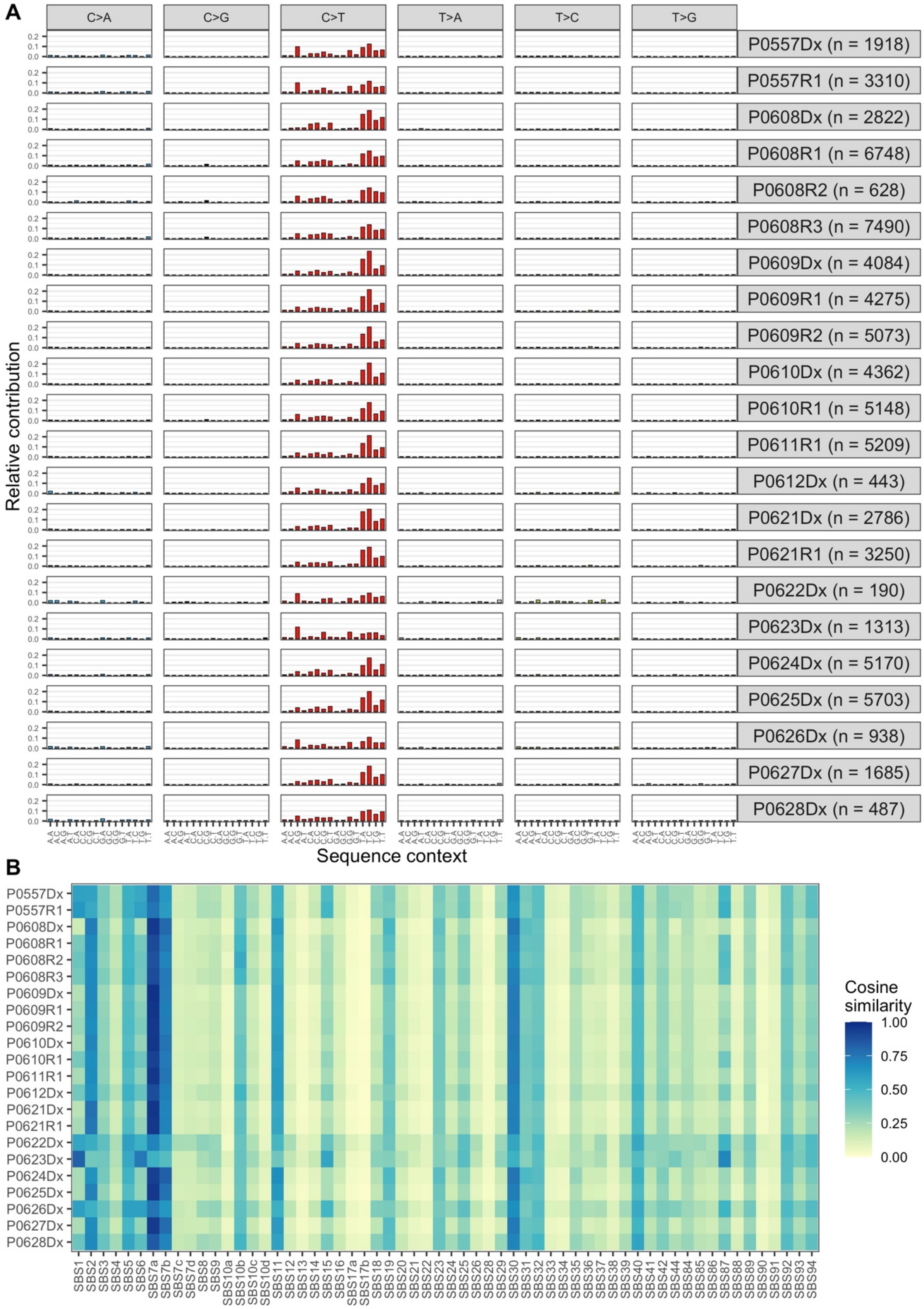
43. Yaacov, A. *et al.* Cancer Mutational Processes Vary in Their Association with Replication Timing and Chromatin Accessibility. *Cancer Res* **81**, 6106–6116 (2021).

44. Lacks, S. A. Repair Mechanisms. in *Brenner's Encyclopedia of Genetics (Second Edition)* (eds. Maloy, S. & Hughes, K.) 134–141 (Academic Press, 2001). doi:10.1016/B978-0-12-374984-0.01295-X.

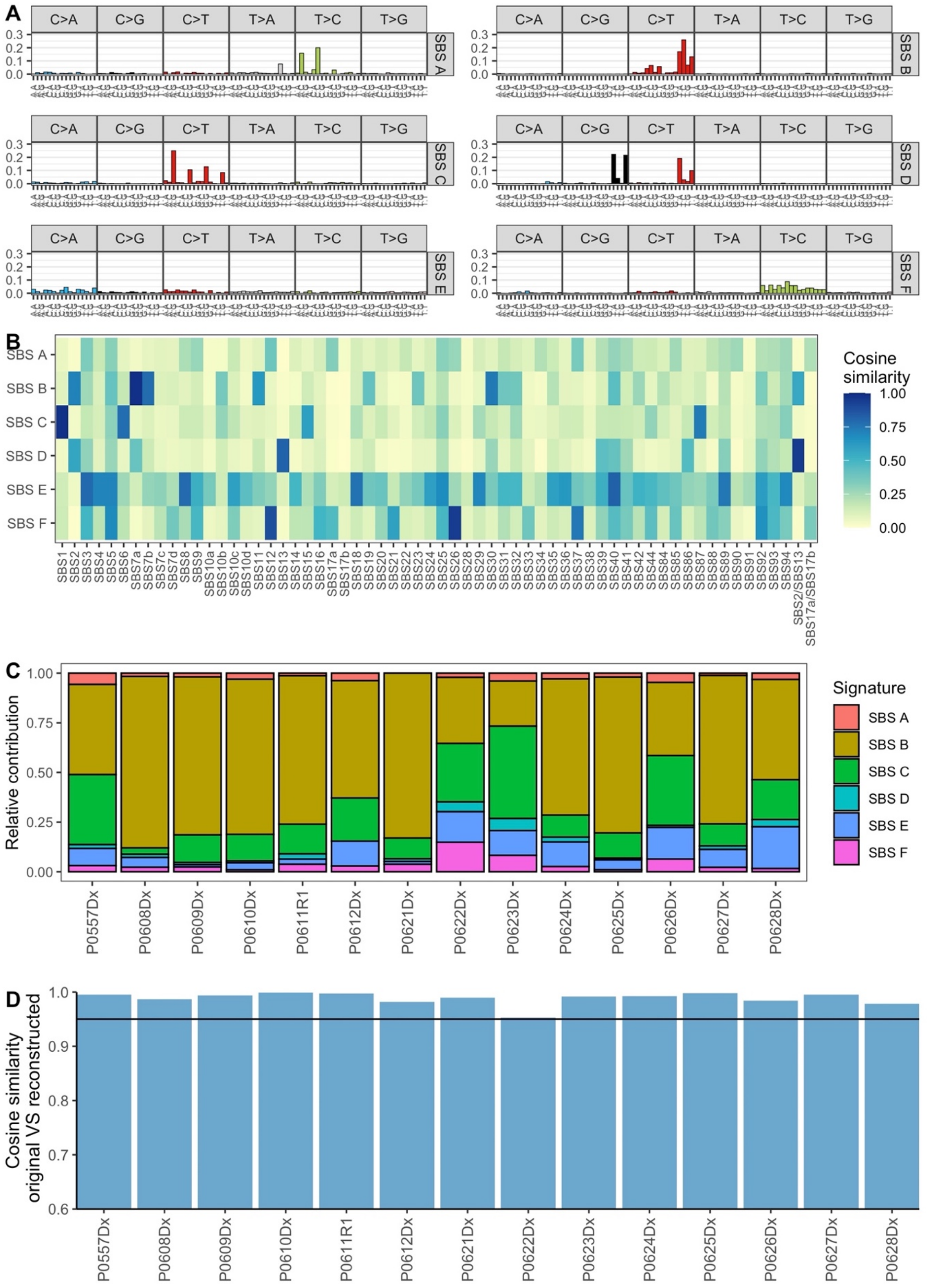
45. Boot, A. *et al.* In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res* **28**, 654–665 (2018).

46. Otlu, B. *et al.* Topography of mutational signatures in human cancer. *bioRxiv* 2022.05.29.493921 (2022) doi:10.1101/2022.05.29.493921.

