RETURN OF THE CLONES

Mechanisms driving clonal evolution and relapse development in pediatric Acute Lymphoblastic Leukemia

Željko Antić

Return of the clones

Mechanisms driving clonal evolution and relapse development in pediatric Acute Lymphoblastic Leukemia

Željko Antić

ISBN 978-90-393-7447-4
DOI 10.33540/1140
https://doi.org/10.33540/1140

Copyright © 2022 Željko Antić. All rights reserved.
No part of this thesis may be reproduced, stored in retrieval system, or transmitted in any form or by any means, without prior permission of the author, or publishers holding copyrights for the published articles.

Cover design Željko Antić
Layout and design PUBLISS | publiss.nl
Printing RidderPrint | ridderprint.nl

Return of the clones
Mechanisms driving clonal evolution and relapse development in pediatric Acute Lymphoblastic Leukemia

Terugkeer van de klonen
Drijvende mechanismen van klonale evolutie en recidiefontwikkeling bij pediatrische acute lymfatische leukemie
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 12 mei 2022 des ochtends te 10.15 uur

door

Željko Antić

geboren op 2 februari 1989
te Belgrado, Joegoslavië

Promotor:

Prof. dr. F.C.P. Holstege

Copromotor:

Dr. R.P. Kuiper

Dit proefschrift werd mogelijk gemaakt met financiële steun van Stichting Kinderen Kankervrij (KIKA).

To my dearest

Најмилијима

ASSESSMENT COMMITTEE

Prof. dr. Monique Minnema (University Medical Center Utrecht)
Prof. dr. Monique den Boer (Erasmus University Medical Center)
Prof. dr. Josef Vormoor (University Medical Center Utrecht)
Prof. dr. Hans Kristian Ploos van Amstel (University Medical Center Utrecht)
Dr. Jeroen de Ridder (University Medical Center Utrecht)

TABLE OF CONTENTS

Chapter 1 General introduction and outline of the thesis 9

Chapter 2 Accurate detection of low-level mosaic mutations in pediatric acute lymphoblastic leukemia using single molecule tagging and deep-sequencing 33
Leukemia & Lymphoma; July 2018; 59(7):1690-1699

Chapter 3 Multiclonal complexity of pediatric acute lymphoblastic leukemia and the prognostic relevance of subclonal mutations 65
Haematologica; December 2021; 106(12):3046-3055

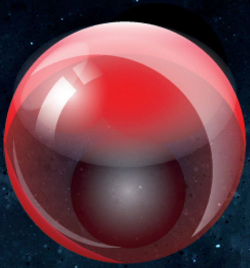
Chapter 4 Clonal dynamics in pediatric B-cell precursor acute lymphoblastic leukemia with very early relapse 93
Pediatric Blood & Cancer; January 2022; 69(1):e29361

Chapter 5 Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia 117
Blood Cancer Discovery; July 2020; 1(1):96-111

Chapter 6 Unravelling the sequential interplay of mutational mechanisms during clonal evolution in relapsed pediatric acute lymphoblastic leukemia 175
Genes (Basel); February 2021; 12(2):214

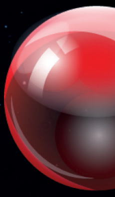
Chapter 7 General discussion and future perspectives 197

Appendix Summary 224
Samenvatting 227
Сажетак 231
Curriculum Vitae 235
List of publications 236
Acknowledgments 238



CHAPTER 1

GENERAL INTRODUCTION AND OUTLINE OF THE THESIS



CANCER - ORIGIN, PATHOGENESIS AND TREATMENT

Cancer represents a group of chronic diseases characterized by abnormal cell growth and the ability to invade or spread to other parts of the body[1]. This acquired capacity of cancer cells to sustain growth, invade and disseminate, is considered one of the most important hallmarks of cancer, in contrast to benign tumors which have limited and localized growth capabilities[1-3].

Approximately 90% of all cancers occur due to somatic mutations caused by environmental factors and lifestyle[4]. Therefore, cancer is usually a late onset disease that requires stepwise acquisition of multiple genetic alterations, which progressively lead to transformation of a normal cell into a malignant one. Although all cancer cells arise from a single malignant cell, individual cancer cells continue to acquire additional genetic lesions, resulting in a continuous process towards heterogeneity and selection of cancer clones with higher proliferative potential or better adaptation to the circumstances in order to survive. Over time, new clones may outgrow existing cancer clones and become the dominant population of cells in the tumor[2,3,5].

Treatment of cancer is dependent on tumor type and on individual characteristics of a patient. In the past, cancer treatment was largely based on the use of multidrug chemotherapy regimens, radiation therapy and surgery[6]. The discovery of frequently altered genes and their roles in cancer initiation and progression during the last decades have facilitated the introduction of innovative therapies[7-9]. Novel treatment options, improved risk stratification, and cancer screening programs have all resulted in improved survival rates in the last 40 years[10,11]. However, due to acquired resistance and other causes for treatment failure, cancer still remains one of the leading causes of death[7,12-17]. Treatment resistant cancer cells often originate from a minor clone present already at the time of diagnosis. These (sub)clones harbor genetic lesions which maintain cell fitness during treatment and can eventually cause treatment failure[13,18-20]. Therefore, better understanding of the mechanism driving acquisition of genetic lesions associated with treatment resistance, as well as their interactions within the cancer cells and with the host genome, may lead to better therapeutic strategies, improved risk stratification and overall survival.

Cancer can occur in all age groups, but generally is a late-onset disease. Therefore, the prevalence increases with age, particularly after the age of 50, and in children cancer is very rare[5,21]. Despite low prevalence, cancer represents the second most common cause of death in children after injuries[22]. The majority of childhood cancers are caused by spontaneous somatic mutations which are acquired early in life, sometimes even before birth, and, less often, heritable germline cancer predisposing mutations. However, childhood cancers do not have the strong link with lifestyle and environmental risk factors as adult cancers do, and cancers that are most common in adults, e.g., lung, colon, breast, are rare in childhood. The most common pediatric malignancies include cancers of blood, brain and liver, with acute lymphoblastic leukemia being the most prevalent[5,21].

PEDIATRIC ACUTE LYMPHOBLASTIC LEUKEMIA

Once Alfred Velpeau in 1827 and Rudlof Virchow in 1845 gave the first accurate descriptions of leukemia, this disease was considered incurable[23,24]. More than a century later, the Boston pathologist Sidney Farber reported the first "temporary remissions" in acute pediatric leukemia induced by aminopterin, marking the advent of chemotherapy for this disease[24-26]. Already in 1961 Emil Frei, Emil Freireich and James Holland introduced the first combinatorial treatment of methotrexate and mercaptopurine, achieving complete remission in 59% of the children[25,26]. Only a year later Pinkel and colleagues introduced the first structured clinical trial with a novel multidrug therapy approach "total therapy" in St. Jude research hospital. One of these trials (Total Therapy Study V, 1967-1968) achieved remarkable results, with almost half of all the enrolled patients in a remission. The success of the St. Jude treatment protocols inspired similar worldwide clinical trials, paving the way for further improving the outcome over the years[25,26]. In 1975, another milestone was achieved with the first risk group stratification based on clinical parameters, e.g., white blood cell count, age and central nervous system involvement[27,28]. In parallel, research on the biology of leukemia was ongoing, which led to recognition of the first cytogenetic parameters for risk stratification, e.g., the presence of the specific chromosomal translocations and aneuploidies[29,30]. Finally, introduction of the first cell surface antibodies for T cell differentiation in 1981 marked a new era in risk group stratification with the introduction of minimal disease monitoring (MRD)[31]. Improvements in molecular genetics technology and introduction of Next Generation Sequencing (NGS) approaches has opened new horizons in the research of leukemia biology with the discovery of novel genetic alterations[32-36], and paved the way for the introduction of novel targeted therapy in leukemia treatment, e.g., tyrosine kinase inhibitor (imatinib) for the treatment of *BCR-ABL1*-positive leukemia[37].

Modern treatment of ALL consists of three phases: remission induction, consolidation (intensification) and maintenance (continuation). The complete course of all three treatment phases lasts 2-3 years, depending on the treatment protocol. Based on the clinical and genetic features, patients are stratified into different treatment arms, with the goal to provide more intensive treatment for patients with less favorable prognostic features and to limit the treatment-associated toxicity for patients with good prognosis[38,39]. Male gender, infancy or adolescence, central nervous system (CNS) involvement, a T-cell immunophenotype, white blood cell (WBC) count at diagnosis higher or equal to $50x10^9/l$ are associated with less favorable prognosis of ALL. Minimal residual disease (MRD), representing the burden of leukemic blast in 10^3 - 10^6 cells at the days 33 and 78 from the start of the treatment, can accurately predict patients at risk to develop relapse and represents another important treatment stratification marker.

Remission induction is composed of corticosteroids (prednisone and dexamethasone), vincristine, asparaginase and anthracyclines administered over 4-6 weeks. This treatment phase is able to induce remission in over 90% of patients and is followed by intensification which consists of cyclophosphamide, cytarabine and mercaptopurine. Finally, maintenance treatment relies on the use of mercaptopurine, methotrexate (sometimes combined with vincristine) and dexamethasone pulses. Although most of the drugs contained in the backbone of ALL therapy have been used for more than five decades, dosage and applications were adjusted based on clinical studies involving different clinical characteristics, MRD response, genetic subtypes, pharmacogenomic studies as well as their tolerability and toxicity[39-43].

Clinical characteristics of pediatric ALL

ALL is the most common pediatric malignancy accounting for up to 20% of all cancers diagnosed in children[44]. Currently, on average 130 children per year are diagnosed with ALL in the Netherlands[45]. Perturbation of normal lymphocyte development leads to uncontrolled proliferation of immature cells, called blasts. These cells have impaired function and their infiltration of healthy bone marrow causes anemia, neutropenia and thrombocytopenia due to suppression of healthy cells[46]. Furthermore, infiltration of secondary organs can give rise to a variety of additional clinical signs and symptoms, which complicates diagnosis and treatment[47-49].

Origins and differentiation of B- and T-cells

In order to understand how leukemia develops, we need to understand normal immune cell development. B- and T-cells originate from aorta-gonad-mesonephros, a descendent of mesoderm, which colonizes developing fetal liver early in embryogenesis. Differentiated hematopoietic stem cells (HPSC) are slowly seeded from fetal liver into the bone marrow, which becomes the main compartment of hematopoiesis in adulthood, where the development of B- and T-cell lymphoid progenitors takes place[50-52] (**Figure 1.1**).

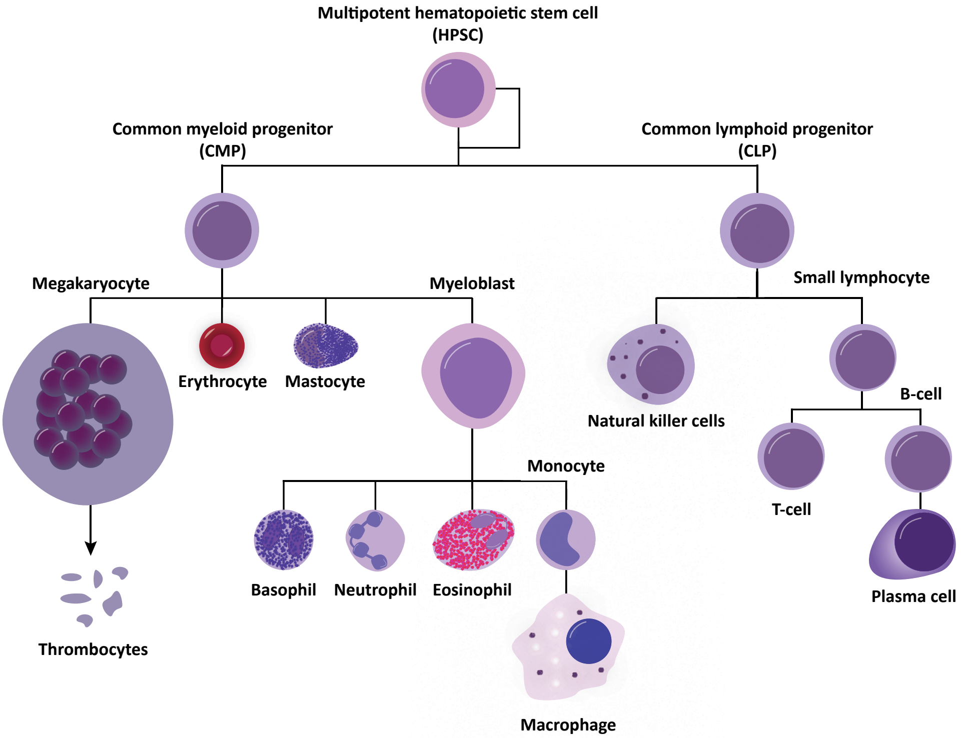

Multipotent hematopoietic stem cell
(HPSC)
Common myeloid progenitor
(CMP)
Common lymphoid progenitor
(CLP)
Megakaryocyte
Myeloblast
Small lymphocyte
Erythrocyte
Mastocyte
Natural killer cells
B-cell
Monocyte
T-cell
Basophil
Neutrophil
Eosinophil
Plasma cell
Thrombocytes
Macrophage

Figure 1.1. Schematic representation of hematopoiesis.

B-cells form an essential component of the adaptive humoral immune system, through secretion of antibodies, cytokines and antigen presentation. Early B-cell development is bone marrow dependent, and includes successful VDJ recombination of heavy chains, as well as VJ rearrangement of the light chains by *RAG1/2* (recombination activated gene). Upon successful rearrangement of µ heavy chains, immature B-cells leave the bone marrow in order to inhabit secondary lymphoid organs where the B-cells mature[50,53]. These B-cells represent a population of cells that express clonally diverse immunoglobulin (Ig) receptors with the ability to recognize diverse epitopes[50,53]. Antigen recognition-mediated B-cell activation causes changes in gene expression, activation of AID (activation-induced cytidine deaminase), and clonal proliferation. AID activation drives localized somatic hypermutation (SHM) in variable Ig loci resulting in formation of clones each carrying different Ig receptors with affinity for a wide variety of epitopes. B-cells that show high affinity for recognized epitopes further proliferate and differentiate into antibody secreting plasma cells. These antibodies retain the unique ability to recognize the same antigen epitopes as B-cell Ig receptors and, together with cytokines produced by activated B-cells, they represent essential components of adaptive and humoral immunity[50,53].

Unlike B-cells, the development of early T-cell progenitors is thymus dependent. Upon commitment and differentiation, common lymphoid progenitors (CLP) leave the bone marrow and inhabit the thymus where RAG-mediated VDJ recombination of the T-cell receptor (TCR) α and β chains occurs. Following successful rearrangement and expression of TCRs, T-cells undergo positive and negative selection, a process during which the majority of maturing T-cells die. T-cells that survive this process represent naive T-cells, which are then released into the bloodstream[54-56].

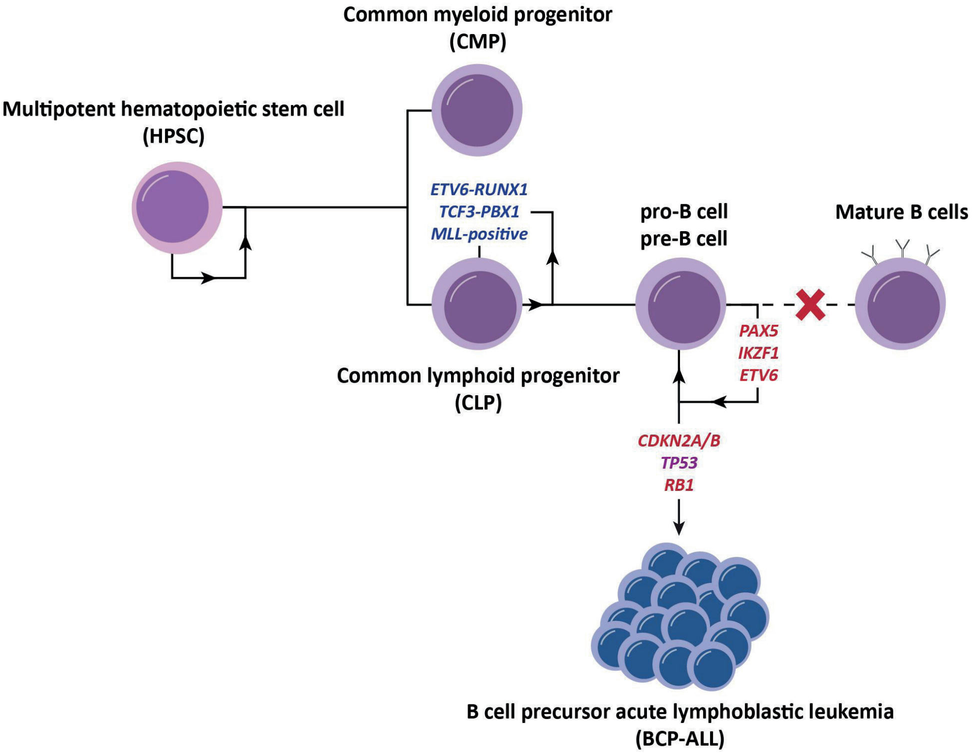

Common myeloid progenitor
(CMP)
Multipotent hematopoietic stem cell
(HPSC)
ETV6-RUNX1
TCF3-PBX1
MLL-positive
pro-B cell
pre-B cell
Mature B cells
PAX5
IKZF1
ETV6
Common lymphoid progenitor
(CLP)
CDKN2A/B
TP53
RB1
B cell precursor acute lymphoblastic leukemia
(BCP-ALL)

Figure 1.2. Schematic representation of leukemogenesis with key genetic alterations affecting maturation and differentiation pathways. Genes frequently affected with gain of function (GoF) alterations are shown in blue, genes with loss of function (LoF) mutations in red, while genes that can be affected with both GoF and LoF alterations are shown in purple.

During B- and T-cell differentiation, maturation and proliferation, several transcription factors become selectively active in different developmental stages. For example, *EBF1*, *PAX5*, *TCF3*, *FOXO1*, *PU1* and *IKZF1* (Ikaros) play a key role in the commitment and differentiation of B-cells, whereas the transcription factors *FOXP3*, *GATA3*, *TCF7*, *BCL11B*, *RUNX1* and *ERG* determine T-cell development[50,51,53,55-58]. Dysregulation of these transcription factors can lead to altered B- and T-cell development and function, and eventually lead to malignancies, including leukemia[59] (**Figure 1.2**).

Leukemogenesis

The pathophysiology of pediatric ALL is complex and involves acquisition of multiple genetic lesions in a stepwise process, leading to transformation of lymphoid blasts into malignant cells[60]. Genomic alterations, such as aneuploidies (hyperdiploidy and hypodiploidy) and chromosomal translocations (e.g., *ETV6-RUNX1*, *TCF3-PBX1*, *BCR-ABL1*, and multiple rearrangements involving *KMT2A*), are early mutually exclusive events in leukemia development and are referred to as primary alterations. In a subset of cases none of the well-defined primary genomic alterations can be identified, and this subgroup has been referred to as B-other[39,61,62] (**Figures 1.2** and **1.3**). The primary alterations can have prenatal origin, but are usually not sufficient to drive development of the disease[62,63]. Therefore, pre-leukemic cells need to acquire additional driving alterations in oncogenes and tumor suppressor genes in order to overcome the intrinsic proliferative block and increase their potential for clonal expansion[62,63] (**Figure 1.2**). Furthermore, although not all secondary alterations directly contribute to disease development, they may be associated with treatment resistance[18,19,33]. The first genome-wide studies unraveled remarkable genomic heterogeneity of pediatric ALL[32,33,64] and revealed the importance of B-cell differentiation-specific transcription factor alterations, like *PAX5*, *IKZF1*, and *EBF1*, in disease development[33,65]. Deletions of the *IKZF1* gene were even shown to be an important predictor for treatment failure in various clinical studies[65-72], leading to specific adaptations in some treatment protocols[66,73].

With improvements in the NGS technologies over the past decade, new ALL subtypes emerged, particularly from the B-others group. First among them were *BCR-ABL1*-like cases, characterized by heterogeneous genomic alterations and an expression profile similar to what is observed in the *BCR-ABL1*-positive ALL[72,74]. Later studies unraveled a plethora of novel primary alterations, including *MEF2D-BCL9*, *ZNF384* and *NUTM1* gene rearrangements, and even new groups characterized by heterogeneous genomic alterations and expression profiles similar to one of the known subtypes, e.g., *ETV6-RUNX1*-like, *KMT2A*-like and *ZNF384*-like[75-79] (**Figure 1.3**). Identification of these novel entities in ALL have prognostic and therapeutic significance and may aid further treatment optimization, as was previously demonstrated with *BCR-ABL1*-like cases[80-82]. Furthermore, improved detection of secondary alterations revealed their association with primary alterations, e.g., *CREBBP* and RAS pathway mutations with hyperdiploidy[83,84], as well as their importance for relapse development, e.g., *CREBBP*, *TP53*, *WHSC1*, *KRAS*, *NRAS*, *NT5C2*[18,19,83,85-88].

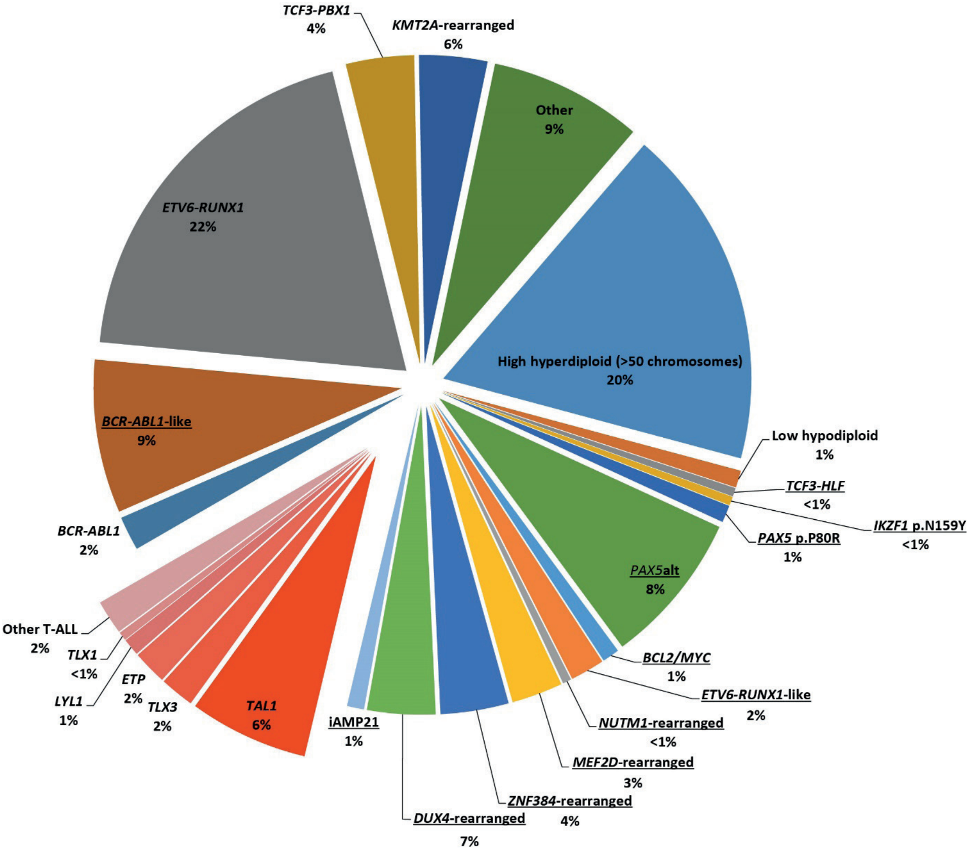


Figure 1.3. Frequency of primary driving alteration in pediatric ALL. Adapted from Inaba H. *et al.*, 2013, Iacobucci I. *et al.*, 2017, and Inaba H. *et al.*, 2020[39,61,89]. Alterations previously assigned to the B-other subgroup are underlined.

With improved genetic characterization of ALL, certain somatic alterations were found to be associated with less favorable prognosis, which could explain previous association with observed clinical factors. Subtypes such as *ETV6-RUNX1* and high hyperdiploid ALL

(>50 chromosomes) have better prognosis compared to hypodiploid (<44 chromosomes), *KMT2A*-rearranged, *BCR-ABL1* and *BCR-ABL1*-like leukemias, which are generally associated with worse prognosis in different treatment protocols[32,38,39,41,46,61,66,90,91]. Acknowledging distinct clinical and biological features of different ALL subtypes, special international treatment studies have been initiated for leukemias for the more rare leukemia subtypes, with particularly unfavorable prognosis, e.g., infant ALL and *BCR-ABL1*-positive ALL[41,42,82,92].

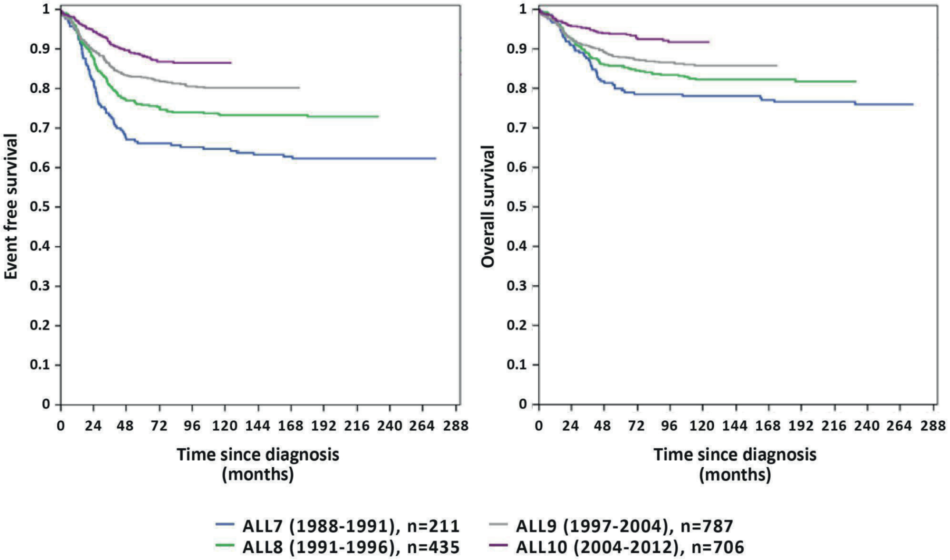


Figure 1.4 - Kaplan-Meier curves depicting event free and overall survival in four consecutive Dutch Childhood Oncology Group (DCOG) ALL treatment protocols (ALL7-ALL10). Analyses were limited to patients between 1-16 years of age. Older patients, infants, *BCR-ABL1*-positive and patients with Down syndrome are excluded from the analyses because of the large differences in inclusion criteria between compared protocols. Furthermore, the high-risk patients who were treated according to institutional protocols (patients with T-ALL treated in Amsterdam Medical Center and cases with CNS involvement treated in ErasmusMC) in the protocols ALL7, ALL8 and ALL9, were not included in the analyses. Figure courtesy of DCOG.

Improved genetic characterization of ALL over the past decades has led to discovery of new alterations with unfavorable prognosis, like *BCR-ABL1*-like subtype and *IKZF1* deletions, resulting in the treatment adjustment for cases with a high risk for treatment failure. In addition, introduction of novel therapeutics, such as first- and second-generation tyrosine kinase inhibitors (e.g., imatinib and dasatinib) in the treatment of *BCR-ABL1*-positive ALL, further underpinned the importance of targeted treatment and the need for improvements in genetic characterization of ALL. Novel genetic findings, improved diagnostic and personalized treatment protocols, coupled with introduction of the new drugs, resulted in a gradual improvement of overall survival of patients diagnosed with ALL, now reaching 90% in developed countries[40,93] (**Figure 1.4**). In the remaining 10% of the cases, treatment is failing due to relapse, refractory disease, infections or drug toxicity[39,40,46,93]. The main goal of further clinical trials remains reduction of treatment intensity, in order to alleviate treatment associated toxicity, while still achieving stable long-term remission.

Relapse in ALL

Relapse is the most common cause of treatment failure in childhood ALL, and outcome among these cases remains poor[94]. Treatment of relapsed ALL is complicated, because these leukemic cells have survived the exposure to multiple drugs and, therefore, require the use of more intensive therapeutic options compared to frontline ALL treatment, including harsher chemotherapeutic therapy regimens and HPSC transplantation. Furthermore, novel experimental therapies are often initially tested at this stage of disease. Examples are targeted drugs like Janus kinase inhibitors, or proteasome inhibitors, and the immunotherapy-based modalities using the CD19/CD22 inhibitors blinatumomab, inotuzumab ozogamicin, or chimeric antigen receptor (CAR) T-cell therapy[94-96]. However, relapse treatment often bears risks of increased toxicity[94] and may cause severe long-term side effects[97,98], indicating the need to develop better treatment strategies in order to prevent relapse of ALL.

Relapse development is a complex process which may be influenced by several factors, including therapy-induced side-effects, interactions between leukemic cells and local microenvironment in which tumor cells reside, and genomic aberrations present in leukemic cells. Toxicity and infections caused by the treatment may lead to relapse due to treatment interruptions or suboptimal levels of chemotherapeutics. Furthermore, recent studies have shown that leukemic cells may be protected from the treatment by the surrounding microenvironment[99], e.g., by hiding in a protective extramedullary niche where drugs commonly used in the treatment cannot penetrate, like the eye and the testis[100,101], or through the formation of tunneling nanotubes[102-104]. These relapses may develop into systemic disease long after treatment has finished, or present as localized relapses. The extraordinary heterogeneity of ALL, represented by a genetically diverse population of leukemic cells that constantly competes for the resources, may give rise to a relapse by enabling survival of leukemic clones during treatment[18,19,85]. Clones with genetic composition that enable treatment resistance, or increased cell fitness, may be present already at the time of diagnosis, as has been shown for clones with *IKZF1* deletions[19,65,66,72,105,106]. Furthermore, they may be present even on a low, subclonal level, as is the case with clones harboring mutations in *WHSC1*[18,19,86] and *CREBBP*[18,19]. In other cases, genetic alterations associated with treatment resistance may not be present already at the time of diagnosis, but are actually acquired during treatment[87,88], and in some cases by frequently used chemotherapeutics, e.g., *NT5C2* and *TP53* mutations during treatment with thiopurines[87,107]. Despite good initial treatment response, leukemic cells with the newly acquired *NT5C2* mutations can grow out, giving rise to a very early, on-treatment, relapse. Furthermore, genetic polymorphisms, that affect drug metabolism, are frequently retained in the leukemic cells and can lead to treatment resistance[108]. Overall, relapsed ALL is a heterogeneous disease, both with respect to the mechanisms leading to its development as well as the genomic landscape. Therefore, relapses should not be considered as a single entity.

Clonal evolution in ALL

Neoplasms like ALL are genetically heterogeneous and can thus be considered as a large population of cells which are genetically subtly different. Genetic alterations can be beneficial for neoplastic cells by inducing cellular expansion and/or increasing fitness of leukemic clones, or they could confer evolutionary disadvantages leading to their contraction[2,18,20,87,88,109]. However, the majority of genetic alterations are neutral for the clone, meaning that they do not confer an evolutionary advantage or disadvantage[110,111]. Nevertheless, these alterations, known as passenger or hitchhiker mutations, together with primary and secondary disease drivers, increase the pool of genetic diversity within the tumor, and may improve tumor adaptation under changing environmental conditions, e.g., during chemotherapy[18,19,65,105,112,113]. Following the Darwinian model of survival of the fittest, these cells may proliferate during treatment, giving rise to a relapse[18,19].

Although the idea of cancer as an evolutionary problem is not new, historically little attention has been given to it. One of the first studies that interrogated the tumor heterogeneity in ALL made use of multiplex fluorescence *in situ* hybridization (FISH) probes. This approach enabled the elucidation of genetic heterogeneity at the cellular level and revealed a remarkable diversity in evolutionary trajectories between individual patients[109], very similar to what had been observed in epithelial cancers[114-117]. Introduction of the first genome-wide technologies improved our understanding of the mutational landscape in cancer[118]. Albeit limited in their resolution to capture tumor heterogeneity, in combination with backtracking studies, they gave the first insights into the composition of rare populations of cells in ALL and the influence of treatment on clonal selection[106]. In order to overcome the problem of examining the genetic landscape of predominant clones, many studies in recent years have focused on comparing tumor evolution by obtaining spatio-temporally separated samples, e.g., samples at diagnosis and relapse[18,19,88], samples from primary tumor and metastasis[119,120], or even by employing limited dilution xenotransplantation of primary tumor samples[121,122]. These studies unraveled unprecedented intra-tumor heterogeneity and even very complex patterns of tumor evolution. Finally, the advent of single cell approaches now offers promise for ultimate resolution to study tumor complexity at the cellular level, by capturing even the rarest cell populations and bottlenecks preceding the rise of complex (epi)genetic landscapes in ALL.

The use of NGS approaches, such as whole exome and whole genome sequencing (WES and WGS) enabled the identification of a number of pathogenic mutations driving development, progression and therapy resistance in ALL[18,19,33,65,68-70,72,75,83,86-88,113,123,124]. Studies examining similarities and differences between diagnosis and relapse have identified several key alterations frequently enriched in relapsed ALL, including alteration affecting RAS pathway genes (*KRAS*, *NRAS*, *PTPN11*, *FLT3*), chromatin remodelers (*CREBBP*, *EP300*, *WHSC1*), glucocorticoid receptor gene *NR3C1*, transcription factor *IKZF1* and purine hydrolase *NT5C2*[18,19,65,66,72,83,85,86,113,123,124]. Alterations in many of these genes were frequently found to be preserved between diagnosis and relapse, while others, e.g., *NT5C2,* are almost exclusively found in relapse samples[88,113].

MUTATIONAL PROCESSES IN ALL

The somatic DNA sequence alterations in ALL genomes can find their origin in several endogenous and exogenous mutational processes or repair defects. The “scars'' or “footprints” of these alterations may be very typical for the underlying mechanisms, which can give highly relevant insights into the etiology of leukemia development. For example, many of the (focal) copy number aberrations that have been observed in ALL show hallmarks of cryptic RAG-mediated recombination. Physiological function of RAG recombinase involves double strand DNA breaks near RSS sequences in order to initiate V(D)J recombinations and represent evolutionary trade-off between genome integrity maintenance and adaptive immunity. Due to their aberrant activity, RAG recombinase family enzymes initiate double strand DNA breaks near cryptic RSS motifs, thus causing acquisition of secondary leukemia-driving lesions in genes essential for normal B-cell development, such as *IKZF1* deletions[125-127].

Another example is a very complex genomic rearrangement, intrachromosomal amplification of chromosome 21 (iAMP21), which arises through breakage-fusion-bridge (BFB) cycles[128,129]. BFB is a mutational mechanism caused by double strand break of telomeres. Upon DNA replication, sister chromatids fuse in the G2 phase by non-homologous end-joining, forming a dicentric chromosome. During mitosis, centromeres are being pulled in the opposite directions forming an anaphase bridge. Breakage of the anaphase bridge during cytokinesis once again results in a loss of telomeres at the end of the amplified chromosome, causing the cycle to repeat until the affected chromosome finally receives a telomere[129-131]. As with other mutational processes, BFB leaves a unique footprint, a fold-back inversion, which can be recognized by closely aligned read pairs that are in inverted orientation[130].

Similar to large chromosomal aberrations caused by BFB cycles or RAG recombinase activity, small mutations like substitution, insertions and deletions can also show typical patterns that can be linked to underlying mutational processes[132-134]. Historically, analysis of six single-base substitution types have unraveled simple mutational patterns in tumors, but showed full potential only after the revelation that adjacent 5' and 3' nucleotides also contribute to mutational rates in the genome[132,134]. Based on adjacent nucleotides, each of six possible single-base substitution types can occur in one of the 16 possible trinucleotide contexts, resulting in much higher complexity with 96 mutation possibilities. These 96 mutation possibilities represent footprints of underlying mutational processes, also known as single-base substitution (SBS) signatures[132-134] (**Figure 1.5**). Mutational processes driving these signatures can be initiated by intrinsic factors, like aging and DNA repair defects, or environmental factors, like UV light exposure, or tobacco smoking. On the level of individual samples, the mutational profile may originate from several mechanisms, each represented by different SBS signatures, which indicates the presence of multiple distinct mutational mechanisms[132-134]. By comparing individual mutational profiles with known, curated profiles of SBS signatures the processes operative in individual samples can be identified. So far, more than 70 SBS signatures have been reported in COSMIC[133].

In addition to single-base substitution biases, some mutational processes can show bias for acquisition of other genomic alterations, e.g., small indels in mismatch repair deficiency, or for specific genomic regions. Combining this information allows the identification of mutational processes present in individual samples without previous knowledge of the individual environmental exposure or the presence of specific genomic lesions[133].

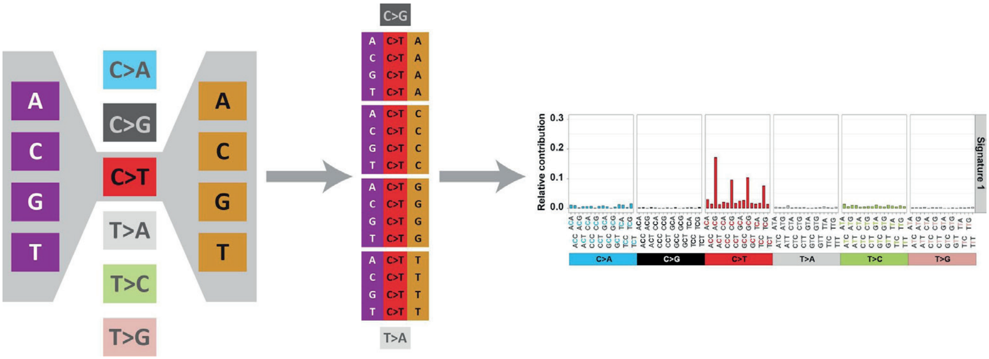


Figure 1.5. Biases towards mutations occurring in specific trinucleotide contexts represent patterns which enable recognition of underlying biological processes driving mutagenesis.

Mutational processes were intensively studied in different cancers, including pediatric ALL[132,133,135,136]. However, data about mutational mechanisms in the context of relapsed ALL remains limited and their better understanding may aid further improvement in treatment outcome.

OUTLINE OF THE THESIS

Despite improvements in the treatment of pediatric ALL over the past decades and the introduction of innovative treatment approaches, relapse still remains the leading cause of treatment failure among children with ALL, and the outcome of relapse is poor. Therefore, the aim of the work described in this thesis is to unravel the genomic background of relapsed ALL and to identify biomarkers that may be used for risk stratification and treatment adjustment. In **chapter 2** we present a novel approach to reliably detect low-level mosaic mutations in leukemia samples and compare it with conventional DNA capture and sequencing techniques. We subsequently applied this approach in **chapter 3** to a cohort of 503 Dutch ALL patients, in which we examine the clinical and prognostic relevance of subclonal alterations for relapse development. In **chapter 4** we present genomic alterations in cases that develop a relapse already before the end of treatment and discuss our findings at the level of individual patients. In **chapter 5** we report on a comprehensive analysis of a large cohort of relapsed ALL patients, which was studied in collaboration with colleagues from the St. Jude Children's research hospital and the University of Toronto. In this study we demonstrate that hypermutation is a frequent phenomenon in relapsed ALL and that different mutational mechanisms, sometimes in combination, can drive this hypermutation phenotype. In **chapter 6** we present a strategy to study the spatial and temporal timing of these mechanisms in patients with multiple relapses and show that mutational mechanisms can be active during the entire process of leukemia development and progression, or switched on or off at later time points. **Chapter 7** provides a general discussion of the work presented in this thesis, highlighting the impact of our findings as well as some future perspectives. Summaries of this thesis (in English, Dutch and Serbian) are provided in the **appendix**.

REFERENCES

1. Cifone MA. In vitro growth characteristics associated with benign and metastatic variants of tumor cells. Cancer Metastasis Rev. 1982;1(4):335-347.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-674.
3. Blagosklonny MV. Molecular theory of cancer. Cancer Biol Ther. 2005;4(6):621-627.
4. Anand P, Kunnumakkara AB, Sundaram C, Harikumar KB, Tharakan ST, Lai OS, et al. Cancer is a preventable disease that requires major lifestyle changes. Pharm Res. 2008;25(9):2097-2116.
5. Kattner P, Strobel H, Khoshnevis N, Grunert M, Bartholomae S, Pruss M, et al. Compare and contrast: pediatric cancer versus adult malignancies. Cancer Metastasis Rev. 2019;38(4):673-682.
6. Arruebo M, Vilaboa N, Sáez-Gutierrez B, Lambea J, Tres A, Valladares M, et al. Assessment of the evolution of cancer treatment therapies. Cancers (Basel). 2011;3(3):3279-3330.
7. Jackson SE, Chester JD. Personalised cancer medicine. Int J Cancer. 2015;137(2):262-266.
8. Helmy KY, Patel SA, Nahas GR, Rameshwar P. Cancer immunotherapy: accomplishments to date and future promise. Ther Deliv. 2013;4(10):1307-1320.
9. Tsimberidou AM, Fountzilas E, Nikanjam M, Kurzrock R. Review of precision cancer medicine: Evolution of the treatment paradigm. Cancer Treat Rev. 2020;86:102019.
10. Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-2011: a population-based study. Lancet. 2015;385(9974):1206-1218.
11. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B, et al. Annual report to the nation on the status of cancer, 1975-2014, featuring survival. J Natl Cancer Inst. 2017;109(9):djx030.
12. Pellicano F, Mukherjee L, Holyoake TL. Concise review: cancer cells escape from oncogene addiction: understanding the mechanisms behind treatment failure for more effective targeting. Stem Cells. 2014;32(6):1373-1379.
13. Gerlinger M, Swanton C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. Br J Cancer. 2010;103(8):1139-1143.
14. Alexander S, Friedl P. Cancer invasion and resistance: interconnected processes of disease progression and therapy failure. Trends Mol Med. 2012;18(1):13-26.
15. Bukowinski AJ, Burns KC, Parsons K, Perentesis JP, O'Brien MM. Toxicity of cancer therapy in adolescents and young adults (AYAs). Semin Oncol Nurs. 2015;31(3):216-226.
16. Palmieri DJ, Carlino MS. Immune checkpoint inhibitor toxicity. Curr Oncol Rep. 2018;20(9):72.
17. Feliu J, Heredia-Soto V, Gironés R, Jiménez-Munarriz B, Saldaña J, Guillén-Ponce C, et al. Management of the toxicity of chemotherapy and targeted therapies in elderly cancer patients. Clin Transl Oncol. 2020;22(4):457-467.
18. Ma X, Edmonson M, Yergeau D, Muzny DM, Hampton OA, Rusch M, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun. 2015;6(1):6604.
19. Waanders E, Gu Z, Dobson SM, Antić Ž, Crawford JC, Ma X, et al. Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. Blood Cancer Discov. 2020;1(1):96-111.

20. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481(7382):506-510.
21. Murphy MFG, Bithell JF, Stiller CA, Kendall GM, O'Neill KA. Childhood and adult cancers: contrasts and commonalities. Maturitas. 2013;76(1):95-98.
22. Cunningham RM, Walton MA, Carter PM. The major causes of death in children and adolescents in the United States. The N Engl J Med. 2018;379(25):2468-2475.
23. Kampen KR. The discovery and early understanding of leukemia. Leuk Res. 2012;36(1):6-13.
24. Piller G. Leukaemia - a brief historical review from ancient times to 1950. Br J Haematol. 2001;112(2):282-292.
25. Simone JV. History of the treatment of childhood ALL: a paradigm for cancer cure. Best Pract Res Clin Haematol. 2006;19(2):353-359.
26. Pui CH, Evans WE. A 50-year journey to cure childhood acute lymphoblastic leukemia. Semin Hematol. 2013;50(3):185-196.
27. Zippin C, Cutler SJ, Reeves WJ, Jr., Lum D. Variation in survival among patients with acute lymphocytic leukemia. Blood. 1971;37(1):59-72.
28. Simone JV, Verzosa MS, Rudy JA. Initial features and prognosis in 363 children with acute lymphocytic leukemia. Cancer. 1975;36(6):2099-2108.
29. Propp S, Lizzi FA. Philadelphia chromosome in acute lymphocytic leukemia. Blood. 1970;36(3):353-360.
30. Secker-Walker LM, Lawler SD, Hardisty RM. Prognostic implications of chromosomal findings in acute lymphoblastic leukaemia at diagnosis. Br Med J. 1978;2(6151):1529-1530.
31. Bradstock KF, Janossy G, Tidman N, Papageorgiou ES, Prentice HG, Willoughby M, et al. Immunological monitoring of residual disease in treated thymic acute lymphoblastic leukaemia. Leuk Res. 1981;5(4-5):301-309.
32. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell. 2002;1(2):133-143.
33. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature. 2007;446(7137):758-764.
34. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. Nat Genet. 2009;41(9):1006-1010.
35. Treviño LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nat Genet. 2009;41(9):1001-1005.
36. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012;481(7380):157-163.
37. Carpiuc KT, Stephens JM, Botteman MF, Feng W, Hay JW. A review of the clinical and economic outcomes of imatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia. Expert Opin Pharmacother. 2007;8(16):2775-2787.
38. Hunger SP, Mullighan CG. Acute lymphoblastic leukemia in children. N Engl J Med. 2015;373(16):1541-1552.

39. Inaba H, Mullighan CG. Pediatric acute lymphoblastic leukemia. Haematologica. 2020;105(11):2524-2539.
40. Pieters R, Groot-Kruseman Hd, Velden VVd, Fiocco M, Berg Hvd, Bont Ed, et al. Successful therapy reduction and intensification for childhood acute lymphoblastic leukemia based on minimal residual disease monitoring: study ALL10 from the Dutch Childhood Oncology Group. J Clin Oncol. 2016;34(22):2591-2601.
41. Pui C-H, Yang JJ, Hunger SP, Pieters R, Schrappe M, Biondi A, et al. Childhood acute lymphoblastic leukemia: progress through collaboration. J Clin Oncol. 2015;33(27):2938-2948.
42. Brown P, Pieters R, Biondi A. How I treat infant leukemia. Blood. 2019;133(3):205-214.
43. Jeha S, Pei D, Choi J, Cheng C, Sandlund JT, Coustan-Smith E, et al. Improved CNS control of childhood acute lymphoblastic leukemia without cranial irradiation: St Jude total therapy study 16. J Clin Oncol. 2019;37(35):3377-3391.
44. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30.
45. Pieters R. Acute lymphoblastic leukaemia in children and adolescents: chance of cure now higher than 80%. Ned Tijdschr Geneeskd. 2010;154:A1577.
46. Mitchell C, Hall G, Clarke RT. Acute leukaemia in children: diagnosis and management. BMJ. 2009;338:b2285.
47. Vadillo E, Dorantes-Acosta E, Pelayo R, Schnoor M. T cell acute lymphoblastic leukemia (T-ALL): New insights into the cellular origins and infiltration mechanisms common and unique among hematologic malignancies. Blood Rev. 2018;32(1):36-51.
48. Shroff GS, Truong MT, Carter BW, Benveniste MF, Kanagal-Shamanna R, Rauch G, et al. Leukemic involvement in the thorax. Radiographics. 2019;39(1):44-61.
49. Chang CY, Chiou TJ, Hsieh YL, Cheng SN. Leukemic infiltration of the urinary bladder presenting as uncontrollable gross hematuria in a child with acute lymphoblastic leukemia. J Pediatr Hematol Oncol. 2003;25(9):735-739.
50. Pieper K, Grimbacher B, Eibel H. B-cell biology and development. J Allergy Clin Immunol. 2013;131(4):959-971.
51. Mosaad YM. Hematopoietic stem cells: an overview. Transfus Apher Sci. 2014;51(3):68-82.
52. Wilkinson AC, Göttgens B. Transcriptional regulation of haematopoietic stem cells. Adv Exp Med Biol. 2013;786:187-212.
53. LeBien TW, Tedder TF. B lymphocytes: how they develop and function. Blood. 2008;112(5):1570-1580.
54. Germain RN. T-cell development and the CD4-CD8 lineage decision. Nat Rev Immunol. 2002;2(5):309-322.
55. Naito T, Tanaka H, Naoe Y, Taniuchi I. Transcriptional control of T-cell development. Int Immunol. 2011;23(11):661-668.
56. Hosokawa H, Rothenberg EV. Cytokines, transcription factors, and the initiation of T-cell development. Cold Spring Harb Perspect Biol. 2018;10(5):a028621.
57. Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E, et al. Transcriptional diversity during lineage commitment of human blood progenitors. Science. 2014;345(6204):1251033.
58. Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, et al. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nat Immunol. 2010;11(7):635-643.

59. Somasundaram R, Prasad MAJ, Ungerbäck J, Sigvardsson M. Transcription factor networks in B-cell differentiation link development to acute lymphoid leukemia. Blood. 2015;126(2):144-152.
60. Bowman RL, Busque L, Levine RL. Clonal hematopoiesis and evolution to hematopoietic malignancies. Cell Stem Cell. 2018;22(2):157-170.
61. Iacobucci I, Mullighan CG. Genetic basis of acute lymphoblastic leukemia. J Clin Oncol. 2017;35(9):975-983.
62. Huang F-L, Liao E-C, Li C-L, Yen C-Y, Yu S-J. Pathogenesis of pediatric B-cell acute lymphoblastic leukemia: Molecular pathways and disease treatments. Oncol Lett. 2020;20(1):448-454.
63. Chan LN, Chen Z, Braas D, Lee J-W, Xiao G, Geng H, et al. Metabolic gatekeeper function of B-lymphoid transcription factors. Nature. 2017;542(7642):479-483.
64. Kuiper RP, Schoenmakers EF, van Reijmersdal SV, Hehir-Kwa JY, van Kessel AG, van Leeuwen FN, et al. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. Leukemia. 2007;21(6):1258-1266.
65. Kuiper RP, Waanders E, van der Velden VH, van Reijmersdal SV, Venkatachalam R, Scheijen B, et al. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. Leukemia. 2010;24(7):1258-1264.
66. van der Veer A, Waanders E, Pieters R, Willemse ME, Van Reijmersdal SV, Russell LJ, et al. Independent prognostic value of BCR-ABL1-like signature and IKZF1 deletion, but not high CRLF2 expression, in children with B-cell precursor ALL. Blood. 2013;122(15):2622-2629.
67. Hinze L, Möricke A, Zimmermann M, Junk S, Cario G, Dagdan E, et al. Prognostic impact of IKZF1 deletions in association with vincristine–dexamethasone pulses during maintenance treatment of childhood acute lymphoblastic leukemia on trial ALL-BFM 95. Leukemia. 2017;31(8):1840-1842.
68. Clappier E, Grardel N, Bakkus M, Rapion J, De Moerloose B, Kastner P, et al. IKZF1 deletion is an independent prognostic marker in childhood B-cell precursor acute lymphoblastic leukemia, and distinguishes patients benefiting from pulses during maintenance therapy: results of the EORTC Children's Leukemia Group study 58951. Leukemia. 2015;29(11):2154-2161.
69. Ofverholm I, Tran AN, Heyman M, Zachariadis V, Nordenskjöld M, Nordgren A, et al. Impact of IKZF1 deletions and PAX5 amplifications in pediatric B-cell precursor ALL treated according to NOPHO protocols. Leukemia. 2013;27(9):1936-1939.
70. Dörge P, Meissner B, Zimmermann M, Möricke A, Schrauder A, Bouquin JP, et al. IKZF1 deletion is an independent predictor of outcome in pediatric acute lymphoblastic leukemia treated according to the ALL-BFM 2000 protocol. Haematologica. 2013;98(3):428-432.
71. Asai D, Imamura T, Suenobu S, Saito A, Hasegawa D, Deguchi T, et al. IKZF1 deletion is associated with a poor outcome in pediatric B-cell precursor acute lymphoblastic leukemia in Japan. Cancer Med. 2013;2(3):412-419.
72. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009;360(5):470-480.
73. Yeoh AEJ, Lu Y, Chin WHN, Chiew EKH, Lim EH, Li Z, et al. Intensifying treatment of childhood B-lymphoblastic leukemia with IKZF1 deletion reduces relapse and improves overall survival: results of Malaysia-Singapore ALL 2010 study. J Clin Oncol. 2018;36(26):2726-2735.
74. Den Boer ML, van Slegtenhorst M, De Menezes RX, Cheok MH, Buijs-Gladdines JG, Peters ST, et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. Lancet Oncol. 2009;10(2):125-134.

75. Gu Z, Churchman M, Roberts K, Li Y, Liu Y, Harvey RC, et al. Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. Nat Commun. 2016;7:13331.
76. Hormann FM, Hoogkamer AQ, Beverloo HB, Boeree A, Dingjan I, Wattel MM, et al. NUTM1 is a recurrent fusion gene partner in B-cell precursor acute lymphoblastic leukemia associated with increased expression of genes on chromosome band 10p12.31-12.2. Haematologica. 2019;104(10):e455-e459.
77. Lilljebjörn H, Henningsson R, Hyrenius-Wittsten A, Olsson L, Orsmark-Pietras C, von Palffy S, et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. Nat Commun. 2016;7:11790.
78. Liu YF, Wang BY, Zhang WN, Huang JY, Li BS, Zhang M, et al. Genomic profiling of adult and pediatric B-cell acute lymphoblastic leukemia. EBioMedicine. 2016;8:173-183.
79. Gu Z, Churchman ML, Roberts KG, Moore I, Zhou X, Nakitandwe J, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. Nat Genet. 2019;51(2):296-307.
80. Pui CH. Precision medicine in acute lymphoblastic leukemia. Front Med. 2020;14(6):689-700.
81. Roberts KG, Reshmi SC, Harvey RC, Chen IM, Patel K, Stonerock E, et al. Genomic and outcome analyses of Ph-like ALL in NCI standard-risk patients: a report from the Children's Oncology Group. Blood. 2018;132(8):815-824.
82. Roberts KG, Pei D, Campana D, Payne-Turner D, Li Y, Cheng C, et al. Outcomes of children with BCR-ABL1–like acute lymphoblastic leukemia treated with risk-directed therapy based on the levels of minimal residual disease. J Clin Oncol. 2014;32(27):3012-3020.
83. Malinowska-Ozdowy K, Frech C, Schönegger A, Eckert C, Cazzaniga G, Stanulla M, et al. KRAS and CREBBP mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia. Leukemia. 2015;29(8):1656-1667.
84. Paulsson K, Lilljebjörn H, Biloglav A, Olsson L, Rissler M, Castor A, et al. The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. Nat Genet. 2015;47(6):672-676.
85. Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011;471(7337):235-239.
86. Jaffe JD, Wang Y, Chan HM, Zhang J, Huether R, Kryukov GV, et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nat Genet. 2013;45(11):1386-1391.
87. Li B, Brady SW, Ma X, Shen S, Zhang Y, Li Y, et al. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. Blood. 2020;135(1):41-55.
88. Meyer JA, Wang J, Hogan LE, Yang JJ, Dandekar S, Patel JP, et al. Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. Nat Genet. 2013;45(3):290-294.
89. Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. Lancet. 2013;381(9881):1943-1955.
90. Kato M, Manabe A. Treatment and biology of pediatric acute lymphoblastic leukemia. Pediatr Int. 2018;60(1):4-12.
91. Irving JA, Enshaei A, Parker CA, Sutton R, Kuiper RP, Erhorn A, et al. Integration of genetic and clinical risk factors improves prognostication in relapsed childhood B-cell precursor acute lymphoblastic leukemia. Blood. 2016;128(7):911-922.
92. Pieters R, De Lorenzo P, Ancliffe P, Aversa LA, Brethon B, Biondi A, et al. Outcome of infants younger than 1 year with acute lymphoblastic leukemia treated with the Interfant-06 protocol: results from an international phase III randomized study. J Clin Oncol. 2019;37(25):2246-2256.

93. Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the Children's Oncology Group. J Clin Oncol. 2012;30(14):1663-1669.
94. Cooper SL, Brown PA. Treatment of pediatric acute lymphoblastic leukemia. Pediatr Clin North Am. 2015;62(1):61-73.
95. Paul S, Rausch CR, Nasnas PE, Kantarjian H, Jabbour EJ. Treatment of relapsed/refractory acute lymphoblastic leukemia. Clin Adv Hematol Oncol. 2019;17(3):166-175.
96. Schwab C, Harrison CJ. Advances in B-cell precursor acute lymphoblastic leukemia genomics. HemaSphere. 2018;2(4):e53.
97. Bruzzi P, Bigi E, Predieri B, Bonvicini F, Cenciarelli V, Felici F, et al. Long-term effects on growth, development, and metabolism of ALL treatment in childhood. Expert Rev Endocrinol Metab. 2019;14(1):49-61.
98. Krull KR, Hardy KK, Kahalley LS, Schuitema I, Kesler SR. Neurocognitive outcomes and interventions in long-term survivors of childhood cancer. J Clin Oncol. 2018;36(21):2181-2189.
99. Ayala F, Dewar R, Kieran M, Kalluri R. Contribution of bone microenvironment to leukemogenesis and leukemia progression. Leukemia. 2009;23(12):2233-2241.
100. Gaudichon J, Jakobczyk H, Debaize L, Cousin E, Galibert M-D, Troadec M-B, et al. Mechanisms of extramedullary relapse in acute lymphoblastic leukemia: Reconciling biological concepts and clinical issues. Blood Rev. 2019;36:40-56.
101. Scharff BFSS, Modvig S, Marquart HV, Christensen C. Integrin-mediated adhesion and chemoresistance of acute lymphoblastic leukemia cells residing in the bone marrow or the central nervous system. Front Oncol. 2020;10:775.
102. Roehlecke C, Schmidt MHH. Tunneling nanotubes and tumor microtubes in cancer. Cancers. 2020;12(4):857.
103. de Rooij B, Polak R, Stalpers F, Pieters R, den Boer ML. Tunneling nanotubes facilitate autophagosome transfer in the leukemic niche. Leukemia. 2017;31(7):1651-1654.
104. Wang J, Liu X, Qiu Y, Shi Y, Cai J, Wang B, et al. Cell adhesion-mediated mitochondria transfer contributes to mesenchymal stem cell-induced chemoresistance on T cell acute lymphoblastic leukemia cells. J Hematol Oncol. 2018;11(1):11.
105. Antić Ž, Yu J, Van Reijmersdal SV, Van Dijk A, Dekker L, Segerink WH, et al. Multiclonal complexity of pediatric acute lymphoblastic leukemia and the prognostic relevance of subclonal mutations. Haematologica. 2021;106(12):3046-3055.
106. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science. 2008;322(5906):1377-1380.
107. Yang F, Brady SW, Tang C, Sun H, Du L, Barz MJ, et al. Chemotherapy and mismatch repair deficiency cooperate to fuel TP53 mutagenesis and ALL relapse. Nat Cancer. 2021;2(8):819-834.
108. Pui CH, Nichols KE, Yang JJ. Somatic and germline genomics in paediatric acute lymphoblastic leukaemia. Nat Rev Clin Oncol. 2019;16(4):227-240.
109. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature. 2011;469(7330):356-361.
110. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458(7239):719-724.

111. Merlo LMF, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. Nat Rev Cancer. 2006;6(12):924-935.

112. Kumar S, Warrell J, Li S, McGillivray PD, Meyerson W, Salichos L, et al. Passenger mutations in more than 2,500 cancer genomes: overall molecular functional impact and consequences. Cell. 2020;180(5):915-927.e916.

113. Tzoneva G, Perez-Garcia A, Carpenter Z, Khiabanian H, Tosello V, Allegretta M, et al. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. Nat Med. 2013;19(3):368-371.

114. Stoecklein NH, Hosch SB, Bezler M, Stern F, Hartmann CH, Vay C, et al. Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer. Cancer Cell. 2008;13(5):441-453.

115. Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, et al. Molecular definition of breast tumor heterogeneity. Cancer Cell. 2007;11(3):259-273.

116. Aubele M, Mattis A, Zitzelsberger H, Walch A, Kremer M, Hutzler P, et al. Intratumoral heterogeneity in breast carcinoma revealed by laser-microdissection and comparative genomic hybridization. Cancer Genet Cytogenet. 1999;110(2):94-102.

117. Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nat Genet. 2006;38(4):468-473.

118. Kashiwagi H, Uchida K. Genome-wide profiling of gene amplification and deletion in cancer. Hum Cell. 2000;13(3):135-141.

119. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575(7781):210-216.

120. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic evolution of breast cancer metastasis and relapse. Cancer Cell. 2017;32(2):169-184.e167.

121. Dobson SM, García-Prat L, Vanner RJ, Wintersinger J, Waanders E, Gu Z, et al. Relapse-fated latent diagnosis subclones in acute B lineage leukemia are drug tolerant and possess distinct metabolic programs. Cancer Discov. 2020;10(4):568-587.

122. Shlush LI, Mitchell A, Heisler L, Abelson S, Ng SWK, Trotman-Grant A, et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. Nature. 2017;547(7661):104-108.

123. Irving J, Matheson E, Minto L, Blair H, Case M, Halsey C, et al. Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. Blood. 2014;124(23):3420-3430.

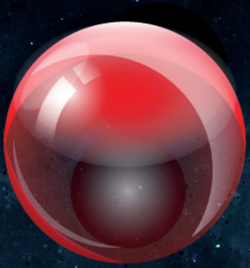
124. Jerchel IS, Hoogkamer AQ, Ariës IM, Steeghs EMP, Boer JM, Besselink NJM, et al. RAS pathway mutations as a predictive biomarker for treatment adaptation in pediatric B-cell precursor acute lymphoblastic leukemia. Leukemia. 2018;32(4):931-940.

125. Waanders E, Scheijen B, van der Meer LT, van Reijmersdal SV, van Emst L, Kroeze Y, et al. The origin and nature of tightly clustered BTG1 deletions in precursor B-cell acute lymphoblastic leukemia support a model of multiclonal evolution. PLoS Genet. 2012;8(2):e1002533.

126. Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet. 2014;46(2):116-125.

127. Kuiper RP, Waanders E. A RAG driver on the road to pediatric ALL. Nat Genet. 2014;46(2):96-98.

128. Robinson HM, Harrison CJ, Moorman AV, Chudoba I, Strefford JC. Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage-fusion-bridge cycle. Genes Chromosomes Cancer. 2007;46(4):318-326.
129. Li Y, Schwab C, Ryan SL, Papaemmanuil E, Robinson HM, Jacobs P, et al. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. Nature. 2014;508(7494):98-102.
130. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature. 2010;467(7319):1109-1113.
131. Murnane JP. Telomere dysfunction and chromosome instability. Mutat Res. 2012;730(1-2):28-36.
132. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-421.
133. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101.
134. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet. 2014;15(9):585-598.
135. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. Nature. 2018;555(7696):321-327.
136. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature. 2018;555(7696):371-376.



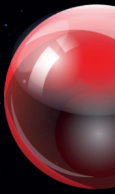
CHAPTER 2

ACCURATE DETECTION OF LOW-LEVEL MOSAIC MUTATIONS IN PEDIATRIC ACUTE LYMPHOBLASTIC LEUKEMIA USING SINGLE MOLECULE TAGGING AND DEEP-SEQUENCING

Jiangyan Yu[1,2], *Željko Antić*[1], *Simon V. van Reijmersdal*[1,2], *Alexander Hoischen*[2,3], *Edwin Sonneveld*[4], *Esmé Waanders*[1] *and Roland P. Kuiper*[1]

[1]Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands
[2]Department of Human Genetics, Radboud University Medical Center and Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands
[3]Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands
[4]Dutch Childhood Oncology Group, The Hague, The Netherlands

Leukemia & Lymphoma; July 2018; 59(7):1690-1699



ABSTRACT

Pathogenic mutations in relapse-associated genes in pediatric acute lymphoblastic leukemia may improve risk stratification when detected at subclonal levels at primary diagnosis. However, to detect subclonal mutations upfront, a deep-sequencing approach with high specificity and sensitivity is required. Here, we performed a proof-of-principle study to detect low-level mosaic RAS pathway mutations by deep sequencing using random tagging-based single molecule Molecular Inversion Probes (smMIPs). The smMIP-based approach could sensitively detect variants with allele frequency as low as 0.4%, which could all be confirmed by other techniques. In comparison, with standard deep-sequencing techniques we reached a detection threshold of only 2.5%, which hampered detection of 7 low-level mosaic mutations representing 24% of all detected mutations. We conclude that smMIP-based deep-sequencing outperforms standard deep-sequencing techniques by showing lower background noise and high specificity, and is the preferred technology for detecting mutations upfront, particularly in genes in which mutations show limited clustering in hotspots.

INTRODUCTION

Acute lymphoblastic leukemia (ALL) is the most common cancer in children. Although long-term overall survival reaches 90% in most Western countries, relapse occurs in at least 10% of children with ALL[1,2]. Recent studies have demonstrated that relapses in pediatric ALL can arise from minor subclones present at diagnosis and genetic alterations identified at relapse can be traced back to subclonal levels in samples taken at diagnosis, sometimes in just a few percent of the tumor cells[3-8]. This raises the interesting possibility to use subclonal mutations in relapse-associated genes for upfront risk stratification at time of diagnosis and during treatment. To determine the prognostic value for risk of relapse of subclonal mutations in candidate genes, unselected cohorts of diagnosis samples need to be screened for (low-level) mosaic mutations.

Next-generation sequencing (NGS) provides the opportunity to rapidly sequence genes of interest at high read depth and commonly uses PCR-based strategies for targeted enrichment. The error rate introduced by most sequencing platforms as well as by the DNA polymerase during amplification is reasonably low and can be compensated by creating more sequencing depth. However, when applying this approach for the detection of low-level mosaic mutations, sensitivity and accuracy of variant calling is challenged[9,10]. Particularly when applying prospective screening of subclonal mutations, where mutation status and location is not known in advance, the reliable detection of mutations is problematic and validations are demanding. Therefore, it is essential to use an approach that minimizes sequencing errors and accurately quantifies subclonal burden of targeted genes of interests.

Molecular Inversion Probes (MIPs) have been used to perform targeted deep-sequencing using multiplex targeted capture of selected genomic regions of interest[11]. Recently, this technology was adapted with single molecule tagging to enable the reliable detection of low-level mosaic mutations. This so-called single molecule Molecular Inversion Probe (smMIP) technology allows the clustering of reads amplified from the same template DNA molecule, which is subsequently used to generate single molecule consensus (smc) reads[12]. Variant calls derived from these smc-reads contain much lower per-base substitution error rates, hence creating higher specificity for detecting low variant allele frequency (VAF) mutations in targeted genes[12]. As such, this approach may be highly suitable for determination of clinically relevant low-level mosaic mutations in leukemia samples.

In this study, we aimed to develop a smMIP-based NGS assay with a lowest possible false-positive rate to perform quantitative detection of subclonal mutations in ALL diagnosis samples. As a pilot study, we made use of the fact that mutations in RAS pathway genes are common in relapsed ALL, and frequently originate from a minor or major clone at initial diagnosis. We applied the smMIP approach to detect subclonal *KRAS*, *NRAS* and *PTPN11* mutations in 21 ALL diagnosis samples from patients with mutations in these genes at time of relapse. We evaluated the performance of the smMIP approach, and assessed its specificity. Our data indicate that smMIP-based NGS is able to accurately detect low-level mosaic mutations in leukemia samples down to a level of 0.4%.

MATERIAL AND METHODS

Patient materials and DNA isolation

A total of 21 primary diagnosis ALL samples from children who later developed relapse were included in this study. These patients carried previously identified mutations in *KRAS* ($n = 11$), *NRAS* ($n = 8$) and *PTPN11* ($n = 2$) at the time of relapse (**Table S2.1**). Genomic DNA of the diagnosis sample was extracted from mononuclear cells and purified using a QIAamp purification kit (Qiagen, Valencia, CA, USA), followed by sequencing of targeted amplicons in *KRAS*, *NRAS* or *PTPN11* using two alternative deep sequencing approaches.

smMIP-based sequencing

The smMIP technology (**Figure 2.1**) was applied according to a previously published protocol.[12] These smMIP probes carry two 5-nt stretches of degenerate bases, which offers the theoretical possibility to distinguish over 1 million captured molecules (**Figure 2.1A**). The smMIP probes targeting exons in RAS pathway genes that carry hotspot mutations were designed (**Figure 2.1B** and **Table S2.2**), ordered (Integrated DNA Technologies, Leuven, Belgium), pooled and phosphorylated as previously described[12]. The genomic regions of interest were captured using 100 ng of genomic DNA and 43.8 picomole of each probe (resulting 800 smMIP molecules per DNA molecule) for each reaction (Steps 1-3 in **Figure 2.1C**). Subsequently, the uniquely-tagged circular molecules were amplified in the presence of sample-specific barcodes, using PCR containing 23 cycles of 10 seconds at 98°C, 30 seconds at 60°C and 30 seconds at 72°C (Steps 4-5 in **Figure 2.1C**). PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter Genomic, Brea, CA, USA). Paired-end sequencing (2x150 bp) was performed using the Illumina NextSeq 500 Desktop Sequencer (Illumina, San Diego, CA, USA; Step 6 in **Figure 2.1C**). Reads were mapped to the human reference genome (hg19) using SeqNext software (v4.2.5, JSI, Ettenheim, Germany).

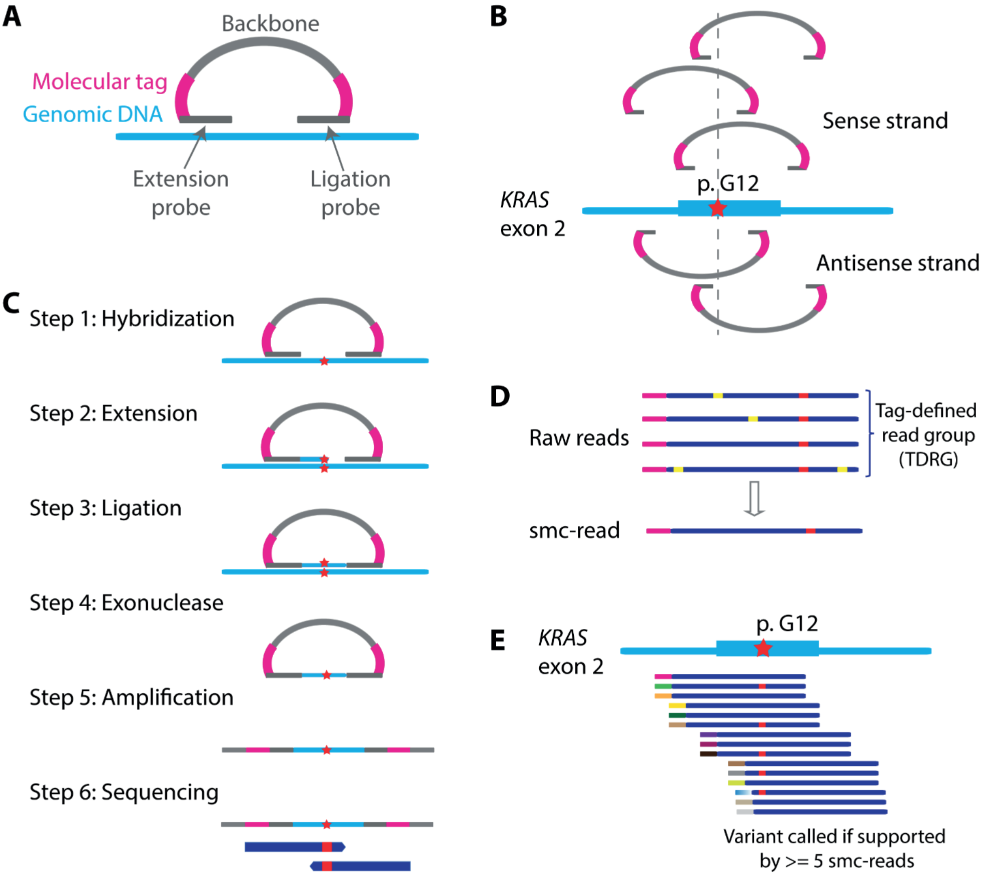

A
Backbone
Molecular tag
Genomic DNA
Extension probe
Ligation probe
B
Sense strand
p. G12
KRAS exon 2
Antisense strand
C
Step 1: Hybridization
Step 2: Extension
Step 3: Ligation
Step 4: Exonuclease
Step 5: Amplification
Step 6: Sequencing
D
Raw reads
Tag-defined read group (TDRG)
smc-read
E
KRAS exon 2
p. G12
Variant called if supported by >= 5 smc-reads

Figure 2.1. Overview of the smMIP approach[12]. **(A)** A smMIP consists of two 16-24 nucleotides (nt) targeting arms, the extension and the ligation probe, joined by a constant 30-nt backbone. Each smMIP molecule contains two unique random 5 nt degenerate molecular tags, which could theoretically distinguish over 1 million captured molecules. **(B)** smMIPs are designed to target both the sense and the antisense DNA strand. **(C)** Schematic of the smMIP approach. Phosphorylated smMIP probes hybridize to the targeted genomic region. The gap is filled through extension and ligation reactions. After exonuclease treatment, only circularized smMIP probes are left, which are subsequently amplified by PCR and sequenced. **(D)** Schematic of the formation of a single molecule consensus read (smc-read) from a tag-defined read group (TDRG). Raw reads from the same probe and sharing the same molecule tag form a TDRG in which random errors (yellow) are filtered out and true variants (red) are kept based on the presence in >70% of the raw reads. **(E)** The smc-reads are used to call variants. Variants are called if they are supported by at least 5 smc-reads.

In order to evaluate the efficiency of random tags in eliminating amplification and sequencing errors during data analysis, we analyzed sequencing data in two alternative ways: (i) a standard variant calling setting using raw sequencing reads (after trimming the 10-nt random tags) and, (ii) a smMIP-based consensus variant calling setting that makes use of the random tags to generate smc-reads. In the first method, sequence reads were mapped to the human reference genome (hg19) and analyzed using SeqNext software (v4.2.5, JSI, Ettenheim, Germany). Variants were called when present in at least 200 raw reads with a VAF>0.5%. In the latter method, the raw sequencing reads that align to the same genomic region and share the unique molecular tag form a tag-defined read group (TDRG; **Figures 2.1D** and **S2.1**). Each TDRG composed of at least 5 raw sequencing reads was used to form a single molecule consensus read (smc-read). Within a TDRG, a variant was established to be valid (and was thus maintained in the smc-read) when 70% of the raw reads contained the same nucleotide substitution. Other variants were determined random errors and were excluded from the analysis. The smc-reads were used to call variants occurring in the samples. Variants were called when present in at least 5 smc-reads (**Figure 2.1E**).

Ampliseq-based deep sequencing

As an alternative method for deep sequencing, we used a previously developed Ion Ampliseq sequencing panel[13]. Primers targeting the three genes were designed using Ion Ampliseq designer (version 1.2.9, Life Technologies, Carlsbad, CA, USA; **Table S2.3**). To minimize amplification bias, four independent PCR reactions with each 15 ng of genomic DNA as input (60 ng in total) were performed for each amplicon. After barcoding and adaptor ligation, products were purified with Agencourt AMPure XP beads (Beckman Coulter Genomics, Brea, CA, USA). Emulsion PCR was performed using OneTouch 200 Template kit (Life Technologies, Carlsbad, CA, USA). Polymerase and sequencing primers were added before loading onto the chip. Sequencing was performed on the Ion Personal Genome Machine (PGM) System (Life Technologies, Carlsbad, CA, USA). Sequence reads were mapped to the human reference genome (hg19) and analyzed using SeqNext software (v4.2.5, JSI, Ettenheim, Germany). Variants were called when present in at least 200 raw reads with a VAF>0.5%.

Sanger sequencing

Sanger sequencing was performed in order to validate one mutation (*KRAS* p.G12A) in patient P0098 observed by the smMIP-based consensus variant calling approach (VAF = 16%), but missed by standard deep sequencing approaches. The following PCR primers were used for amplification and sequencing: forward 5'-TCATGAAAATGGTCAGAGAAACC-3' and reverse 5'-AAAAGGTACTGGTGGAGTATTTGA-3'.

RNase H2-based PCR

In order to validate low-level mosaic variants identified by the smMIP-based consensus variant calling approach, RNase H2-based PCR (rhPCR) assays were performed (**Figure S2.2**)[14]. The allele-specific primers (Integrated DNA Technologies, Leuven, Belgium) contain the mutant RNA base at position -6 and a 3' mismatch that prevents extension during PCR (**Figure S2.2A** and **Table S2.4**). Extension can only occur when the DNA-bound primer is cleaved by the RNase H2 enzyme to remove the 3' mismatched sequence (**Figure S2.2B**). PCR was performed in 10 µl reactions, each containing 10 ng of genomic DNA, 200 nM of each primer, 5ul 2x Bio-Rad iQ SYBR Green Supermix (Bio-Rad, Hercules, CA, USA), and 200 mU RNase H2 enzyme (Integrated DNA Technologies, Leuven, Belgium). Following 2 minutes at 95°C, the PCR included 40 cycles of 10 seconds at 95°C and 30 seconds at the optimal annealing temperature for each primer pair (range 60-65°C). PCR products were visualized on a 2% gel or the Agilent 4200 Tapestation system (Agilent, Santa Clara, CA, USA).

RESULTS

Mutations in the RAS pathway genes *KRAS*, *NRAS* and *PTPN11* are relatively common in relapsed ALL, and frequently originate from a minor or major clone at initial diagnosis[5,15,16]. To increase the chance to detect low-level mosaic mutations, we made use of this phenomenon and selected 21 ALL diagnosis samples in which we previously identified mutations in one of these RAS genes in the major clones at relapse (**Table S2.1**). Next to the smMIP-based enrichment of the mutation hotspot regions in these RAS pathway genes, we used for comparison an Ampliseq enrichment panel that we previously designed for the same regions. Subsequently, we evaluated the performance of low frequency variant calling between routine variant calling pipelines applied to both datasets, and a sensitive consensus-based variant calling strategy.

Variant calling from deep sequencing raw reads

The smMIP-based and Ampliseq-based amplicons carry linkers for Illumina and IonTorrent PGM sequencing platforms, respectively. Although both platforms sequence DNA by monitoring the addition of nucleotides during DNA synthesis, they use different chemistry, which may result in differences in sequencing performance. Therefore, we evaluated the overall performance of the two sequencing strategies, by first applying similar analysis and variant calling pipelines to the two raw datasets. Average read depth around the mutation hotspot regions was 50,526 for the Ampliseq-based amplification and 103,763 for the smMIP-based amplification (**Figure S2.3**). To reveal the level of noise we analyzed the results obtained after sequencing of two DNA samples from healthy donors and defined that, for both methods, 99.8% of the called variants had a VAF<0.5%, and that the number of called variants raised rapidly below a VAF of 0.5%, suggesting that these variants predominantly represent sequence errors (**Figure S2.4**). Based on this observation, we only considered variant calls in the leukemia samples with VAF>0.5%.

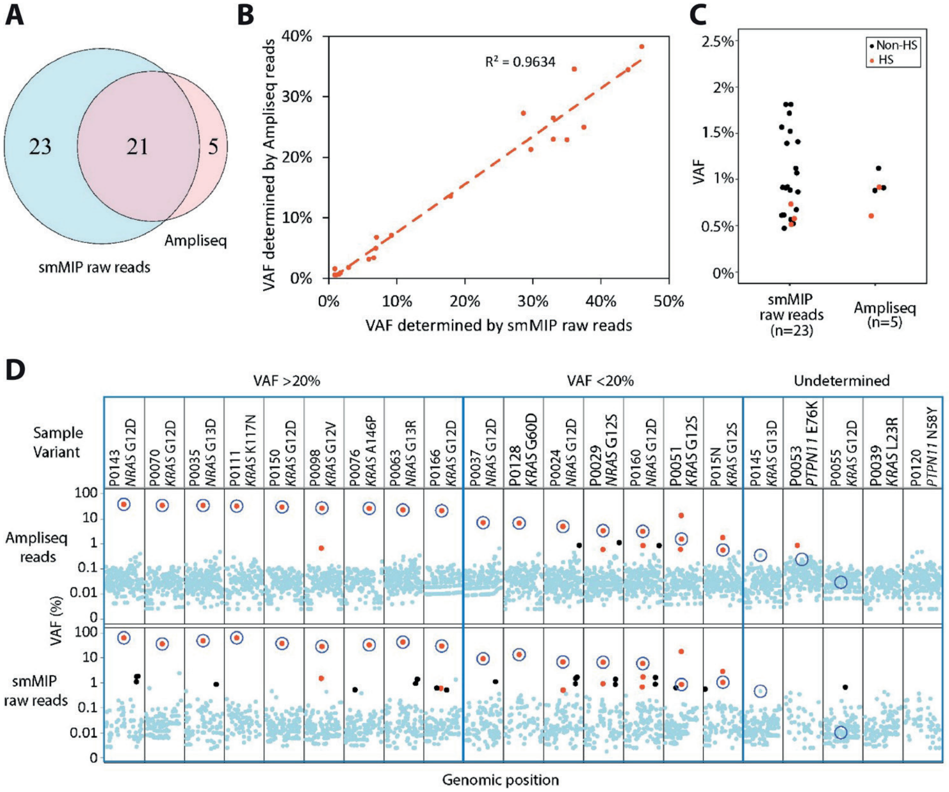


Figure 2.2. Overview of RAS mutations detected in the diagnosis samples. **(A)** The total number of SNVs observed by Ampliseq and smMIP raw reads. **(B)** The VAF of 21 variants determined by the two approaches correlates well. All of these variants were located in hotspot regions. **(C)** VAF of variants that were only identified by one method. Out of 28 variants, five were located in hotspot regions (HS; Orange). **(D)** RAS mutations detected in the diagnosis samples by analysis of Ampliseq reads and smMIP raw reads. Top panel shows sample IDs with the RAS pathway mutation detected in the relapse sample. All called variants (before filtering) are depicted as dots, ranked by their genomic position in each vertical panel. Variants that did not pass filtering are colored in blue, while variants that passed the detection limit are colored in orange (known hotspot mutation) or black (non-hotspot mutation). The cut off used for calling of the variants using Ampliseq and smMIP raw reads analysis was VAF >0.5% and >200 variant reads. We detected 26 variants using Ampliseq and 44 variants using smMIP raw reads. Blue circles mark variants previously identified in the paired relapse sample (**Table S2.1**). Details of all identified mutations are displayed in **Table S2.5**.