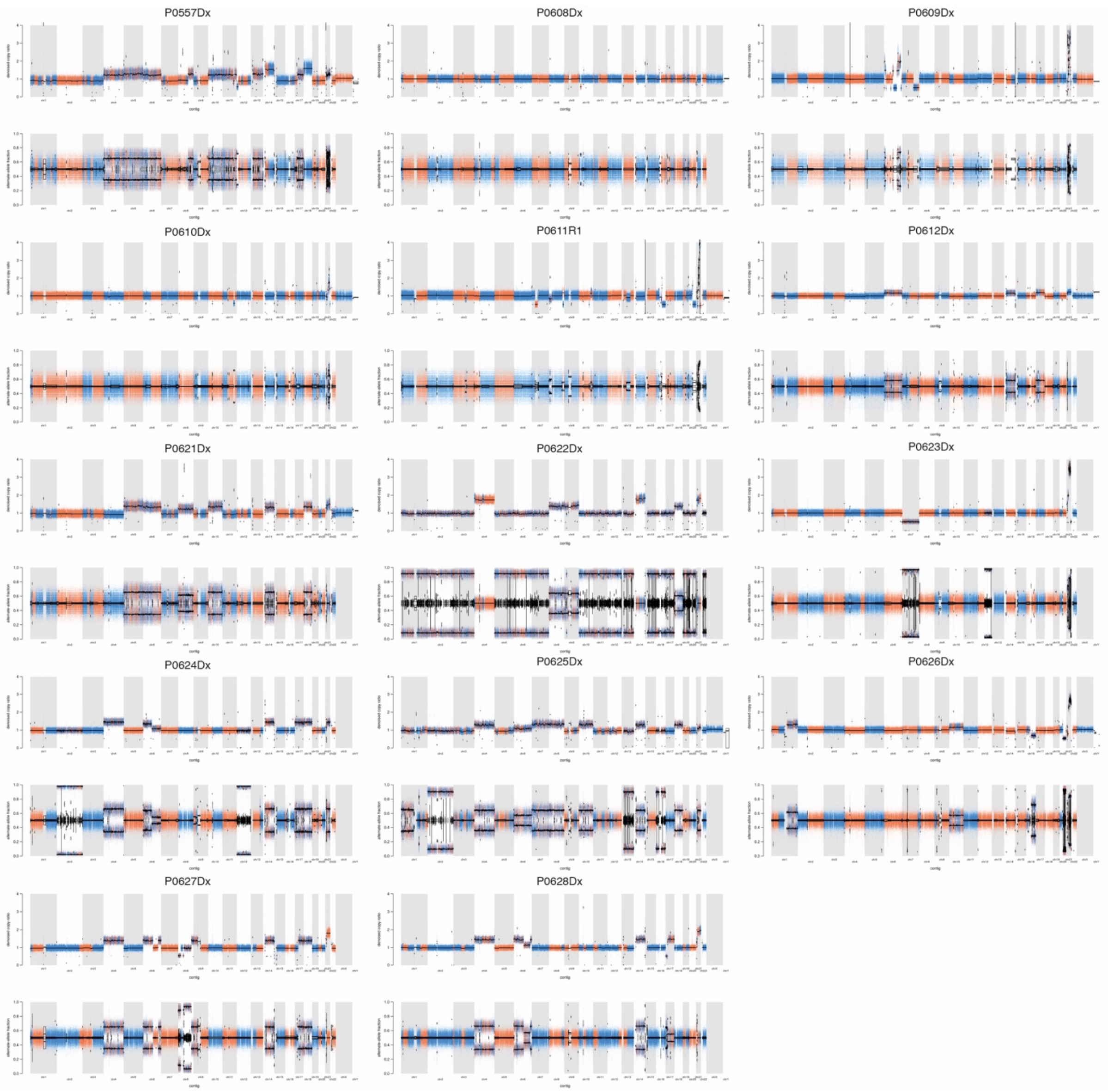
Supplementary figures



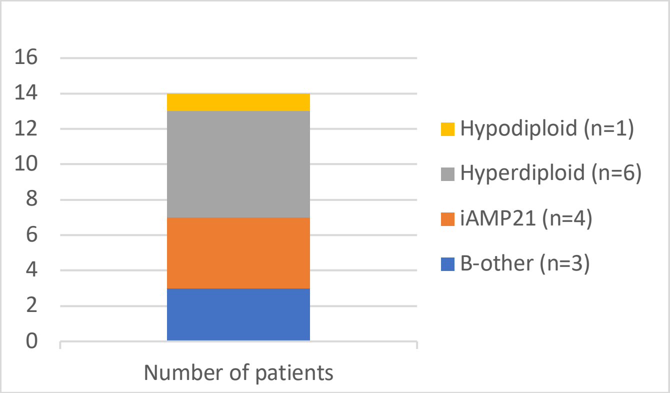
***Supplementary Figure 1: 14 patients with SBS7a-positive ALL. a** Mutational profiles of 14 patients with BCP-ALL. **b** Heatmap of cosine similarity of mutational profiles to COSMIC signatures. All mutational profiles show a high cosine similarity to signature SBS7a. Dx: initial diagnosis, R: relapse*



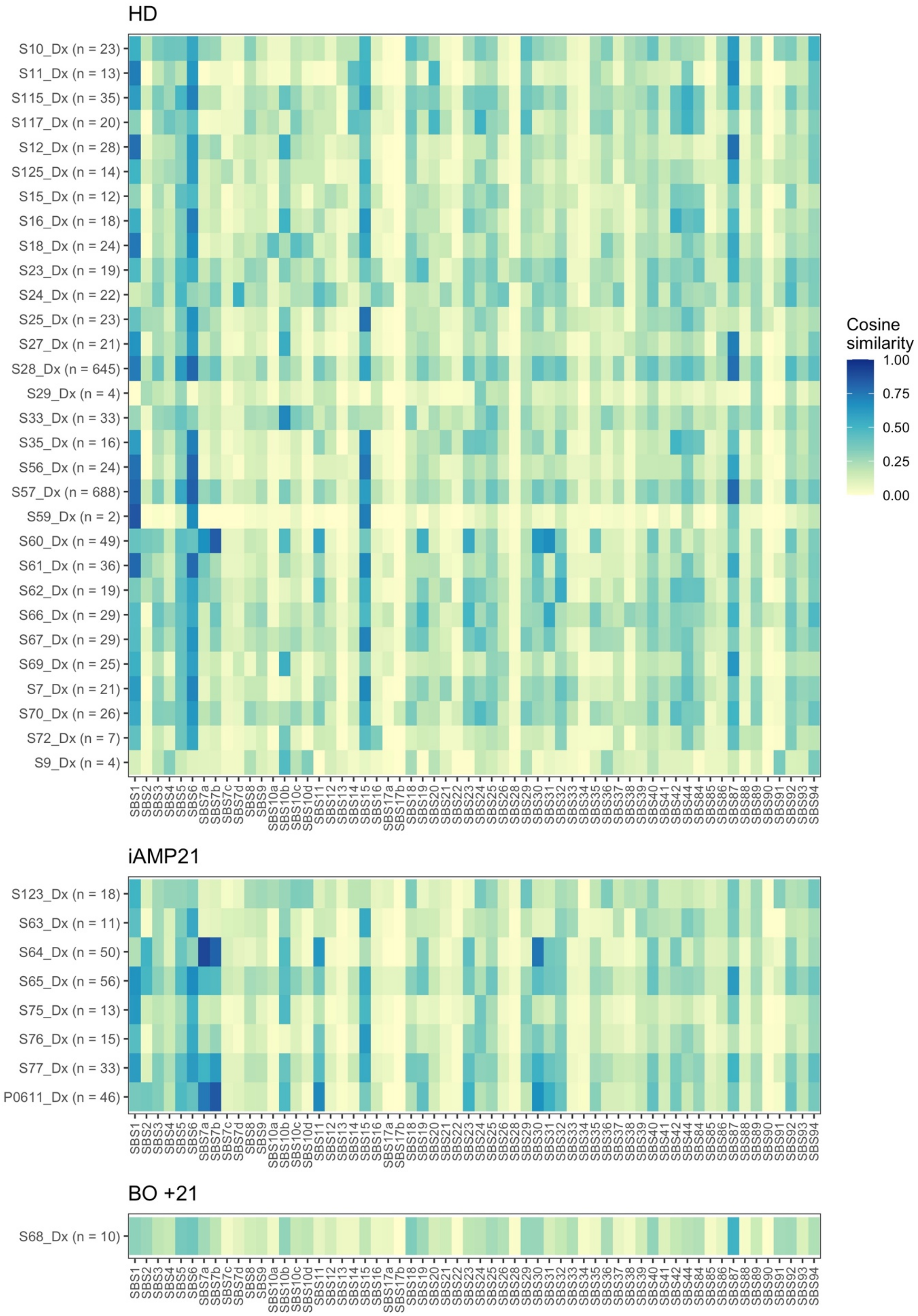
Supplementary Figure 2: 14 patients show a contribution of SBS7a. a *6 signatures were extracted from the combined data of the first timepoints of 14 BCP-ALL patients and 214 external patients.* ***b*** *Cosine similarity of extracted signatures with signatures from the COSMIC database. Extracted signature SBS B is very similar to COSMIC signature SBS7a.* ***c*** *Contributions of extracted signatures to 14 samples of patients with BCP-ALL. All patients have a contribution of signature B (SBS7a-like) of at least 0.2.* ***d*** *Cosine similarity of the original mutational profile with the reconstructed profile based on the extracted signatures. Dx: initial diagnosis, R: relapse*



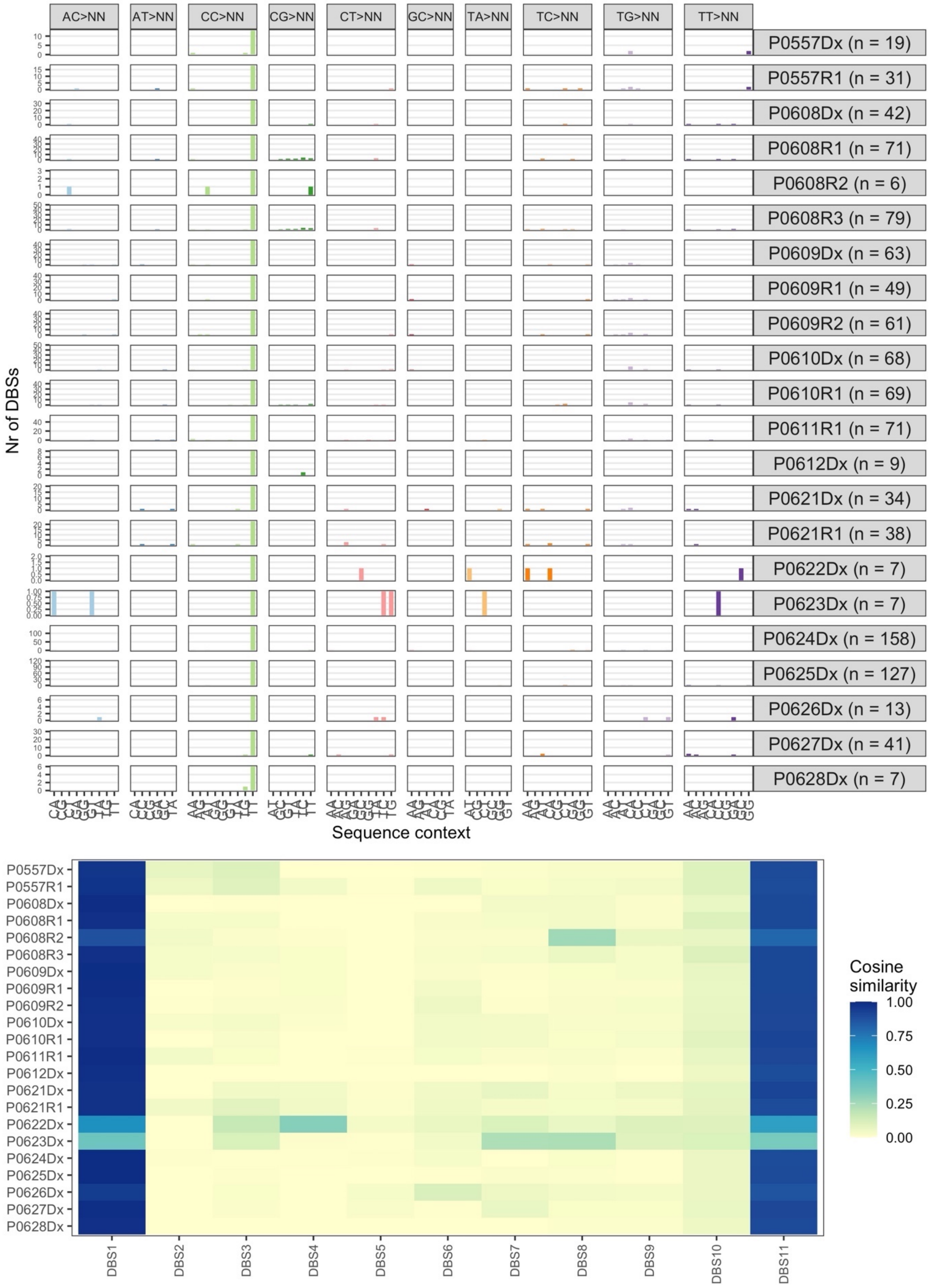
Supplementary Figure 3: CNA plots for all 14 patients with signature SBS7a. *Apart for patient P0608, all patients show CNAs of chromosome 21. Dx: initial diagnosis, R: relapse*



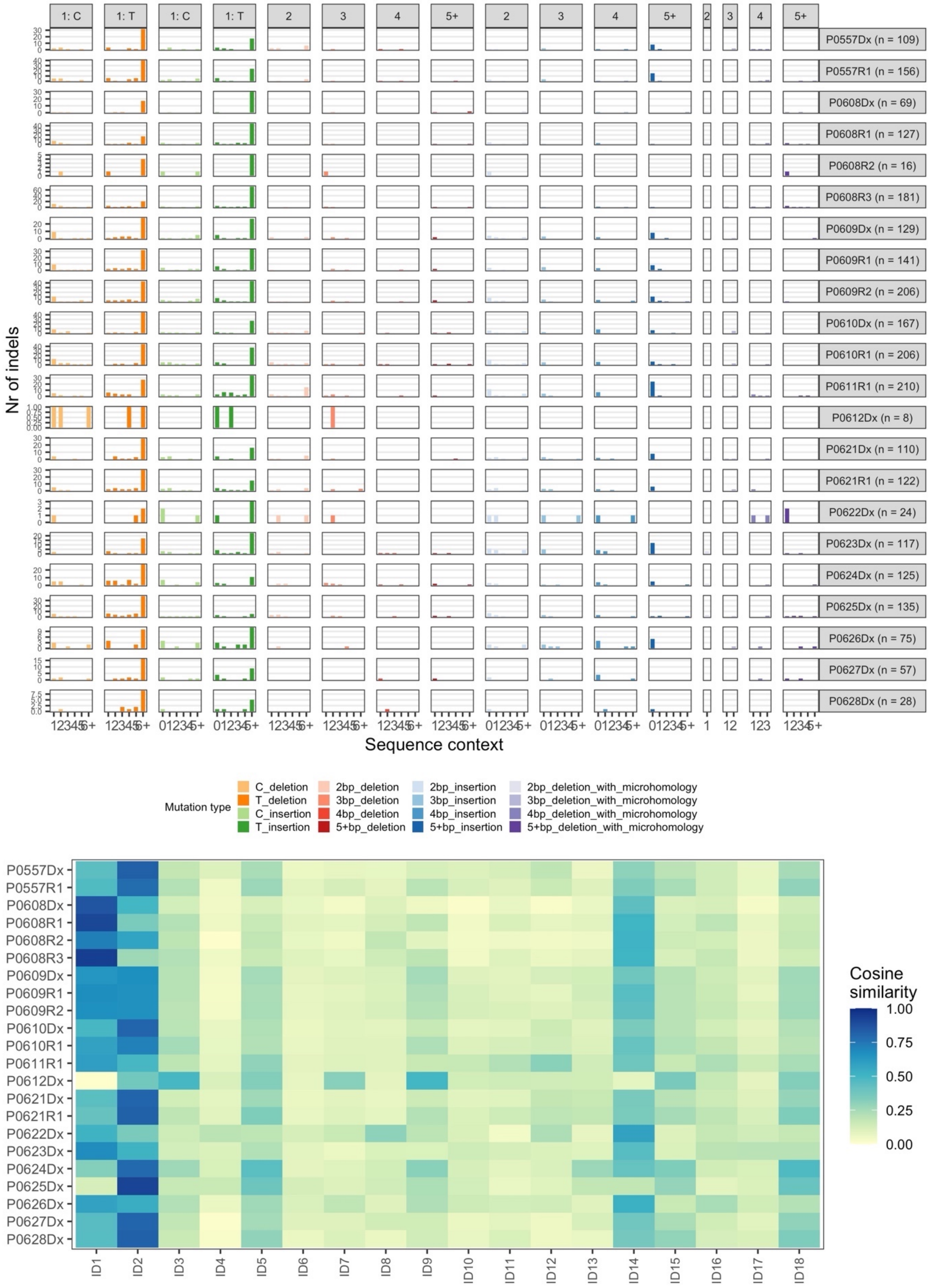
Supplementary Figure 4: Subtypes with amplifications of chromosome 21 are overrepresented in the group of ALL patients with signature SBS7a. *Bar graph of the distribution of subtypes within the cohort of BCP-ALL patients with signature SBS7a.*



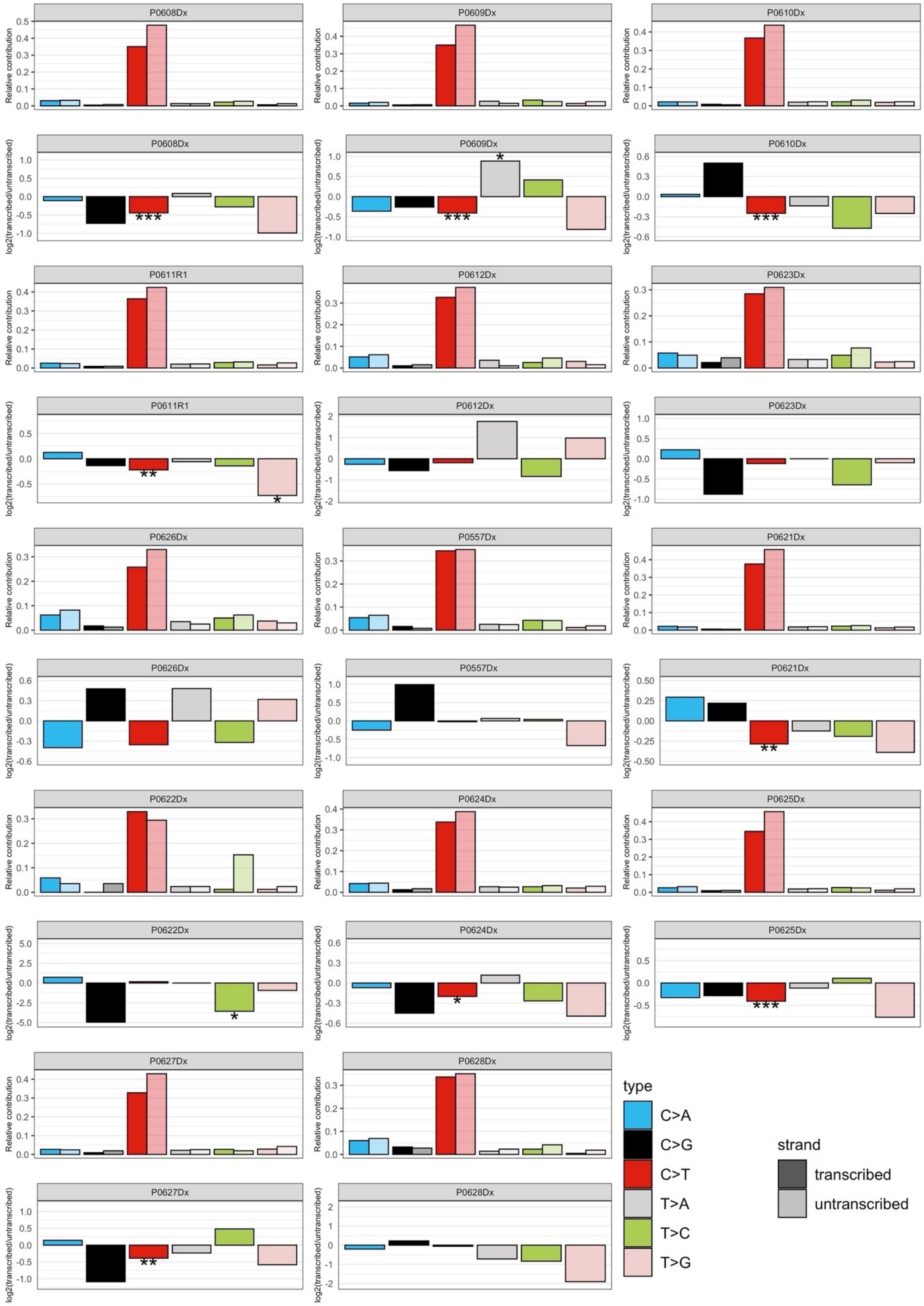
Supplementary Figure 5: Heatmap of cosine similarity between COSMIC signatures and mutational profiles of BCP-ALL patients with CNAS of chromosome 21. *Mutational profiles are based on WES data. HD: hyperdiploid, BO: B-other*