In the 21 diagnosis samples, a total of 26 single nucleotide variants (SNVs) were called in the Ampliseq reads, and 44 SNVs were found in the smMIP raw reads (**Table S2.5**). Of these variants, 21 were identified by both approaches with highly comparable allele frequencies (**Figures 2.2A** and **2.2B**), all of which were located in mutation hotspots as expected for RAS pathway mutations (**Tables S2.5** and **S2.6**). Discordancy between the two methods was exclusively observed for variants called in the low-mosaic range (VAF<2.5%; **Figure 2.2C**), and the majority of these were located outside the hotspots (**Figure 2.2C**), indicating high false-positive rates in Ampliseq reads (3/5; 60%) and smMIP raw reads (20/23; 87%). Furthermore, none of these variants reappeared in the relapse sample, whereas 16 variants that were identified by both approaches, including 7 with low VAF at time of diagnosis (range 0.8-13%), were present at relapse (**Figure 2.2D**). Collectively, we conclude that the two methods perform similarly in detecting single nucleotide substitutions that are present in at least 5% of the leukemic cells (VAF>2.5%), but generate many variants with lower confidence below this threshold.

smMIP-based consensus variant calling

The smMIP raw sequence reads carry random single-molecule tags that allow an additional analysis step that corrects for post-capture amplification bias and sequence errors by generating single molecule consensus (smc) reads (**Figures 2.1** and **S2.1**). This provides the opportunity to obtain lower numbers of false-positive variant calls and better estimates of clonality. Smc-reads had an average 142 underling raw smMIP reads (range 105-174; **Figure S2.5A**) and each probe, was covered with an average of 336 smc-reads (**Figure S2.5B**). Up to 75% of the targeted genomic regions was covered with at least 2 independent probes, mostly in two orientations. All hotspot regions were targeted with more than 3 probes, reaching a total average smc-read coverage of 1,432x (range 512-2,284x; **Figure S2.6**). Variants were included only if they were supported by at least 5 smc-reads (**Figure 2.1E**).

We assessed whether this smMIP-based consensus calling approach could sensitively determine the presence of RAS mutations at primary diagnosis. The total number of variants detected in smc-reads (301 variants) was drastically lower compared to Ampliseq (36,391) and smMIP raw reads (4,258). In total, 29 variants in *NRAS* (*n* = 13), *KRAS* (*n* = 13) and *PTPN11* (*n* = 3) passed filtering, of which 2 were with a VAF below 0.5% (**Table 2.1** and **Figure 2.3A**). All mutations were located in known hotspot regions in these genes[5,15,17]. Two samples were determined to carry 3 hotspot mutations in *KRAS* (p.G12A, p.G12V, p.G13D) and *NRAS* (p.G12S, p.G12D, p.G13D), respectively. For each of these two samples, one hotspot mutation was detected as major clone at time of relapse (Cases P0098 and P0160; **Table 2.1** and **Figure 2.3A**). Of the 21 mutations identified at relapse, 18 could be detected in the paired diagnosis samples, including 2 mutations with a low VAF (**Figure S2.7**). This indicates that the smMIP-based consensus calling approach could sensitively detect low-level mosaic mutations.

To confirm low-level mosaic variants called by smc-reads we applied allele-specific RNase H2-based PCR (rhPCR), a cost-effective mutation-specific amplification strategy (**Figure S2.2**)[14]. We analyzed eight low-level mosaic variants including all variants with VAF<1% and a variant with a low VAF (1.8%) uniquely detected by the smMIP-based consensus approach. Each variant assay was highly specific and did not show cross reactivity for variants located in adjacent codons (**Figures S2.8A** and **S2.8B** and data not shown). The rhPCR technique confirmed all eight variants, including five variants with an allele frequency of ≤0.6% (**Figures 2.3B** and **S2.8C**).

We next evaluated whether the 29 variants called by the smMIP-based consensus approach were also found by the above-mentioned strategies, which was the case for 21 mutations (72%). One variant was missed by the smMIP raw read analysis, two were missed by the Ampliseq-based sequencing analysis and 5 variants were not detected by either of the two approaches (**Table 2.1**). One of these variants was a *KRAS* p.G12A variant with a VAF of 16%. This, however, was most likely caused by the fact that this variant overlapped with another *KRAS* mutation at the same position (p.G12V; VAF = 32%). Both variants were confirmed by Sanger sequencing (**Figure S2.9**). Interestingly, of the 13 variants with VAF<2.5%, 6 and 5 were missed by Ampliseq and smMIP raw read analysis pipelines, respectively (**Figures 2.3C** and **S2.10**). Furthermore, whereas all variants detected by the smMIP-based consensus approach were hotspot mutations, 23 of the 25 low-frequency variants only detected by the standard deep sequencing techniques were located outside hotspot regions (**Figures 2.3C** and **S2.10**), suggesting that they mainly represent false-positives. However, since validation of these 25 variants was not performed, they may include some true variants that have been excluded by our stringent filtering approach, and should thus be considered as false-negatives of our smMIP-based consensus approach. Taken together, we conclude that the random-tagged smMIP approach outperforms standard deep sequencing techniques by detecting more low-level mosaic mutations and less false-positive calls.

Table 2.1. Variants called from smMIP smc-reads

Sample	Gene	Variant	Amino acid change	Number of smc-reads supporting the variant	VAF (%)[1]	Present in major clone at relapse	Detected by Ampliseq and/ or smMIP raw reads	Confirmed by other technique
P0053	*PTPN11*	c.226G>A	p.E76K	7	0.4	Yes	No	rhPCR
P0120	*PTPN11*	c.205G>A	p.E69K	6	0.4	No	No	rhPCR
P0160	*NRAS*	c.38G>A	p.G13D	7	0.6	No	Yes (smMIP raw)	rhPCR
P0024	*NRAS*	c.34G>T	p.G12C	6	0.6	No	Yes (smMIP raw)	rhPCR
P0145	*KRAS*	c.38G>A	p.G13D	6	0.6	Yes	No	rhPCR
P0029	*NRAS*	c.35G>A	p.G12D	9	0.8	No	Yes (both)	rhPCR
P0053	*PTPN11*	c.205G>A	p.E69K	13	0.9	No	Yes (Ampliseq)	rhPCR
P0151	*NRAS*	c.34G>A	p.G12S	14	1.0	Yes	Yes (both)	no
P0098	*KRAS*	c.38G>A	p.G13D	12	1.3	No	Yes (both)	no
P0160	*NRAS*	c.34G>A	p.G12S	21	1.5	No	Yes (both)	no
P0166	*KRAS*	c.35G>T	p.G12V	17	1.8	No	No	rhPCR
P0151	*NRAS*	c.35G>A	p.G12D	29	2.2	No	Yes (both)	no
P0051	*KRAS*	c.34G>A	p.G12S	13	2.4	Yes	Yes (both)	no
P0160	*NRAS*	c.35G>A	p.G12D	59	4.3	Yes	Yes (both)	no
P0029	*NRAS*	c.34G>A	p.G12S	69	5.7	Yes	Yes (both)	no
P0024	*NRAS*	c.35G>A	p.G12D	73	6.4	Yes	Yes (both)	no
P0128	*KRAS*	c.179G>A	p.G60D	91	6.5	Yes	Yes (both)	no
P0037	*NRAS*	c.35G>A	p.G12D	60	8.8	Yes	Yes (both)	no
P0098	*KRAS*	c.35G>C	p.G12A	152	16.1	No	No	Sanger
P0051	*KRAS*	c.35G>A	p.G12D	109	19.0	No	Yes (both)	no
P0166	*KRAS*	c.35G>A	p.G12D	295	28.4	Yes	Yes (both)	no
P0098	*KRAS*	c.35G>T	p.G12V	302	31.9	Yes	Yes (both)	no
P0111	*KRAS*	c.351A>T	p.K117N	608	33.1	Yes	Yes (both)	no
P0076	*KRAS*	c.436G>C	p.A146P	218	33.3	Yes	Yes (both)	no
P0150	*KRAS*	c.35G>A	p.G12D	562	34.7	Yes	Yes (both)	no
P0063	*NRAS*	c.37G>C	p.G13R	386	35.5	Yes	Yes (both)	no
P0070	*KRAS*	c.35G>A	p.G12D	305	40.3	Yes	Yes (both)	no
P0035	*NRAS*	c.38G>A	p.G13D	527	42.3	Yes	Yes (both)	no
P0143	*NRAS*	c.35G>A	p.G12D	601	45.4	Yes	Yes (both)	no

[1] VAF: variant allele frequency

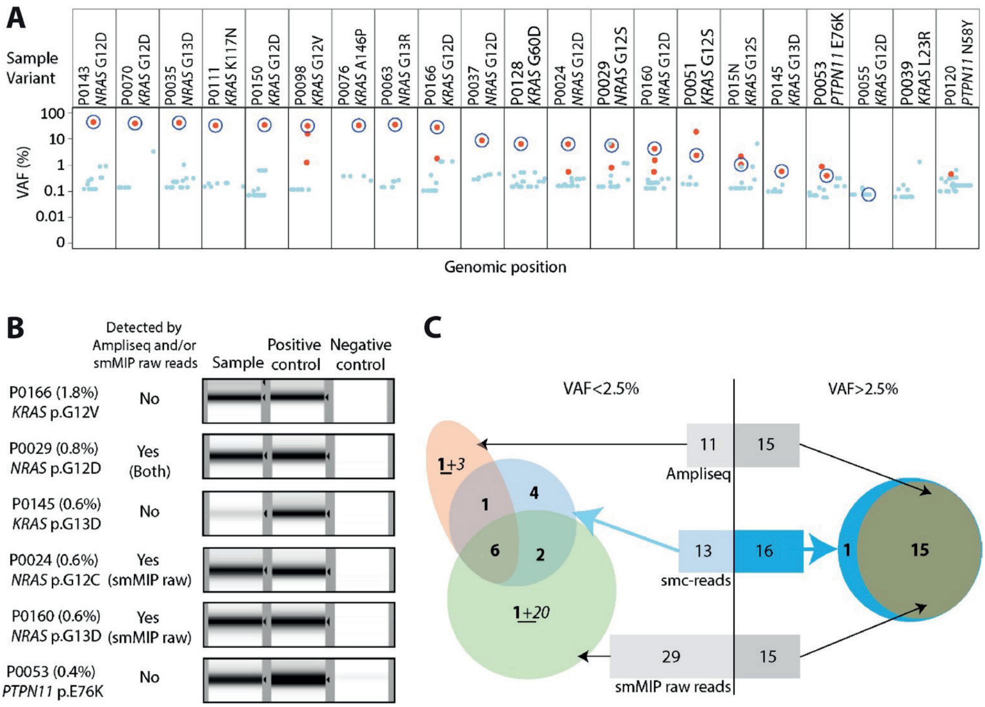


Figure 2.3. Overview of RAS mutations detected in paired diagnosis samples by smMIP single molecule consensus reads (smc-reads). **(A)** A total of 29 mutations and a dramatically decreased background were detected in the diagnosis samples by consensus-based variant calling using smc-reads. The top panel shows sample IDs with the RAS pathway mutation detected in the relapse sample. All called variants (before filtering) are depicted as dots, ranked by their genomic position in each vertical panel. Variants that did not pass filtering are colored in blue, while variants that passed the detection limit are colored in orange (known hotspot mutation) or black (non-hotspot mutation). The cut off used was presence in >= 5 smc-reads. Blue circles mark variants previously identified in the paired relapse sample (**Table S2.1**). Details of all identified mutations are displayed in **Table S2.5**. **(B)** Variants with a low VAF identified by smMIP consensus variant calling were confirmed by RNase H2 based PCR (rhPCR). The black arrow on the right of the gel indicates the expected amplicon size. **(C)** Comparison of the variants detected by smMIP based consensus calling approach and standard deep sequencing approaches. To achieve this, we divided all mutations identified from these approaches into two groups according to the VAF. Comparison of variants with a VAF<2.5% and VAF>2.5% are shown in Venn diagrams on the left and right. The number of hotspot mutations are shown in bold and non-hotspot mutations are underlined and italic. Details for variants are displayed in **Figure S2.10**.

DISCUSSION

In this study, we have developed a smMIP panel to sensitively and accurately detect low-level mosaic mutations in leukemia samples. We compared the widely applied deep sequencing strategy with a smMIP-based approach, which generates consensus reads for variant calling. To test these approaches, we chose to target three commonly mutated RAS pathway genes. Driver mutations in these genes cluster in hotspots, which facilitates the discrimination of true mutations from background noise. All methods had high sensitivity rates, making them suitable for backtracking of relapse-associated mutations. However, variant calling on smMIP-based consensus reads resulted in drastically reduced false-positive rates, thereby creating much higher specificity in determining the presence of subclonal mutations in primary diagnosis samples from ALL patients, irrespective of the applied sequence depth. Therefore, this smMIP approach is very suitable for upfront detection of subclonal mutations in ALL in an unbiased manner, particularly in genes where mutations are more randomly distributed.

The reliable detection of low-level mosaic mutations by NGS sequencing is hampered by the occurrence of amplification-based errors that arise during the preparation and sequencing processes[9,10]. We here demonstrate that common deep sequencing approaches can reliably estimate the VAF of variants, but cannot correct for random errors resulting in high error rates. Even with very high sequencing depth (>50,000x), the common NGS approach could not accurately determine mosaic mutations with a VAF below 2.5%. In comparison, the smMIP approach uses a unique tag in each molecule with which we could eliminate random errors. This results in a notable reduction of background noise and thus allows for the sensitive detection of mutations in leukemia samples with VAF as low as 0.4%.

We experienced a relatively low capture efficiency with our smMIP-based approach, which leaves room for improvement. Using 100 ng genomic DNA as input (which equals approximately ~15,000 haploid copies) and an average smc-read coverage of 1,432x in the hotspot regions (which represents 1,432 unique DNA molecules in the patient sample), we detected around 10% of total input, which creates a limit to the detection of very low-level mosaic alterations (VAF>0.4% when using a 5-smc read threshold). Actual capture efficiency may be higher, because inefficiently amplified fragments may have been missed with sequencing. This can be easily optimized by deeper sequencing to reach saturation of the capture. In addition, the efficacy can be further improved by increasing the DNA input during capture, and by applying multiple PCR reactions in parallel. Such improvements will directly improve detection variants below 0.4%.

For further increase in sensitivity, towards levels that enable the detection of minimal residual disease detection (i.e., the presence of residual tumor cells <0.01% during or after treatment), alternative strategies are required like duplex sequencing or droplet PCR[18,19].

A problem with prospective screening of subclonal mutations is that they are more difficult to validate. This is particularly problematic in genes, such as *TP53* and *CREBBP*, where mutations are scattered throughout the coding sequence, because the region where mutations are expected is much larger. In this study, we performed allele-specific rhPCR to validate variants with low VAF, a cost-effective technique with improved selectivity between the mutant and wild-type allele. In addition, we made use of the fact that mutations in RAS pathway genes specifically occur in hotspot regions, which provides an alternative and indirect way to discriminate between true mutations and noise. The combined use of these two approaches allowed us to establish that all variants detected by the smMIP-based consensus method between 0.4-1% VAF are indeed true. Therefore, with the smMIP-based consensus method and stringent filtering process, we can reduce the number of false-positive calls to nearly zero, which largely decreases time and costs for validation. The consequence of the stringent filtering process is that some true variants might be missed. In fact, among the 25 variants detected by the standard deep-sequencing strategies that were missed or excluded by filtering in the smMIP-based consensus approach, may be true variants present in very small subclones. Therefore, it is crucial to establish the desired sensitivity of low-level variants and determine the required smMIP sequencing depth. Together, we believe that smMIP based sequencing can be applied to detect subclonal mutations in a prospective manner, also in genes like *TP53* and *CREBBP*, where true pathogenic mutations show much less clustering.

Common alterations identified in relapsed ALL are frequently present in minor or major clones at initial diagnosis. This raises questions about the impact of subclonal alterations at the time of initial diagnosis and their prognostic significance on the risk of relapse. Prospective screening of leukemia-associated alterations in large unselected ALL cohorts can reveal their prognostic relevance. However, this requires a technique that can accurately and reliably determine low-level mosaic mutations with low rates of false-positives. Since all low-level mosaic variants, with VAFs down to 0.4%, could be confirmed independently by other techniques, the smMIP-based consensus variant calling approach is highly suitable to study the prognostic value of (sub)clonal alterations identified at the time of diagnosis.

In clinical practice, it is important to track multiple genetic markers in follow-up samples to monitor leukemia development. Commonly used sensitive techniques like quantitative real-time PCR and digital droplet PCR are limited in use due to obstacles for multiplexing in one experiment[20]. In contrast, a single smMIP panel is capable of sensitively and accurately identifying subclonal mutations and copy number variants in a large set of cancer genes[21-23], as was recently demonstrated for leukemia-associated mutations[24,25]. In addition, probes for new interesting regions can easily be added into an optimized smMIP panel. smMIP offers a scalable and rapid workflow to track genetic alterations in leukemia samples and monitor clonal changes during leukemia development in clinical practice.

In conclusion, we have shown that the random tagging based smMIP approach is sensitive and can accurately identify subclonal mutations in leukemia samples. Our data indicate that the smMIP approach can eliminate mistakes that likely occur during PCR and sequencing, thereby drastically reducing the background noise. The technique provides opportunities to study prognostic value of low-level mosaic mutations in relapse development, and monitor leukemia development in routine clinical practice.

SUPPLEMENTARY DATA

Supplementary Tables

Table S2.1. RAS mutations identified in relapse samples

Sample ID	Gene	Variant	Amino acid change
P0024	*NRAS*	c.35G>A	p.G12D
P0029	*NRAS*	c.34G>A	p.G12S
P0035	*NRAS*	c.38G>A	p.G13D
P0037	*NRAS*	c.35G>A	p.G12D
P0039	*KRAS*	c.68T>G	p.L23R
P0051	*KRAS*	c.34G>A	p.G12S
P0053	*PTPN11*	c.226G>A	p.E76K
P0055	*KRAS*	c.35G>A	p.G12D
P0063	*NRAS*	c.37G>C	p.G13R
P0070	*KRAS*	c.35G>A	p.G12D
P0076	*KRAS*	c.436G>C	p.A146P
P0098	*KRAS*	c.35G>T	p.G12V
P0111	*KRAS*	c.351A>T	p.K117N
P0120	*PTPN11*	c.172A>T	p.N58Y
P0128	*KRAS*	c.179G>A	p.G60D
P0143	*NRAS*	c.35G>A	p.G12D
P0145	*KRAS*	c.38G>A	p.G13D
P0150	*KRAS*	c.35G>A	p.G12D
P0151	*NRAS*	c.34G>A	p.G12S
P0160	*NRAS*	c.35G>A	p.G12D
P0166	*KRAS*	c.35G>A	p.G12D

Table S2.2. smMIP sequences

Gene	Exon	Chromosome	hg19 start	hg19 stop	smMIP probe sequence*
NRAS	2	1	115258658	115258769	gtactcagtcatttcacaccagcNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtcaggtcagcgggctac
NRAS	2	1	115258731	115258842	cagtactttaaagctttctataatNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNgctggattgtcagtgc
NRAS	2	1	115258683	115258794	cccactatagaggtgaggccNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtactggtttccaacaggttc
NRAS	2	1	115258683	115258794	cccaccatagaggtgaggccNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtactggtttccaacaggttc
NRAS	2	1	115258717	115258828	gctaatccagaaccactttgtagaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtgtagatgtggctcgc
KRAS	2	12	25398181	25398292	actaccacaagattatattcagtcNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNaatggtcctgcaccag

Table S2.2. Continued

Gene	Exon	Chromosome	hg19 start	hg19 stop	smMIP probe sequence*
KRAS	2	12	25398181	25398292	actaccacaagtttatattcagtcNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNaatggtcctgcaccag
KRAS	2	12	25398200	25398311	cagtcattttcagcaggcctNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNgtaatatgcatattaaaaca
KRAS	2	12	25398261	25398372	cacacataaggttaatacacNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNattctgaattagctgtatcg
KRAS	2	12	25398178	25398289	gtgcaggaccattctttgatNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtataatcttgtggtagttgg
KRAS	2	12	25398178	25398289	gtgcaggaccattctttgatNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtataaacttgtggtagttgg
KRAS	2	12	25398247	25398358	cagaatcattttgtggatgaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtgtgtgacatgttctaatat
KRAS	2	12	25398247	25398358	cagaatcattttgtggacgaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtgtgtgacatgttctaatat
KRAS	3	12	25380186	25380297	ccaagagacaggtttctccaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNacctataatggtgaatatct
KRAS	3	12	25380226	25380337	gtaggaatcctgagaagggaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtatggcaaatacacaaagaa
KRAS	3	12	25380242	25380353	agaactggggagggctttctNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNatccagactgtgtttctccc
KRAS	3	12	25380242	25380353	aggactggggagggctttctNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNatccagactgtgtttctccc
KRAS	3	12	25380260	25380371	atgagggaccagtacataagNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNaaaaggtgcactgtaataat
KRAS	3	12	25380260	25380371	atgagggaccagtacatgagNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNaaaaggtgcactgtaataat
KRAS	4	12	25378485	25378596	gctaagtcctgagcctgtttNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtgagagaaaaactgatatat
KRAS	4	12	25378561	25378672	catcttcagagtccttaactctNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNcttacctgtcttgtcttt
KRAS	4	12	25378640	25378751	atatctttcaaaacctgtccacNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNctactgttctagaaggca
KRAS	4	12	25378549	25378660	ggtaagtaacactgaaataaaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNgaagatgtacctatggtcc
KRAS	4	12	25378557	25378668	acaagacaggtaagtaacactgNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNggactctgaagatgtacc
KRAS	4	12	25378613	25378724	aggctcaggacttagcaagaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNactaatgactgtgctataac
PTPN11	3	12	112888079	112888190	gaaatttgccactttggctgaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtgattttactggtgtaact
PTPN11	3	12	112888099	112888210	aggcaaggggacaaataaagaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNatgtaatactggaccaact
PTPN11	3	12	112888152	112888263	atcttgatgtgggtgacagcNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtatttaagctcaatgacatc
PTPN11	3	12	112888191	112888302	cctacctctgaaaggtcagtaaNNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNtgacctgtatggagggga

* N refers to random nucleotides in the unique tags; lower case indicates ligation and extension probes, upper case indicates backbone sequence.

Table S2.3. Ampliseq panel sequencing primers

Gene	Exon	Chromosome	hg19 start	hg19 stop	Forward primer	Reverse primer
KRAS	2	chr12	25398051	25398267	GGTTACATATAACTTGAAACCCAAGGTA	GCCTTGACGATACAGCTAATTCAGAA
KRAS	2	chr12	25398214	25398393	TGTTGGATCATATTCGTCCACAAAATGA	AGTGTATTAACCTTATGTGTGACATGTT
KRAS	3	chr12	25380223	25380385	GAAAGCCCTCCCCAGTCC	TGCACTGTAATAATCCAGACTGTGTTT
KRAS	3	chr12	25380135	25380262	AATGTCAGCTTATTATATTCAATTTAAACCCAC	GCAATGAGGGACCAGTACATGA
KRAS	4	chr12	25378390	25378605	GGATTAAGAAGCAATGCCCTCTCA	GGACTTAGCAAGAAGTTATGGAATTCCT
KRAS	4	chr12	25378548	25378679	CTGTCTTGTCTTTGCTGATGTTTCAATA	GAAGATGTACCTATGGTCCTAGTAGGAA
KRAS	4	chr12	25378624	25378773	ACTGTTCTAGAAGGCAAATCACATTTA	GTGGACAGGTTTTGAAAGATATTTGTGT
NRAS	2	chr1	115258622	115258845	CGACAAGTGAGAGACAGGATCA	CTGTAGATGTGGCTCGCCAA
PTPN11	3	chr12	112887956	112888177	ACATGTGGTTATTTCACCCATCGT	GGTCATAGTAATCACCAGTGTTCTGAAT
PTPN11	3	chr12	112888129	112888350	GGAGCTGTCACCCACATCAAG	CAGCAGACTTTGTGGTCACTAAAATG

Primer sequences are depicted 5’-3’.

Table S2.4. Primers for RNase H2 based PCR assay

Gene	Variant	Amino acid change	Forward primer*	Reverse primer*
NRAS	chr1:115258747C>T	p.G12D	GCTTTTCCCAACACCArUC/SpC3//SpC3/TA	TAGATGTGGCTCGCCAATTrAA/SpC3//SpC3/CT
NRAS	chr1:115258748C>A	p.G12C	CGCTTTTCCCAACACCACrAT/3SpC3//3SpC3/GA	CTCVCCMATTAACCCTGATTAC
NRAS	chr1:115258744C>T	p.G13D	CACCTCTATGGTGGGATCATATTC	GGTGGTTGGAGCAGGTGrAT/3SpC3//3SpC3/GC
KRAS	chr12:25398281C>T	p.G13D	TGGTCCTGCACCAGTAATATG	TGGTAGTTGGAGCTGGTGrAC/iSpC3//iSpC3/GA
KRAS	chr12:25398284C>A	p.G12V	CACTCTTGCCTACGCCArAC/3SpC3//3SpC3/AC	AACCTTATGTGTGACATGTTCT
PTPN11	chr12:112888210G>A	p.E76K	AGAAATTTGCCACTTTGGCTrAA/3SpC3//3SpC3/GC	GACCTTTCAGAGGTAGGATCTG
PTPN11	chr12:112888189G>A	p.E69K	GAGCTGTCACCCACATCAA	AGCCAAAGTGGCAAATTTCTrUC/3SpC3//3SpC3/CA

* /SpC3/ is a C3 propanediol spacer. rN (A, C, G, U) is a ribonucleotide linkage and the cleavage site of RNase H2 which activates the amplification after cleavage

Table S2.5. Variants detected by standard deep sequencing

Sample	Gene	Variant[1]	Amino acid change	Hotspot[2]	VAF (%)	
					Ampliseq reads	smMIP raw reads
P0024	*NRAS*	c.-17-11G>A		No	0	0.9
P0024	*NRAS*	c.-17-12G>A		No	0	1.5
P0024	*NRAS*	c.-17-19G>T		No	0	1.7
P0024	*NRAS*	c.-17-30C>A		No	0.9	0
P0024	*NRAS*	c.34G>T	p.G12C	Yes	0	0.5
P0024	*NRAS*	c.35G>A	p.G12D	Yes	5.0	6.9
P0029	*NRAS*	c.-17-11G>A		No	0	0.9
P0029	*NRAS*	c.-17-12G>A		No	0	1.4
P0029	*NRAS*	c.-17-30C>A		No	1.1	0
P0029	*NRAS*	c.34G>A	p.G12S	Yes	3.4	6.6
P0029	*NRAS*	c.35G>A	p.G12D	Yes	0.6	0.9
P0035	*NRAS*	c.-17-11G>A		No	0	0.9
P0035	*NRAS*	c.38G>A	p.G13D	Yes	34.5	44.0
P0037	*NRAS*	c.-17-11G>A		No	0	1.1
P0037	*NRAS*	c.35G>A	p.G12D	Yes	7.1	9.2
P0051	*KRAS*	c.34G>A	p.G12S	Yes	1.6	0.9
P0051	*KRAS*	c.35G>A	p.G12D	Yes	13.6	17.9
P0051	*KRAS*	c.38G>A	p.G13D	Yes	0.6	0
P0051	*KRAS*	c.58A>G	p.T20A	No	0	0.6
P0053	*PTPN11*	c.205G>A	p.E69K	Yes	0.9	0
P0055	*KRAS*	c.16C>T	p.L6F	No	0	0.7
P0063	*NRAS*	c.-17-11G>A		No	0	0.9
P0063	*NRAS*	c.-17-19G>T		No	0	1.4
P0063	*NRAS*	c.37G>C	p.G13R	Yes	22.9	35.0
P0070	*KRAS*	c.35G>A	p.G12D	Yes	34.6	36.1
P0076	*KRAS*	c.436G>C	p.A146P	Yes	26.5	33.0
P0076	*KRAS*	c.450+32T>C		No	0	0.5
P0098	*KRAS*	c.35G>T	p.G12V	Yes	27.3	28.6
P0098	*KRAS*	c.38G>A	p.G13D	Yes	0.7	1.5
P0111	*KRAS*	c.351A>T	p.K117N	Yes	23.0	33.0
P0128	*KRAS*	c.179G>A	p.G60D	Yes	6.8	7.0
P0143	*NRAS*	c.-17-11G>A		No	0	1.1
P0143	*NRAS*	c.-17-12G>A		No	0	1.8
P0143	*NRAS*	c.-17-19G>T		No	0	1.8
P0143	*NRAS*	c.35G>A	p.G12D	Yes	38.3	46.0
P0150	*KRAS*	c.35G>A	p.G12D	Yes	25.0	37.5
P0151	*NRAS*	c.111+12T>C		No	0	0.6
P0151	*NRAS*	c.34G>A	p.G12S	Yes	0.6	1.1
P0151	*NRAS*	c.35G>A	p.G12D	Yes	1.8	2.9
P0160	*NRAS*	c.-17-11G>A		No	0	0.9

Table S2.5. Continued

Sample	Gene	Variant[1]	Amino acid change	Hotspot[2]	VAF (%)	
					Ampliseq reads	smMIP raw reads
P0160	*NRAS*	c.-17-12G>A		No	0	1.6
P0160	*NRAS*	c.-17-30C>A		No	0.9	0
P0160	*NRAS*	c.34G>A	p.G12S	Yes	0.9	1.7
P0160	*NRAS*	c.35G>A	p.G12D	Yes	3.2	5.9
P0160	*NRAS*	c.38G>A	p.G13D	Yes	0	0.7
P0166	*KRAS*	c.14A>G	p.K5R	No	0	0.5
P0166	*KRAS*	c.35G>A	p.G12D	Yes	21.3	29.7
P0166	*KRAS*	c.39C>T	p.G13G	Yes	0	0.6
P0166	*KRAS*	c.56T>C	p.L19S	No	0	0.6

[1] *NRAS* mutations at position c.-17 are located in intron 1, upstream of the start codon in exon 2; [2] Recurrently observed in COSMIC (see **Table S2.6**).

Table S2.6. Mutation hotspot regions in the three RAS genes

Gene	Exon	Substitution*
KRAS	2	G12D, G12V, G12A, G12S, G12C, G12R, G12G, G13D, G13C, G13G, G13A, V14I, A18D, L19F, Q22K, L23R, D33E
KRAS	3	T58I, A59G, A59E, G60D, Q61H, Q61R, Q61P, Q61L, E63K, Y64N
KRAS	4	K117N, A146T, A146V, A146P
NRAS	2	G12D, G12S, G12V, G12C, G12A, G12R, G12N, G13D, G13R, G13V, G13C, G13A, Q22K
PTPN11	3	T52S, N58Y, G60V, G60R, G60A, D61Y, D61V, D61N, D61H, D61G, E69K, F71L, A72V, A72T, A72D, A72S, T73I, E76K, E76G, E76A, E76Q, E76V

*Only substitutions that have been recurrently identified in hematologic malignancies are included (based on COSMIC database, version 81).

Supplementary Figures

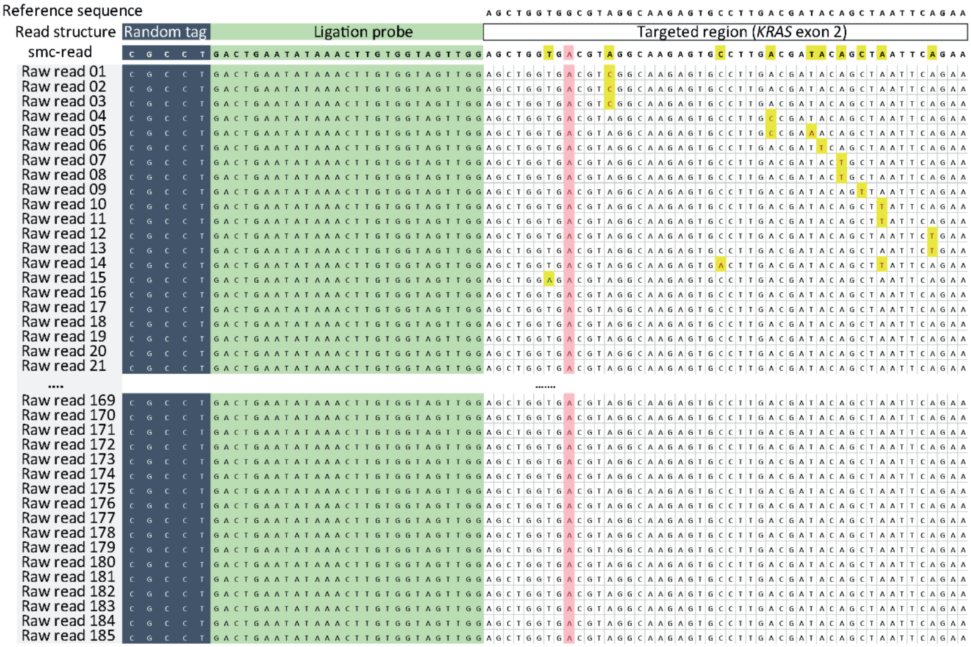

Reference sequence
Read structure
Random tag
Ligation probe
Targeted region (KRAS exon 2)
smc-read
Raw read 01
Raw read 02
Raw read 03
Raw read 04
Raw read 05
Raw read 06
Raw read 07
Raw read 08
Raw read 09
Raw read 10
Raw read 11
Raw read 12
Raw read 13
Raw read 14
Raw read 15
Raw read 16
Raw read 17
Raw read 18
Raw read 19
Raw read 20
Raw read 21
....
Raw read 169
Raw read 170
Raw read 171
Raw read 172
Raw read 173
Raw read 174
Raw read 175
Raw read 176
Raw read 177
Raw read 178
Raw read 179
Raw read 180
Raw read 181
Raw read 182
Raw read 183
Raw read 184
Raw read 185
AGCTGGTGGCGTAGGCAAGAGTGCCTTGACGATACAGCTAATTCAGAA
CGCCT
GACTGAATATAAACTTGTGGTAGTTGG

Figure S2.1. Schematic of raw reads forming a tag-defined read group. Groups of raw smMIP reads that share the unique molecular tag plus sample index and align to the same genomic region form a tag-defined read group (TDRG). Random errors (yellow) that occur during library construction and sequencing may be present in some reads within the TDRG at some positions, but are eliminated in the analysis pipeline, while true variants (red) that occur in >70% of the raw reads are kept.

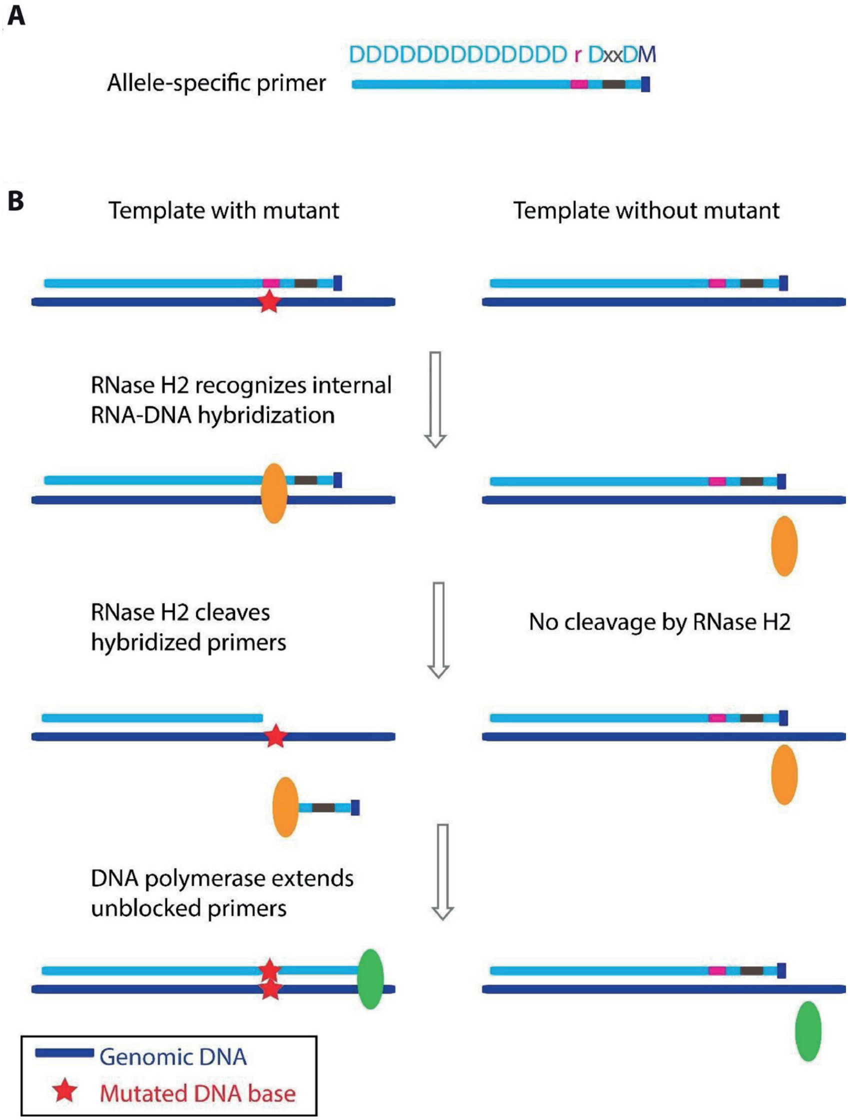

A
DDDDDDDDDDDDD r DxxDM
Allele-specific primer
B
Template with mutant
Template without mutant
RNase H2 recognizes internal
RNA-DNA hybridization
RNase H2 cleaves
hybridized primers
No cleavage by RNase H2
DNA polymerase extends
unblocked primers
Genomic DNA
Mutated DNA base

Figure S2.2. A schematic of the allele-specific RNase H2 based PCR (Adapted from Dobosy *et al*, 2011)[14]. **(A)** Structure of the blocked cleavable primer used for RNase H2 based PCR (rhPCR). D and M represent DNA bases that match and mismatch to the target genomic region, respectively. The RNA base that matches to the mutant is depicted by r. X depicts a C3 propanediol spacer. **(B)** Schematic flow of rhPCR approach. RNase H2 enzyme is an endoribonuclease, which prefers to cleave a single RNA residue within the context of RNA-DNA duplex. After cleavage of 3' RNA residue, the primer is activated to amplify the targeted genomic region. However, the blocked primer is not able to support extension and amplification during PCR.

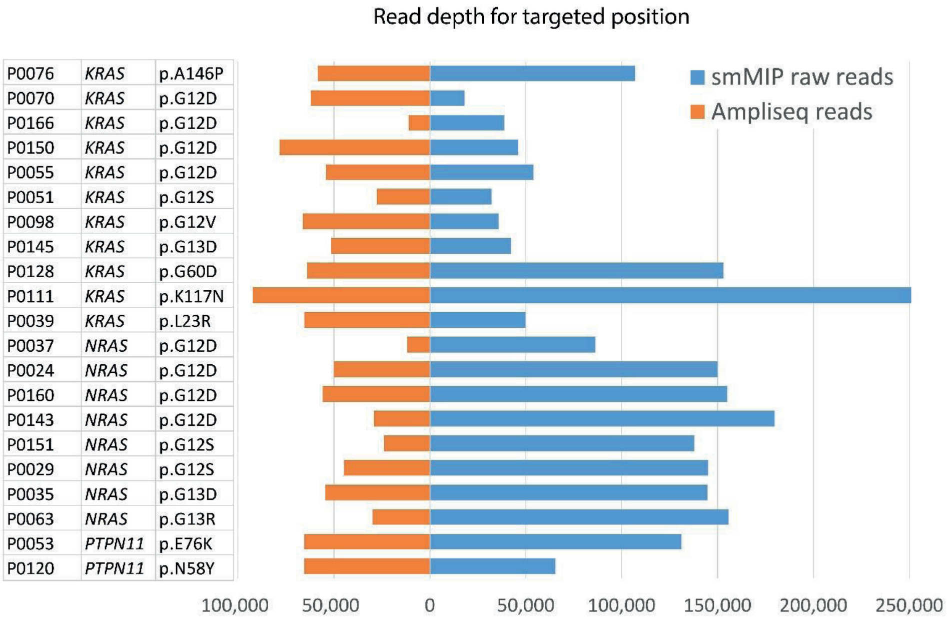


Figure S2.3. Read depth per sample at the targeted position of the mutation for the two deep sequencing approaches.