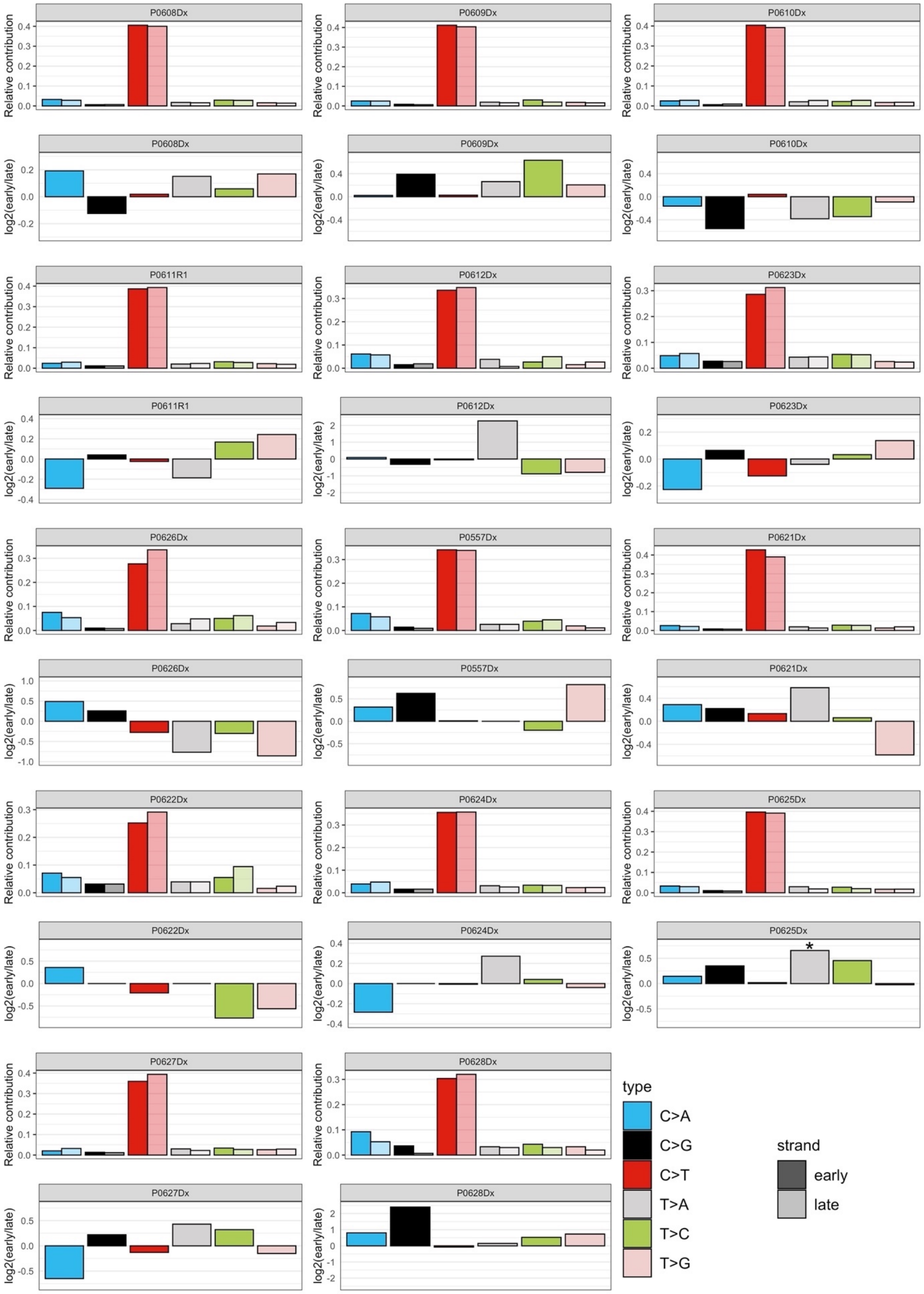
Supplementary Figure 6: DBS profiles of 14 patients with SBS7a show the presence of signature DBS1. *Top panel shows the DBS mutational profiles of BCP-ALL patients with SBS7a. Bottom panel shows the cosine similarity between the mutational profiles and COSMIC signatures. The mutational profiles of most patients show a high cosine similarity to DBS1. The other patients have a low number of DBSs. Most samples also show a high cosine similarity to DBS11, which is similar to DBS1 and probably caused by APOBEC activity. Dx: initial diagnosis, R: relapse*



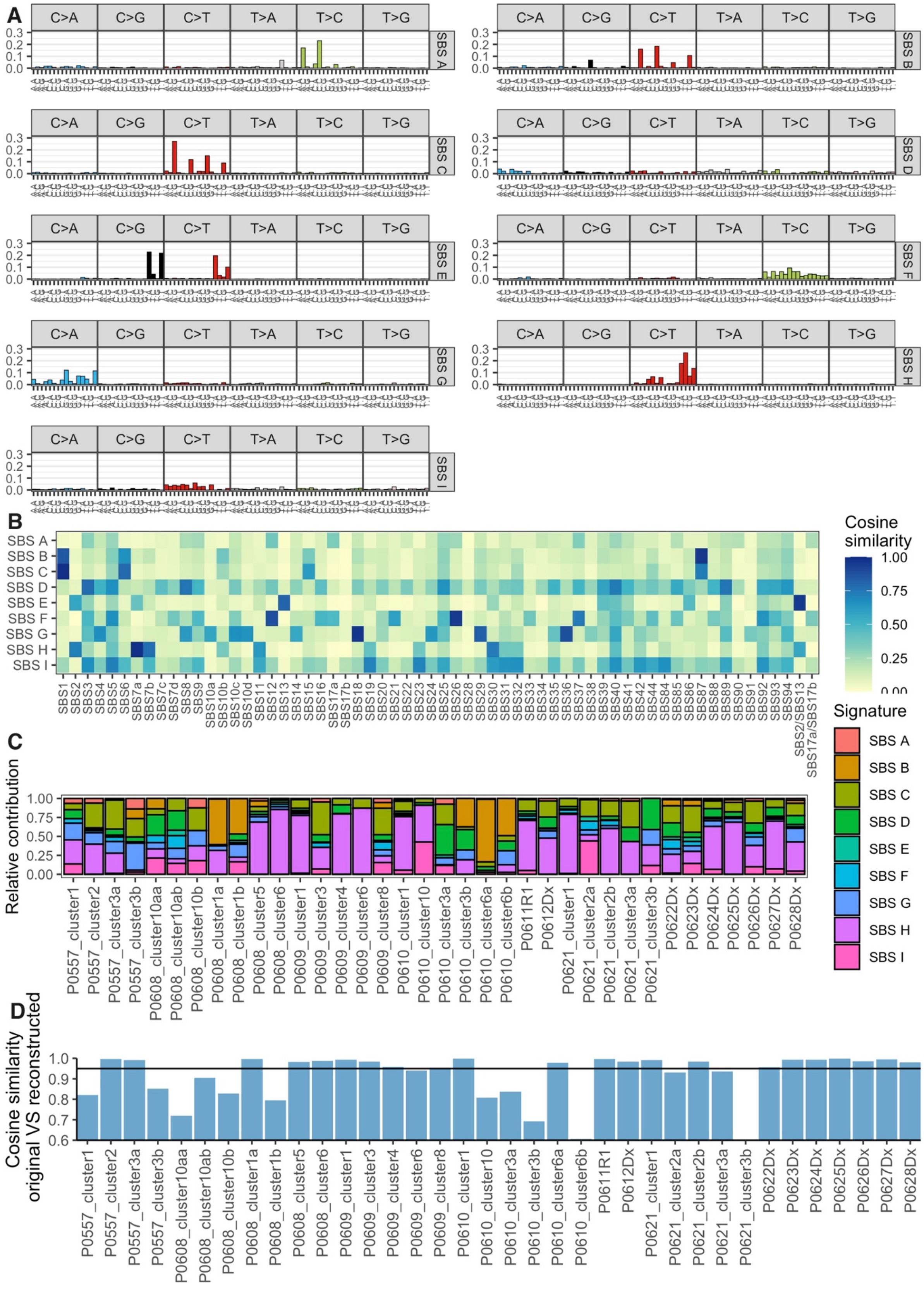
Supplementary Figure 7: Indel profiles for 14 BCP-ALL patients with SBS7a do not confirm the presence of signature ID13. *Top panel shows the ID mutational profiles of BCP-ALL patients with SBS7a. Bottom panel shows the cosine similarity between the mutational profiles and COSMIC signatures. The mutational profiles show no similarity to UV-related signature ID13. Dx: initial diagnosis, R: relapse*



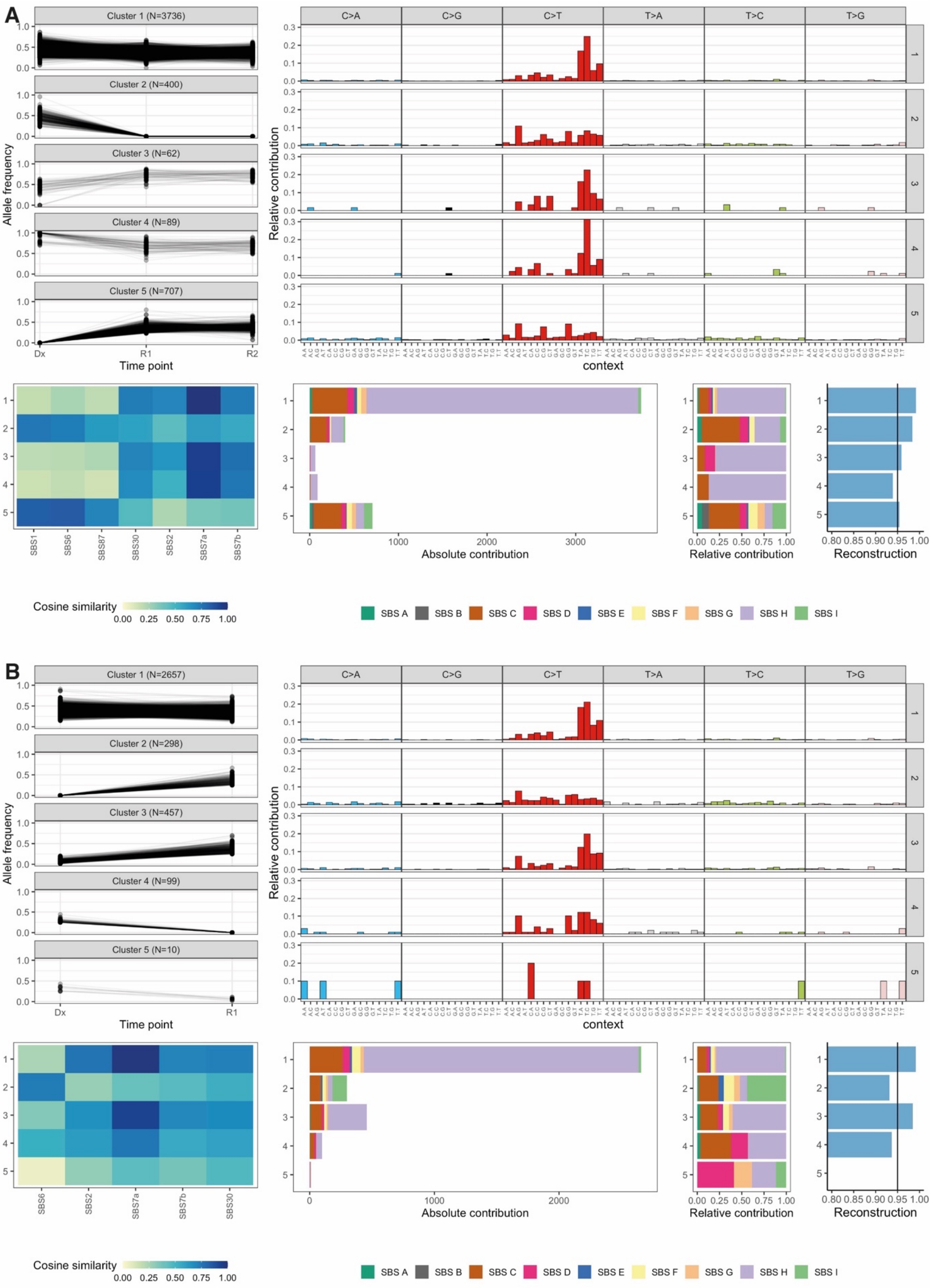
Supplementary Figure 8: C>T mutations in 8/14 patients show a significant bias towards the untranscribed strand. *For each patient, the relative contributions of both strands are shown for each mutation type (top panel). Additionally, the ratio of SNVs on the transcribed and untranscribed strand is shown for each patient (bottom panel). *: FDR<0.05, **: FDR<0.01, ***: FDR<0.005, Dx: initial diagnosis, R: relapse*



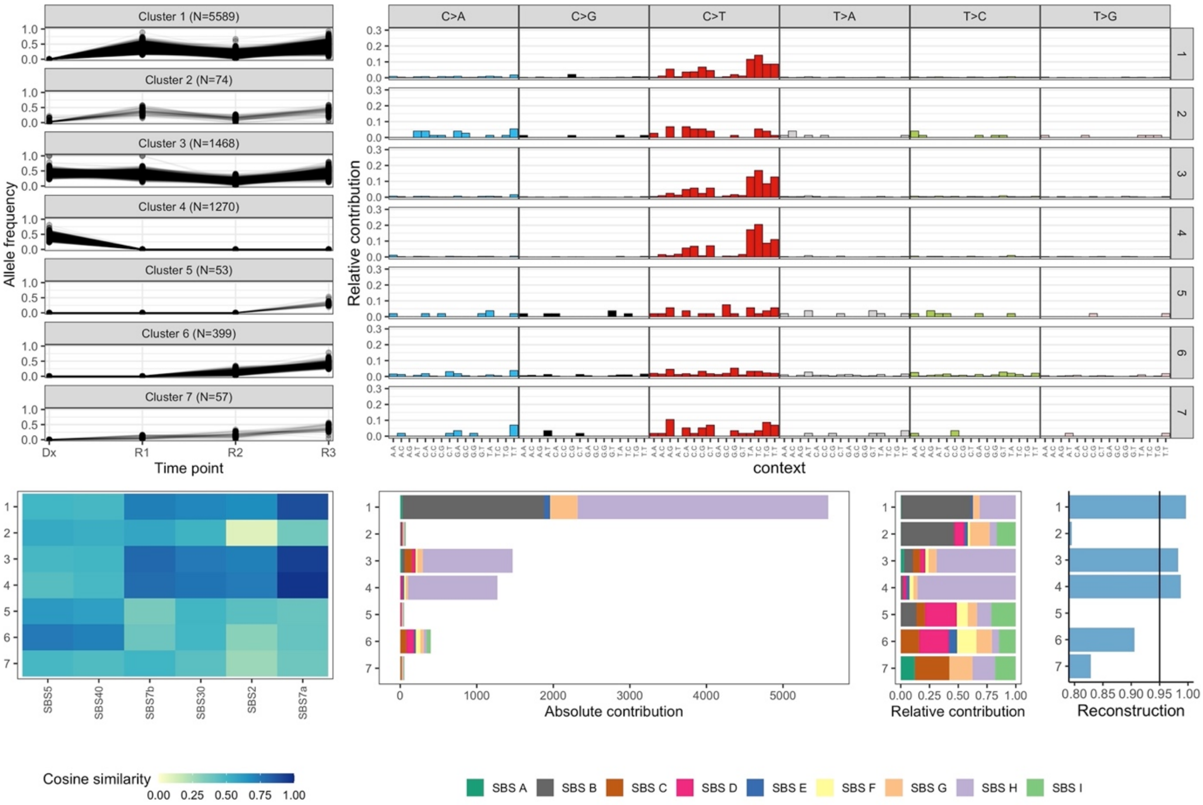
Supplementary Figure 9: No replication bias is seen in 14 BCP-ALL patients with signature SBS7a. *For each patient, the relative contributions of each mutation type are shown for both early and late replicating regions (top panel). Additionally, the ratio of SNVs on the early and late replicating regions is shown for each patient (bottom panel). *: FDR<0.05, **: FDR<0.01, ***: FDR<0.005, Dx: initial diagnosis, R: relapse*



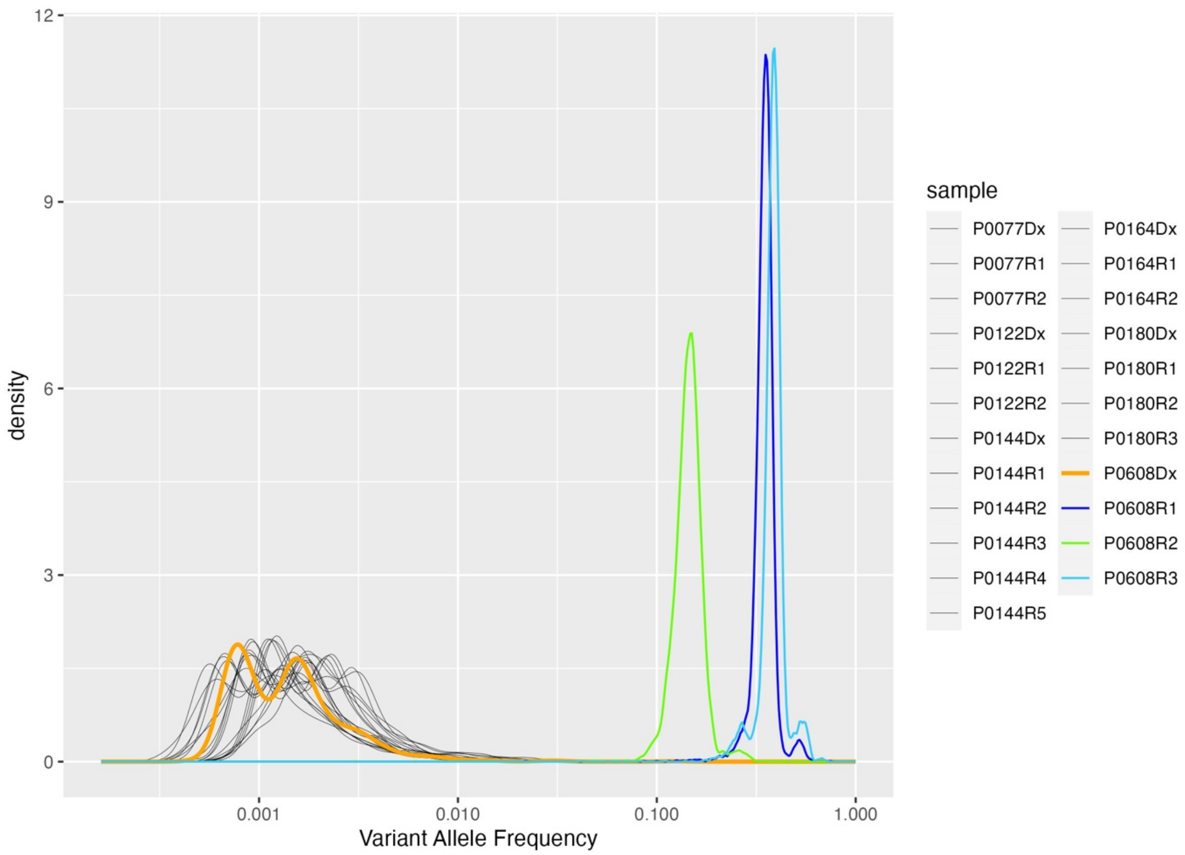
Supplementary Figure 10: 9 signatures were extracted from separate clusters of SNVs. ***a*** *De novo extraction of separate clusters combined with other SBS7a-positive patients and 214 external patients results in 9 signatures.* ***b*** *Cosine similarity of extracted signatures with signatures from the COSMIC database. While signatures A, D and I seem to mostly consist of noise, the other signatures show a great resemblance to known COSMIC signatures. Signature H shows resemblance to signature SBS7a.* ***c*** *Contributions of extracted signatures to separate clusters of SNVs. Many clusters show a large contribution of signature H.* ***d*** *Cosine similarity of the original mutational profile with the reconstructed profile based on the extracted signatures. Small clusters tend to have a lower cosine similarity between the original mutational profile and the reconstructed profile. Dx: initial diagnosis, R: relapse*