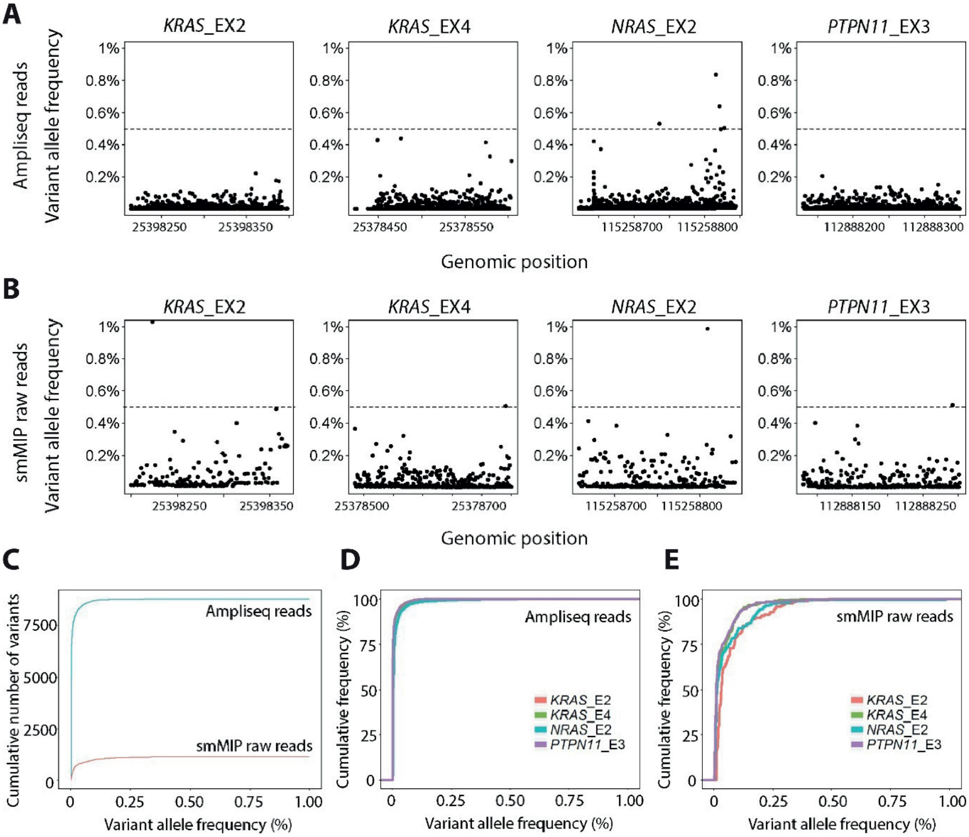


Figure S2.4. Detection limit of deep sequencing approaches using raw reads. In order to define the background noise levels, we sequenced the targeted amplicons in DNA from healthy donors. For further analysis, we set the detection limit as variant allele frequency (VAF) >0.5% with >200 variant reads, which excludes 99.8% of the variants detected in the healthy samples. (**A**-**B**) VAF of all variants called from reference samples in targeted regions, using Ampliseq-based NGS approach sequenced on the Ion Personal Genome Machine platform **(A)** and the smMIP approach sequenced on the Illumina platform **(B)**. The dashed line indicates VAF = 0.5%. Genomic position is based on hg19. **(C)** Cumulative number of variants identified from the reference samples by the two approaches. (**D-E**) Cumulative frequency of variants detected in reference samples by the Ampliseq **(D)** and smMIP raw reads **(E)**.

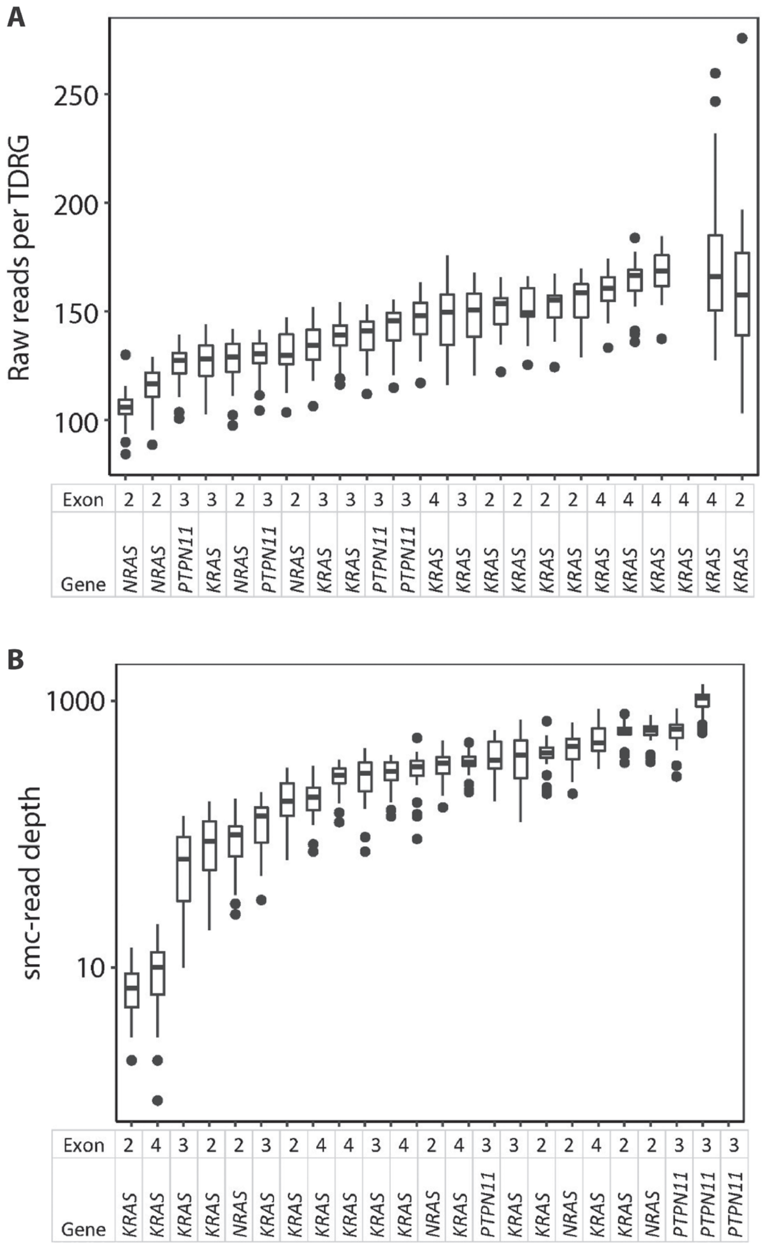


Figure S2.5. The smMIP-based variant calling performance. **(A)** Number of raw reads in each tag-defined read group (TDRG) for each probe. **(A)** TDRG is composed of all raw smMIP sequence reads amplified from the same unique molecule present in the patient material (**Figure 2.1D**). On average, each TDRG was comprised of 142 individual raw sequencing reads (range 105-174). **(B)** Number of single molecule consensus (smc) reads for each probe. Each smc-read is derived from a TDRG (**Figure 2.1D**).

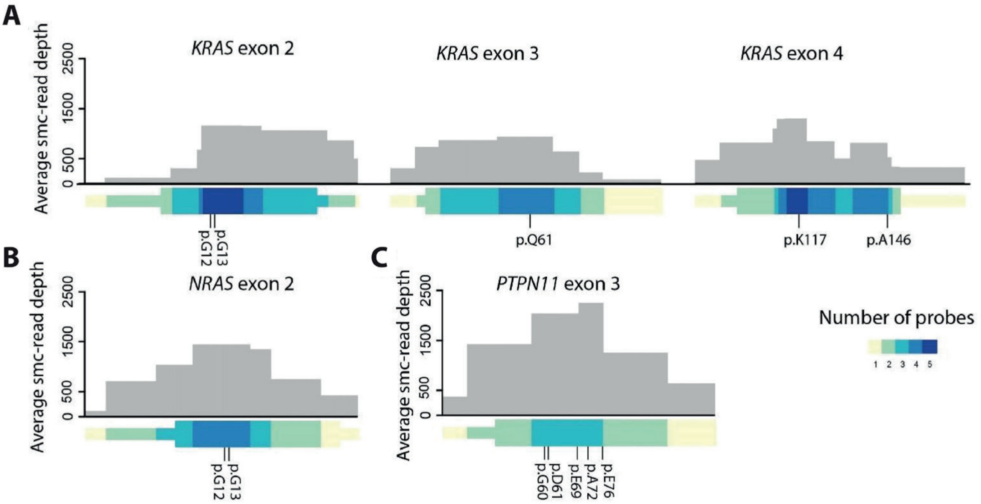
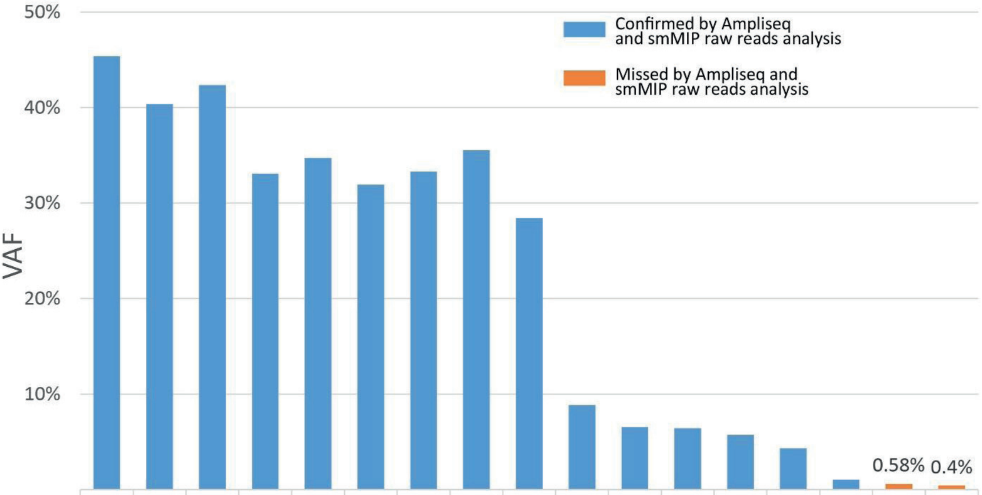


Figure S2.6. Average smc-read depth of the targeted regions in *KRAS* **(A)**, *NRAS* **(B)** and *PTPN11* **(C)**. The x-axis shows the genomic position for the targeted region. The y-axis shows average single molecule consensus (smc) read depth of the targeted position. The structure of the targeted regions is shown below, where bigger boxes represent exons, smaller boxes represent introns and black lines indicate positions for hotspots and known RAS mutations. Gradient color represents the number of smMIP probes covering the position.



Sample	P0143	P0070	P0035	P0111	P0150	P0098	P0076	P0063	P0166	P0037	P0128	P0024	P0029	P0160	P0151	P0145	P0053
Gene	*NRAS*	*KRAS*	*NRAS*	*KRAS*	*KRAS*	*KRAS*	*KRAS*	*NRAS*	*KRAS*	*NRAS*	*KRAS*	*NRAS*	*NRAS*	*NRAS*	*NRAS*	*KRAS*	*PTPN11*
Mutation	p.G12D	p.G12D	p.G13D	p.K117N	p.G12D	p.G12V	p.A146P	p.G13R	p.G12D	p.G12D	p.G60D	p.G12D	p.G12S	p.G12D	p.G12S	p.G13D	p.E76K
smc-read depth	1325	756	1245	1838	1620	946	655	1086	1038	680	1393	1139	1207	1375	1340	1036	1770

Figure S2.7. RAS mutations detected in relapse samples were determined in the paired diagnosis samples. The y-axis shows the VAF determined by the smMIP consensus variant calling approach. Details of each mutation are shown below the x-axis.

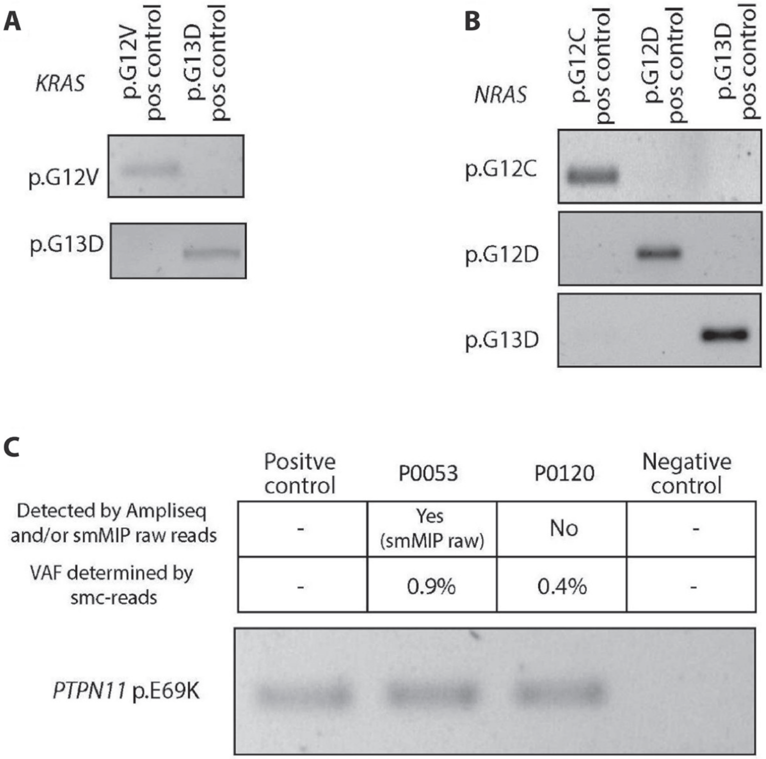

A
KRAS
p.G12V pos control
p.G13D pos control
p.G12V
p.G13D
B
NRAS
p.G12C pos control
p.G12D pos control
p.G13D pos control
p.G12C
p.G12D
p.G13D
C
Positve control
P0053
P0120
Negative control
Detected by Ampliseq and/or smMIP raw reads
-
Yes (smMIP raw)
No
-
VAF determined by smc-reads
-
0.9%
0.4%
-
PTPN11 p.E69K

Figure S2.8. Variants called by smc-reads were analyzed by RNase H2 based PCR. To assess the specificity of the assay, we performed RNase H2 based PCR (rhPCR) using allele specific primers on the sample with a variant located on the adjacent position in *KRAS* **(A)** and *NRAS* **(B)**. Failure to detect the product indicates that concurrent variants within the primer annealing region do not interfere with the specificity of the reaction. **(C)** *PTPN11* p.E69K detected by the smMIP consensus variant calling in two samples was confirmed by the rhPCR assay and gel electrophoresis. Genomic DNA from a healthy donor was used for the negative control.

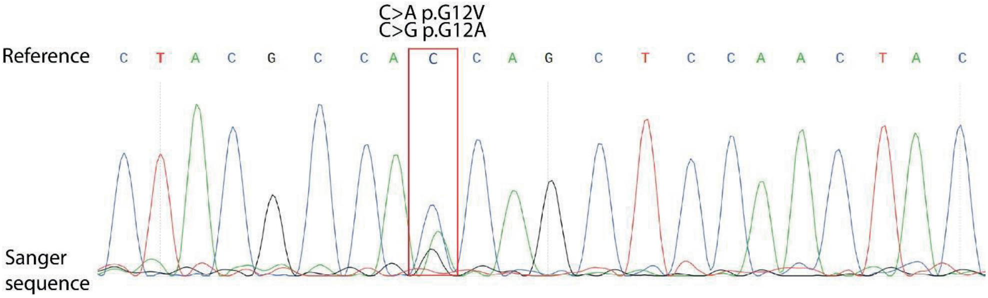


Figure S2.9. Variant called from smMIP-based smc-reads validated by Sanger sequencing. Two *KRAS* mutations are observed at the same position in sample P0098 by smMIP-based consensus variant calling approach: chr12: 25398284C>A (p.G12V, VAF = 32%) and chr12: 25398284C>G (p. G12A, VAF = 16%). The latter was not called in Ampliseq, but was confirmed by Sanger sequencing.

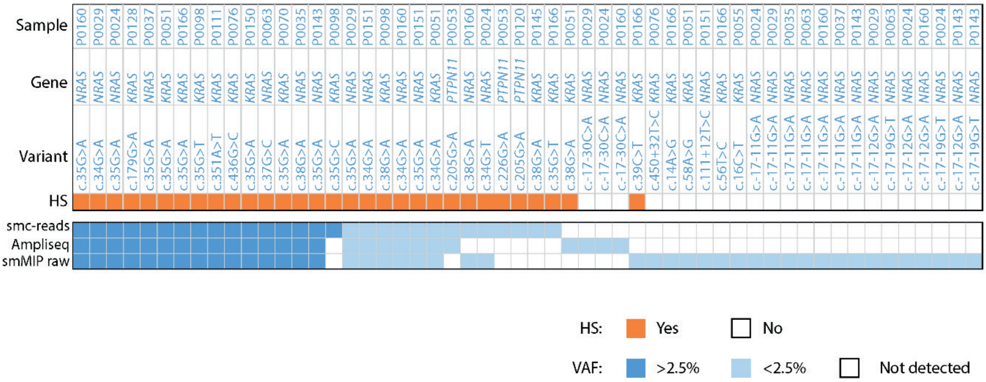


Figure S2.10. Overview of variants identified from 21 diagnosis samples by approaches studied. Details of samples and variants are displayed on top. Variants located in known hotspot regions are colored in orange. Variants identified by each deep sequencing approach are shown in bottom. Variants with a VAF<2.5% and VAF>2.5% are colored in light blue and dark blue, respectively.

FUNDING

This work was supported by grants from the Dutch Cancer Society (KWF; KUN2009-4298 to RPK, KUN2012-5366 to EW), Stichting Kinderen Kankervrij (KiKa 150 to RPK and AH), and the China Scholarship Council (CSC201304910347 to JY).

CONFLICTS OF INTEREST

No conflicts of interest to disclose.

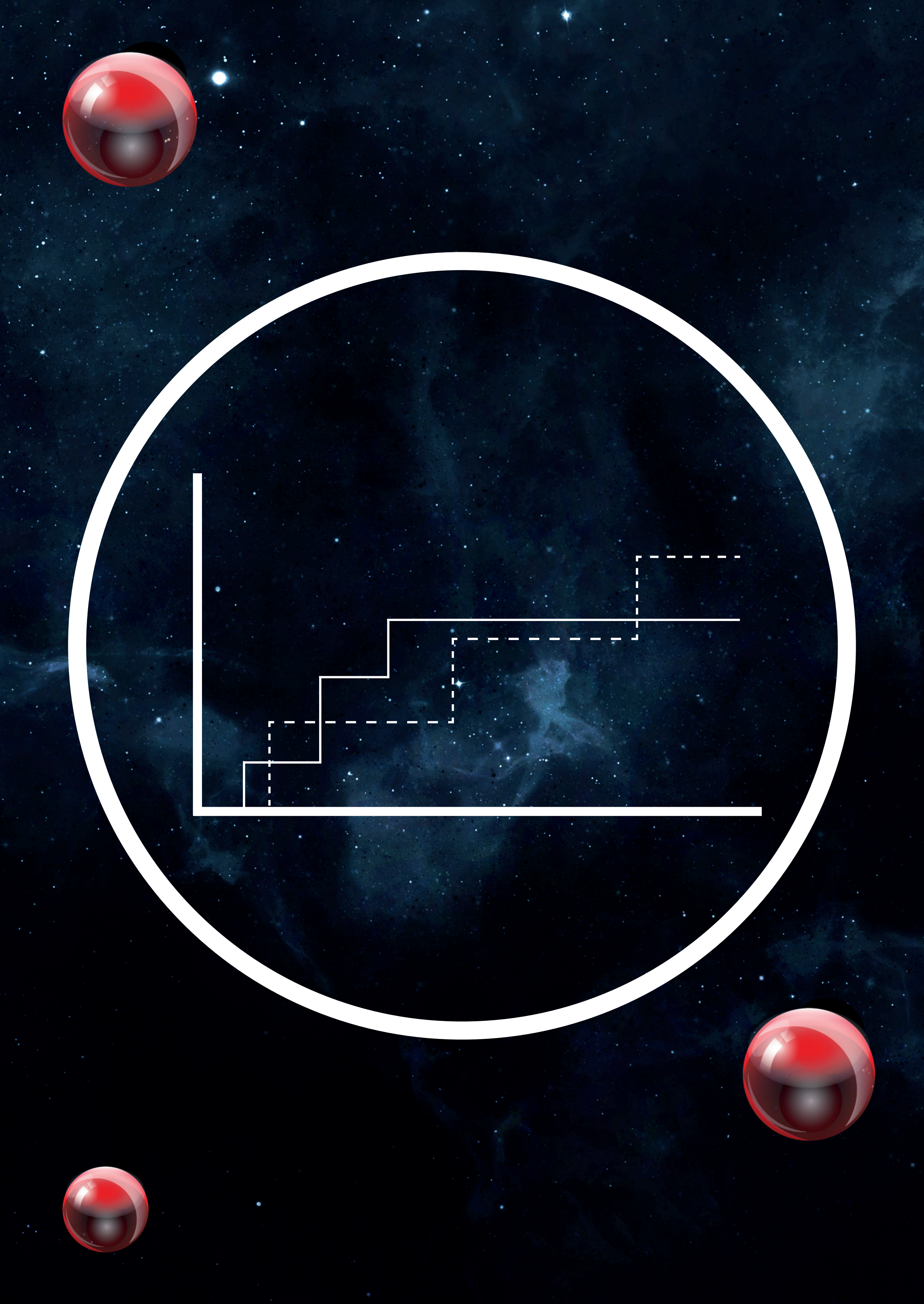
CONTRIBUTIONS

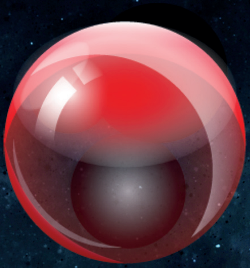
RPK, EW, and AGvK conceived the study. RPK, EW and JY designed the study. JY, ŽA and SVvR performed experiments and analyzed the data. JY and RPK wrote the manuscript and created the figures and tables. ES provided samples and clinical information. AH provided technical support in designing, performing and interpreting results obtained through smMIP-based sequencing. All authors critically reviewed the manuscript and approved the final submitted manuscript.

REFERENCES

1. Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the Children's Oncology Group. J Clin Oncol. 2012;30(14):1663-1669.
2. Pieters R, Groot-Kruseman Hd, Velden VVd, Fiocco M, Berg Hvd, Bont Ed, et al. Successful therapy reduction and intensification for childhood acute lymphoblastic leukemia based on minimal residual disease monitoring: study ALL10 from the Dutch Childhood Oncology Group. J Clin Oncol. 2016;34(22):2591-2601.
3. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science. 2008;322(5906):1377-1380.
4. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature. 2011;469(7330):356-361.
5. Irving J, Matheson E, Minto L, Blair H, Case M, Halsey C, et al. Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. Blood. 2014;124(23):3420-3430.
6. Gawad C, Koh W, Quake SR. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. Proc Natl Acad Sci U S A. 2014;111(50):17947-17952.
7. Ma X, Edmonson M, Yergeau D, Muzny DM, Hampton OA, Rusch M, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun. 2015;6(1):6604.
8. Bashford-Rogers RJ, Nicolaou KA, Bartram J, Goulden NJ, Loizou L, Koumas L, et al. Eye on the B-ALL: B-cell receptor repertoires reveal persistence of numerous B-lymphoblastic leukemia subclones from diagnosis to relapse. Leukemia. 2016;30(12):2312-2321.
9. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364.
10. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genom. 2012;13(1):341.
11. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012;338(6114):1619-1622.
12. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 2013;23(5):843-854.
13. Yu J, Waanders E, van Reijmersdal SV, Sonneveld E, Hoogerbrugge PM, van Leeuwen FN, et al. The majority of RAS pathway mutations detected in relapsed B-Cell precursor acute lymphoblastic leukemia are present in major or minor subclones at diagnosis. Blood. 2014;124(21):2397-2397.
14. Dobosy JR, Rose SD, Beltz KR, Rupp SM, Powers KM, Behlke MA, et al. RNase H-dependent PCR (rhPCR): improved specificity and single nucleotide polymorphism detection using blocked cleavable primers. BMC Biotechnol. 2011;11(1):80.
15. Zhang J, Mullighan CG, Harvey RC, Wu G, Chen X, Edmonson M, et al. Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. Blood. 2011;118(11):3080-3087.

16. Malinowska-Ozdowy K, Frech C, Schönegger A, Eckert C, Cazzaniga G, Stanulla M, et al. KRAS and CREBBP mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia. Leukemia. 2015;29(8):1656-1667.
17. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. Cancer Res. 2012;72(10):2457-2467.
18. Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. Nat Protoc. 2014;9(11):2586-2606.
19. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, et al. Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat Methods. 2013;10(10):1003-1005.
20. Della Starza I, Nunes V, Cavalli M, De Novi LA, Ilari C, Apicella V, et al. Comparative analysis between RQ-PCR and digital-droplet-PCR of immunoglobulin/T-cell receptor gene rearrangements to monitor minimal residual disease in acute lymphoblastic leukaemia. Br J Haematol. 2016;174(4):541-549.
21. Eijkelenboom A, Kamping EJ, Kastner-van Raaij AW, Hendriks-Cornelissen SJ, Neveling K, Kuiper RP, et al. Reliable next-generation sequencing of formalin-fixed, paraffin-embedded tissue using single molecule tags. J Mol Diagn. 2016;18(6):851-863.
22. Neveling K, Mensenkamp AR, Derks R, Kwint M, Ouchene H, Steehouwer M, et al. BRCA testing by single-molecule molecular inversion probes. Clin Chem. 2017;63(2):503-512.
23. Weren RD, Mensenkamp AR, Simons M, Eijkelenboom A, Sie AS, Ouchene H, et al. Novel BRCA1 and BRCA2 tumor test as basis for treatment decisions and referral for genetic counselling of patients with ovarian carcinomas. Hum Mutat. 2017;38(2):226-235.
24. Acuna-Hidalgo R, Sengul H, Steehouwer M, van de Vorst M, Vermeulen SH, Kiemeney L, et al. Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. Am J Hum Genet. 2017;101(1):50-64.
25. Waalkes A, Penewit K, Wood BL, Wu D, Salipante SJ. Ultrasensitive detection of acute myeloid leukemia minimal residual disease using single molecule molecular inversion probes. Haematologica. 2017;102(9):1549-1557.





CHAPTER 3

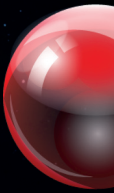
MULTICLONAL COMPLEXITY OF PEDIATRIC ACUTE LYMPHOBLASTIC LEUKEMIA AND THE PROGNOSTIC RELEVANCE OF SUBCLONAL MUTATIONS

Željko Antić[1], Jiangyan Yu[1,2*], Simon V. van Reijmersdal[1,2], Anke van Dijk[2], Linde Dekker[1], Wouter H. Segerink[1], Edwin Sonneveld[1,3], Marta Fiocco[1,4,5], Rob Pieters[1,3], Peter M. Hoogerbrugge[1,3], Frank N. van Leeuwen[1], Ad Geurts van Kessel[2], Esmé Waanders[1,6] and Roland P. Kuiper[1]*

[*]These authors contributed equally to this study as lead authors

[1]Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands
[2]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
[3]Dutch Childhood Oncology Group, Utrecht, The Netherlands
[4]Medical Statistics, Department of Biomedical Data Science, Leiden University Medical Center, Leiden, The Netherlands.
[5]Mathematical Institute, Leiden University, The Netherlands.
[6]Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

Haematologica; December 2021; 106(12):3046-3055



ABSTRACT

Genomic studies of pediatric acute lymphoblastic leukemia (ALL) have shown remarkable heterogeneity in initial diagnosis, with multiple (sub)clones harboring lesions in relapse-associated genes. However, the clinical relevance of these subclonal alterations remains unclear. We assessed the clinical relevance and prognostic value of subclonal alterations in the relapse-associated genes *IKZF1*, *CREBBP*, *KRAS*, *NRAS*, *PTPN11*, *TP53*, *NT5C2*, and *WHSC1* in 503 ALL cases. Using molecular inversion probe sequencing and breakpoint-spanning polymerase chain reaction analysis we reliably detected alterations with an allele frequency below 1%. We identified 660 genomic alterations in 285 diagnostic samples of which 495 (75%) were subclonal. RAS pathway mutations were common, particularly in minor subclones, and comparisons between RAS hotspot mutations revealed differences in their capacity to drive clonal expansion in ALL. We did not find an association of subclonal alterations with unfavorable outcome. Particularly for *IKZF1*, an established prognostic marker in ALL, all clonal but none of the subclonal alterations were preserved at relapse. We conclude that, for the genes tested, there is no basis to consider subclonal alterations detected at diagnosis for risk group stratification of ALL treatment.

INTRODUCTION

Improvements in the treatment of pediatric acute lymphoblastic leukemia (ALL) have resulted in high overall survival rates, now approaching 90%[1,2]. Nevertheless, relapse still remains the most common cause of treatment failure and death in children with ALL, and better recognition of individuals at risk of developing relapse will likely aid further improvements in outcome. Recent studies describing the genomic landscape of relapsed ALL have shown that relapse often originates from a minor (sub)clone at diagnosis, at a cellular fraction often undetectable by routine diagnostic methods[3-5]. These minor (sub)clones harbor genomic alterations acquired later during leukemia development, which could potentially contribute to clonal drift, but are unlikely to be essential for initiation of the primary disease. However, selective pressure of the upfront treatment may provide a competitive advantage to subclones that harbor alterations in cancer genes, enabling their selective survival, eventually leading to treatment failure. Both the number and clonal burden of the alterations in these genes are expected to be increased at the time of relapse, compared to initial diagnosis. Indeed, mutations in relapse-associated genes, such as those in the histone acetyltransferase (HAT) domain of the histone methyltransferase *CREBBP*, can often be traced back to minor subclones in the diagnostic sample[4,6,7].

Genomic characterization of relapsed pediatric ALL has revealed multiple alterations that are enriched compared to diagnosis, including activating mutations in RAS pathway genes, HAT domain mutations in *CREBBP* and deletions or mutations in the B-cell transcription factor *IKZF1*[6-13]. The presence of these aberrations at the time of diagnosis can be of potential prognostic relevance, as has been demonstrated extensively for *IKZF1* in many different treatment protocols[12,14-19] and can even lead to adjustments in stratification and treatment[14,20]. However, it remains unclear whether mutations in relapse-associated genes when present in a minor subclone at initial diagnosis are also clinically relevant.

Subclonal mutations can be identified using deep targeted, next-generation sequencing techniques[21,22]. Despite the sensitivity of these techniques, both amplification and sequencing can easily lead to errors that hamper the reliable detection of low-level mosaic mutations. We previously demonstrated that single molecule Molecular Inversion Probes (smMIP), which use unique molecular identifiers to barcode each DNA copy, can correct for sequencing and amplification artefacts, resulting in a reliable detection of low-level mosaic mutations, down to a variant allele frequency of 0.4%[23].

In this study we used the smMIP-based sequencing approach to perform deep targeted sequencing of seven relapse-associated genes in a cohort of 503 pediatric

ALL samples taken at initial diagnosis, resulting in the detection of 141 clonal and 469 subclonal mutations. In addition, we performed real-time quantitative polymerase chain reaction (PCR) to sensitively detect subclonal *IKZF1* exon 4-7 deletions (del4-7), which were found at a similar frequency as full-clonal deletions. Subsequently, we estimated their potential as drivers of clonal expansion and prognostic markers for relapse development.

METHODS

In this study we analyzed two cohorts of diagnostic samples from B-cell precursor ALL patients treated according to the Dutch Childhood Oncology Group (DCOG) protocols DCOG-ALL9 (*n* = 131)[12,24] and DCOG-ALL10 (*n* = 245) (**Table S3.1**). Both cohorts were representative selections of the total studies[12,24] (**Table S3.2**). The median age at diagnosis of the patients in these cohorts was 4 and 5 years, and the median follow-up time, estimated with a reverse Kaplan-Meier method, was 138 and 104 months, respectively[25]. Relapse occurred in 18% (24/131) and 11% (27/245) of the patients, while 0.7% (1/131) and 2.8% (7/245) died during the follow-up. DNA was isolated from mononuclear cells obtained from bone marrow or peripheral blood. The median blast percentage of the samples was 92% (**Table S3.3**). To increase the number of patients for the comparisons between relapsed and non-relapsed cases, we used an extended cohort of diagnostic samples from 127 additional ALL patients treated according to the DCOG-ALL9 (*n* = 76) or DCOG-ALL10 (*n* = 51) protocols; this cohort was enriched for patients who had a relapse and also contained 55 patients with T-cell ALL. This latter cohort was not included in the survival analyses. In order to detect mutations preserved in major clones at relapse, we performed Sanger sequencing (73/171) or used previously published Ampliseq-based deep-sequencing data (98/171) to verify alterations observed at diagnosis[26]. In accordance with the Declaration of Helsinki, written informed consent was obtained from all patients and/or their legal guardians before enrollment in the study, and the DCOG institutional review board approved the use of excess diagnostic material for this study (OC2017-024).

In order to accurately detect subclonal alterations in diagnostic samples, 166 smMIP were designed in *CREBBP*, *PTPN11*, *NT5C2*, *WHSC1*, *TP53*, *KRAS* and *NRAS*, seven genes that are frequently mutated in relapsed ALL (**Table S3.4** and **Supplementary Materials and Methods**). *IKZF1* and *ERG* deletion status was assessed using the multiplex ligation-dependent probe amplification assay (MLPA) SALSA P335 ALL-IKZF1 and P327 iAMP-ERG kits, respectively (MRC-Holland, Amsterdam, the Netherlands), according to the manufacturer's instructions and as described before[12,24]. Additionally, *IKZF1* 4-7 deletions were assessed with Sanger sequencing and real-time quantitative PCR, using an IQ SYBR Green supermix (Bio-Rad, Hercules, CA, USA). For detailed descriptions of the smMIP-based sequencing, *IKZF1* deletion detection and data analysis, see the **Supplementary Materials and Methods** (**Figures S3.1**, **S3.2**, and **Tables S3.4-S3.6**).

To test continuous and categorical variables, the nonparametric Wilcoxon signed rank and Fisher exact tests were used, respectively (R packages ggpubr version 0.2 and stats version 3.5.1). Cumulative incidence of relapse (CIR) was estimated by employing a competing-risk model with death as a competing event[27]. To assess the statistical difference between CIR, the Gray test[28] was applied. To investigate the effect of prognostic factors on relapse, univariate and multivariate Cox proportional hazard regression models were estimated. Competing risk analysis was performed with the R packages cmprsk (version 2.2-7) and survminer (version 0.4.3). Univariate and multivariate Cox models were estimated using R package survival (version 3.1-12). Data were visualized using the R package ggplot2 (version 3.2.1) and cBioPortal MutationMapper[29,30].

RESULTS

A total of 503 diagnostic samples from children with ALL (**Table S3.1**) was subjected to targeted deep sequencing of the relapse-associated genes *TP53*, *CREBBP* (HAT domain), *KRAS*, *NRAS*, *PTPN11*, *NT5C2* and *WHSC1* using smMIP, which contain random molecular tags to accurately detect low-level mosaic variants[23]. Each targeted region was covered with an average of 308 unique capture-based consensus reads (**Figures 3.1**, **S3.1A** and **S3.1B**), enabling the reliable detection of alterations with allele frequencies even below 1%. A total of 7,836 quality-filtered variants was detected, of which 610 were absent in public and private variant databases and were predicted as pathogenic. The allele frequency of these mutations ranged from 0.03-100% (**Figure 3.2A** and **Table S3.3**). The majority of the mutations (473/610; 78%) was found in one of the three RAS pathway genes (*KRAS*, *NRAS*, *PTPN11*), of which 418 (88%) were known hotspot mutations.

In addition to sequencing the seven relapse-associated genes, we performed sensitive screening for *IKZF1* deletions, which are strongly associated with the occurrence of relapse. We chose to focus on exon 4-7 deletions, which represent 25% of all *IKZF1* deletions, have a similar unfavorable outcome as other *IKZF1* deletions[31], and show the strongest clustering of deletion breakpoints, thus enabling their sensitive upfront detection by breakpoint-spanning semi-quantitative PCR[32]. Applying this strategy to the 503 diagnostic samples revealed all 22 *IKZF1* exon 4-7 deletions previously identified using a standard MLPA method, as well as 28 additional cases carrying deletions that were missed with the MLPA technique. All breakpoints were sequenced to determine their unique breakpoint-spanning sequences (**Table S3.6**). Using a dilution series of a control sample with a full-clonal *IKZF1* exon 4-7 deletion, we determined the level of clonality of the deletions, which ranged from 100% down to 0.32% (**Figures 3.1**, **3.2** and **S3.1C**). All but one of the subclonal *IKZF1* exon 4-7 deletions had allele frequencies below 10% (**Table S3.7**).

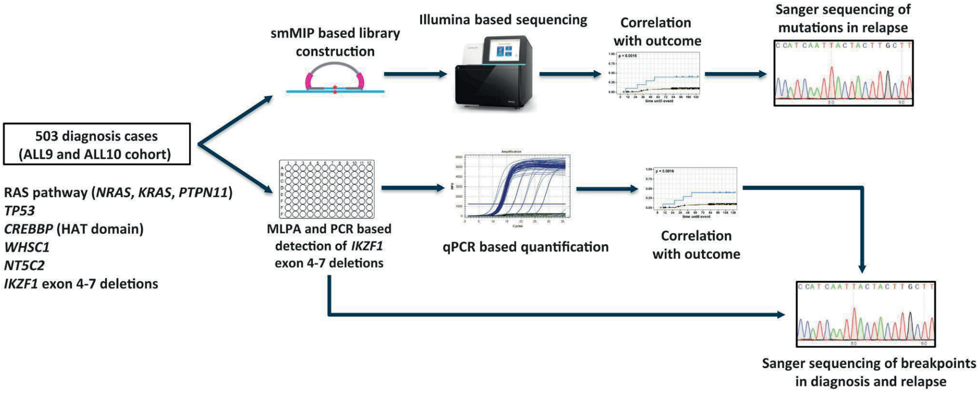
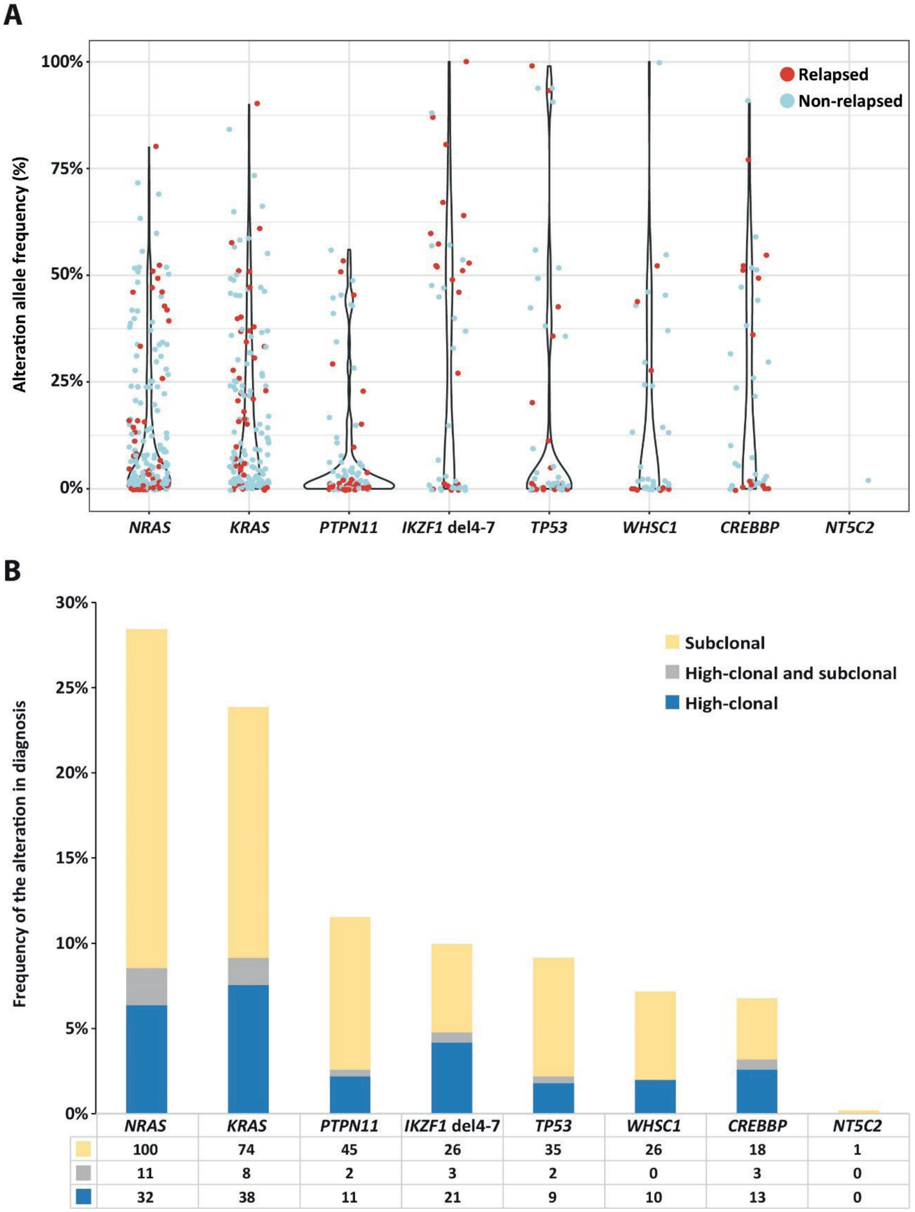

503 diagnosis cases
(ALL9 and ALL10 cohort)
RAS pathway (*NRAS, KRAS, PTPN11*)
TP53
CREBBP (HAT domain)
WHSC1
NT5C2
IKZF1 exon 4-7 deletions
smMIP based library construction
Illumina based sequencing
Correlation with outcome
Sanger sequencing of mutations in relapse
MLPA and PCR based detection of *IKZF1* exon 4-7 deletions
qPCR based quantification
Correlation with outcome
Sanger sequencing of breakpoints in diagnosis and relapse

Figure 3.1. Schematic representation of the study design. Single molecule Molecular Inversion Probe-based sequencing approach and real-time quantitative polymerase chain reaction were used in order to detect alterations in known relapse-associated genes in a large cohort of diagnostic samples from patients with acute lymphocytic leukemia. Detected alterations were correlated with outcome and Sanger sequencing was performed on available relapse samples in order to confirm that exactly the same alteration was present in the major clone in relapse. smMIP: single molecule molecular inversion probe; MLPA: multiplex ligation-dependent probe amplification assay; PCR: polymerase chain reaction; qPCR: real-time quantitative polymerase chain reaction.