Supplementary Figure 11: Two patients show no new SBS7a mutations after the initial diagnosis. ***a*** *Patient P0609 has developed 2 relapses and shows 5 clusters. The rising cluster (cluster 5) contains mainly SBS1 mutations.* ***b*** *Patient P0621 has developed 1 relapse and shows 5 clusters. While cluster 3, which is subclonal at the first timepoint and clonal at the second timepoint, shows a high contribution of SBS7a mutations, the rising cluster (cluster 2) contains a combination of SBS1 and noise mutations. Dx: initial diagnosis, R: relapse*



Supplementary Figure 12: Patient P0609 acquires new SBS7a mutations between the initial diagnosis and the first relapse. *This patient has developed 3 relapses and shows 7 clusters. The rising cluster (cluster 1) contains a combination of SBS87 and SBS7a mutations. Dx: initial diagnosis, R: relapse*



Supplementary Figure 13: Deep sequencing results from patient P0608 could not confirm the presence of cluster 1 mutations at the timepoint of initial diagnosis. *The distribution of the VAF is shown for variants in cluster 1 of patient P608. Dx: initial diagnosis, R: relapse*

Supplementary tables

Supplementary Table 1: achieved sequencing depth of patient P0608

Sample		Median depth	Percentage ≥20X coverage	Estimated blast percentage
Control 1	CR1	30X	93%	
Control 2	CR2	46X	97%	
Diagnosis	Dx	43X	97%	±88%
Relapse 1	R1	36X	96%	±80%
Relapse 2	R2	42X	97%	±34%
Relapse 3	R3	37X	96%	±88%

Supplementary Table 2: List of known ALL driver genes

Gene Symbol	Source	Ensembl Gene ID	Gene Length
AFF4	Cosmic	ENSG00000072364	11281
FCGR2B	Cosmic	ENSG00000072694	6655
JAK1	Cosmic	ENSG00000162434	12620
NCKIPSD	Cosmic	ENSG00000213672	3431
CDK6	Cosmic	ENSG00000105810	12239
EPS15	Cosmic	ENSG00000085832	7378
FOXP1	Cosmic	ENSG00000114861	32800
HLF	Cosmic	ENSG00000108924	6776
MLLT11	Cosmic	ENSG00000213190	2483
MLLT3	Cosmic	ENSG00000171843	8081
ZNF384	Cosmic	ENSG00000126746	5181
ZNF521	Cosmic	ENSG00000198795	6225
CREBBP	Cosmic	ENSG00000005339	15479
JAK2	Cosmic	ENSG00000096968	7908
IKZF1	Cosmic	ENSG00000185811	10921
IL7R	Cosmic	ENSG00000168685	6020
AFF3	Cosmic	ENSG00000144218	15701
FLT3	Cosmic	ENSG00000122025	4159
KMT2A	Cosmic	ENSG00000118058	31469
NUP214	Cosmic	ENSG00000126883	24733
RUNX1	Cosmic	ENSG00000159216	15574
PML	Cosmic	ENSG00000140464	12672
BCL9	Cosmic	ENSG00000116128	6331
ELN	Cosmic	ENSG00000049540	6572
IKZF3	Cosmic	ENSG00000161405	10129
P2RY8	Cosmic	ENSG00000182162	4528
CRLF2	Cosmic	ENSG00000205755	1983
LEF1	Cosmic	ENSG00000138795	5361
CCND1	Cosmic	ENSG00000110092	4761
BCR	Cosmic	ENSG00000186716	12234
ABL1	Cosmic	ENSG00000097007	7018
LYN	Cosmic	ENSG00000254087	6164
PHF6	Cosmic	ENSG00000156531	16534
ECT2L	Cosmic	ENSG00000203734	4663
EWSR1	Cosmic	ENSG00000182944	10326
IGH	Cosmic	NA	NA

SH2B3	Cosmic	ENSG00000111252	5707
PAX5	Cosmic	ENSG00000196092	9570
BCL11B	Cosmic	ENSG00000127152	8528
CNOT3	Cosmic	ENSG00000088038	6821
LMO2	Cosmic	ENSG00000135363	2896
RAP1GDS1	Cosmic	ENSG00000138698	5421
STIL	Cosmic	ENSG00000123473	7821
SET	Cosmic	ENSG00000119335	5940
TRA	Cosmic	NA	NA
TRB	Cosmic	NA	NA
TAL2	Cosmic	ENSG00000186051	668
TLX1	Cosmic	ENSG00000107807	2626
TLX3	Cosmic	ENSG00000164438	1534
CCNC	Cosmic	ENSG00000112237	5610
DNM2	Cosmic	ENSG00000079805	17565
LYL1	Cosmic	ENSG00000104903	3119
LCK	Cosmic	ENSG00000182866	3339
OLIG2	Cosmic	ENSG00000205927	3262
PTPRC	Cosmic	ENSG00000081237	9168
RPL10	Cosmic	ENSG00000147403	4602
RPL5	Cosmic	ENSG00000122406	3687
PICALM	Cosmic	ENSG00000073921	6342
NOTCH1	Cosmic	ENSG00000148400	12531
LMO1	Cosmic	ENSG00000166407	1705
JAK3	Cosmic	ENSG00000105639	6813
EP300	Cosmic	ENSG00000100393	11692
FBXW7	Cosmic	ENSG00000109670	10979
ETV6	Cosmic	ENSG00000139083	7230
TAF15	Cosmic	ENSG00000270647	13607
IRS4	Cosmic	ENSG00000133124	16618
TCF3	Cosmic	ENSG00000071564	6090
PBX1	Cosmic	ENSG00000185630	21370
TFPT	Cosmic	ENSG00000105619	1215
NT5C2	Cosmic	ENSG00000076685	9137
KDM6A	Cosmic	ENSG00000147050	27742
ETV6	StJude	ENSG00000139083	7230
PAX5	StJude	ENSG00000196092	9570
RUNX1	StJude	ENSG00000159216	15574
IKZF1	StJude	ENSG00000185811	10921
NRAS	StJude	ENSG00000213281	4326
KRAS	StJude	ENSG00000133703	9230
TCF3	StJude	ENSG00000071564	6090
CDKN2A	StJude	ENSG00000147889	3885
ERG	StJude	ENSG00000157554	7568
PBX1	StJude	ENSG00000185630	21370
CRLF2	StJude	ENSG00000205755	1983
TP53	StJude	ENSG00000141510	5676
ABL1	StJude	ENSG00000097007	7018
FLT3	StJude	ENSG00000122025	4159
CREBBP	StJude	ENSG00000005339	15479
DGKB	StJude	ENSG00000136267	9491
KMT2A	StJude	ENSG00000118058	31469
TBL1XR1	StJude	ENSG00000177565	11689
BCR	StJude	ENSG00000186716	12234
JAK2	StJude	ENSG00000096968	7908
CDKN2B	StJude	ENSG00000147883	4000
SETD2	StJude	ENSG00000181555	14308
KMT2D	StJude	ENSG00000167548	25436

NF1	StJude	ENSG00000196712	41150
MTAP	StJude	ENSG00000099810	14527
RYR2	StJude	ENSG00000198626	17775
PTPN11	StJude	ENSG00000179295	13456
DUX4	StJude	ENSG00000260596	2036
NSD2	StJude	NA	NA
P2RY8	StJude	ENSG00000182162	4528
ADD3	StJude	ENSG00000148700	6539
GRM1	StJude	ENSG00000152822	7408
BTG1	StJude	ENSG00000133639	4855
AFF1	StJude	ENSG00000172493	10953
KDM6A	StJude	ENSG00000147050	27742
IGH	StJude	NA	NA
EBF1	StJude	ENSG00000164330	6333
ZCCHC7	StJude	ENSG00000147905	3957
CDKN2B-AS1	StJude	ENSG00000240498	20420
CD200	StJude	ENSG00000091972	2991
MIR99AHG	StJude	ENSG00000215386	43467
EP300	StJude	ENSG00000100393	11692
GRIN2A	StJude	ENSG00000183454	18663
ARID2	StJude	ENSG00000189079	12086
CTCF	StJude	ENSG00000102974	10106
PLD5	StJude	ENSG00000180287	9546
TBC1D30	StJude	ENSG00000111490	9877
SYT16	StJude	ENSG00000139973	15237
RB1	StJude	ENSG00000139687	6434
TOX	StJude	ENSG00000198846	4076
UBA2	StJude	ENSG00000126261	4815
ZNF384	StJude	ENSG00000126746	5181
DCAF8L2	StJude	ENSG00000189186	5608
ANK3	StJude	ENSG00000151150	25759
ELF1	StJude	ENSG00000120690	4441
LAMA5	StJude	ENSG00000130702	14291
MLLT3	StJude	ENSG00000171843	8081
ATRX	StJude	ENSG00000085224	22740
DNAH8	StJude	ENSG00000124721	15348
FAM30A	StJude	ENSG00000277059	9884
USP9X	StJude	ENSG00000124486	14152
ASCC1	StJude	ENSG00000138303	5916
ATF7IP	StJude	ENSG00000171681	13141
FAT3	StJude	ENSG00000165323	19640
CSMD1	StJude	ENSG00000183117	18700
DNAH9	StJude	ENSG00000007174	16941
NT5C2	StJude	ENSG00000076685	9137
SLX4IP	StJude	ENSG00000149346	6486
UNC80	StJude	ENSG00000144406	19623
ZFHX4	StJude	ENSG00000091656	15851
CBL	StJude	ENSG00000110395	11718
ZEB2	StJude	ENSG00000169554	26151
MEF2D	StJude	ENSG00000116604	6351
NOTCH1	StJude	ENSG00000148400	12531
SH2B3	StJude	ENSG00000111252	5707
SLC35F4	StJude	ENSG00000151812	4272
STAG2	StJude	ENSG00000101972	15987
ARID5B	StJude	ENSG00000150347	8760
CSMD3	StJude	ENSG00000164796	20407
APC	StJude	ENSG00000134982	12440
PTCH1	StJude	ENSG00000185920	17281

TENM2	StJude	ENSG00000145934	21297
ATM	StJude	ENSG00000149311	39184
BCL9	StJude	ENSG00000116128	6331
IL7R	StJude	ENSG00000168685	6020
KIAA0368	StJude	ENSG00000136813	8456
LEF1	StJude	ENSG00000138795	5361
NCOR1	StJude	ENSG00000141027	15392
SHROOM2	StJude	ENSG00000146950	8218
SLC24A2	StJude	ENSG00000155886	10993
BRCA2	StJude	ENSG00000139618	12778
LEMD3	StJude	ENSG00000174106	5578
MSH6	StJude	ENSG00000116062	10770
NRXN1	StJude	ENSG00000179915	26509
EPOR	StJude	ENSG00000187266	3077
LRP1B	StJude	ENSG00000168702	16355
NCAM2	StJude	ENSG00000154654	8286
PRDM16	StJude	ENSG00000142611	9767
PTPRT	StJude	ENSG00000196090	12943
ZFP36L2	StJude	ENSG00000152518	3693
DSCAM	StJude	ENSG00000171587	8596
FGF14	StJude	ENSG00000102466	14059
HLF	StJude	ENSG00000108924	6776
INO80	StJude	ENSG00000128908	6940
PAG1	StJude	ENSG00000076641	11023
PKHD1	StJude	ENSG00000170927	17443
RNF213	StJude	ENSG00000173821	28592
XBP1	StJude	ENSG00000100219	3028
ZNF654	StJude	ENSG00000175105	6800
ASXL1	StJude	ENSG00000171456	13676
CNTNAP5	StJude	ENSG00000155052	11381
FCGBP	StJude	ENSG00000275395	12961
FLNB	StJude	ENSG00000136068	29144
LRP2	StJude	ENSG00000081479	16109
NCOA6	StJude	ENSG00000198646	11133
PCDH15	StJude	ENSG00000150275	15397
SLCO1C1	StJude	ENSG00000139155	4883
ZMYM2	StJude	ENSG00000121741	13611
BRCA1	StJude	ENSG00000012048	9306
COL12A1	StJude	ENSG00000111799	18860
FAT1	StJude	ENSG00000083857	16158
FAT4	StJude	ENSG00000196159	21269
GRID2	StJude	ENSG00000152208	10617
IGF2R	StJude	ENSG00000197081	22613
MPDZ	StJude	ENSG00000107186	11373
NIPBL	StJude	ENSG00000164190	12519
ROBO1	StJude	ENSG00000169855	11050
SPTBN5	StJude	ENSG00000137877	12312
ARPP21	StJude	ENSG00000172995	10156
IFTAP	StJude	NA	NA
CASP12	StJude	ENSG00000204403	1786
CHD4	StJude	ENSG00000111642	12713
EGFR	StJude	ENSG00000146648	12507
LCOR	StJude	ENSG00000196233	29854
MGA	StJude	ENSG00000174197	19590
MXRA5	StJude	ENSG00000101825	9804
MYC	StJude	ENSG00000136997	4584
NYNRIN	StJude	ENSG00000205978	7733
PDZD2	StJude	ENSG00000133401	15922

PHF6	StJude	ENSG00000156531	16534
TENM3	StJude	ENSG00000218336	12010
XIST	StJude	ENSG00000229807	25266
ALK	StJude	ENSG00000171094	7382
ARMCX4	StJude	ENSG00000196440	13347
AUTS2	StJude	ENSG00000158321	24887
COL19A1	StJude	ENSG00000082293	11018
COL6A5	StJude	ENSG00000172752	9581
FLG	StJude	ENSG00000143631	12793
FOXO1	StJude	ENSG00000150907	5957
GRIK2	StJude	ENSG00000164418	26797
HMCN1	StJude	ENSG00000143341	18739
JAK1	StJude	ENSG00000162434	12620
KIAA1217	StJude	ENSG00000120549	12281
MEF2C	StJude	ENSG00000081189	11755
NOTCH2	StJude	ENSG00000134250	16744
NR3C1	StJude	ENSG00000113580	9161
ODF2	StJude	ENSG00000136811	6393
PCDH7	StJude	ENSG00000169851	12903
SPEG	StJude	ENSG00000072195	17066
TFCP2L1	StJude	ENSG00000115112	9367
UHRF1	StJude	ENSG00000276043	5145
WAC	StJude	ENSG00000095787	12780
APC2	StJude	ENSG00000115266	12635
C10ORF67	StJude	NA	NA
CEP112	StJude	ENSG00000154240	7899
CSMD2	StJude	ENSG00000121904	18114
GRIN2B	StJude	ENSG00000273079	31134
KCNT2	StJude	ENSG00000162687	7192
KSR2	StJude	ENSG00000171435	18301
PCM1	StJude	ENSG00000078674	14321
PTPRD	StJude	ENSG00000153707	11468
SALL3	StJude	ENSG00000256463	6849
SPTB	StJude	ENSG00000070182	14796
TACC2	StJude	ENSG00000138162	18855
TSC22D1	StJude	ENSG00000102804	8586
UBR4	StJude	ENSG00000127481	24661
WBSCR17	StJude	ENSG00000185274	4895
ADGRB3	StJude	NA	NA
ANKS1B	StJude	ENSG00000185046	11335
ARID1A	StJude	ENSG00000117713	15939
ASXL3	StJude	ENSG00000141431	13636
BNC2	StJude	ENSG00000173068	15323
BTLA	StJude	ENSG00000186265	3610
CCDC168	StJude	ENSG00000175820	21470
COL5A2	StJude	ENSG00000204262	7319
DOCK9	StJude	ENSG00000088387	14808
DOT1L	StJude	ENSG00000104885	11488
DYSF	StJude	ENSG00000135636	8545
ELN	StJude	ENSG00000049540	6572
FBN2	StJude	ENSG00000138829	13218
GRIK1	StJude	ENSG00000171189	4242
JMJD1C	StJude	ENSG00000171988	11155
KIAA1549	StJude	ENSG00000122778	12498
MED12	StJude	ENSG00000184634	17341
MEG3	StJude	ENSG00000214548	23224
MSH2	StJude	ENSG00000095002	13674
PIK3R1	StJude	ENSG00000145675	10731

PKHD1L1	StJude	ENSG00000205038	20621
POLRMT	StJude	ENSG00000099821	5060
PXDNL	StJude	ENSG00000147485	5239
SSBP2	StJude	ENSG00000145687	11501
TET2	StJude	ENSG00000168769	16474
TLN2	StJude	ENSG00000171914	16247
TMEM132D	StJude	ENSG00000151952	7434
ABL2	StJude	ENSG00000143322	16689
ADAMTSL1	StJude	ENSG00000178031	13446
ARFGEF3	StJude	NA	NA
ARID1B	StJude	ENSG00000049618	45931
BCOR	StJude	ENSG00000183337	9965
BCORL1	StJude	ENSG00000085185	7984
BRWD3	StJude	ENSG00000165288	14234
C1ORF112	StJude	NA	NA
CIITA	StJude	ENSG00000179583	22553
CNTN6	StJude	ENSG00000134115	5507
COL6A2	StJude	ENSG00000142173	5334
CTNNA3	StJude	ENSG00000183230	16452
DCLK1	StJude	ENSG00000133083	14451
DSCAML1	StJude	ENSG00000177103	7036
EZH2	StJude	ENSG00000106462	10774
GRIK3	StJude	ENSG00000163873	10497
ITPR2	StJude	ENSG00000123104	13909
JAKMIP3	StJude	ENSG00000188385	14669
KIF2B	StJude	ENSG00000141200	2267
KMT2C	StJude	ENSG00000055609	53322
MBD3	StJude	ENSG00000071655	7909
STIL	digital MLPA list PMC	ENSG00000123473	7821
TAL1	digital MLPA list PMC	ENSG00000162367	5705
IKZF2	digital MLPA list PMC	ENSG00000030419	12243
CD200	digital MLPA list PMC	ENSG00000091972	2991
BTLA	digital MLPA list PMC	ENSG00000186265	3610
TBL1XR1	digital MLPA list PMC	ENSG00000177565	11689
FHIT	digital MLPA list PMC	ENSG00000189283	7472
LEF1	digital MLPA list PMC	ENSG00000138795	5361
NR3C2	digital MLPA list PMC	ENSG00000151623	6607
SPARC	digital MLPA list PMC	ENSG00000113140	4641
EGR1	digital MLPA list PMC	ENSG00000120738	3137
PDGFRB	digital MLPA list PMC	ENSG00000113721	7135
FLT4	digital MLPA list PMC	ENSG00000037280	8852
CTNNA1	digital MLPA list PMC	ENSG00000044115	8830
EBF1	digital MLPA list PMC	ENSG00000164330	6333
NR3C1	digital MLPA list PMC	ENSG00000113580	9161
RPS14	digital MLPA list PMC	ENSG00000164587	5082
CASP8AP2	digital MLPA list PMC	ENSG00000118412	8147
MYB	digital MLPA list PMC	ENSG00000118513	5497
KCNH2	digital MLPA list PMC	ENSG00000055118	7207
EPHA1	digital MLPA list PMC	ENSG00000146904	4697
EZH2	digital MLPA list PMC	ENSG00000106462	10774
IKZF1	digital MLPA list PMC	ENSG00000185811	10921
TOX	digital MLPA list PMC	ENSG00000198846	4076
CDKN2B	digital MLPA list PMC	ENSG00000147883	4000
CDKN2A	digital MLPA list PMC	ENSG00000147889	3885
PAX5	digital MLPA list PMC	ENSG00000196092	9570
ABL1	digital MLPA list PMC	ENSG00000097007	7018
MLLT3	digital MLPA list PMC	ENSG00000171843	8081
NUP214	digital MLPA list PMC	ENSG00000126883	24733