Subclonal alterations in relapse-associated genes are common at diagnosis

Combining sequence mutations and *IKZF1* exon 4-7 deletions, we detected 660 genomic alterations in 285 diagnostic samples, of which 165 (25%) were present in the major fraction of cells (allele frequency ≥25%), which were referred to as high-clonal. The remaining 495 mutations (75%), most of which had an allele frequency <10% were referred to as subclonal (**Figure S3.2** and **Table S3.7**). A total of 147/285 patients carried at least one alteration in a major clone, while 138/285 (48%) patients carried exclusively subclonal alterations. *NRAS* and *KRAS* were the most frequently affected genes, showing major clone mutations in 6% and 8% of the cases and subclonal mutations in 20% and 15% of the cases, respectively (**Figures 3.2A** and **3.2B**). The proportion of subclonal alterations, relative to major clone alterations, was variable among different genes, ranging from 59% for *IKZF1* exon 4-7 deletions to 86% for *PTPN11* mutations (**Figure 3.2B**).



	NRAS	*KRAS*	*PTPN11*	*IKZF1* del4-7	*TP53*	*WHSC1*	*CREBBP*	*NT5C2*
🟨	100	74	45	26	35	26	18	1
▨	11	8	2	3	2	0	3	0
🟦	32	38	11	21	9	10	13	0

Figure 3.2. Prevalence and distribution of alterations in eight relapse-associated genes. **(A)** Violin plot showing the variability in mutation allele frequency at diagnosis in the genes studied. The color of the dots indicates whether the mutation was detected in a case without relapse (blue) or with relapse (red). **(B)** Bar plot showing the frequencies of major clone (high-clonal) and subclonal alterations per case in the genes studied. RAS pathway genes (*NRAS*, *KRAS*, *PTPN11*) were the most frequently mutated. Subclonal mutations (yellow bar) were highly prevalent in all genes tested. A subset of cases had both clonal and subclonal alterations in the same gene (gray bar).

Only one thus far unknown (subclonal) *NT5C2* mutation (p.Arg507Trp) was identified in a leukemia sample from a patient who did not relapse (**Figures 3.2A** and **3.2B**). Subclonal mutations were relatively common in hyperdiploid ALL (184 cases), particularly for mutations in RAS pathway genes (190/256; 74%), *WHSC1* (22/26; 85%) and *CREBBP* (13/27; 48%) (**Table S3.8**). Major clone *WHSC1* mutations were mostly identified in *ETV6-RUNX1*-positive cases (4/10, 40%).

Potency of RAS pathway genes as drivers of clonal expansion

We identified 473 RAS pathway mutations in 225/503 (45%) cases, of which 78% were subclonal (median allele frequency = 3.5%). Over half of the RAS-affected cases were hyperdiploid (>47 chromosomes), in line with previous studies indicating that RAS mutations are associated with hyperdiploidy at diagnosis[10,33]. The abundance of these mutations in major and minor clones suggests that these mutations drive clonal expansion during the development of leukemia. Major clone RAS pathway mutations (n = 102; all being known hotspots) were found to be mutually exclusive, and 52/102 (51%) of these RAS-mutated cases had at least one additional subclonal mutation in one of the three RAS pathway genes. The mutations mostly affected codons 12 and 13 of *KRAS* and *NRAS* (**Figures 3.3A**-**3.3C**), and considerable variability in the level of clonality was observed between the different RAS hotspot mutations at the time of diagnosis. For example, *NRAS* p.G12A (10 cases), *NRAS* p.G12V (7 cases), and *PTPN11* p.E76K (7 cases) were never found to be present in a major clone, whereas 55% (n = 11) of the *KRAS* p.G13D and 27% (n = 9) of the *KRAS* p.G12D mutations were found in major clones. With these high numbers of RAS mutations, the variability in clonal burden between hotspot mutations may provide an opportunity to compare the capacity of different hotspots to drive clonal expansion of ALL. In order to test this hypothesis we compared allele frequencies and performed statistical analyses. We found that *KRAS* hotspot mutations had a significantly higher allele frequency compared to both *NRAS* and *PTPN11* mutations (Wilcoxon signed-rank test, $P < 0.01$) (**Figure 3.3D**). When comparing the different hotspot mutations within *KRAS*, p.A146V showed the lowest allele frequency, indicating a weaker potential of this hotspot to drive clonal expansion compared to the other *KRAS* hotspots. Furthermore, the allele frequency of *KRAS* p.G13D was significantly higher than that of the *NRAS* hotspot mutations p.G13D, p.G12D and *KRAS* p.A146V (**Figure 3.3E**). This finding indicates that some RAS hotspot mutations (e.g., *KRAS* p.G12D, p.G13D, p.A146T) may result in a stronger expansion potential compared to others (e.g., *KRAS* p.A146V, *NRAS* p.G12D, p.G13D), and further illustrates the complex heterogeneity of RAS hotspot mutations in their potential to drive clonal expansion.

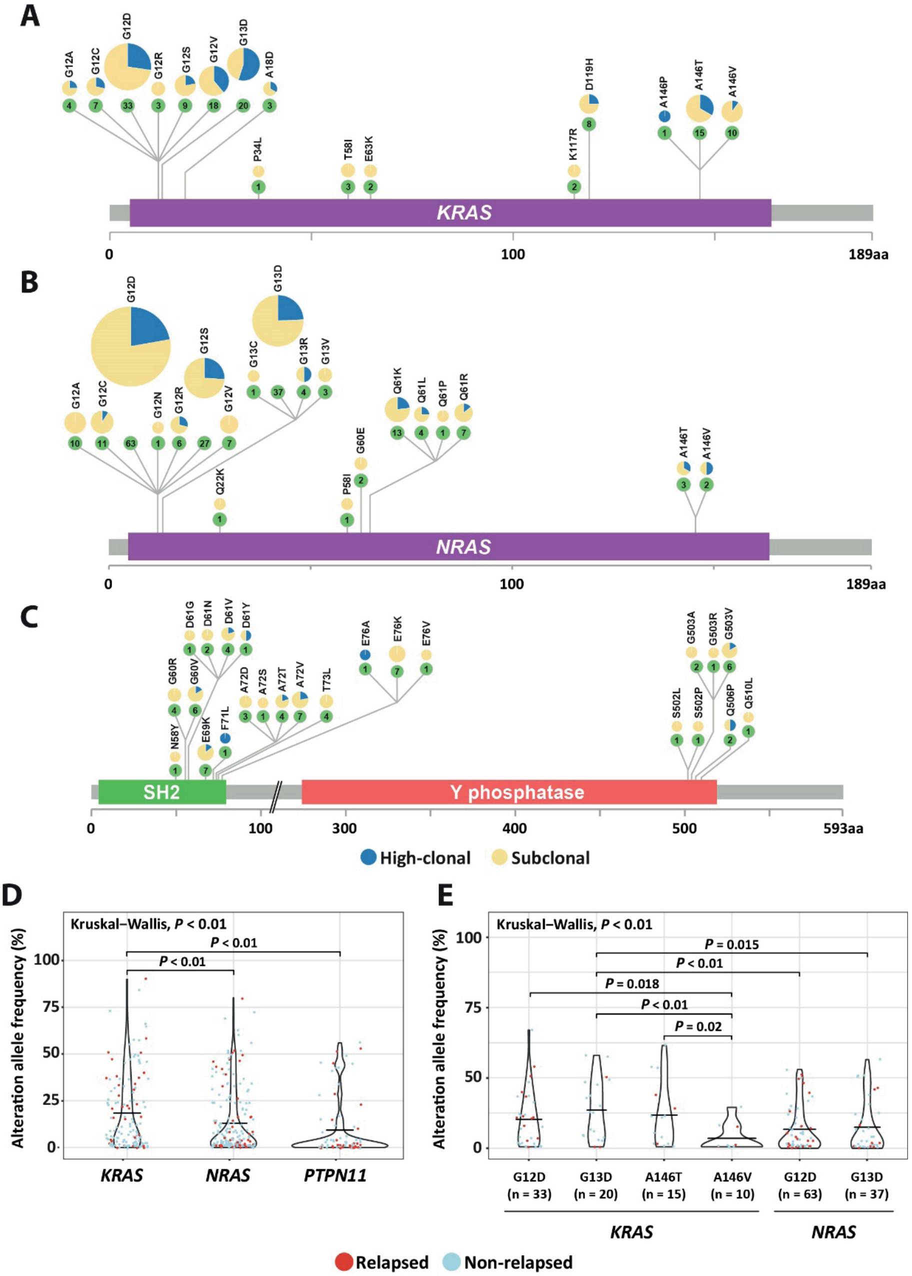


Figure 3.3. Potency of RAS pathway mutations to drive clonal expansion. **(A-C)** Schematic representation of *KRAS* **(A)**, *NRAS* **(B)** and *PTPN11* **(C)** indicating the prevalence of common hotspot mutations. **(D)** Violin plot showing allele frequency in hotspot mutations of three investigated RAS pathway genes. The median allele frequency was significantly higher in *KRAS*, indicating a high potential of *KRAS* hotspot mutations to drive clonal expansion. **(E)** Violin plot showing allele frequencies in frequent *KRAS* and *NRAS* hotspot mutations. The median allele frequency was significantly higher in *KRAS* p.G13D, suggesting their higher potential to drive clonal expansion compared to other RAS hotspot mutations.

Relevance of gene alterations to relapse development

The high number of alterations in these relapse-associated genes at the time of diagnosis triggers the hypothesis that these could be used as prognostic biomarkers for relapse development, even when present at subclonal levels. To test this hypothesis, we first explored whether alterations in each of the eight genes were enriched in diagnostic samples from patients who subsequently relapsed compared to diagnostic samples from patients who did not relapse. In general, subclonal alterations were very common at primary diagnosis in patients who relapsed (60/82; 73%) as well as in patients who did not (165/203; 81%). For high-clonal alterations, we only observed a higher percentage of relapse development in cases with *IKZF1* deletions compared to wild-type cases, whereas an association with relapse development was not observed for diagnostic samples with subclonal alterations in any of the genes, including *IKZF1* (**Figure 3.4**). Furthermore, patients with high-clonal *IKZF1* 4-7 deletions were more often classified as having high minimal residual disease (MRD; $>5\times10^{-4}$ at day 79 or 84 after start of the treatment) in both representative ALL9 and ALL10 cohorts (Fisher exact test, $P < 0.01$ and $P < 0.05$, respectively), compared to patients without an *IKZF1* deletion (**Table S3.10**).

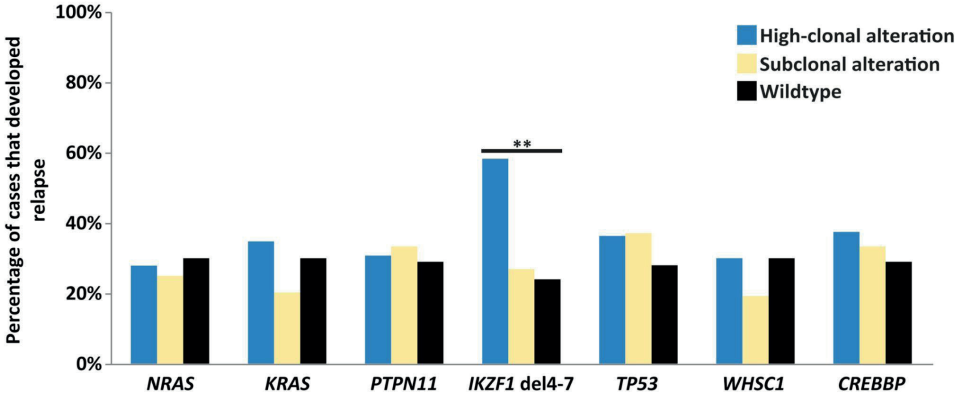
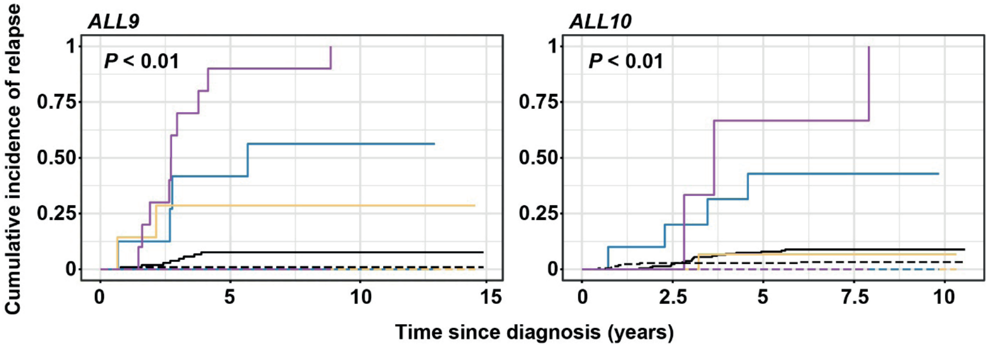


Figure 3.4. Prevalence of relapse-associated genomic alterations at diagnosis. Bar plot showing the percentage of relapses in cases with high-clonal (blue) or subclonal (yellow) mutations in seven relapse-associated genes, and in cases that were wild-type (black) for these genes. Only cases with high-clonal *IKZF1* 4-7 deletions showed a significantly higher percentage of relapse development compared to wild-type cases (Fisher exact test, $P < 0.01$) (**Table S3.9**).

The CIR at 5 years was 41.7% (SE 0.04%) and 42.9% (SE 0.03%) in patients with high-clonal *IKZF1* 4-7 deletions treated according to the ALL9 and ALL10 protocols, respectively (**Figure 3.5**). The cause-specific hazard ratio (HR_{CS}) in the two representative cohorts ($n = 376$), estimated with a univariate Cox proportional hazards regression model, revealed an association of high-clonal *IKZF1* exon 4-7 deletions with relapse (HR = 7.22; 95% CI: 3.27-15.95; $P < 0.01$). In the multivariate Cox model, in which age at diagnosis, gender and MRD status were included, the adjusted HR_{CS} was 3.6 (95% CI: 1.38-9.55; $P < 0.01$) (**Tables 3.1** and **S3.11**). These data are in line with those from earlier studies on these cohorts in which all *IKZF1* deletions were included[12,24]. However, when we assessed the clinical relevance of subclonal alterations for relapse development in *IKZF1*, or any of the other genes, Cox regression analysis revealed no significant associations in the combined ALL9 and ALL10 cohorts compared to wild-type cases (**Tables 3.1** and **S3.11**), and the CIR was similar in the two groups (**Figure 3.5**). Furthermore, patients with subclonal *IKZF1* 4-7 deletions did not have significantly different levels of MRD compared to *IKZF1* wild-type patients (**Table S3.10**). Since previous studies have shown a lack of association of *IKZF1* deletion with relapse in patients who carry a deletion in *ERG*[34,35], we used MLPA to test whether there was an enrichment of *ERG* deletions in cases with subclonal *IKZF1* exon 4-7 deletions compared to those with clonal *IKZF1* exon 4-7, but these deletions were infrequent in both groups (**Table S3.12**).



Number at risk

⬛	106	96	87	6	217	206	187	144	31
🟦	8	4	3	0	10	8	5	4	0
🟨	7	5	4	0	15	15	14	12	2
🟪	10	1	0	0	3	3	1	1	0

Figure 3.5. Cumulative incidence of relapse for high-clonal and subclonal *IKZF1* deletions. The cumulative incidence of relapse (CIR) was estimated using a competing-risk model with death as a competing event. CIR plots are presented for the representative ALL9 (left) and ALL10 (right) cohorts. Lines represent the *IKZF1* deletion status and include wild-type (black line), subclonal exon 4-7 deletion (yellow), other high-clonal deletion (purple), and high-clonal exon 4-7 deletion (blue). Straight lines depict relapses and dotted lines death. *P*-values shown are obtained by employing the Gray test to compare CIR curves. The 5-year CIR was higher in cases with high-clonal *IKZF1* deletions, compared to wild-type cases in both representative cohorts.

Table 3.1. Cox regression analysis in combined representative ALL9 and ALL10 cohort (*n* = 376).

	Status	Number of patients	Univariate Cox regression[1]	Multivariate Cox regression[1]
IKZF1	High-clonal[3]	18	$P < 0.01$; HR = 7.22 [3.27-15.95]	$P < 0.01$; HR = 3.6 [1.38-9.55]
	Subclonal	22	$P = 0.39$; HR = 1.69 [0.51-5.57]	$P = 0.34$; HR = 1.8 [0.53-6.24]
	Other high-clonal[4]	13	$P < 0.01$; HR = 19.92 [9.76-40.66]	$P < 0.01$; HR = 13 [5.68-29.79]
MRD	Low	111	1 (Ref)	1 (Ref)
	Medium	227	$P = 0.1$; HR = 2 [0.88-4.61]	$P = 0.28$; HR = 1.6 [0.67-3.78]
	High	22	$P < 0.01$; HR = 10.85 [4.16-28.28]	$P < 0.01$; HR = 5 [1.64-15.23]
Age at diagnosis	0-4	173	1 (Ref)	1 (Ref)
	5-9	133	$P = 0.34$; HR = 1.34 [0.73-2.46]	$P = 0.6$; HR = 1.2 [0.64-2.25]
	10-14	47	$P = 0.83$; HR = 0.9 [0.34-2.38]	$P = 0.6$; HR = 0.8 [0.27-2.12]
	15-18	23	$P = 0.3$; HR = 1.76 [0.6-5.12]	$P = 0.3$; HR = 0.34 [0.04-2.66]
Gender	Male	215	1 (Ref)	1 (Ref)
	Female	161	$P = 0.06$; HR = 0.56 [0.31-1.03]	$P = 0.08$; HR = 0.53 [0.26-1.1]

[1]The hazard ratio is given with a 95% confidence interval. Multivariate Cox regression analysis included gender, age at diagnosis and minimal residual disease status as covariates; [2]Analysis was done on the combined cohort stratified on treatment protocol. Representative ALL9 and ALL10 cohorts are outlined in **Table S3.1**. [3]Clonality status based on detection at initial diagnosis; high-clonal: AF ⩾ 25%, subclonal: AF < 25%. [4]Detectable using multiplex ligation-dependent probe amplification assay. A full overview of all comparisons is given in **Table S3.11**. MRD: minimal residual disease.

Tracing of major and minor clone mutations at the time of relapse

To obtain further insight into the clinical relevance of the identified alterations in relapse development, we investigated whether these were preserved in the cases that relapsed. For this analysis, we used all 146 cases that later developed a relapse, of which 82 carried alterations in a major or minor clone in one or more of the genes (**Tables S3.13** and **S3.14**). Overall, we found that for most genes at the time of diagnosis the frequency of subclonal alterations was similar or slightly higher compared to that of the alterations detected in a major clone (**Figure S3.3A**, **Tables S3.13** and **S3.14**).

We collected 73 relapse samples from patients who carried these major or minor clone alterations at the time of diagnosis (89%), which enabled us to trace 171 of the 185 sequence mutations, and 25 of the *IKZF1* exon 4-7 deletions. We did not assess whether mutations detected at diagnosis were still preserved in minor clones at relapse, since these clones were unlikely to be true relapse drivers. Overall, 56% (22/39) of the tested major clone mutations were found to be preserved in the major clone at relapse, whereas the value for the subclonal mutations was 7% (9/132) (Fisher exact test, $P < 0.01$) (**Figures 3.6**, **S3.3B** and **Table S3.7**). For *IKZF1* exon 4-7 deletions, the difference was even more striking. Here, the presence of deletions was studied in 19 available relapse samples using breakpoint-spanning PCR, followed by Sanger sequencing to confirm that the breakpoint sequences were identical at diagnosis and relapse (**Figures 3.6**, **S3.3B**, **Tables S3.6** and **S3.7**). All major clone *IKZF1* exon 4-7 deletions were found to be preserved in the major clone at the time of relapse (*n* = 12), which is in agreement with earlier findings and illustrates their relevance to relapse development in these treatment protocols[12,24]. In contrast, none of the subclonal exon 4-7 deletions in *IKZF1* (*n* = 13) was preserved in either the major or a minor clone at relapse. Collectively, the data from the present study indicate that these deletions, when present at initial diagnosis at a subclonal level, do not drive relapse in pediatric ALL.

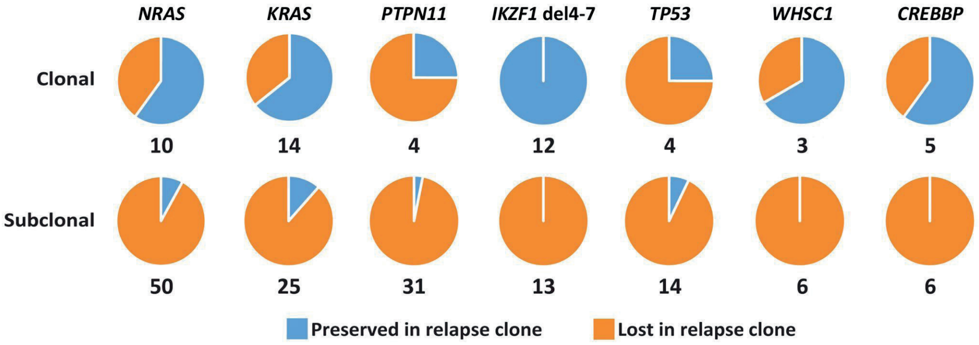


Figure 3.6. Preservation of clonal and subclonal mutations at the time of relapse. Tracing of major clone (top) and subclonal (bottom) alterations detected in initial diagnosis samples from relapsed patients in the matched relapse samples. The pie charts depict the fractions of preserved (blue) and lost (orange) alterations at the time of relapse.

DISCUSSION

ALL is a heterogeneous disease in which specific genomic alterations show strong associations with relapse risk and outcome. In this study, we assessed the clinical relevance and prognostic value of subclonal alterations in eight genes frequently mutated in relapsed B-cell precursor ALL in a cohort of 503 diagnostic samples. Our data demonstrate that subclonal alterations in these genes are very common at the time of diagnosis, but that these mutations do not provide a basis for risk stratification in pediatric ALL. This finding is particularly relevant for *IKZF1* alterations, which are currently used or implemented for treatment stratification in multiple upfront treatment protocols[14,20].

The selection of these genes was made based on enriched mutation frequencies in relapse found in previous studies. Of all alterations identified in this study, 75% were subclonal at diagnosis, suggesting that these relapse-associated gene mutations accumulate during progression of the leukemia before the initial diagnosis, thereby increasing the clonal complexity. Whereas seven of the genes selected in our study showed this high mutational burden at diagnosis, both in terms of numbers and level of clonality, we identified only a single, not previously reported, subclonal *NT5C2* mutation in a non-relapsed case (follow-up time 9.5 years). *NT5C2* encodes the cytosolic nucleotidase, which is responsible for inactivating cytotoxic thiopurine monophosphate nucleotides, and activating mutations in this gene are recurrently found in relapsed ALL, mainly T-cell ALL[4,9,36-38]. One explanation for the low number of activating *NT5C2* mutations at diagnosis is that these mutations decrease cell fitness, and only obtain their selective advantage during treatment with thiopurine[36]. If already present at the time of initial diagnosis, these mutations are usually detectable in only a very small subset of cells, far below the detection level of our smMIP analysis[36].

Hotspot RAS pathway mutations have been detected in nearly half of the cases, often of the hyperdiploid subtype, and their frequency and clonal burden varied between the different mutations. In our study, we used this variability to compare the potential of different hotspot mutations to drive clonal expansion under physiological conditions. Compared to diagnosis, we observed a less diverse spectrum of *KRAS* and *NRAS* hotspot mutations in relapse, with p.G12D, p.G12V and p.G13D together accounting for two-thirds of *KRAS* and *NRAS* hotspot mutations found in relapse-fated clones. Studies in other cancers have demonstrated that the prevalence of different RAS pathway mutations varies depending on the type of cancer and tissue of origin, with *KRAS* mutations p.G12D, p.G12V, p.G13D and p.G12C being among the most common ones[39,40]. Comparison of oncogenic capacities of different RAS hotspots has also been performed using *in vitro* and *in vivo* modeling studies, focusing primarily on *KRAS*. These studies identified *KRAS* mutations p.G12D, p.G12V and p.G13D as having higher proliferative and transforming potential compared to other common hotspots in various tumors of epithelial origin[39,41,42]. Our data indicate that in competition of multiple RAS hotspot mutations, some of these not only confer a proliferative advantage but can also more effectively sustain a treatment-induced selective sweep[4,10].

The presence of *IKZF1* deletions has been shown to be associated with relapse and survival in multiple clinical ALL studies[12,14-19], and these deletions have been described to play a role in resistance to tyrosine kinase inhibitors and glucocorticoids[43-46]. Therefore, with the advance of more sensitive detection techniques, the question of whether subclonal alterations are also associated with relapse is very relevant, both from biological and from clinical perspectives. We here demonstrate that, in contrast to major clone *IKZF1* exon 4-7 deletions, cases that carry this deletion only in a subset of the cells do not show an association with relapse. Moreover, whereas all major clone exon 4-7 deletions were preserved in cases that relapsed, none of the relapses from cases with subclonal exon 4-7 deletions at diagnosis carried this deletion. Importantly, the majority of subclonal deletions had allele frequencies below 10% (**Table S3.7**). Therefore, since a threshold to distinguish subclonal from major clone deletions is difficult, deletions closer to our threshold of 25% should be evaluated with caution. Nevertheless, the difference between major and minor clone *IKZF1* 4-7 deletions is striking, and the reason behind this remains unclear. Possibly, the functional impact of full-clonal *IKZF1* deletions, which arise early during leukemia development, is different from that of deletions that occur in later stages when the leukemia has already expanded. Other deletions in *IKZF1* show much less clustering in their breakpoints and, therefore, screening for these subclonal deletions in diagnostic samples is much less efficient. We did not, therefore, directly assess the stability and potential prognostic importance of whole gene and rare intragenic *IKZF1* deletions. However, a previous study showed that other *IKZF1* deletion subtypes have similar prognostic relevance as exon 4-7 deletions[31], suggesting that subclonal alterations in these other *IKZF1* deletions may show the same lack of association.

In summary, we show that subclonal alterations in the relapse-associated genes *IKZF1*, *CREBBP*, *KRAS*, *NRAS*, *PTPN11*, *TP53*, and *WHSC1* in pediatric ALL are frequently present at initial diagnosis, often at a subclonal level. At relapse, however, most of these subclonal mutations are lost, suggesting that their selective advantages over wild-type clones during treatment is limited. This finding has direct implications for clinical practice, particularly in the case of *IKZF1*, where deletion status is used for routine risk stratification. We conclude that, at least for the investigated set of genes, there is no basis for the use of subclonal alterations at initial diagnosis as a prognostic marker.

SUPPLEMENTARY DATA

Supplementary material and methods

smMIP-based sequencing and variant calling

In order to accurately detect subclonal alterations in diagnosis samples, a total of 166 smMIP probes was designed to cover the hotspot regions of the genes *CREBBP*, *PTPN11*, *NT5C2,* and *WHSC1*, and coding regions of *TP53, KRAS* and *NRAS*, seven genes that are frequently mutated in relapsed ALL (**Table S3.4**). All genomic regions of interest were covered by at least two probes, preferably covering both the sense and antisense strands. smMIP-based sequencing was performed as previously described using paired-end sequencing on an Illumina NextSeq 500 Desktop Sequencer (Illumina, San Diego, CA, USA), after which smMIP-based consensus variant calling was performed using SeqNext software (JSI) version 4.2.5, as previously described[23]. Mutant allele frequencies were corrected based on the blast percentage of the sample. We achieved an average on-target read depth of 22,985 raw reads per probe (**Figure S3.1A**). After removing random errors present in less than 70% of raw reads, consensus reads were formed from reads with the same unique molecular identifier. We achieved an average on-target depth of 308 (median 141) unique-capture-based consensus reads per probe (**Figure S3.1B**), with multiple probes overlapping on hotspot regions[23]. After exclusion of the variants called from repetitive regions, poorly performing probes, variants called from less than 2 independent probes and variants called by less than 5 unique reads in one consensus read, a total of 7,836 variants remained with an average on-target depth of 1,419 consensus reads. We further filtered out variants present in an in-house database of the Radboud University Medical Center (Nijmegen, the Netherlands) containing exomes from 20,000 individuals[47] and variants predicted as non-pathogenic (synonymous, phyloP < 2.5, CADD score < 15) (**Table S3.3**). For the final list of variants, a correction of the mutant allele frequency was made based on the percentage of blast cells determined at the time of diagnosis, which was high for the majority of cases (>70% for 93% of cases). In a low number of cases an aneuploidy required a correction of mutant allele frequency for copy number as well, which involved *WHSC1* (17 mutations), *TP53* (10 mutations), *KRAS* (6 mutations), *NRAS* (5 mutations), *PTPN11* (3 mutations) and *CREBBP* (1 mutation) (**Table S3.5**).

IKZF1 deletion detection

IKZF1 status was assessed using the Multiplex ligation probe assay (MLPA) SALSA P335 ALL-IKZF1 kit (MRC-Holland, Amsterdam, The Netherlands), according to manufacturer's instructions and as described before[12,24]. Additionally, *IKZF1* 4-7 deletions were assessed using real-time quantitative PCR. Primers covering the breakpoint clusters in introns 3 and 7 were designed using Primer3 software version 0.4.0. Quantitative PCR (qPCR) was performed using an IQ SYBR Green supermix (Bio-Rad, Hercules, CA, USA) according to manufacturer's instructions. The primer sequences used for qPCR were: 5'-CTCCCAGCCCATAGGGTATAA-3' (forward) and 5'-GTTAAATAAAGAACCCTCAGGCATT-3' (reverse). The sensitivity of the qPCR assay was tested using dilution series of a sample with a high tumor load (96% blasts) and a full-clonal *IKZF1* 4-7 deletions (detected by MLPA). All qPCR reactions were performed in duplicate, and the percentage of cells with *IKZF1* exon 4-7 deletions was calculated based on the dilution series of the control sample, and a correction was made based on the percentage of blast cells determined at the time of diagnosis. For every sample with a clonal or subclonal *IKZF1* exon 4-7 deletion, PCR products were sequenced with both forward and reverse primers using Sanger sequencing, after which the sequences were mapped to the reference genome (hg19) to determine the exact breakpoint positions and unique interstitial sequences (**Figure S3.1C** and **Table 3.6**).

Defining clonal and subclonal alterations

Previous studies have used allele frequency (AF) thresholds to define mutations as being present in a minor subclonal or major clone (clonal) ranging between 20% and 30%[4,10,48,49]. In this study, we used an AF threshold of 25% to separate major clonal from subclonal alterations, since mutations below 25% represent a minor cell fraction. The same threshold was used for defining clonal *IKZF1* 4-7 deletions, which correlates with deletions that are detectable using MLPA.

Cox regression analysis

Multivariate Cox regression model was estimated including all covariates significant in the univariate model, as well as age at diagnosis, gender and MRD as clinically relevant covariates. Multivariable model including combined ALL9 and ALL10 cohorts was stratified based on the treatment protocol. Proportional hazard assumption was checked by visual inspection of Schoenfeld Residuals. The score test was used to test violation of the proportional hazard assumption for each variable[50]. We did not identify violation of the proportional hazard assumption for any of the tested covariates. Potential multicolinearity was inspected using variance inflation factor (VIF). Univariate interactions were inspected for each of the tested covariates and reported in the **Table S3.11**.

Supplementary Tables

Supplementary Tables are available online using the following link: https://doi.org/10.3324/haematol.2020.259226.

Supplementary Figures

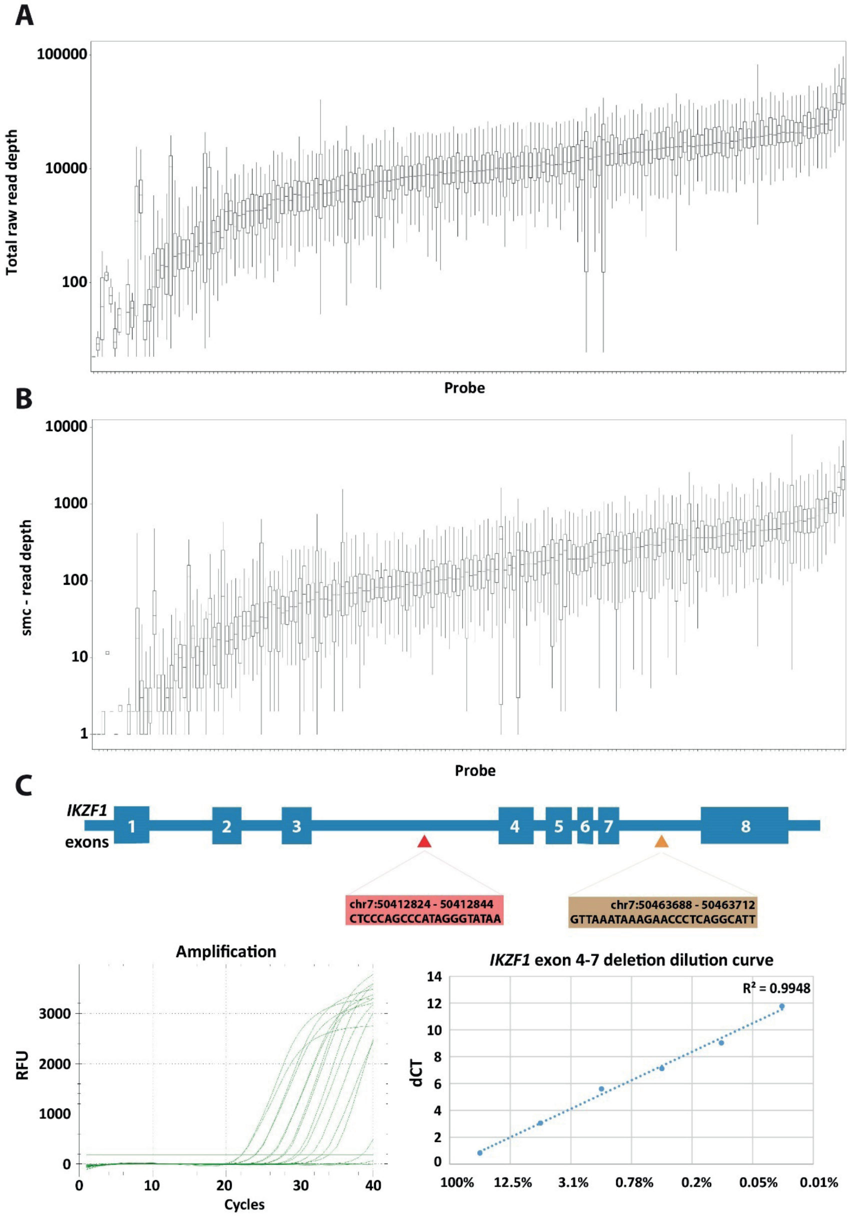


Figure S3.1. (A) Box plot showing total raw read depth per probe. **(B)** Box plots showing unique-capture-based consensus reads (smc) depth per probe. **(C)** Schematic representation of the *IKZF1* gene, indicating the position of the common breakpoint cluster in introns 3 and 7 (top panel). The bottom left panel shows a representative example of qPCR amplification curves of dilution series and the bottom right panel displays the dilution curve from the same experiment indicating high correlation ($R^2 = 0.9948$) for quantifiable samples.

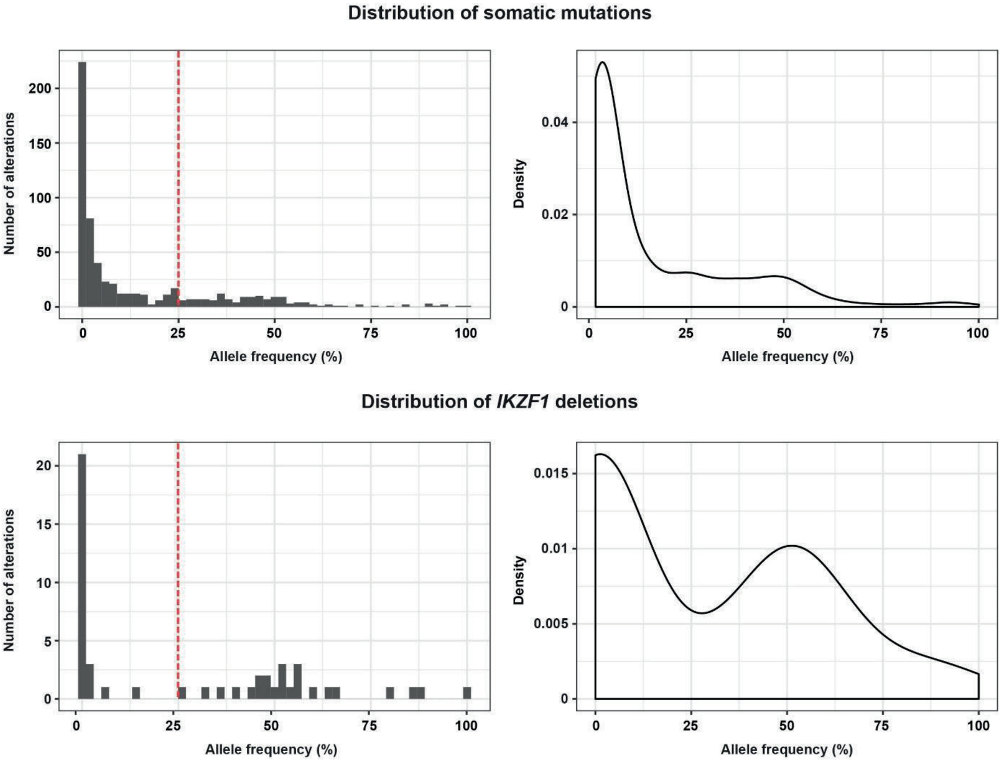


Figure S3.2. Histogram (left panel) and density plot (right panel) showing the distribution of the mutant allele frequency in cases with mutations and *IKZF1* 4-7 deletions. Allele frequency (AF) of 25% was chosen as the threshold for clonal versus subclonal mutations (indicated in red). Mutations above this threshold are present in more than 50% of the cells and thus always represent a major clone. As depicted in these panels the majority of (subclonal) mutations and deletions have AF below 10%.

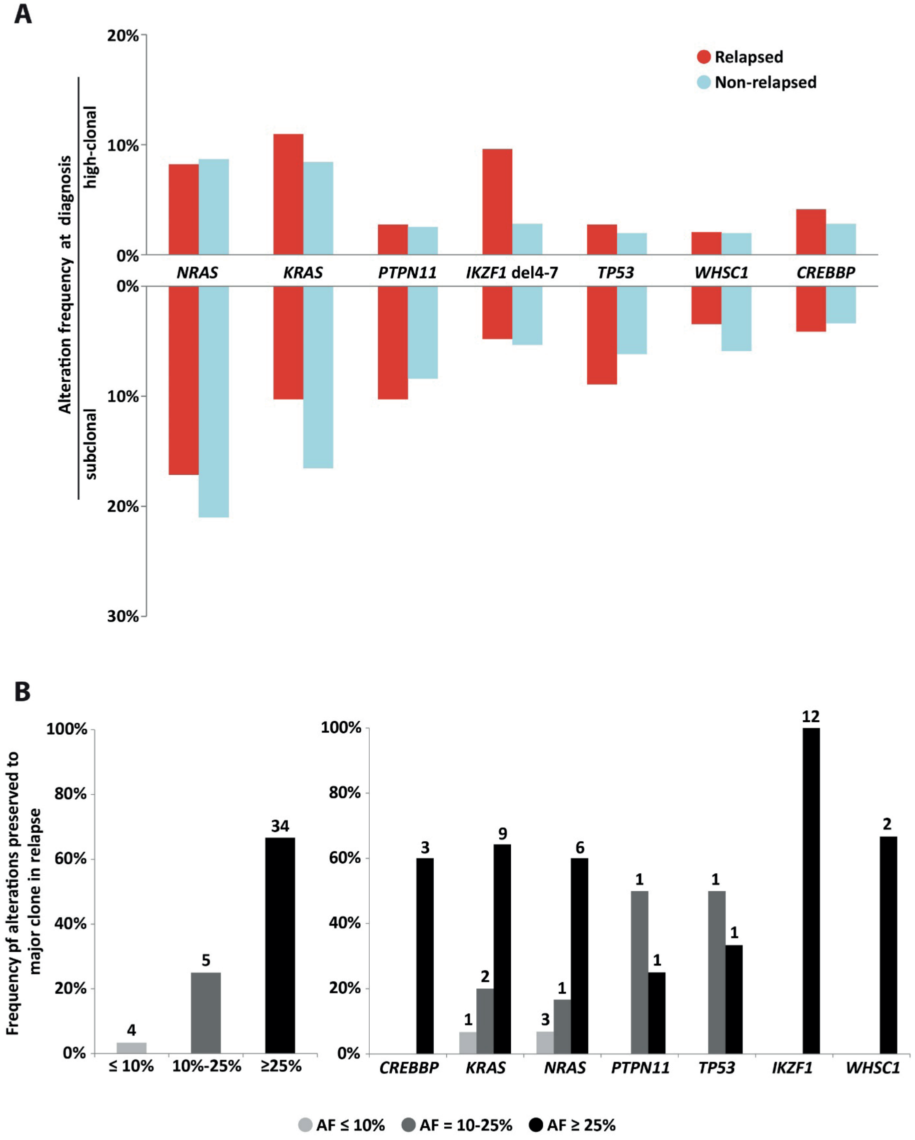


Figure S3.3. (A) Bar plot showing alteration frequencies at diagnosis for relapsed (red) and non-relapsed cases (blue). Major clone alterations are shown upwards, while subclonal are presented downwards. **(B)** Bar plot indicating fraction of alterations preserved to major clones in available relapse samples for all genes tested (left panel) and per gene (right panel). Samples are grouped based on the alteration allele frequency detected in diagnosis. Absolute numbers for each group are shown on the top of the bars (**Table S3.7**).