MTAP	digital MLPA list PMC	ENSG00000099810	14527
NOTCH1	digital MLPA list PMC	ENSG00000148400	12531
DMD	digital MLPA list PMC	ENSG00000198947	55038
PHF6	digital MLPA list PMC	ENSG00000156531	16534
GYG2	digital MLPA list PMC	ENSG00000056998	3621
AKAP17A	digital MLPA list PMC	ENSG00000197976	4070
IL3RA	digital MLPA list PMC	ENSG00000185291	1710
CD99	digital MLPA list PMC	ENSG00000002586	4858
ZBED1	digital MLPA list PMC	ENSG00000214717	4764
ASMT	digital MLPA list PMC	ENSG00000196433	1788
P2RY8	digital MLPA list PMC	ENSG00000182162	4528
CSF2RA	digital MLPA list PMC	ENSG00000198223	4093
CRLF2	digital MLPA list PMC	ENSG00000205755	1983
SHOX	digital MLPA list PMC	ENSG00000185960	9145
SRY	digital MLPA list PMC	ENSG00000184895	828
ADD3	digital MLPA list PMC	ENSG00000148700	6539
PTEN	digital MLPA list PMC	ENSG00000171862	12547
SLC1A2	digital MLPA list PMC	ENSG00000110436	22800
LMO1	digital MLPA list PMC	ENSG00000166407	1705
LMO2	digital MLPA list PMC	ENSG00000135363	2896
CD44	digital MLPA list PMC	ENSG00000026508	9992
RAG2	digital MLPA list PMC	ENSG00000175097	3502
BTG1	digital MLPA list PMC	ENSG00000133639	4855
ETV6	digital MLPA list PMC	ENSG00000139083	7230
RB1	digital MLPA list PMC	ENSG00000139687	6434
IGHM	digital MLPA list PMC	ENSG00000211899	1871
SPRED1	digital MLPA list PMC	ENSG00000166068	7750
CREBBP	digital MLPA list PMC	ENSG00000005339	15479
CTCF	digital MLPA list PMC	ENSG00000102974	10106
BRIP1	digital MLPA list PMC	ENSG00000136492	19639
TP53	digital MLPA list PMC	ENSG00000141510	5676
SUZ12	digital MLPA list PMC	ENSG00000178691	6551
IKZF3	digital MLPA list PMC	ENSG00000161405	10129
NF1	digital MLPA list PMC	ENSG00000196712	41150
PTPN2	digital MLPA list PMC	ENSG00000175354	8718
TMPRSS15	digital MLPA list PMC	ENSG00000154646	4362
ADAMTS5	digital MLPA list PMC	ENSG00000154736	9765
HSPA13	digital MLPA list PMC	ENSG00000155304	4153
BACH1	digital MLPA list PMC	ENSG00000156273	7666
TIAM1	digital MLPA list PMC	ENSG00000156299	9198
KCNE2	digital MLPA list PMC	ENSG00000159197	1061
SIM2	digital MLPA list PMC	ENSG00000159263	7670
TFF1	digital MLPA list PMC	ENSG00000160182	492
RUNX1	digital MLPA list PMC	ENSG00000159216	15574
COL6A2	digital MLPA list PMC	ENSG00000142173	5334
PSMG1	digital MLPA list PMC	ENSG00000183527	2152
TMPRSS2	digital MLPA list PMC	ENSG00000184012	6589
RIPK4	digital MLPA list PMC	ENSG00000183421	4016
OLIG2	digital MLPA list PMC	ENSG00000205927	3262
HLCS	digital MLPA list PMC	ENSG00000159267	10549
APP	digital MLPA list PMC	ENSG00000142192	6316
BTG3	digital MLPA list PMC	ENSG00000154640	2304
PRMT2	digital MLPA list PMC	ENSG00000160310	7603
ETS2	digital MLPA list PMC	ENSG00000157557	4977
MIR99A	digital MLPA list PMC	ENSG00000207638	81
MIR155	digital MLPA list PMC	ENSG00000275402	NA
ITGB2	digital MLPA list PMC	ENSG00000160255	6869
ERG	digital MLPA list PMC	ENSG00000157554	7568

NCAM2	digital MLPA list PMC	ENSG00000154654	8286
SAMSN1	digital MLPA list PMC	ENSG00000155307	5185
SLC19A1	digital MLPA list PMC	ENSG00000173638	11235
KCNJ6	digital MLPA list PMC	ENSG00000157542	19979
DYRK1A	digital MLPA list PMC	ENSG00000157540	26763
CYYR1	digital MLPA list PMC	ENSG00000166265	3271
IGLL1	digital MLPA list PMC	ENSG00000128322	898
VPREB1	digital MLPA list PMC	ENSG00000169575	650
NT5C2	Li et al. blood (2020)	ENSG00000076685	9137
PRPS1	Li et al. blood (2020)	ENSG00000147224	3390
KRAS	Li et al. blood (2020)	ENSG00000133703	9230
FPGS	Li et al. blood (2020)	ENSG00000136877	3931
MSH2	Li et al. blood (2020)	ENSG00000095002	13674
MSH6	Li et al. blood (2020)	ENSG00000116062	10770
PMS2	Li et al. blood (2020)	ENSG00000122512	6231
WHSC1	Li et al. blood (2020)	ENSG00000109685	20352
NRAS	UENO et al. Blood Adv. (2020)	ENSG00000213281	4326
ATF7IP	UENO et al. Blood Adv. (2020)	ENSG00000171681	13141
FLT3	UENO et al. Blood Adv. (2020)	ENSG00000122025	4159
SETD2	UENO et al. Blood Adv. (2020)	ENSG00000181555	14308
KMT2D	UENO et al. Blood Adv. (2020)	ENSG00000167548	25436
PTPN11	UENO et al. Blood Adv. (2020)	ENSG00000179295	13456
RAG1	UENO et al. Blood Adv. (2020)	ENSG00000166349	7839
ATM	UENO et al. Blood Adv. (2020)	ENSG00000149311	39184
DOT1L	UENO et al. Blood Adv. (2020)	ENSG00000104885	11488
MGA	UENO et al. Blood Adv. (2020)	ENSG00000174197	19590
ASXL1	UENO et al. Blood Adv. (2020)	ENSG00000171456	13676
SH2B3	UENO et al. Blood Adv. (2020)	ENSG00000111252	5707
KDM6A	UENO et al. Blood Adv. (2020)	ENSG00000147050	27742
CFTR	UENO et al. Blood Adv. (2020)	ENSG00000001626	12116
FWXB7	UENO et al. Blood Adv. (2020)	NA	NA
CDH2	UENO et al. Blood Adv. (2020)	ENSG00000170558	8583
TRRAP	UENO et al. Blood Adv. (2020)	ENSG00000196367	19972
USP9X	UENO et al. Blood Adv. (2020)	ENSG00000124486	14152

Supplementary Table 3: ALL-associated genes which are altered in their protein coding sequence in patients with SBS7a.

chr	start	ref	alt	Consequence	IMPACT	SYMBOL	Gene	Probability of being caused by SBS H	sample
chr4	105272871	C	T	missense_variant	MODERATE	TET2	ENSG00000168769	0.88	P0608
chr5	143298725	G	A	missense_variant	MODERATE	NR3C1	ENSG00000113580	0.88	P0608
chr5	143399684	G	A	stop_gained	HIGH	NR3C1	ENSG00000113580	0.25	P0608
chr6	101626391	C	A	missense_variant	MODERATE	GRIK2	ENSG00000164418	0.03	P0608
chr7	14841244	C	A	missense_variant	MODERATE	DGKB	ENSG00000136267	0.06	P0608
chr10	54023128	G	A	missense_variant	MODERATE	PCDH15	ENSG00000150275	0.25	P0608
chr14	99175727	A	T	missense_variant	MODERATE	BCL11B	ENSG00000127152	0.04	P0608
chr2	209912640	C	T	synonymous_variant	LOW	UNC80	ENSG00000144406	0.38	P0609
chr2	140776173	T	C	missense_variant	MODERATE	LRP1B	ENSG00000168702	0.02	P0609
chr12	112489096	C	A	missense_variant	MODERATE	PTPN11	ENSG00000179295	0.04	P0609
chr6	160040670	G	C	missense_variant	MODERATE	IGF2R	ENSG00000197081	0,00E+00	P0610
chr10	60114300	T	A	missense_variant	MODERATE	ANK3	ENSG00000151150	0,00E+00	P0610
chr4	186618872	C	T	missense_variant	MODERATE	FAT1	ENSG00000083857	0.99	P0611
chr7	98937789	C	T	missense_variant	MODERATE	TRRAP	ENSG00000196367	0.99	P0611
chr8	76854830	C	T	missense_variant	MODERATE	ZFHX4	ENSG00000091656	0.95	P0611
chr12	112450362	A	T	missense_variant	MODERATE	PTPN11	ENSG00000179295	0.25	P0611
chr8	109412300	C	T	synonymous_variant	LOW	PKHD1L1	ENSG00000205038	0.03	P0623
chr12	111418215	C	T	missense_variant	MODERATE	SH2B3	ENSG00000111252	0.02	P0623
chr9	13121870	C	T	synonymous_variant	LOW	MPDZ	ENSG00000107186	0.06	P0557
chr1	237674774	C	T	stop_gained	HIGH	RYR2	ENSG00000198626	0.11	P0621
chr2	219482817	C	T	missense_variant	MODERATE	SPEG	ENSG00000072195	0.98	P0621
chr8	76863998	C	T	synonymous_variant	LOW	ZFHX4	ENSG00000091656	0.85	P0621
chr16	3736796	A	T	missense_variant	MODERATE	CREBBP	ENSG00000005339	0.23	P0621
chr4	186620720	C	T	missense_variant	MODERATE	FAT1	ENSG00000083857	0.69	P0621
chr2	140716042	G	A	missense_variant	MODERATE	LRP1B	ENSG00000168702	0.93	P0624
chr2	209826045	C	T	stop_gained	HIGH	UNC80	ENSG00000144406	0.29	P0624
chr16	3736739	G	C	missense_variant	MODERATE	CREBBP	ENSG00000005339	0,00E+00	P0624
chr16	3736742	C	T	missense_variant	MODERATE	CREBBP	ENSG00000005339	0.95	P0624
chr20	42678048	T	G	missense_variant	MODERATE	PTPRT	ENSG00000196090	0.1	P0624
chr7	117652916	G	A	stop_gained	HIGH	CFTR	ENSG00000001626	0.99	P0625
chr10	122084853	G	A	missense_variant	MODERATE	TACC2	ENSG00000138162	0.99	P0625
chr12	129699851	A	G	synonymous_variant	LOW	TMEM132D	ENSG00000151952	0.3	P0627
chr21	36937138	G	A	synonymous_variant	LOW	HLCS	ENSG00000159267	0.94	P0627