FUNDING

This work was supported by grants from Stichting Kinderen Kankervrij (KIKA 150 to RPK), Stichting Bergh in het Zadel (to RPK) and the China Scholarship Council (CSC201304910347 to JY).

CONFLICTS OF INTEREST

No conflicts of interest to disclose.

CONTRIBUTIONS

RPK, EW, FvL, and AGvK conceived the study. RPK, EW, ŽA and JY designed the study. PMH and RP are responsible for the clinical outcome data of patients included in the study. ŽA, JY, SVvR, AvD, LD and WHS performed experiments and analyzed the data. ŽA and MF performed statistical analyses. ŽA, JY and RPK wrote the manuscript and created the figures and tables. ES provided samples and clinical information. All authors critically reviewed the manuscript and approved the final submitted manuscript.

ACKNOWLEDGMENTS

We thank Radboud University Medical Center, Department of Human Genetics for bioinformatic support in data analysis.

REFERENCES

1. Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the Children's Oncology Group. J Clin Oncol. 2012;30(14):1663-1669.
2. Pieters R, Groot-Kruseman Hd, Velden VVd, Fiocco M, Berg Hvd, Bont Ed, et al. Successful therapy reduction and intensification for childhood acute lymphoblastic leukemia based on minimal residual disease monitoring: study ALL10 from the Dutch Childhood Oncology Group. J Clin Oncol. 2016;34(22):2591-2601.
3. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature. 2011;469(7330):356-361.
4. Ma X, Edmonson M, Yergeau D, Muzny DM, Hampton OA, Rusch M, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun. 2015;6(1):6604.
5. Waanders E, Gu Z, Dobson SM, Antić Ž, Crawford JC, Ma X, et al. Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. Blood Cancer Discov. 2020;1(1):96-111.
6. Malinowska-Ozdowy K, Frech C, Schönegger A, Eckert C, Cazzaniga G, Stanulla M, et al. KRAS and CREBBP mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia. Leukemia. 2015;29(8):1656-1667.
7. Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011;471(7337):235-239.
8. Irving J, Matheson E, Minto L, Blair H, Case M, Halsey C, et al. Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. Blood. 2014;124(23):3420-3430.
9. Tzoneva G, Perez-Garcia A, Carpenter Z, Khiabanian H, Tosello V, Allegretta M, et al. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. Nat Med. 2013;19(3):368-371.
10. Jerchel IS, Hoogkamer AQ, Ariës IM, Steeghs EMP, Boer JM, Besselink NJM, et al. RAS pathway mutations as a predictive biomarker for treatment adaptation in pediatric B-cell precursor acute lymphoblastic leukemia. Leukemia. 2018;32(4):931-940.
11. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009;360(5):470-480.
12. Kuiper RP, Waanders E, van der Velden VH, van Reijmersdal SV, Venkatachalam R, Scheijen B, et al. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. Leukemia. 2010;24(7):1258-1264.
13. Jaffe JD, Wang Y, Chan HM, Zhang J, Huether R, Kryukov GV, et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nat Genet. 2013;45(11):1386-1391.
14. van der Veer A, Waanders E, Pieters R, Willemse ME, Van Reijmersdal SV, Russell LJ, et al. Independent prognostic value of BCR-ABL1-like signature and IKZF1 deletion, but not high CRLF2 expression, in children with B-cell precursor ALL. Blood. 2013;122(15):2622-2629.
15. Hinze L, Möricke A, Zimmermann M, Junk S, Cario G, Dagdan E, et al. Prognostic impact of IKZF1 deletions in association with vincristine–dexamethasone pulses during maintenance treatment of childhood acute lymphoblastic leukemia on trial ALL-BFM 95. Leukemia. 2017;31(8):1840-1842.

16. Clappier E, Grardel N, Bakkus M, Rapion J, De Moerloose B, Kastner P, et al. IKZF1 deletion is an independent prognostic marker in childhood B-cell precursor acute lymphoblastic leukemia, and distinguishes patients benefiting from pulses during maintenance therapy: results of the EORTC Children's Leukemia Group study 58951. Leukemia. 2015;29(11):2154-2161.
17. Ofverholm I, Tran AN, Heyman M, Zachariadis V, Nordenskjöld M, Nordgren A, et al. Impact of IKZF1 deletions and PAX5 amplifications in pediatric B-cell precursor ALL treated according to NOPHO protocols. Leukemia. 2013;27(9):1936-1939.
18. Dörge P, Meissner B, Zimmermann M, Möricke A, Schrauder A, Bouquin JP, et al. IKZF1 deletion is an independent predictor of outcome in pediatric acute lymphoblastic leukemia treated according to the ALL-BFM 2000 protocol. Haematologica. 2013;98(3):428-432.
19. Asai D, Imamura T, Suenobu S, Saito A, Hasegawa D, Deguchi T, et al. IKZF1 deletion is associated with a poor outcome in pediatric B-cell precursor acute lymphoblastic leukemia in Japan. Cancer Med. 2013;2(3):412-419.
20. Yeoh AEJ, Lu Y, Chin WHN, Chiew EKH, Lim EH, Li Z, et al. Intensifying treatment of childhood B-lymphoblastic leukemia with IKZF1 deletion reduces relapse and improves overall survival: results of Malaysia-Singapore ALL 2010 study. J Clin Oncol. 2018;36(26):2726-2735.
21. Absalan F, Ronaghi M. Molecular inversion probe assay. Methods Mol Biol. 2007;396:315-330.
22. Berglund EC, Lindqvist CM, Hayat S, Övernäs E, Henriksson N, Nordlund J, et al. Accurate detection of subclonal single nucleotide variants in whole genome amplified and pooled cancer samples using HaloPlex target enrichment. BMC Genom. 2013;14(1):856.
23. Yu J, Antić Ž, van Reijmersdal SV, Hoischen A, Sonneveld E, Waanders E, et al. Accurate detection of low-level mosaic mutations in pediatric acute lymphoblastic leukemia using single molecule tagging and deep-sequencing. Leuk Lymphoma. 2018;59(7):1690-1699.
24. Waanders E, van der Velden VH, van der Schoot CE, van Leeuwen FN, van Reijmersdal SV, de Haas V, et al. Integrated use of minimal residual disease classification and IKZF1 alteration status accurately predicts 79% of relapses in pediatric acute lymphoblastic leukemia. Leukemia. 2011;25(2):254-258.
25. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Control Clin Trials. 1996;17(4):343-346.
26. Yu J, Waanders E, van Reijmersdal SV, Antić Ž, van Bosbeek CM, Sonneveld E, et al. Upfront treatment influences the composition of genetic alterations in relapsed pediatric B-cell precursor acute lymphoblastic leukemia. Hemasphere. 2020;4(1):e318.
27. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med. 2007;26(11):2389-2430.
28. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat. 1988;16(3):1141-1154.
29. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1.
30. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401-404.
31. Boer JM, van der Veer A, Rizopoulos D, Fiocco M, Sonneveld E, de Groot-Kruseman HA, et al. Prognostic value of rare IKZF1 deletion in childhood B-cell precursor acute lymphoblastic leukemia: an international collaborative study. Leukemia. 2016;30(1):32-38.

32. Caye A, Beldjord K, Mass-Malo K, Drunat S, Soulier J, Gandemer V, et al. Breakpoint-specific multiplex polymerase chain reaction allows the detection of IKZF1 intragenic deletions and minimal residual disease monitoring in B-cell precursor acute lymphoblastic leukemia. Haematologica. 2013;98(4):597-601.
33. Paulsson K, Lilljebjörn H, Biloglav A, Olsson L, Rissler M, Castor A, et al. The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. Nat Genet. 2015;47(6):672-676.
34. Clappier E, Auclerc MF, Rapion J, Bakkus M, Caye A, Khemiri A, et al. An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. Leukemia. 2014;28(1):70-77.
35. Zaliova M, Potuckova E, Hovorkova L, Musilova A, Winkowska L, Fiser K, et al. ERG deletions in childhood acute lymphoblastic leukemia with DUX4 rearrangements are mostly polyclonal, prognostically relevant and their detection rate strongly depends on screening method sensitivity. Haematologica. 2019;104(7):1407-1416.
36. Tzoneva G, Dieck CL, Oshima K, Ambesi-Impiombato A, Sánchez-Martín M, Madubata CJ, et al. Clonal evolution mechanisms in NT5C2 mutant-relapsed acute lymphoblastic leukaemia. Nature. 2018;553(7689):511-514.
37. Meyer JA, Wang J, Hogan LE, Yang JJ, Dandekar S, Patel JP, et al. Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. Nat Genet. 2013;45(3):290-294.
38. Pieters R, Huismans DR, Loonen AH, Peters GJ, Hählen K, van der Does-van den Berg A, et al. Relation of 5'-nucleotidase and phosphatase activities with immunophenotype, drug resistance and clinical prognosis in childhood leukemia. Leuk Res. 1992;16(9):873-880.
39. Miller M, Miller L. RAS mutations and oncogenesis: not all RAS mutations are created equally. Front genet. 2012;2:100.
40. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. Cancer Res. 2012;72(10):2457-2467.
41. Winters IP, Chiou S-H, Paulk NK, McFarland CD, Lalgudi PV, Ma RK, et al. Multiplexed in vivo homology-directed repair and tumor barcoding enables parallel quantification of Kras variant oncogenicity. Nat Commun. 2017;8(1):2053.
42. Stolze B, Reinhart S, Bulllinger L, Fröhling S, Scholl C. Comparative analysis of KRAS codon 12, 13, 18, 61 and 117 mutations using human MCF10A isogenic cell lines. Sci Rep. 2015;5(1):8535.
43. Marke R, Havinga J, Cloos J, Demkes M, Poelmans G, Yuniati L, et al. Tumor suppressor IKZF1 mediates glucocorticoid resistance in B-cell precursor acute lymphoblastic leukemia. Leukemia. 2016;30(7):1599-1603.
44. Imamura T, Yano M, Asai D, Moriya-Saito A, Suenobu SI, Hasegawa D, et al. IKZF1 deletion is enriched in pediatric B-cell precursor acute lymphoblastic leukemia patients showing prednisolone resistance. Leukemia. 2016;30(8):1801-1803.
45. Churchman ML, Low J, Qu C, Paietta EM, Kasper LH, Chang Y, et al. Efficacy of retinoids in IKZF1-mutated BCR-ABL1 acute lymphoblastic leukemia. Cancer Cell. 2015;28(3):343-356.
46. van der Veer A, Zaliova M, Mottadelli F, De Lorenzo P, Te Kronnie G, Harrison CJ, et al. IKZF1 status as a prognostic feature in BCR-ABL1-positive childhood ALL. Blood. 2014;123(11):1691-1698.
47. Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, Mensenkamp AR, et al. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. Hum Mutat. 2013;34(12):1721-1726.

48. Spinella JF, Richer C, Cassart P, Ouimet M, Healy J, Sinnett D. Mutational dynamics of early and late relapsed childhood ALL: rapid clonal expansion and long-term dormancy. Blood Adv. 2018;2(3):177-188.
49. Agraz-Doblas A, Bueno C, Bashford-Rogers R, Roy A, Schneider P, Bardini M, et al. Unraveling the cellular origin and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis. Haematologica. 2019;104(6):1176-1188.
50. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515-526.

CHAPTER 4

CLONAL DYNAMICS IN PEDIATRIC B-CELL PRECURSOR ACUTE LYMPHOBLASTIC LEUKEMIA WITH VERY EARLY RELAPSE

*Željko Antić[1,2], Jiangyan Yu[1,3], Beat C. Bornhauser[4], Stefan H. Lelieveld[1], Cedric G. van der Ham[1], Simon V. van Reijmersdal[1,3], Lionel Morgado[1], Sarah Elitzur[5], Jean-Pierre Bourquin[4], Giovanni Cazzaniga[6], Cornelia Eckert[7], Mireia Camós[8,9,10], Rosemary Sutton[11], Hélène Cavé[12,13], Anthony V. Moorman[14], Edwin Sonneveld[1,15], Ad Geurts van Kessel[3], Frank N. van Leeuwen[1], Peter M. Hoogerbrugge[1,15], Esmé Waanders[1,16] and Roland P. Kuiper[1,16]**

[1]Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands
[2]Department of Human Genetics, Hannover Medical School, Hannover, Germany
[3]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
[4]Department of Oncology and Children's Research Centre, University Children's Hospital Zurich, Zurich, Switzerland
[5]Pediatric Hematology-Oncology, Schneider Children's Medical Center and Sackler Faculty of Medicine, Tel Aviv University, Israel
[6]Centro Ricerca Tettamanti, Fondazione Tettamanti, University of Milan Bicocca, Monza, Italy
[7]Pediatric Oncology/Hematology, Charité - Universitätsmedizin Berlin, Berlin, Germany
[8]Leukemia and other pediatric hemopathies, Developmental Tumor Biology Group, Institut de Recerca Sant Joan de Déu, Santa Rosa 39-57, 08950, Esplugues de Llobregat, Spain.
[9]Hematology Laboratory, Hospital Sant Joan de Deu Barcelona, Passeig Sant Joan de Déu 2, 08950, Esplugues de Llobregat, Spain.
[10]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain
[11]Molecular Diagnostics, Children's Cancer Institute, University of New South Wales, Sydney, NSW, Australia
[12]Department of Genetics, Robert Debré Hospital, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France
[13]INSERM U1131, Saint-Louis Research Institute, University of Paris, Paris, France
[14]Wolfson Childhood Cancer Centre, Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, District of Columbia, United Kingdom
[15]Dutch Childhood Oncology Group, Utrecht, The Netherlands
[16]Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

Pediatric Blood & Cancer; January 2022; 69(1):e29361

ABSTRACT

Introduction

One-quarter of the relapses in children with B-cell precursor acute lymphoblastic leukemia (BCP-ALL) occur very early (within 18 months, before completion of treatment), and prognosis in these patients is worse compared to cases that relapse after treatment has ended.

Methods

In this study, we performed a genomic analysis of diagnosis-relapse pairs of 12 children who relapsed very early, followed by a deep-sequencing validation of all identified mutations. In addition, we included one case with a good initial treatment response and on-treatment relapse at the end of upfront therapy.

Results

We observed a dynamic clonal evolution in all cases, with relapse almost exclusively originating from a subclone at diagnosis. We identified several driver mutations that may have influenced the outgrowth of a minor clone at diagnosis to become the major clone at relapse. For example, a minimal residual disease (MRD)-based standard-risk patient with *ETV6-RUNX1*-positive leukemia developed a relapse from a *TP53*-mutated subclone after loss of the wild-type allele. Furthermore, two patients with *TCF3-PBX1*-positive leukemia that developed a very early relapse carried p.E1099K *WHSC1* mutations at diagnosis, a hotspot mutation that was recurrently encountered in other very early *TCF3-PBX1*-positive leukemia relapses as well. In addition to alterations in known relapse drivers, we found two cases with truncating mutations in the cohesin gene *RAD21*.

Conclusion

Comprehensive genomic characterization of diagnosis-relapse pairs shows that very early relapses in BCP-ALL frequently arise from minor subclones at diagnosis. A detailed understanding of the therapeutic pressure driving these events may aid the development of improved therapies.

INTRODUCTION

B-cell precursor acute lymphoblastic leukemia (BCP-ALL) is the most common pediatric malignancy[1-4]. Despite treatment improvements over the past decades, relapse still represents the most common cause of therapy failure[5,6]. Studies describing the genomic landscape of relapsed ALL have identified alterations in several genes associated with treatment resistance, including *IKZF1*, *CREBBP*, and *NT5C2*[7-12]. Selective pressure during upfront treatment can give competitive advantage to leukemic cells harboring these alterations, eventually leading to the rise of relapse. For example, activating mutations in the *NT5C2* gene, encoding cytosolic nucleosidase, accelerate depletion of intracellular purine nucleotides and lower proliferative potential of the tumor cells harboring these mutations[10,13,14]. During treatment with thiopurines, *NT5C2* mutations decrease cell vulnerability, thus giving the leukemic blasts a competitive and proliferative advantage compared to cells without these mutations. Nevertheless, exact mechanisms in which genomic alterations enable leukemic cells to develop resistance against chemotherapeutics are not fully elucidated. Furthermore, recent studies have shown that treatment failure may not always be the consequence of genomic alterations conferring treatment resistance and proliferative advantage, but rather persistence of leukemic cells in a protective niche[15-17].

Over 40% of BCP-ALL relapses present late, 6 months after the end of treatment (>30 months from initial diagnosis)[18,19]. These relapses may originate from leukemic (sub)clones that remained in a quiescent state during treatment or that could not be reached by the chemotherapeutics, for example, due to presence of physiological blood-organ barriers. On the other side of the spectrum are patients with very early relapses (<18 months from initial diagnosis), which represents approximately 25% of relapses that occur during treatment, and which are associated with dismal outcome[18-20]. Very early relapses may be different in that they display clonal outgrowth in the presence of chemotherapeutics. Alternatively, relapses may be the result of a suboptimal treatment, for example, due to toxicity-related discontinuations or omissions. Therefore, mechanisms driving relapses during treatment are not yet fully understood.

In this study, we explored the genomic abnormalities in 12 children with BCP-ALL who experienced very early relapse, and included one additional case with good initial treatment response and on-treatment relapse (within 24 months from diagnosis). We subsequently characterized the genomic alterations at diagnosis and relapse using whole exome sequencing (WES) and digital multiplex ligation-dependent probe amplification (MLPA) assay, and investigated the clonal dynamics of these leukemias. These different cases revealed diverse but relevant lessons of on-treatment relapses in BCP-ALL.

METHODS

Patients and samples

In this study, we selected and analyzed a cohort of 12 children diagnosed with BCP-ALL, who experienced very early bone marrow relapse (<18 months after diagnosis), according to the definition of Berlin-Frankfurt-Münster (BFM) study group, and for which samples from diagnosis, complete remission, and relapse were available. These children were treated according to a Dutch Childhood Oncology Group (DCOG) protocol ALL9 (*n* = 6), ALL10 (*n* = 2), ALL11 (*n* = 3), or Interfant-06 (*n* = 1). In addition, we included a child with favorable cytogenetics (*ETV6-RUNX1*-positive leukemia) and good initial treatment response, who developed an unexpected bone marrow relapse at the end of DCOG-ALL10-based standard-risk treatment (23 months). Median age at diagnosis for the total cohort was 8 years (range 1-14 years), and median remission time was 13 months (range 6-23 months) (**Table S4.1**). Deoxyribonucleic acid (DNA) was isolated from mononuclear cells obtained from bone marrow or peripheral blood at the time of diagnosis, complete remission, and relapse using QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Remission status of remission samples was determined according to the criteria of the respective study protocols, and all 13 patients reached complete remission. Informed written consent was obtained from all patients and/or their legal guardians before enrollment in the study. In addition, we collected 16 relapse samples from patients diagnosed with *TCF3-PBX1*-positive BCP-ALL. Primer sequences used to perform screening for mutations in *WHSC1* hotspots are given in **Table S4.2**.

Whole-genome amplifications (WGA)

Due to limited amounts of genomic DNA in 17 samples from six patients, we performed WGA using the REPLI-g mini kit (Qiagen, Hilden, Germany). Master mix was prepared using 50 ng of genomic DNA. In order to limit unspecific amplification and random errors, incubation was done in four independent reactions of 12.5 µl. After an 8-hour incubation, the four reactions for each patient were pooled and purified. WGA material was used only for the discovery of somatic mutations using WES and all detected mutations were subsequently validated using targeted deep sequencing on non-amplified genomic DNA.

WES and whole-genome sequencing (WGS)

WES was performed by the Beijing Genome Institute (BGI) in Copenhagen and Hong Kong using 3 µg of genomic DNA or whole-genome amplified material. DNA was randomly fragmented using Covaris technology and exon capture was done using Agilent SureSelectXT Human All Exon v4 or v5. Sequencing was performed using the Illumina HiSeq4000 platform. We achieved an average on-target sequencing depth of 93.3 (range 52-152.8; **Table S4.3**).

WGS was performed by BGI in Hong Kong using genomic DNA obtained from primary patient material. Genomic DNA was randomly fragmented using Covaris technology and sequenced on the BGISEQ-500 platform. We achieved an average sequencing depth of 45.7 (range 43.2-47.3; **Table S4.4**).

After excluding low-quality reads, clean reads were mapped to GRCh38 using Burrows-Wheeler aligner (BWA). Variant calling was done using HaplotypeCaller and Mutect2 of GATK. We excluded the following variants from further analysis: (i) likely germline variants (allele frequency >5% in the germline sample and variants detected in multiple independent germline samples); (ii) variants supported by only a single read; (iii) variants with multiple alternative alleles; and (iv) variants called within homopolymer regions. All remaining candidate somatic variants were manually curated using Integrated Genome Viewer (IGV) before subjecting them to deep-sequencing validation.

Targeted deep-sequencing validation of somatic variants

In order to exclude possible false-positive somatic mutations and to obtain accurate estimates of the mutant allele frequency (MAF), we performed deep targeted sequencing using the SeqCap EZ HyperCap workflow (Roche, Rotkreuz, Switzerland). Library preparation was done according to the manufacturer's instructions using 100 ng of genomic DNA from diagnosis, relapse, and complete remission. One patient (P0135) underwent a stem cell transplantation and for this patient remission samples before and after stem cell transplantation were used for deep-sequencing validation. Following enzymatic fragmentation, adapter ligation, and pre-capture amplification of ~300 bp fragments, DNA capture was done at 47°C for 48 hours. After post-capture amplification, paired-end 2 × 150 sequencing was performed on Illumina NextSeq 500 sequencer.

Digital MLPA assay

Copy number status of the most frequently affected genes in BCP-ALL was assessed using SALSA D007-X3 ALL digital MLPA (MRC-Holland, Amsterdam, The Netherlands)[21]. In brief, 100 ng of input DNA was denatured for 10 minutes at 98°C, followed by probe hybridization for 16 hours at 60°C. Ligation master mix was added at 48°C, and after incubation for 30 minutes at the same temperature, PCR reactions were performed (1 minute at 65°C, 45 cycles of 30 seconds at 95°C, 40 seconds at 65°C, and 90 seconds at 72°C). In total, 5 µl of each reaction was pooled and diluted to 10 nM amplicon library using 5 µl of mixture and 95 µl of water. Single-end 1 × 150 sequencing was performed on Illumina MiSeq sequencer.

Bioinformatics analysis

Fish plots were made using the R package fishplot (version 0.5)[22]. Mutational profiles were analyzed using R package MutationalPatterns (version 1.8.0)[23]. Data visualization was done using R package ggplot2 (version 3.2.1) and cBioPortal MutationMapper[24,25].

RESULTS

A total of 13 cases (nine males and four females) with bone marrow relapse during upfront treatment (<2 years of initial diagnosis) were included in this sequencing study (**Table S4.1**). Of these 13 cases, 12 relapsed very early (<18 months from diagnosis), while one *ETV6-RUNX1*-positive case with good initial treatment response experienced a relapse after 23 months while under standard-risk maintenance therapy. The total cohort consisted of patients with *ETV6-RUNX1* (n = 3), *TCF3-PBX1* (n = 2), hyperdiploid (n = 3), hypodiploid (n = 1), and B-other (n = 4) ALL subtypes, the latter including one patient with Down syndrome (DS) (**Table S4.1**). Median time between initial and relapse diagnosis of the total cohort was 13 months. Treatment initiated according to various upfront protocols was discontinued due to infections in three patients and stopped due to severe toxicity in one patient (**Table S4.1**).

Following WES of samples taken at diagnosis, complete remission, and relapse, all identified somatic variant calls (single-nucleotide variants and small insertions and deletions) were subjected to targeted capture and deep-sequencing (average depth 1,658) in order to validate the mutations and obtain accurate estimates of their MAFs at each time point. These approaches resulted in a total of 861 confirmed somatic mutations with mutant allele frequencies varying from 0.1% to 100% (**Tables S4.5** and **S6**). Validation rate was 88% for mutations with MAF ≥ 25%, and for subclonal mutations (MAF <25%) 56% were confirmed. In addition, we identified 99 somatic aneuploidies and copy number alterations (CNAs) using digital MLPA (**Tables S4.5** and **S4.7**). Median number of mutations was 21 (range 13-100) for diagnosis samples and 46 (range 20-227) in those from relapse, while median number of aneuploidies and CNAs was five both at diagnosis (range 0-7) and at relapse (range 0-14) (**Figure 4.1** and **Table S4.5**). We did not find a correlation between the number of mutations and CNAs in relapse and the duration of first complete remission. After filtering for mutations that were predicted to be pathogenic (non-synonymous, phyloP ≥ 2.5, CADD PHRED score ≥ 15), we detected in total 186 mutations in the leukemias of 13 patients (**Table S4.6**). The most commonly affected genes, both at diagnosis and relapse, were *KRAS* and *CDKN2A/B* (**Figure S4.1**).

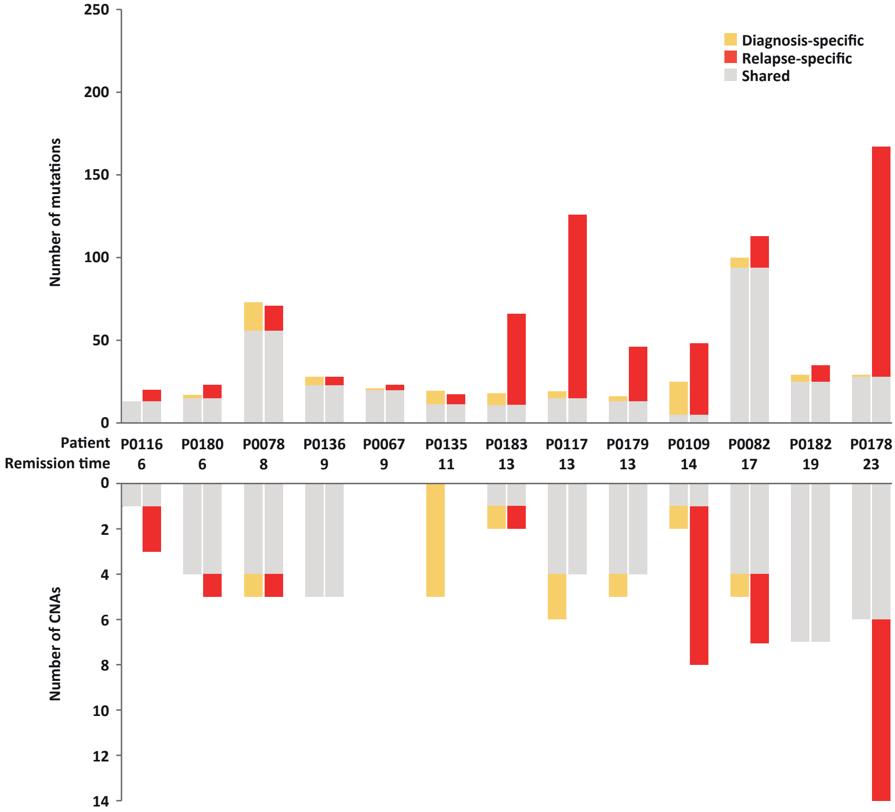


Figure 4.1. Bar plot showing number of mutations (point mutations and indels; upper panel) and copy number alteration (CNAs) (bottom panel) per patient in diagnosis (left bars) and relapse samples (right bars). Gray bars are shared mutations, whereas colored bars represent alterations that are unique at diagnosis (yellow) or relapse (red). Remission time is shown in months. Case P0135 had between 5% and 10% blast cells in the relapse sample, which is below detection limit of the digital multiplex ligation-dependent probe amplification assay.

Clonal dynamics of BCP-ALL relapsing during treatment

Based on the clonal dynamics between diagnosis and relapse, we observed two patterns, which have been recognized in unselected diagnosis-relapse cohorts as well[7,8,26-29]. The first pattern was represented by a group of nine cases (69%), in which the dominant clone detected at relapse emerged from one of the existing subclones at diagnosis while acquiring new relapse-specific alterations (**Figures 4.2A** and **S4.2**).

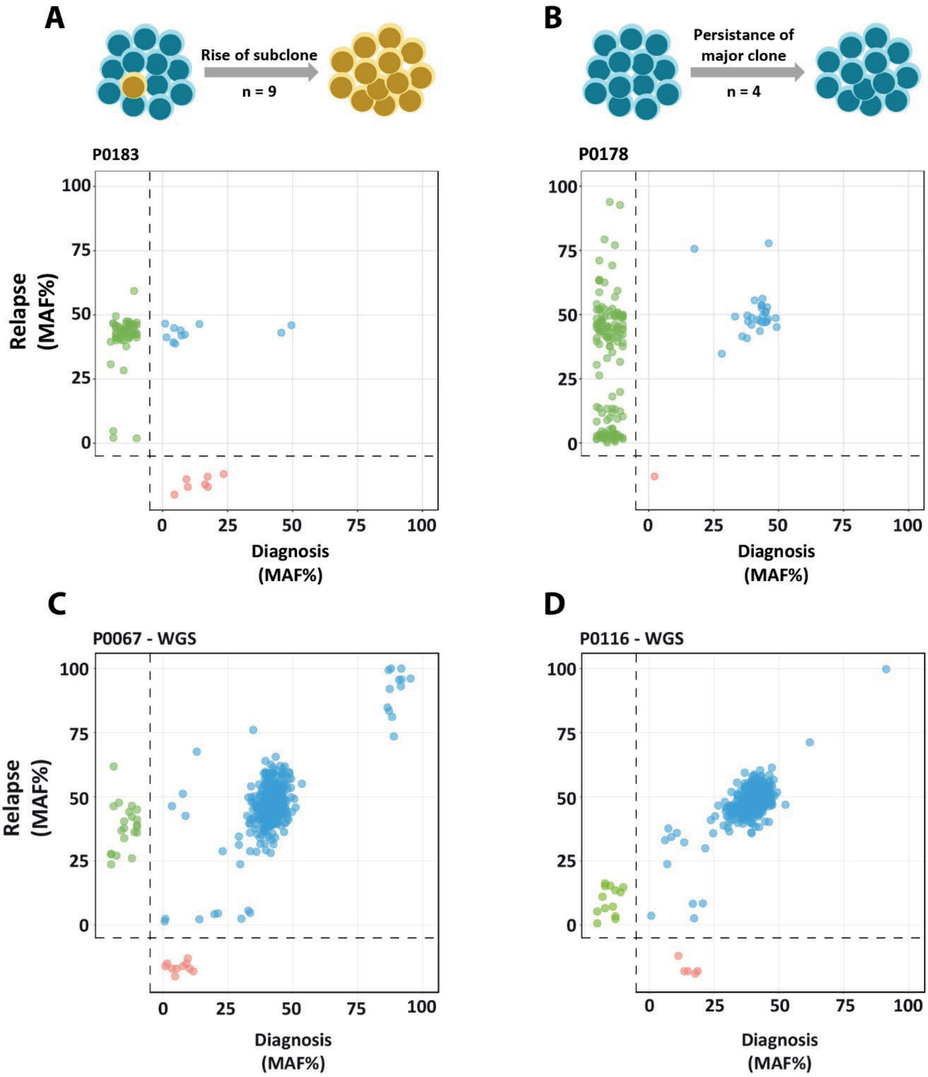


Figure 4.2. Schematic representation of clonal evolution models and scatter plots of representative cases showing clonal dynamics in the two groups of B-cell precursor acute lymphoblastic leukemia (BCP-ALL) that relapsed during treatment. In both models, we observed new mutations at time of relapse. In the first model **(A)** the dominant clone at relapse originated from a subclone at diagnosis (*n* = 9), while in the second model **(B)** the dominant clone at relapse originated from the major clone at diagnosis (*n* = 4). Scatter plots from all cases are presented in **Figure S4.2**. (**C** and **D**) Whole-genome sequencing (WGS) analysis of two patients from the second group revealed that in both patients, there are few mutations that support that the relapse did originate from a minor subclone at diagnosis. Diagnosis and relapse-specific mutations are in red and green, respectively, while alterations shared between two time points are depicted in blue.

A second pattern was observed in a group of four cases (31%), in which the dominant clone at relapse emerged from leukemic cells that were part of the major clone in diagnosis (**Figures 4.2B** and **S4.2**). In both of these groups, we observed a cluster of full-clonal mutations shared between diagnosis and relapse, indicating their origin from the common ancestral clone (**Figures 4.2A**, **4.2B** and **S4.2**). In order to further investigate the second pattern and to identify subgroup-specific mutations, we performed WGS of all samples of the two *TCF3-PBX1*-positive cases from this group. Both cases relapsed very early (6 and 9 months after diagnosis). Upon targeted deep-sequencing validation, somatic mutations were compared between all samples of the same patient, which confirmed that in both patients the far majority of mutations were preserved (**Figures 4.2C** and **4.2D**). However, in both patients we also identified a small number of mutations in the major clone at relapse that were subclonal at diagnosis, suggesting that they represent the first pattern of clonal dynamics as well. Together, these findings indicate that the majority of the very early and on-treatment relapses in our cohort originated from a minor clone at diagnosis that branched off from the major clone.

Recurrently mutated genes in relapse samples

We identified 118 predicted pathogenic mutations in major clones at relapse, including established drivers of BCP-ALL relapse like *NT5C2*, *CREBBP*, *WHSC1*, and *KRAS* (**Figure S4.3**). In addition, in two relapsed cases, we detected nonsense mutations in the cohesin complex component gene *RAD21*, a gene previously not associated with relapse or treatment resistance in BCP-ALL (**Figure S4.3E**). Both mutations resulted in truncation of the C-terminal domain of *RAD21*. *RAD21* is essential for sister chromatid cohesion, chromatin organization, and DNA looping through interaction with other members of the cohesin complex, while the C-terminal domain is involved in promotion of apoptosis in a positive-feedback manner[30-33]. Truncating mutations in *RAD21*, as well as other members of the cohesin complex, were previously reported as disease drivers in solid tumors and myeloid malignancies[34-36], but not in BCP-ALL. Two cases, both with *TCF3-PBX1*-positive tumors, were found to carry a hotspot *WHSC1* p.E1099K mutation at diagnosis that was preserved at relapse. At diagnosis *TCF3-PBX1*-positive leukemia is associated with favorable outcomes[37-39] and relapses are rare. However, if relapses occur, these are often fatal, particularly if they present very early[40]. Therefore, we investigated whether *WHSC1* mutations were more common in *TCF3-PBX1*-positive relapses. In collaboration with eight other centers, we collected 16 additional *TCF3-PBX1*-positive relapses, of which six occurred very early (**Table S4.2**). Upon sequencing of the hotspot region of *WHSC1* in these tumors, one additional patient with very early relapse (1.3 years remission time) was found to carry a p.E1099K mutation in the relapse. In a previous study, which included 376 patients from two representative Dutch ALL cohorts, we showed that full-clonal *WHSC1* mutations occurred in 2% of pediatric ALL cases at diagnosis[41]. In *TCF3-PBX1*-positive ALL, these mutations were shown to be enriched at diagnosis (15%)[7,42], similar to the percentage we found in *TCF3-PBX1*-positive relapses (3/18; 16%). However, when considering only very early relapses, three of eight (38%) carried *WHSC1* mutations among the very early *TCF3-PBX1* relapses.

Mechanisms causing hypermutation in BCP-ALL relapses during treatment

Three cases carried a high mutation burden (≥100), above the threshold of 85 mutations in the coding regions (>1.3 mutation/megabase), which we previously used to define hypermutation in pediatric ALL[8]. This hypermutation phenotype was restricted to relapse in two cases and was already present at the time of diagnosis in a third case. We and others have previously identified several mutational mechanisms being responsible for a high mutation load, including activation-induced deaminase (AID)/apolipoprotein B mRNA editing enzyme (APOBEC) activity[8,43-45]. These processes cause mutational patterns that can be recognized by analyzing single-base substitutions (SBSs) in specific three-nucleotide contexts[46]. Whereas in one case, the mutational profile was highly similar to the previously reported SBS1 mutational signature; we identified AID/APOBEC mutagenesis (SBS2/13) in the remaining two cases. In one of these cases, this mutational process was active at time of diagnosis, with 88% of the mutations being preserved at relapse (**Figure 4.3**), whereas the number of newly acquired relapse-specific mutations was low and unlikely to be driven by AID/APOBEC activity. In contrast, in the relapse sample of patient P0178, the far majority of relapse-specific mutations appeared to be driven by AID/APOBEC activity (**Figure 4.3**).

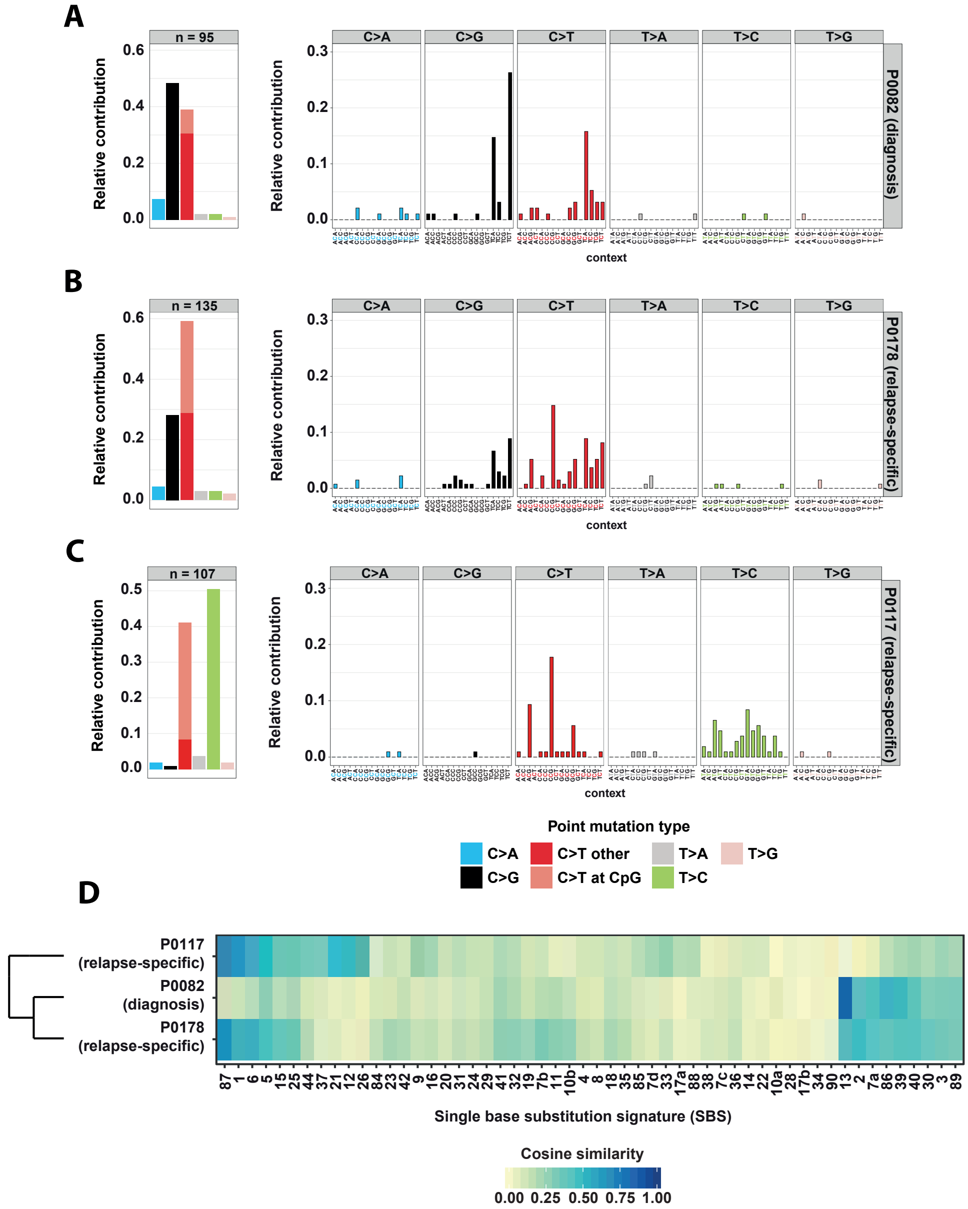


Figure 4.3. Mutational signatures at relapse. (**A**-**C**) Frequency of the substitutions (left panel) and mutational profile (right panel) of the three cases with high mutational load (>1.3 mutation/megabase). **(D)** Comparison of mutational profiles between four cases and known COSMIC signatures revealed high cosine similarity between SBS2 and SBS13 attributed to aberrant AID/APOBEC activity and SBS1 clock-like signature attributed to spontaneous deamination of methylated cytosine in CpGs.

JAK2-mutated clones in leukemia of a DS patient

In a patient with DS, we identified two dominant clones at diagnosis, each harboring different well-known *JAK2* hotspot mutations, p.R683S and p.R683G, with MAF of 24% and 27%, respectively (**Figure 4.4**). This observation indicates that most, if not all, leukemic cells in this patient carry either one of these mutations, illustrating the driving capacity of *JAK2* mutations in DS-ALL as reported by others[47-49]. The clone with *JAK2* p.R683S mutation expanded during treatment and gave rise to the sole dominant clone in relapse, while the clone with *JAK2* p.R683G was preserved as a minor subclone in relapse (MAF = 7%) (**Figure 4.4**).

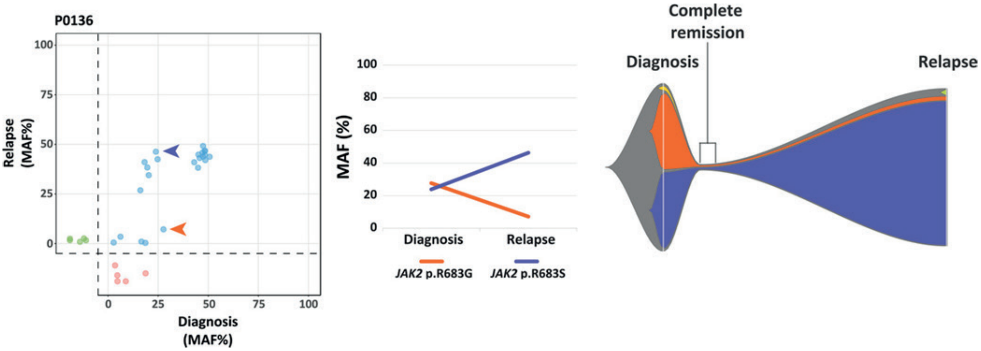


Figure 4. Clonal dynamics in a Down syndrome patient with two codominant clones in leukemia at diagnosis, each harboring *JAK2* mutations. One of the clones expanded and became a dominant clone in relapse, while the other clone decreased to a subclone at relapse.

A standard-risk stratified patient with on-treatment relapse

One patient in our cohort was diagnosed with a low-risk *ETV6-RUNX1* subtype and, based on a good response to induction therapy, stratified as standard risk. Nevertheless, he relapsed just before the end of treatment. Genomic characterization revealed that this patient had a subclonal *TP53* p.R282W hotspot mutation at diagnosis (MAF = 18%). *TP53* mutations are rare in ALL, particularly at diagnosis, but in relapse these aberrations are associated with poor outcome[40,50-52]. Relapse occurred during 6-mercaptopurine and methotrexate courses in which the *TP53*-mutated clone emerged with a whole gene deletion of the second allele. Furthermore, this relapse appeared to be driven by an AID/APOBEC mutational mechanism (**Figure 4.3**).

DISCUSSION

During disease progression, multiple clones emerge, some of which may harbor alterations associated with treatment resistance. Chemotherapy can drive selection of these resistant clones, which may lead to their expansion and to relapse during treatment. Furthermore, intrinsic factors, such as drug toxicity and infections, can cause treatment interruption and subsequently potential relapse. The relapses in our study occurred during treatment and have poor prognosis compared to late relapses, after treatment has ended. In this study, we performed WES and digital MLPA of diagnosis-relapse pairs and complete remission samples of 12 cases who relapsed very early (<18 months from initial diagnosis) and one case with favorable prognosis and relapse at the end of treatment. Most cases (9/13) did not experience any treatment interruptions before the relapse occurred, indicating that treatment-driven clonal selection may be the dominant mechanism.

Our data indicate that clonal evolution in cases that relapse during treatment is a dynamic process, with the majority of relapses evolving from a subclone present at the time of diagnosis, which acquired additional mutations. In the majority of cases, we were able to identify preserved or newly acquired relapse-associated genetic alterations in the rising clone at relapse. This finding supports the model in which early relapses are associated with more dynamic changes in clonal evolution compared to late relapses[8,53].

In this study, we identified several genes recurrently altered in the major clone of relapsed ALL, including *CREBBP*, *WHSC1*, *NT5C2*, *IKZF1*, *CDKN2A*, and *CDKN2B*, which were previously associated with early relapse, in different studies[7-11,14,54]. Alterations in epigenetic modifiers (for example, *ARID2*, *EP300*, *TOX*, and *ERG*) were rare or absent, which is in line with previous studies suggesting their association with late relapse[8,44,53]. In addition, we found two cases with mutations in cohesin complex member *RAD21*, a gene that was previously not associated with relapsed BCP-ALL. Studies have shown that mutations in cohesin complex genes are often truncating and have the potential to disrupt the entire complex[35,36]. However, myeloid leukemias with truncating mutations in cohesin complex genes usually do not have aneuploidies, suggesting that the mechanism in which these genes drive malignant transformation might include one of their non-canonical functions[30,31,33,35,36]. Although previous studies indicate importance of the conserved C-terminal domain for induction of apoptosis[30,31], the relevance of *RAD21*, as well as other cohesin complex mutations, for BCP-ALL development and treatment failure still needs to be fully elucidated.

We have identified a hotspot *WHSC1* p.E1099K mutation in two patients with very early relapse and *TCF3-PBX1*-positive leukemia, a subtype with favorable outcome at initial diagnosis. In both cases, the mutation was preserved from a major clone already present at diagnosis, indicating that this alteration was an early event during disease development. Further screening of 16 relapsed cases with *TCF3-PBX1* leukemia revealed one additional relapse with the p.E1099K *WHSC1* mutation, which interestingly also relapsed very early. Furthermore, Ma et al. reported five very early *TCF3-PBX1*-positive relapses, three of which carried *WHSC1* mutations as well[7,8]. Although patients with *TCF3-PBX1*-positive BCP-ALL have favorable prognosis in modern treatment protocols and rarely relapse[5,55,56], the outcome is poor when relapses occur[57,58]. Clinical diversity and discrepancies in outcome between relapsed and non-relapsed patients suggest the presence of additional prognostic biomarkers[58]. *WHSC1* mutations may drive one of the mechanisms leading to (very early) relapse in *TCF3-PBX1*-positive leukemia. Mutations in the SET domain of methyltransferase *WHSC1* were shown to cause perturbations in epigenetic makeup through chromatin remodeling[42,59]. Among pediatric tumors, the frequency of *WHSC1* mutations is highest in BCP-ALL, especially in *ETV6-RUNX1* (20%) and *TCF3-PBX1* (15%) ALL subtypes[42]. Therefore, a *WHSC1* mutation is unlikely to be a strong risk factor for relapse in these subtypes by itself. However, considering the apparent higher frequency of these hotspot mutations in very early *TCF3-PBX1*-positive relapses, an epigenetic event may be involved, which is worth studying further, for example, by comparing the epigenetic landscape of *WHSC1*-mutated non-relapsed and very early relapsed *TCF3-PBX1*-positive cases.

We did find interesting candidate mutations that could explain the observed clonal dynamics that preceded the relapse. For example, the proliferation of one of the two *JAK2* mutated clones in the patient with DS-ALL is in line with previous functional studies, which showed that *JAK2* p.R683S has higher proliferative advantage compared to p.R683G mutations in cell lines[47,49]. DS-ALL are more prone to treatment-related morbidity and mortality, which is the reason for reduction of treatment intensity in this group[5,20,60]. Our findings emphasize the fact that *JAK2* mutations may enhance proliferative capacities of DS-ALL cells, and support current ongoing trials using JAK inhibitors in DS-ALL to prevent relapses while reducing treatment intensity.

A second example was the patient with minimal residual disease (MRD)-based standard-risk stratification, who nevertheless relapsed while on treatment. This study identified a likely explanation for this particular patient as he carried a hotspot *TP53* mutation in a subclone at diagnosis, which grew out to relapse upon deletion on the wild-type allele. *TP53* alterations are enriched in relapsed BCP-ALL[7,8], and associated with unfavorable outcome even in patients with good initial response to induction therapy[61]. Previously, we have shown that subclonal (likely) pathogenic mutations in *TP53* are relatively common, but do not necessarily indicate a high risk for relapse, even in standard risk-stratified patients (*n* = 9)[41]. Furthermore, mutations on both alleles are associated with poor outcome, as demonstrated in adult ALL[62], but the risk of acquiring a second hit has not been studied in larger cohorts. Nevertheless, as *TP53* mutations are enriched in relapse[8,44,50,51], the detection at time of diagnosis, even at subclonal level, may warrant careful monitoring by MRD analysis, particularly when stratified in low-risk treatment arms.

Taken together, our data reveal highly dynamic clonal evolution in cases that relapse very early, with subclones present at diagnosis emerging as new dominant clones in the majority of cases. While the small cohort size limits our capacity to generalize, specific mutations detected at diagnosis, like the *TP53* mutation in a standard-risk patient or the *WHSC1* p.E1099K mutations in three *TCF3-PBX1*-positive leukemias, are likely to have contributed to these relapses. Furthermore, we identified recurrent mutations in *RAD21* at relapse, a gene previously not associated with relapsed BCP-ALL.

SUPPLEMENTARY DATA

Supplementary Tables

Supplementary Tables are available online using the following link: https://doi.org/10.1002/pbc.29361.

Supplementary Figures

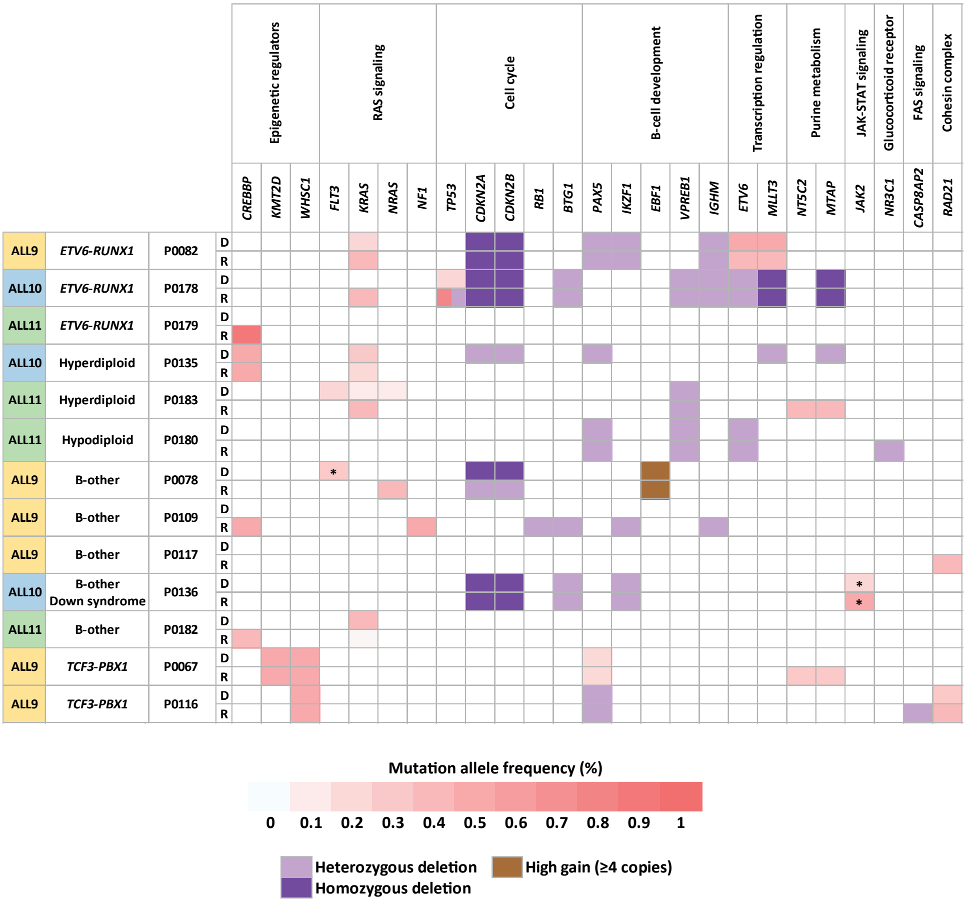


Figure S4.1. Heatmap indicating alterations in cancer-associated genes and genes frequently altered in relapse samples. Diagnosis (D) and Relapse (R) are compared for the 13 patients grouped by trial (ALL9, ALL10, ALL11). Genes with two mutations in the same sample are marked with asterisks. Mutation allele frequencies were normalized for blast percentage in respective samples. Due to low blast percentage in the relapse samples of P0135, copy number changes observed at diagnosis were likely below the detection threshold at relapse (see also **Table S4.7**).

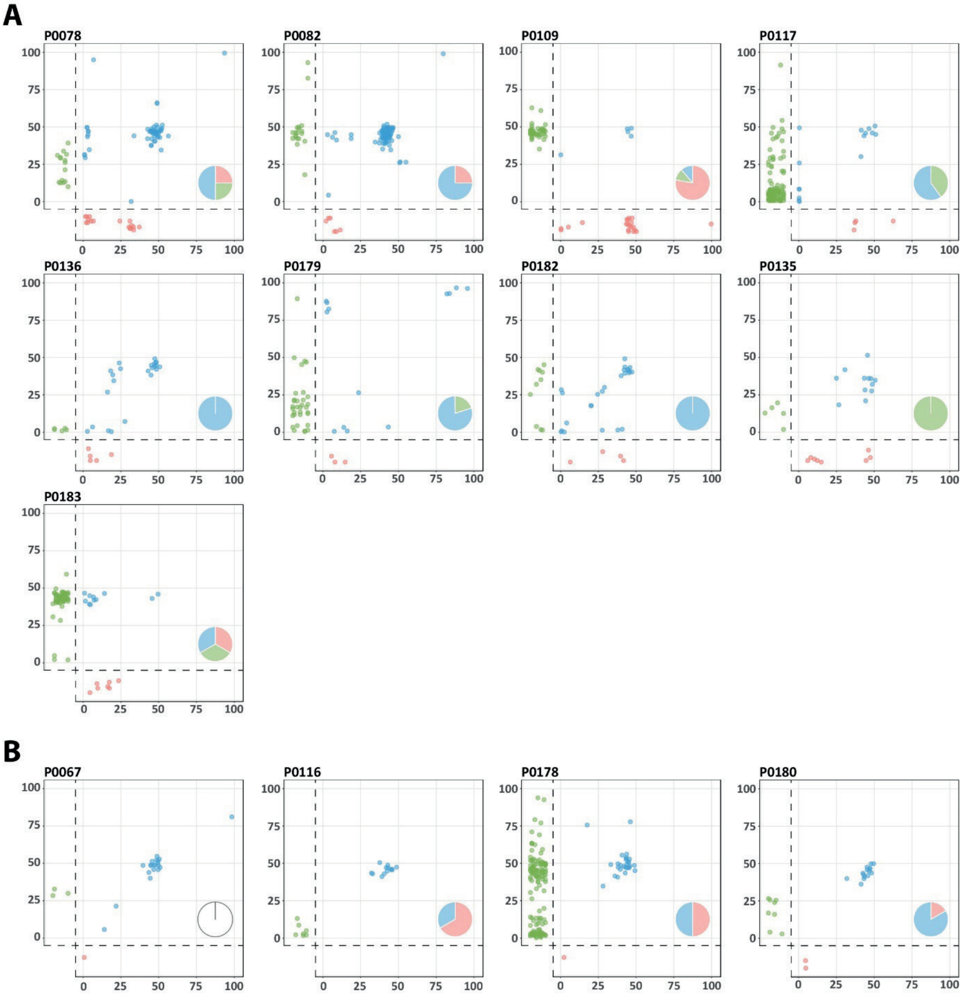


Figure S4.2. Scatter plots showing clonal dynamics of point mutations and indels between diagnosis and relapse in cases that relapsed during treatment. Pie charts depict the number of CNAs, including aneuploidies. Diagnosis and relapse-specific mutations are in red and green, respectively, while alterations shared between two time points are depicted in blue. Mutation allele frequencies were normalized for blast percentage in respective samples. **(A)** Nine cases in which the relapse originated from a prominent subclone at diagnosis. **(B)** Four cases where the shared (preserved) mutations represent the major clone at diagnosis and relapse. Due to low blast percentage in the relapse samples of P0135, copy number changes observed at diagnosis were likely below the detection threshold at relapse (see also **Table S4.7**).

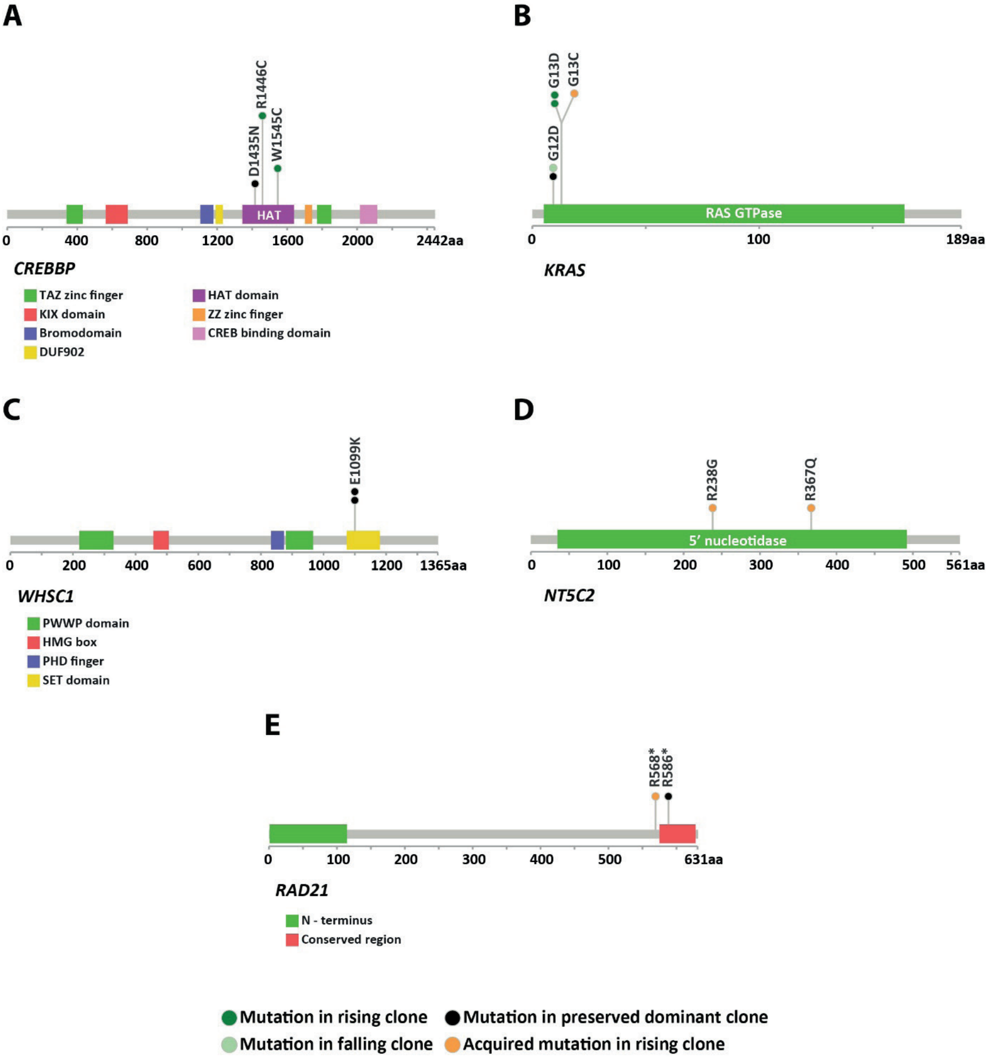

A
D1435N
R1446C
W1545C
HAT
0
400
800
1200
1600
2000
2442aa
CREBBP
TAZ zinc finger
KIX domain
Bromodomain
DUF902
HAT domain
ZZ zinc finger
CREB binding domain
B
G12D
G13D
G13C
RAS GTPase
0
100
189aa
KRAS
C
E1099K
0
200
400
600
800
1000
1200
1365aa
WHSC1
PWWP domain
HMG box
PHD finger
SET domain
D
R238G
R367Q
5' nucleotidase
0
100
200
300
400
500
561aa
NT5C2
E
R568*
R586*
0
100
200
300
400
500
631aa
RAD21
N - terminus
Conserved region
Mutation in rising clone
Mutation in preserved dominant clone
Mutation in falling clone
Acquired mutation in rising clone

Figure S4.3. Schematic representation of the five frequently mutated genes *NT5C2* **(A)**, *WHSC1* **(B)**, *CREBBP* **(C)**, *KRAS* **(D)** and *RAD21* **(E)** in relapse samples indicating mutations and their status relative to the samples at diagnosis.

FUNDING

This work was supported by grants from Stichting Kinderen Kankervrij (KIKA 150 to RPK), the Dutch Cancer society (KWF 12842 to RPK and FNvL), China Scholarship Council (CSC201304910347 to JY), and the Center for Biological Resources (CRB-cancer; BB-0033-00076 to HC) of the Robert Debré Hospital.

DATA AVAILABILITY STATEMENT

The sequence data that support the findings of this study have been deposited in the European Genome-phenome Archive (EGA) under accession number EGAS00001005615.

CONFLICTS OF INTEREST

No conflicts of interest to disclose.

CONTRIBUTIONS

RPK, EW, PMH, FvL, and AGvK conceived the study. RPK, EW, ŽA, JY and BCB designed the study. ŽA, JY, SVvR and CGvdH performed experiments and analyzed the data. SHL and LM performed bioinformatic analyses. ŽA and RPK wrote the manuscript and created the figures and tables. ES, BCB, SE, J-PB, CE, MC, RS, HC and AVM provided samples and clinical information. All authors critically reviewed the manuscript and approved the final submitted manuscript.

ACKNOWLEDGMENTS

We are grateful to MRC-Holland for their support in performing and analyzing the digital MLPA experiments. We thank Radboud University Medical Center, Department of Human Genetics for bioinformatics support in data analysis.

REFERENCES

1. Barr RD, Ferrari A, Ries L, Whelan J, Bleyer WA. Cancer in adolescents and young adults: a narrative review of the current status and a view of the future. JAMA Pediatr. 2016;170(5):495-501.
2. Katanoda K, Shibata A, Matsuda T, Hori M, Nakata K, Narita Y, et al. Childhood, adolescent and young adult cancer incidence in Japan in 2009-2011. Jpn J Clin Oncol. 2017;47(8):762-771.
3. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. CA Cancer J Clin. 2014;64(2):83-103.
4. Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al. International incidence of childhood cancer, 2001-10: a population-based registry study. Lancet Oncol. 2017;18(6):719-731.
5. Pieters R, Groot-Kruseman Hd, Velden VVd, Fiocco M, Berg Hvd, Bont Ed, et al. Successful therapy reduction and intensification for childhood acute lymphoblastic leukemia based on minimal residual disease monitoring: study ALL10 from the Dutch Childhood Oncology Group. J Clin Oncol. 2016;34(22):2591-2601.
6. Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the Children's Oncology Group. J Clin Oncol. 2012;30(14):1663-1669.
7. Ma X, Edmonson M, Yergeau D, Muzny DM, Hampton OA, Rusch M, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun. 2015;6(1):6604.
8. Waanders E, Gu Z, Dobson SM, Antić Ž, Crawford JC, Ma X, et al. Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. Blood Cancer Discov. 2020;1(1):96-111.
9. Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011;471(7337):235-239.
10. Tzoneva G, Perez-Garcia A, Carpenter Z, Khiabanian H, Tosello V, Allegretta M, et al. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. Nat Med. 2013;19(3):368-371.
11. Kuiper RP, Waanders E, van der Velden VH, van Reijmersdal SV, Venkatachalam R, Scheijen B, et al. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. Leukemia. 2010;24(7):1258-1264.
12. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009;360(5):470-480.
13. Tzoneva G, Dieck CL, Oshima K, Ambesi-Impiombato A, Sánchez-Martín M, Madubata CJ, et al. Clonal evolution mechanisms in NT5C2 mutant-relapsed acute lymphoblastic leukaemia. Nature. 2018;553(7689):511-514.
14. Meyer JA, Wang J, Hogan LE, Yang JJ, Dandekar S, Patel JP, et al. Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. Nat Genet. 2013;45(3):290-294.
15. Polak R, de Rooij B, Pieters R, den Boer ML. B-cell precursor acute lymphoblastic leukemia cells use tunneling nanotubes to orchestrate their microenvironment. Blood. 2015;126(21):2404-2414.
16. Iwamoto S, Mihara K, Downing JR, Pui C-H, Campana D. Mesenchymal cells regulate the response of acute lymphoblastic leukemia cells to asparaginase. J Clin Invest. 2007;117(4):1049-1057.

17. Cahu X, Calvo J, Poglio S, Prade N, Colsch B, Arcangeli ML, et al. Bone marrow sites differently imprint dormancy and chemoresistance to T-cell acute lymphoblastic leukemia. Blood Adv. 2017;1(20):1760-1772.
18. Tallen G, Ratei R, Mann G, Kaspers G, Niggli F, Karachunsky A, et al. Long-term outcome in children with relapsed acute lymphoblastic leukemia after time-point and site-of-relapse stratification and intensified short-course multidrug chemotherapy: results of trial ALL-REZ BFM 90. J Clin Oncol. 2010;28(14):2339-2347.
19. Oskarsson T, Söderhäll S, Arvidson J, Forestier E, Frandsen TL, Hellebostad M, et al. Treatment-related mortality in relapsed childhood acute lymphoblastic leukemia. Pediatr Blood Cancer. 2018;65(4).
20. Oskarsson T, Söderhäll S, Arvidson J, Forestier E, Montgomery S, Bottai M, et al. Relapsed childhood acute lymphoblastic leukemia in the Nordic countries: prognostic factors, treatment and outcome. Haematologica. 2016;101(1):68-76.
21. Benard-Slagter A, Zondervan I, de Groot K, Ghazavi F, Sarhadi V, Van Vlierberghe P, et al. Digital multiplex ligation-dependent probe amplification for detection of key copy number alterations in T- and B-cell lymphoblastic leukemia. J Mol Diagn. 2017;19(5):659-672.
22. Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, et al. Visualizing tumor evolution with the fishplot package for R. BMC Genom. 2016;17(1):880.
23. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10(1):33.
24. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1.
25. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401-404.
26. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science. 2008;322(5906):1377-1380.
27. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481(7382):506-510.
28. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature. 2011;469(7330):356-361.
29. van Delft FW, Horsley S, Colman S, Anderson K, Bateman C, Kempski H, et al. Clonal origins of relapse in ETV6-RUNX1 acute lymphoblastic leukemia. Blood. 2011;117(23):6247-6254.
30. Chen F, Kamradt M, Mulcahy M, Byun Y, Xu H, McKay MJ, et al. Caspase proteolysis of the cohesin component RAD21 promotes apoptosis. J Biol Chem. 2002;277(19):16775-16781.
31. Pati D, Zhang N, Plon SE. Linking sister chromatid cohesion and apoptosis: role of Rad21. Mol Cell Biol. 2002;22(23):8267-8277.
32. Zhang N, Jiang Y, Mao Q, Demeler B, Tao YJ, Pati D. Characterization of the interaction between the cohesin subunits Rad21 and SA1/2. PLoS One. 2013;8(7):e69458.
33. Seitan VC, Hao B, Tachibana-Konwalski K, Lavagnolli T, Mira-Bontenbal H, Brown KE, et al. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. Nature. 2011;476(7361):467-471.

34. Thota S, Viny AD, Makishima H, Spitzer B, Radivoyevitch T, Przychodzen B, et al. Genetic alterations of the cohesin complex genes in myeloid malignancies. Blood. 2014;124(11):1790-1798.
35. Hill VK, Kim JS, Waldman T. Cohesin mutations in human cancer. Biochim Biophys Acta. 2016;1866(1):1-11.
36. Fisher JB, McNulty M, Burke MJ, Crispino JD, Rao S. Cohesin mutations in myeloid malignancies. Trends Cancer. 2017;3(4):282-293.
37. Pang L, Liang Y, Pan J, Wang JR, Chai YH, Zhao WL. Clinical features and prognostic significance of TCF3-PBX1 fusion gene in Chinese children with acute lymphoblastic leukemia by using a modified ALL-BFM-95 protocol. Pediatr Hematol Oncol. 2015;32(3):173-181.
38. Felice MS, Gallego MS, Alonso CN, Alfaro EM, Guitter MR, Bernasconi AR, et al. Prognostic impact of t(1;19)/ TCF3-PBX1 in childhood acute lymphoblastic leukemia in the context of Berlin-Frankfurt-Münster-based protocols. Leuk Lymphoma. 2011;52(7):1215-1221.
39. Kager L, Lion T, Attarbaschi A, Koenig M, Strehl S, Haas OA, et al. Incidence and outcome of TCF3-PBX1-positive acute lymphoblastic leukemia in Austrian children. Haematologica. 2007;92(11):1561-1564.
40. Irving JA, Enshaei A, Parker CA, Sutton R, Kuiper RP, Erhorn A, et al. Integration of genetic and clinical risk factors improves prognostication in relapsed childhood B-cell precursor acute lymphoblastic leukemia. Blood. 2016;128(7):911-922.
41. Antić Ž, Yu J, Van Reijmersdal SV, Van Dijk A, Dekker L, Segerink WH, et al. Multiclonal complexity of pediatric acute lymphoblastic leukemia and the prognostic relevance of subclonal mutations. Haematologica. 2021;106(12):3046-3055.
42. Jaffe JD, Wang Y, Chan HM, Zhang J, Huether R, Kryukov GV, et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nat Genet. 2013;45(11):1386-1391.
43. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature. 2018;555(7696):371-376.
44. Li B, Brady SW, Ma X, Shen S, Zhang Y, Li Y, et al. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. Blood. 2020;135(1):41-55.
45. Antić Ž, Lelieveld SH, van der Ham CG, Sonneveld E, Hoogerbrugge PM, Kuiper RP. Unravelling the sequential interplay of mutational mechanisms during clonal evolution in relapsed pediatric acute lymphoblastic leukemia. Genes (Basel). 2021;12(2):214.
46. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-421.
47. Bercovich D, Ganmore I, Scott LM, Wainreb G, Birger Y, Elimelech A, et al. Mutations of JAK2 in acute lymphoblastic leukaemias associated with Down's syndrome. Lancet. 2008;372(9648):1484-1492.
48. Kearney L, Gonzalez De Castro D, Yeung J, Procter J, Horsley SW, Eguchi-Ishimae M, et al. Specific JAK2 mutation (JAK2R683) and multiple gene deletions in Down syndrome acute lymphoblastic leukemia. Blood. 2009;113(3):646-648.
49. Mullighan CG, Zhang J, Harvey RC, Collins-Underwood JR, Schulman BA, Phillips LA, et al. JAK mutations in high-risk childhood acute lymphoblastic leukemia. Proc Natl Acad Sci U S A. 2009;106(23):9414-9418.

50. Yu C-H, Chang W-T, Jou S-T, Lin T-K, Chang Y-H, Lin C-Y, et al. TP53 alterations in relapsed childhood acute lymphoblastic leukemia. Cancer Sci. 2020;111(1):229-238.
51. Hof J, Krentz S, van Schewick C, Körner G, Shalapour S, Rhein P, et al. Mutations and deletions of the TP53 gene predict nonresponse to treatment and poor outcome in first relapse of childhood acute lymphoblastic leukemia. J Clin Oncol. 2011;29(23):3185-3193.
52. Krentz S, Hof J, Mendioroz A, Vaggopoulou R, Dörge P, Lottaz C, et al. Prognostic value of genetic alterations in children with first bone marrow relapse of childhood B-cell precursor acute lymphoblastic leukemia. Leukemia. 2013;27(2):295-304.
53. Spinella JF, Richer C, Cassart P, Ouimet M, Healy J, Sinnett D. Mutational dynamics of early and late relapsed childhood ALL: rapid clonal expansion and long-term dormancy. Blood Adv. 2018;2(3):177-188.
54. Schroeder MP, Bastian L, Eckert C, Gökbuget N, James AR, Tanchez JO, et al. Integrated analysis of relapsed B-cell precursor Acute Lymphoblastic Leukemia identifies subtype-specific cytokine and metabolic signatures. Sci Rep. 2019;9(1):4188.
55. Takahashi H, Kajiwara R, Kato M, Hasegawa D, Tomizawa D, Noguchi Y, et al. Treatment outcome of children with acute lymphoblastic leukemia: the Tokyo Children's Cancer Study Group (TCCSG) Study L04-16. Int J Hematol. 2018;108(1):98-108.
56. Lin A, Cheng FWT, Chiang AKS, Luk CW, Li RCH, Ling ASC, et al. Excellent outcome of acute lymphoblastic leukaemia with TCF3-PBX1 rearrangement in Hong Kong. Pediatr Blood Cancer. 2018;65(12):e27346.
57. Jeha S, Pei D, Raimondi SC, Onciu M, Campana D, Cheng C, et al. Increased risk for CNS relapse in pre-B cell leukemia with the t(1;19)/TCF3-PBX1. Leukemia. 2009;23(8):1406-1409.
58. Moorman VA. New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia. Haematologica. 2016;101(4):407-416.
59. Pierro J, Saliba J, Narang S, Sethia G, Saint Fleur-Lominy S, Chowdhury A, et al. The NSD2 p.E1099K mutation is enriched at relapse and confers drug resistance in a cell context-dependent manner in pediatric acute lymphoblastic leukemia. Mol Cancer Res. 2020;18(8):1153-1165.
60. Buitenkamp TD, Izraeli S, Zimmermann M, Forestier E, Heerema NA, van den Heuvel-Eibrink MM, et al. Acute lymphoblastic leukemia in children with Down syndrome: a retrospective analysis from the Ponte di Legno study group. Blood. 2014;123(1):70-77.
61. Eckert C, Groeneveld-Krentz S, Kirschner-Schwabe R, Hagedorn N, Chen-Santel C, Bader P, et al. Improving stratification for children with late bone marrow B-Cell acute lymphoblastic leukemia relapses with refined response classification and integration of genetics. J Clin Oncol. 2019;37(36):3493-3506.
62. Stengel A, Schnittger S, Weissmann S, Kuznia S, Kern W, Kohlmann A, et al. TP53 mutations occur in 15.7% of ALL and are associated with MYC-rearrangement, low hypodiploidy, and a poor prognosis. Blood. 2014;124(2):251-258.

CHAPTER 5

MUTATIONAL LANDSCAPE AND PATTERNS OF CLONAL EVOLUTION IN RELAPSED PEDIATRIC ACUTE LYMPHOBLASTIC LEUKEMIA

Esmé Waanders[1,2,3], Zhaohui Gu[1*], Stephanie M. Dobson[4,5*], Željko Antić[2*], Jeremy Chase Crawford[6], Xiaotu Ma[7], Michael N. Edmonson[7], Debbie Payne-Turner[1], Maartje van de Vorst[8], Marjolijn CJ. Jongmans[2,3], Irina McGuire[9], Xin Zhou[7], Jian Wang[7], Lei Shi[10], Stanley Pounds[10], Deqing Pei[10], Cheng Cheng[10], Guangchun Song[1], Yiping Fan[7], Ying Shao[7], Michael Rusch[7], Kelly McCastlain[1], Jiangyan Yu[2], Ruben van Boxtel[2], Francis Blokzijl[11,12], Ilaria Iacobucci[1], Kathryn G. Roberts[1], Ji Wen[1], Gang Wu[7], Jing Ma[1], John Easton[7], Geoffrey Neale[13], Scott R. Olsen[13], Kim E. Nichols[14], Ching-Hon Pui[14], Jinghui Zhang[7], William E. Evans[15], Mary V. Relling[15], Jun J. Yang[15], Paul G. Thomas[6], John E. Dick[4,5], Roland P. Kuiper[2#] and Charles G. Mullighan[1#]*

[*]These authors contributed equally to this study as lead authors
[#]These authors contributed equally to this study as senior authors

[1]Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee.
[2]Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands.
[3]Department of Genetics, University Medical Center Utrecht, Utrecht, the Netherlands.
[4]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada.
[5]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.
[6]Department of Immunology, St. Jude Children's Research Hospital, Memphis, Tennessee. [7]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee.
[8]Department of Human Genetics, Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, the Netherlands.
[9]Department of Information Services, St. Jude Children's Research Hospital, Memphis, Tennessee.
[10]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee.
[11]Oncode Institute, University Medical Center Utrecht, Utrecht, the Netherlands.
[12]Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands.
[13]The Hartwell Center for Bioinformatics and Biotechnology, St. Jude Children's Research Hospital, Memphis, Tennessee.
[14]Department of Oncology, St. Jude Children's Research Hospital, Memphis, Tennessee.
[15]Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, Tennessee

Blood Cancer Discovery; July 2020; 1(1):96-111

ABSTRACT

Relapse of acute lymphoblastic leukemia (ALL) remains a leading cause of childhood cancer-related death. Prior studies have shown clonal mutations at relapse often arise from relapse-fated subclones that exist at diagnosis. However, the genomic landscape, evolutionary trajectories, and mutational mechanisms driving relapse are incompletely understood. In an analysis of 92 cases of relapsed childhood ALL incorporating multimodal DNA and RNA sequencing, deep digital mutational tracking, and xenografting to formally define clonal structure, we identified 50 significant targets of mutation with distinct patterns of mutational acquisition or enrichment. *CREBBP*, *NOTCH1*, and RAS signaling mutations arose from diagnosis subclones, whereas variants in *NCOR2*, *USH2A*, and *NT5C2* were exclusively observed at relapse. Evolutionary modeling and xenografting demonstrated that relapse-fated clones were minor (50%), major (27%), or multiclonal (18%) at diagnosis. Putative second leukemias, including those with lineage shift, were shown to most commonly represent relapse from an ancestral clone rather than a truly independent second primary leukemia. A subset of leukemias prone to repeated relapse exhibited hypermutation driven by at least three distinct mutational processes, resulting in heightened neoepitope burden and potential vulnerability to immunotherapy. Finally, relapse-driving sequence mutations were detected prior to relapse using droplet digital PCR at levels comparable with orthogonal approaches to monitor levels of measurable residual disease. These results provide a genomic framework to anticipate and circumvent relapse by earlier detection and targeting of relapse-fated clones.

Significance: This study defines the landscape of mutations that preexist and arise after commencement of ALL therapy and shows that relapse may be propagated from ancestral, major, or minor clones at initial diagnosis. A subset of cases exhibit hypermutation that results in expression of neoepitopes that may be substrates for immunotherapeutic intervention.

INTRODUCTION

Relapsed acute lymphoblastic leukemia (ALL) is the second leading cause of cancer-related death in children[1]. There are few targeted therapeutic approaches for relapsed ALL and outcome is frequently poor[2], even with the advent of immunotherapeutic approaches. ALL typically exhibits a relatively low burden of somatic mutations, which has allowed delineation of the nature and sequence of acquisition of genetic variants that drive treatment failure[3]. These include inherited variants that are often associated with leukemia subtype (e.g., *TP53* mutations and low hypodiploid ALL), founding chromosomal rearrangements (e.g., *BCR-ABL1* and rearrangement of *KMT2A*), secondary genomic alterations (e.g., alteration of *IKZF1*), and somatic alterations that are enriched from minor clones or acquired after initiation of therapy[4]. Mutations targeting signaling pathways, chromatin patterning, tumor suppression, and nucleoside metabolism are enriched at relapse[5]. These can confer resistance to specific drugs, such as mutations in *NT5C2* to thiopurines[4,6-10] and mutations in the glucocorticoid receptor *NR3C1* and acetyltransferase *CREBBP* to glucocorticoids[11], or confer sensitivity to targeted agents, such as RAS pathway mutations and MEK inhibition[12]. Prior studies also suggest that “relapse-fated” clones commonly exist as minor clones at diagnosis; along with the predominant major clone, these originate from a common ancestral clone that undergoes divergent evolution[8].

The early identification and genetic characterization of relapse-fated clones offer the opportunity to improve treatment outcomes by anticipating relapse and adjusting therapy, or by targeting relapse-fated clones prior to the acquisition of additional mutations facilitating leukemic progression. However, prior genomic studies of relapsed ALL have typically been limited in cohort size and the extent of genomic analysis such that a rigorous analysis of the relapse driver mutations, formal delineation of clonal structure and disease progression, and deep sequencing to distinguish preexisting clones from acquired mutations has not been possible.

RESULTS

Patterns of Relapse in ALL

Multiple tools were used to describe mutational landscape, clonal structure, and clonal evolution for 92 children with ALL and sequential diagnosis, remission and relapse samples treated on St Jude Total Therapy Studies[13,14], results of which may be explored at https://stjuderesearch.org/site/data/relapsed-all (**Table S5.1**). Somatic sequence variants detected at diagnosis and relapse were subjected to confirmatory capture-based sequencing at each time point to optimize estimation of mutant allele frequency (MAF) and time acquisition of relapse-associated mutations (**Figures S5.1A**-**S5.1D**). The deep sequencing identified subclonal somatic mutations in a subset of the remission samples (**Table S5.2**), raising the possibility that mutational persistence early in therapy may predict relapse, as observed in acute myeloid leukemia (AML)[15]. However, comparative analysis of germline mutation burden in 12 cases from this relapsed cohort, with samples at days 27 to 49 and 20 cases that did not relapse, showed no correlation between mutational burden early in therapy and likelihood of relapse (**Supplementary Data**).

The burden of single-nucleotide variants (SNV), short insertions/deletions (indels), and copy number alterations (CNA) increased with disease progression (**Figures S5.2A**-**S5.2D**, **S5.3A**-**S5.3C** and **Tables S5.3**-**S5.7**). Across the cohort, the majority of CNAs (60%) were preserved from diagnosis to relapse, whereas the majority of SNV/indels (74%) were acquired (**Table S5.8**). Twenty-seven tumors from 18 patients were hypermutated (>85 mutations per sample, ~1.3 mutations/Mb; **Figures S5.4A**-**S5.4F**), including 9 of 14 second relapses (64%), 6 of which were already hypermutated at first relapse. Apart from an increased mutational burden at early second relapse, no relationship was observed between mutation burden and time to relapse. *CREBBP* mutations (n = 15 cases) were associated with a longer time to relapse (mean 4.3 vs. 2.8 years, Student t test P = 0.019; **Table S5.9**). Notably, 8 cases with outlier early relapse harbored combinations of alterations known to be involved in relapse development (**Supplementary Data**).

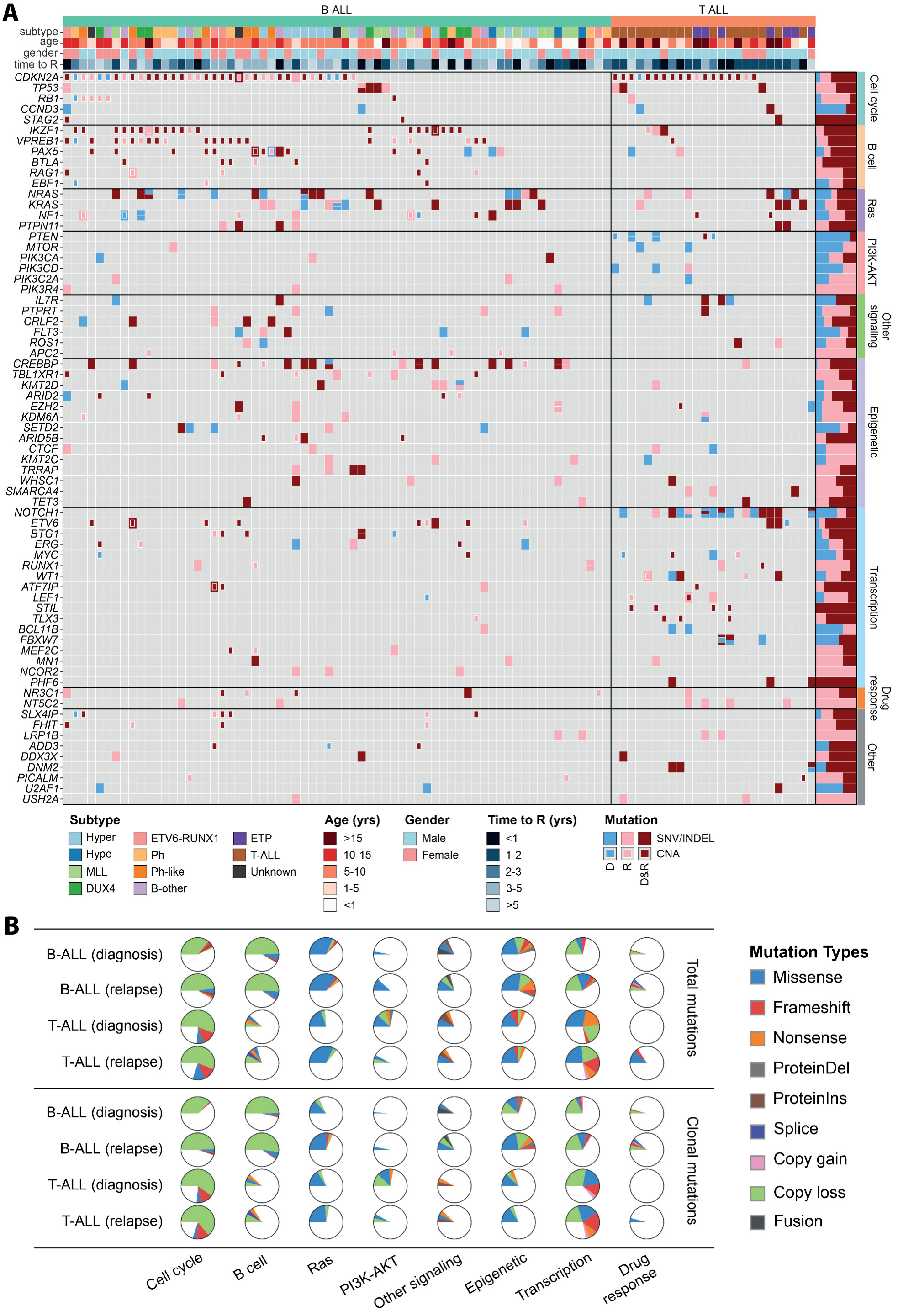
Frequently Mutated Genes and Pathways

A total of 4,509 genes harbored nonsilent sequence mutations at diagnosis (D) or relapse (R; **Tables S5.3** and **S5.4**). Clonal, nonsilent SNV/indels, or focal CNAs were acquired or selected for in 125 genes at first relapse in at least 2 cases and 28 genes in at least 3 cases. Among the recurrent mutated genes (≥2 cases), 38 were known cancer/leukemia genes[16] and 87 had not previously been described (**Table S5.10**). Using GRIN[17], a model that incorporates analysis of multimodal genomic data (Methods; **Tables S5.11** and **S5.12**), 23 genes were significantly mutated (q < 0.1) at diagnosis and relapse, and 50 genes significantly enriched with mutations at R1 (by rising MAF or acquisition of mutation following diagnosis). Most (n = 20, 87%) of the D-R1 shared mutated genes, but only 14 (28%) R1 specific, were known targets of mutation in cancer/leukemia.

B-ALL relapses were enriched with mutations in RAS pathway (relapse 31.3% vs. diagnosis 17.9%) and epigenetic modifiers/regulators, including *PRDM2* (n = 4), *PHF19* (n = 3), *TET3* (n = 3), and *SIN3A* (n = 3), 16 of which had not been reported in ALL

(49.3% vs. 29.9%; **Figures 5.1A**, **5.1B**, **S5.5A**, **S5.5B**, **S5.6** and **Table S5.13**). Of 61 cases with signaling pathway mutations, 31 harbored at least one RAS pathway mutation at diagnosis, with 11 cases having multiple, commonly subclonal RAS pathway mutations at diagnosis (**Table S5.14**). Seven cases showed convergence to one or two clonal RAS pathway mutations. In contrast, only three cases showed acquisition of new RAS pathway mutations at relapse. Thus, multiclonality of signaling pathway mutations is frequent at diagnosis in ALL, indicating that they are secondary lesions in leukemia evolution, and the observed mutational extinction and convergence to clonal dominance supports a selective advantage to RAS pathway mutations in many cases. In contrast, PI3K-AKT pathway mutations were common at diagnosis in T-ALL but were often lost at relapse, suggesting that inhibition of this pathway may not reduce likelihood of relapse.

Of the genes known to play a role in the development of ALL, 29 (including *IKZF1*, *TP53*, *NR3C1*, *TBL1XR1*, and *PTPN11*) showed universal enrichment at relapse, whereby mutations were always retained from diagnosis to relapse (i.e., were truncal variants) or subsequently emerged at relapse (**Table S5.15**). Of these, six were never observed at diagnosis (*NT5C2*, *LRP1B*, *USH2A*, *APC2*, *PIK3R4*, and *NCOR2*). An additional 20 genes showed extinction of mutations present at diagnosis in only a single case (e.g., *VPREB1*, *CREBBP*, *ETV6*, and *KDM6A*). Several genes not previously reported to be mutated in ALL were notable for preservation or clonal selection of mutations from diagnosis to relapse, including *PTPRT* (*n* = 6), *ROBO2* (*n* = 5), and *TRRAP* (*n* = 5), suggesting a role in promoting leukemogenesis and relapse (**Figures S5.7A**, **S5.7B**, **Tables S5.16** and **S5.17**). Mutations in the glucocorticoid receptor (*NR3C1*) and purine/pyrimidine synthesis pathway (*NT5C2*) were frequent in both B- and T-ALL at relapse, but often as subclonal events, suggesting that additional targeting of these drug-specific resistance-driving mutations may eradicate all relapse clones in many cases.


A
B-ALL
T-ALL
subtype
age
gender
time to R
CDKN2A
TP53
RB1
CCND3
STAG2
IKZF1
VPREB1
PAX5
BTLA
RAG1
EBF1
NRAS
KRAS
NF1
PTPN11
PTEN
MTOR
PIK3CA
PIK3CD
PIK3C2A
PIK3R4
IL7R
PTPRT
CRLF2
FLT3
ROS1
APC2
CREBBP
TBL1XR1
KMT2D
ARID2
EZH2
KDM6A
SETD2
ARID5B
CTCF
KMT2C
TRRAP
WHSC1
SMARCA4
TET3
NOTCH1
ETV6
BTG1
ERG
MYC
RUNX1
WT1
ATF7IP
LEF1
STIL
TLX3
BCL11B
FBXW7
MEF2C
MN1
NCOR2
PHF6
NR3C1
NT5C2
SLX4IP
FHIT
LRP1B
ADD3
DDX3X
DNM2
PICALM
U2AF1
USH2A
Cell cycle
B cell
Ras
PI3K-AKT
Other signaling
Epigenetic
Transcription
Drug response
Other
Subtype
Hyper
Hypo
MLL
DUX4
ETV6-RUNX1
Ph
Ph-like
B-other
ETP
T-ALL
Unknown
Age (yrs)
>15
10-15
5-10
1-5
<1
Gender
Male
Female
Time to R (yrs)
<1
1-2
2-3
3-5
>5
Mutation
SNV/INDEL
CNA
D
R
D&R
B
B-ALL (diagnosis)
B-ALL (relapse)
T-ALL (diagnosis)
T-ALL (relapse)
Total mutations
B-ALL (diagnosis)
B-ALL (relapse)
T-ALL (diagnosis)
T-ALL (relapse)
Clonal mutations
Cell cycle
B cell
Ras
PI3K-AKT
Other signaling
Epigenetic
Transcription
Drug response
Mutation Types
Missense
Frameshift
Nonsense
ProteinDel
ProteinIns
Splice
Copy gain
Copy loss
Fusion

◂**Figure 5.1.** Somatic mutation spectrum in ALL at diagnosis and relapse. **(A)** Nonsilent mutations in recurrently affected (⩾3 cases) key genes (COSMIC Cancer Gene Census or reported leukemia relevant genes) in diagnosis (D) and first available relapse (R) sample per case. The B-ALL cases are grouped into well-defined disease subtypes, which include hyperdiploid (Hyper), hypodiploid (Hypo), *KMT2A* (*MLL*)-rearranged, *DUX4*-rearranged (DUX4), *ETV6-RUNX1*, *BCR-ABL1* (Ph), Ph-like, and a group of other B-ALL subtypes including B-other, *PAX5* p.P80R, and iAMP21 ALL. Mutations in the form of SNV/indels and focal CNAs are shown as rectangles with different sizes. Mutations observed only in D, only in R or shared by D and R are shown in blue, pink, and dark red colors, respectively. The prevalence for each gene mutation is shown in bar graph on the right. **(B)** Distribution of recurrent mutations in key pathways. Top, all recurrent mutations; bottom, the clonal (MAF ⩾ 30%) nonsilent mutations. Samples are divided into B-ALL (*n* = 67) and T-ALL (*n* = 25) and the mutation ratio in diagnosis and relapse stages is shown. Detailed mutation types are indicated by different colors.

Integration of Mutational Landscapes with Clonal Structure

Using the rise and fall of CNAs and MAF of somatic SNV/indels, most tumor pairs [*n* = 79 D-R1/secondary tumor (S; 86%), and *n* = 12 R1-R2/S (92%)], showed clonal extinction and evolution of new clones in the subsequent tumor, indicating branching evolution. One quarter of the relapses (*n* = 28, including 2 second relapses) arose from the major clone (MAF > 30%) present at diagnosis or the previous relapse, and half of the relapses (*n* = 53, of which 6 were second relapses) developed from a minor clone (**Figures 5.2A**, **5.2B** and **Table S5.18**). Of the 53 relapses arising from a minor clone, nine (17%) had relapse-enriched mutations already present in that subclone and 26 (49%) acquired new or additional relapse-enriched mutations following diagnosis. Nineteen tumors (18%), exhibited polyclonal evolution in which multiple diagnosis clones persisted at relapse.

We found that second relapses evolved more often in a polyclonal fashion ($\chi^2 P = 0.043$) and with a shorter remission time than first relapses (average 1.5 years vs. 3 years, Student *t* test, $P = 0.0044$). Compared with first relapses, variants in second relapses were more often variants preserved from subclones (10% vs. 4%) or preserved at subclonal levels (11% vs. 2%) and less often acquired (69% vs. 79%), which reflects the polyclonal evolution model ($\chi^2 P < 2.2 \times 10^{-16}$). Thus, initial evolution from diagnosis to relapse is characterized by mutational convergence, and commonly, emergence from a minor clone, but subsequent progression exhibits preservation of the initially selected clones and variants.

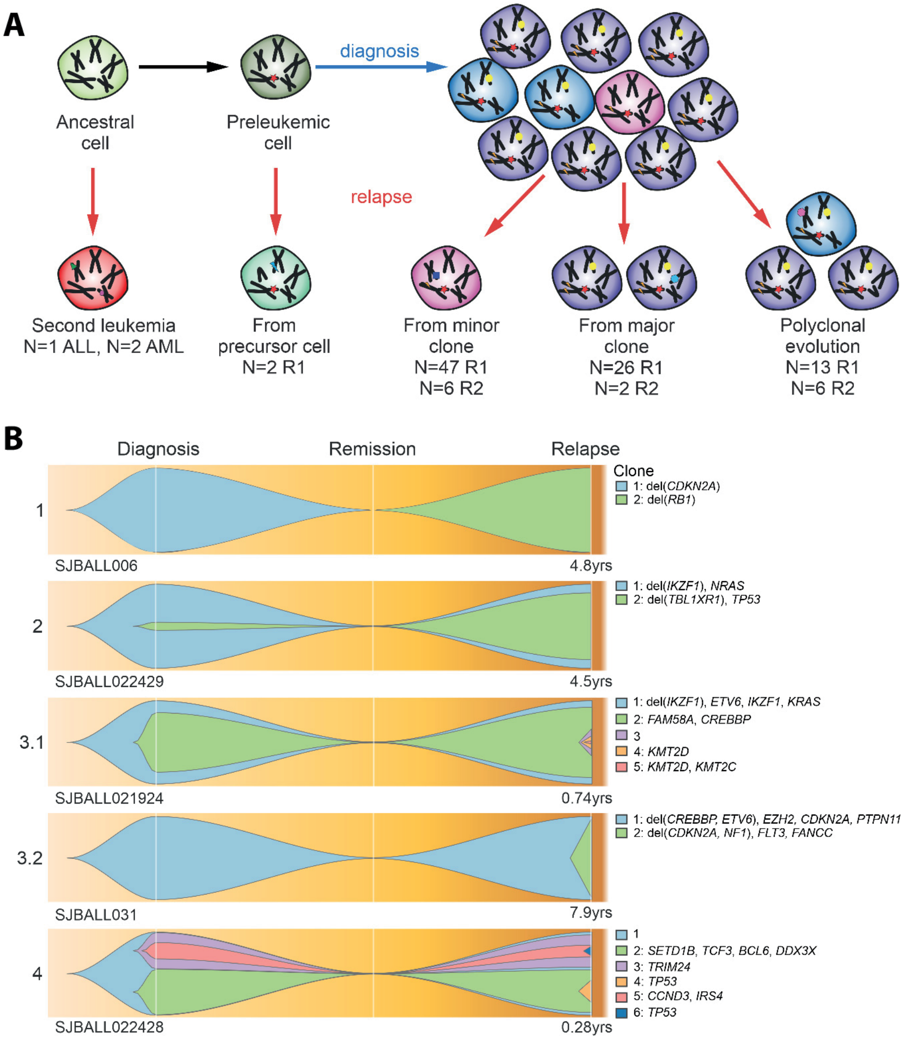

A
Ancestral cell
Preleukemic cell
diagnosis
relapse
Second leukemia
N=1 ALL, N=2 AML
From precursor cell
N=2 R1
From minor clone
N=47 R1
N=6 R2
From major clone
N=26 R1
N=2 R2
Polyclonal evolution
N=13 R1
N=6 R2
B
Diagnosis
Remission
Relapse
Clone
1
1: del(CDKN2A)
2: del(RB1)
SJBALL006
4.8yrs
2
1: del(IKZF1), NRAS
2: del(TBL1XR1), TP53
SJBALL022429
4.5yrs
3.1
1: del(IKZF1), ETV6, IKZF1, KRAS
2: FAM58A, CREBBP
3
4: KMT2D
5: KMT2D, KMT2C
SJBALL021924
0.74yrs
3.2
1: del(CREBBP, ETV6), EZH2, CDKN2A, PTPN11
2: del(CDKN2A, NF1), FLT3, FANCC
SJBALL031
7.9yrs
4
1
2: SETD1B, TCF3, BCL6, DDX3X
3: TRIM24
4: TP53
5: CCND3, IRS4
6: TP53
SJBALL022428
0.28yrs

Figure 5.2. Patterns of relapse in ALL. **(A)** Schematic overview of mechanisms of clonal evolution. Three patients developed a second primary tumor that was not clonally related to the previous tumor occurrence. Two patients developed a tumor that shared only one founding fusion between diagnosis and relapse, indicating the disease relapsed from a preleukemic cell. Further relapses arose through evolution from a minor clone, a major clone, or multiple clones. **(B)** Fish plots of the clonal evolution models inferred from somatic mutations detected in diagnosis and relapse samples. MAF of the somatic mutations was used by the sciClone R package[14] to infer potential clonal clusters (shown in different colors) and visualized by Fishplot[13] (see Methods). Four major clonal evolution models were observed: 1, relapse sample is a second primary leukemia with no somatic mutations shared with diagnosis; 2, a minor clone (somatic mutations' median MAF of the clone is less than 30%) in diagnosis develops into the major clone in relapse; 3, a major clone (somatic mutations' median MAF of the clone is greater than 30%) is preserved from diagnosis to relapse (3.1) or emerges as a major clone at relapse (3.2); 4, multiple subclones in diagnosis develop as multiple subclones in relapse. For each exemplary case, the time from diagnosis to relapse is indicated. Focal deletions and nonsilent somatic mutations on cancer genes (according to the COSMIC Cancer Gene Census and well-known leukemia relevant genes) for each inferred clone are shown on the right side.

Second Primary Leukemia

Four first relapses (SJBALL006, SJTALL142, SJPHALL005, and SJPHALL022425) and one second relapse (SJTALL049) were fully discordant for all genetic alterations (SV, CNA, SNV/Indel, and antigen receptor rearrangements) or shared only a leukemia fusion (**Tables S5.19** and **S5.20**), suggesting distinct second leukemias masquerading as relapse. These scenarios are important to distinguish, as second leukemias may be curable with standard therapy and multiple primary tumors suggest heritable leukemia predisposition. Both tumors in SJBALL006 harbored *MEF2D-BCL9* fusions, but with unique RNA and DNA breakpoints, and evidence of the second *MEF2D-BCL9* fusion at low level in the primary sample (**Tables S5.20** and **S5.21**). In addition, there was a constitutional gain of the chromosome 1q neuroblastoma breakpoint (*NBPF*)[18] region on chromosome 1q that contains *MEF2D* and *BCL9*, and discordant somatic complex genomic amplifications of the *NBPF* region at diagnosis and relapse (**Figures 5.3A**-**5.3D**, **Tables S5.6** and **S5.22**). Thus, while emergence of a tumor with the same fusion partners suggests relapse, this case represents germline structural variation promoting development of multiple tumors with distinct initiating fusions and different latencies of presentation. *BCR-ABL1* cases SJPHALL005 and SJPHALL022425 lacked shared nonsilent variants at diagnosis and relapse, raising the possibility of second leukemia rather than relapse.

However, whole genome sequencing (WGS) of SJPHALL005 identified identical rearrangement breakpoints at diagnosis and relapse and 65 shared somatic noncoding SNVs, demonstrating relapse from a common, ancestral clone (**Table S5.20**). SJTALL049 developed three tumors: *STIL-TAL1* rearranged T-ALL at age 6 that relapsed at age 14 plus an independent *BCR-ABL1*-positive chronic myeloid leukemia (CML) at age 20. Breakpoint spanning PCR revealed unique *STIL-TAL1* rearrangements at diagnosis and relapse, but WGS showed 19 shared noncoding SNVs. WGS showed no shared SNVs between either of the T-ALLs and the CML, indicating that the CML developed as a second primary tumor. Thus, completely genetically distinct second tumors masquerading as relapse of acute leukemia are rare, and even if complete nonsilent mutational discordance is present, they may arise from divergent evolution soon after leukemia initiation. Three relapses were clinically considered second tumors based on a shift to myeloid lineage (SJTALL008, SJTALL124, and SJTALL164). However, these tumors revealed shared mutations between diagnosis and relapse, indicating a common clonal origin (**Supplementary Data**). This recapitulates the lineage plasticity that is independent of genetic variegation we have recently described in acute leukemia of ambiguous lineage[20] and highlights the importance of genomic analysis to accurately interpret the relationship of diagnosis and relapse samples.

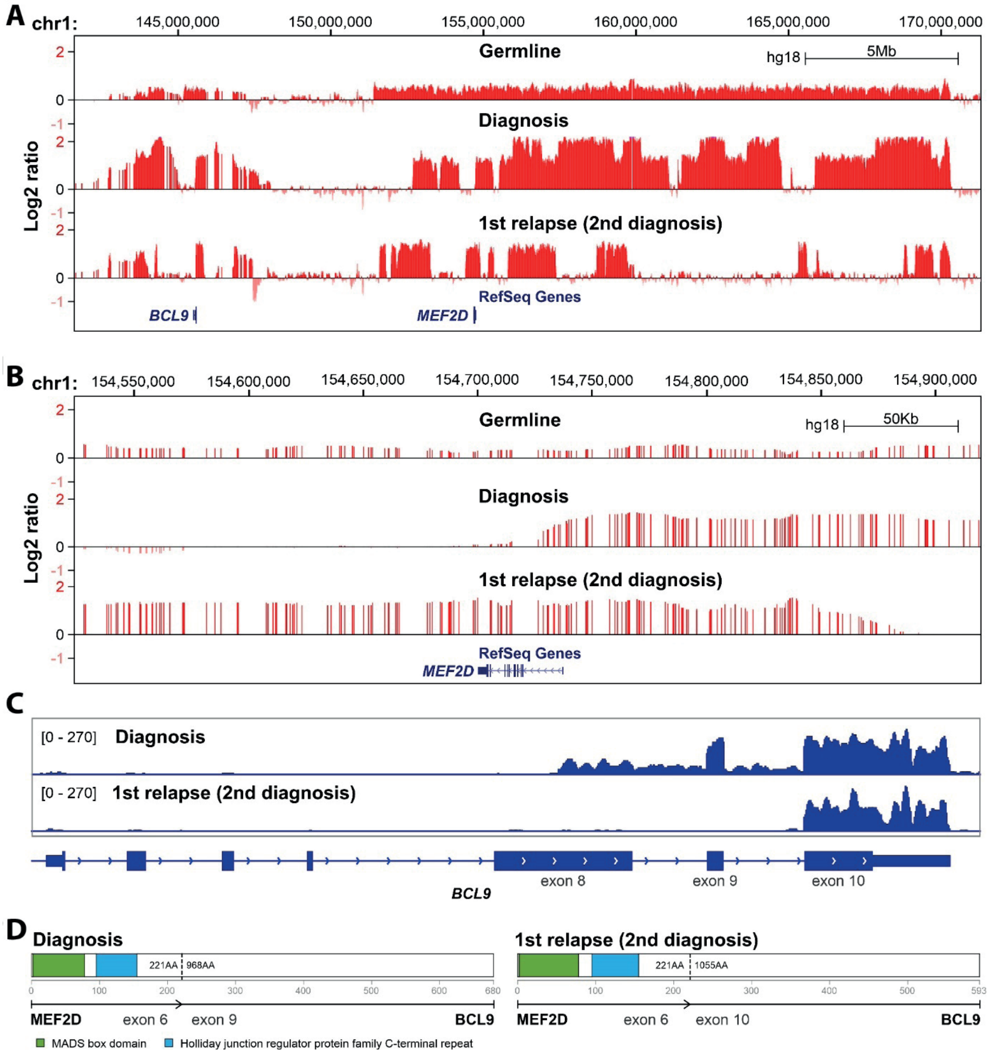


Figure 5.3. CNAs and *MEF2D-BLC9* rearrangements in patient SJBALL006. Signal intensity from Affymetrix SNP6.0 microarrays was normalized to log2 ratio (>0 indicate copy gain; <0 indicate copy loss) and shown in the University of California, Santa Cruz Genome Browser in large **(A)** and focal scale **(B)** to show the distinct alteration patterns between diagnosis and "relapse" (second diagnosis) samples. Constitutional copy number gains were observed in the germline sample. **(C)** RNA-seq depth on exons of *BCL9* gene. The sequencing depth was scaled from 0- to 270-fold for both diagnosis samples. The uptick of expression of exon 9 and 10 was observed for first and second diagnosis samples, respectively, indicating different rearrangement breakpoints on *BCL9*, which was consistent with fusions called from RNA-seq. The RNA-seq library for the first diagnosis sample was total RNA, so the intronic region was covered by sequencing reads. **(D)** Schematic visualization of *MEF2D-BCL9* chimeric protein structure. Two fusion isoforms have been reported as the most recurrent *MEF2D-BCL9* rearrangements[19].

Tracing the Evolution of Relapse

Bulk sequencing data may fail to unambiguously define a clonal evolution model, particularly for mutationally sparse samples and those exhibiting a continuum of variant MAFs. We performed limiting dilution xenografting and sequencing of eight matched diagnosis and relapse samples **(Table S5.1)**. Most (90.5% of 232) of the somatic mutations detected in primary samples were identified in at least one diagnosis or relapse xenograft (MAF ≥ 0.01; **Figure S5.8A**). Of the 22 mutations not captured in xenografts, 20 were observed only in the bulk diagnosis sample and not in the relapse sample, suggesting that such nonxenografted diagnosis mutations must be present in cells that lack clonal propagating potential and are less likely to initiate relapse.

Genomic analysis including xenografting overcame the challenge of assigning mutations to individual subclones and enabled unambiguous delineation of clonal structure. Xenograft analysis of SJBALL036 (*ETV6-RUNX1*-like subtype) identified two linearly related clones (2.1 and 2.2) from mutations originally allocated to a common clone (**Figure 5.4A**). Mutations in relapse-fated clone 2.1 were clonal in all xenografts, including those propagated from diagnosis, whereas the *CREBBP* mutations in clone 2.2 were observed in a subset of xenografts, indicating that subclone 2.2 arose from 2.1. In addition, xenografting demonstrated the selective advantage of clone 2.1 versus clone 5 that was lost at relapse and was not represented in any of the xenografts transplanted with the diagnosis sample. Furthermore, the xenografts derived from the relapse sample captured mutually exclusive variants, providing definitive evidence that clones 3 and 4 were unrelated, and represent branching evolution (**Figures 5.4B** and **5.4C**). Similarly, xenograft data of SJETV010, of which the second relapse sample is hypermutated (1,699 somatic mutations), resolved 13 clones following linear and branching patterns of evolution (**Figures S5.8B** and **S5.8C**). These data support the notion of branching evolution in ALL suggested by FISH and bulk sequencing analysis[21,22], but now with mutational data enabling unambiguous clonal delineation. A subset of xenografts has also been utilized to demonstrate that relapse-fated clones may be detected at diagnosis that already exhibit resistance to therapy[23,24].

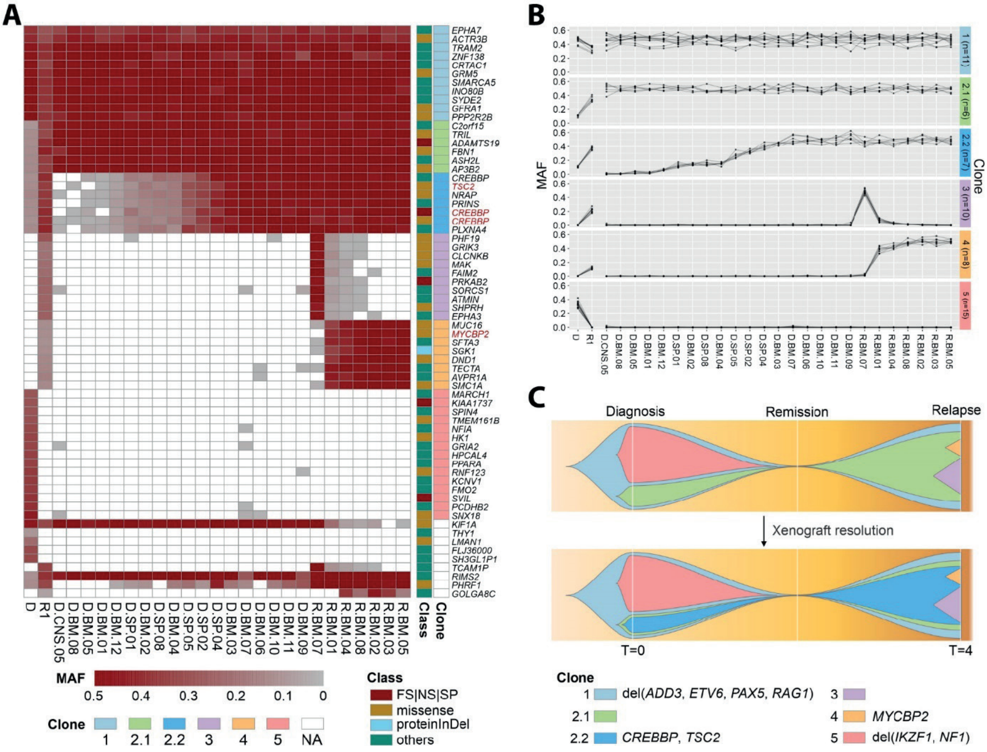


Figure 5.4. Integration of mutational landscape and xenografts resolves clonal structure in ALL. **(A)** Somatic mutation spectrum of diagnosis (D), first relapse (R1), and xenografted leukemia samples. Leukemic cells from D (D.*.#) and R1 (R.*.#) from patient SJBALL036 were xenografted in mice and collected from bone marrow (*.BM.#), central nervous system (*.CNS.#), and spleen (*.SP.#). Cancer genes with nonsilent mutations are highlighted in red. FS, frameshift; NS, nonsense; SP, canonical splice site; proteinInDel, protein insertion/deletion. **(B)** Delineation of clonal model from xenografted samples. MAF of SNV/indels were analyzed by sciClone[14] to infer clonal clusters. On the basis of the MAF in D and R1, clone 2.1 and 2.2 were indistinguishable. Xenograft data shows that clone 2.1 rises as a major clone (MAF = 0.5) in relapse alone or together with clone 2.2, indicating that 2.1 is the parental clone of 2.2. In addition, xenograft data showed variability in MAFs between clones 3 and 4, indicating that clones 3 and 4 were two distinct subclones. The clones are color-coded in the schema as in A. The number of somatic mutations in each clone is shown in parentheses. **(C)** Fishplot of the leukemia evolution model. The top plot shows the original evolution model based on D and R1, and the bottom plot is the refined evolution model after incorporating the information from xenografted samples. The time (T) at diagnosis is defined as 0 and the first relapse was observed 4 years later. Nonsilent mutations and focal deletions (Del) affecting cancer genes are highlighted for each clone.

Tracing Mutation Acquisition Prior to Relapse

As xenografting identified resistance-driving mutations in relapse-fated subclones present at diagnosis, we sought to determine whether these mutations could be detected in patient samples obtained between diagnosis and relapse. We used droplet digital PCR (ddPCR) to track the emergence of relapse-specific mutations in *CREBBP*, *NRAS*, *KRAS*, *NT5C2*, and *WHSC1* in 50 samples from 5 patients (**Table S5.23**). With an input of 500 ng DNA a frequency of >0.005% (>7 copies) could be consistently detected (**Figures S5.9A-S5.9D** and **Table S5.24**). ddPCR identified previously undetected minor clones in three tumor samples (*NRAS* p.G12R MAF = 0.4% in SJBALL013-R1, *KRAS* p.G12S MAF = 0.4% in SJBALL022481-D, *NT5C2* p.R39Q MAF = 0.006% in SJBALL192-R1; **Figure 5.5**), confirming the minor clone evolution model for these tumors.

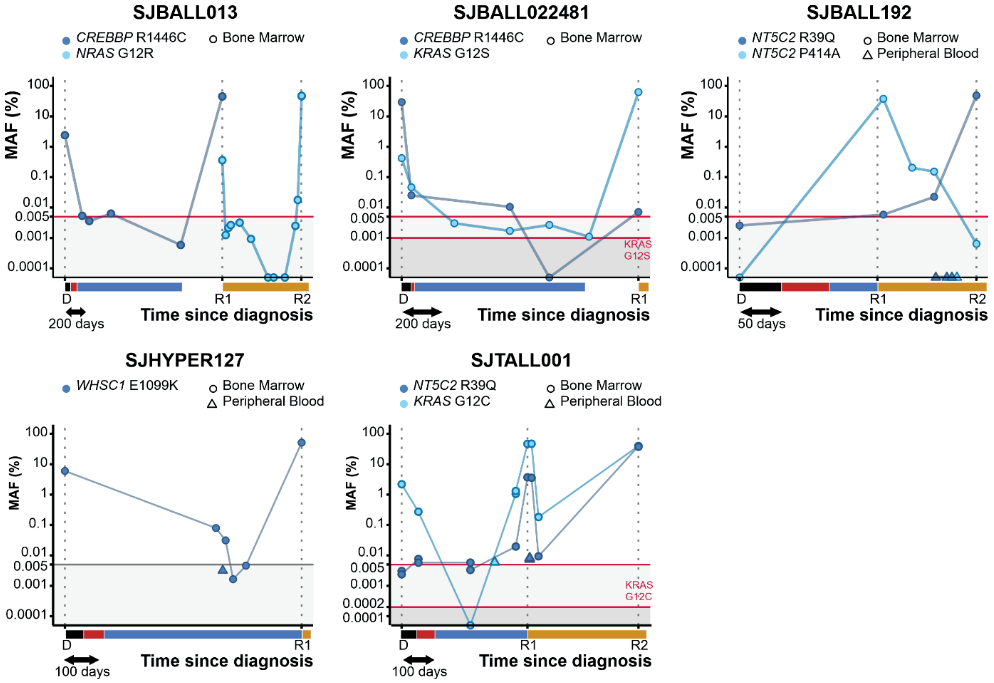


Figure 5.5. ddPCR reveals mutations at low levels in intermediate complete remission samples. MAF of the indicated variants was determined in bone marrow (circle) and peripheral blood (triangle) samples for 5 patients. The time to diagnosis is scaled on the x-axis, with the treatment blocks indicated in black (induction), red (consolidation), blue (maintenance), and orange (relapse treatment). SJBALL192, SJHYPER127, and SJTALL001 relapsed during maintenance treatment. Detection limits are indicated with a red horizontal line and shaded background. Detection limits in gray were extrapolated from the other assays (i.e., not experimentally determined). The MAF at relapse of *WHSC1* in SJHYPER127 was determined in our capture validation analysis as no DNA was available for ddPCR. The y-axis is in logarithmic scale.

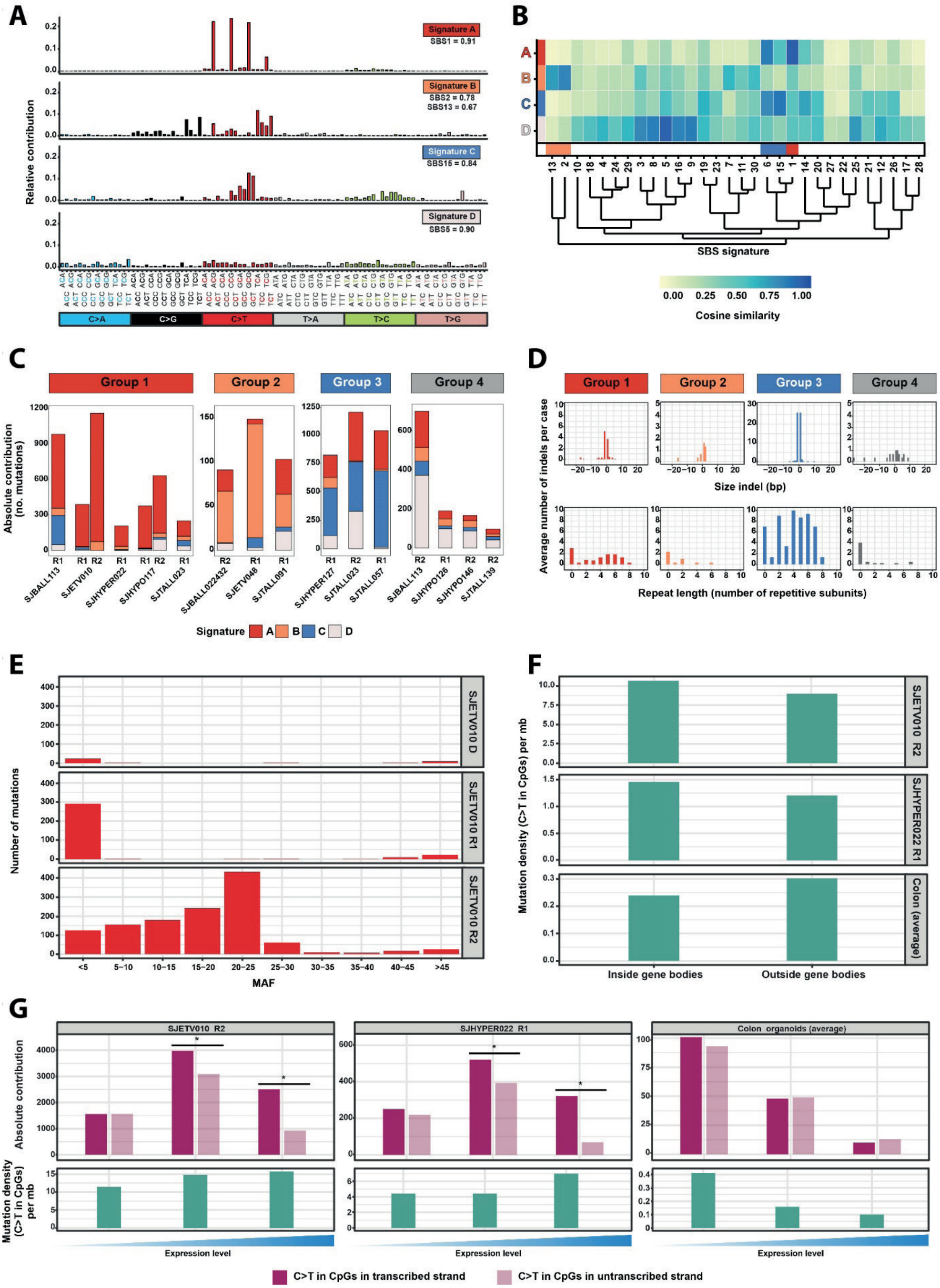
Difference in the temporal dynamics and occurrence of mutations in SJBALL022481 (*CREBBP* and *KRAS*), SJBALL192 (two *NT5C2* mutations), and SJTALL001 (*NT5C2* and *KRAS*) demonstrated clonal exclusivity of these mutations in each case. Despite complete remission by conventional minimal residual disease (MRD) testing, we detected tumor-specific mutations up to 534 days after the diagnosis (SJBALL022481, *CREBBP* p.R1446C), as well as 40 days prior to relapse (SJBALL013 *NRAS* p.G12R) in complete remission bone marrow samples. Moreover, in SJBALL022481, the *KRAS* p.G12S mutation was detectable at regular intervals during the 1,169 day period between diagnosis and relapse, even though the samples were deemed complete remission. Peripheral blood samples obtained from patients with B-lineage (SJBALL192, SJHYPER127) as well as T-lineage ALL (SJTALL001) with eventual bone marrow relapse were negative or contained much lower MAFs compared with the bone marrow. Thus, leukemic cells may persist in bone marrow and may be detected at low levels during complete remission, indicating the utility of this approach for disease monitoring and early relapse detection.

Mutational Drivers of Hypermutation and Neoepitope Expression

Three percent of diagnoses, 17% of first relapse, and 64% of second relapse cases exhibited hypermutation. This was defined by an inflection at 85 mutations, approximately 1.3 mutations/Mb, a burden that was more conservative than the cutoff determined by the Segmented algorithm[25] (**Figure S5.4A**). Hypermutation was observed in cases relapsing from minor or multiple clones of all 3 hypodiploid, 2 of 5 *ETV6-RUNX1*, 5 of 13 hyperdiploid, 1 Ph-like, 2 unclassified B-ALL, 1 ETP, and 4 T-ALL cases. To understand the processes responsible for hypermutation, we used non-negative matrix factorization (NMF)[26] and extracted four single-base substitution (SBS) signatures (**Figures 5.6A**, **5.6B**, **S5.10A** and **S5.10B**). The high prevalence of hypermutation in second relapses suggests that hypermutation may be driven by treatment exposure.

However, we did not uncover a mutational signature associated with treatment, such as temozolomide-associated signatures found in glioblastoma and melanoma[26]. On the basis of the most prominent mutational signature (**Figure S5.10B**), we classified 17 hypermutated relapses from 13 patients into four groups (**Figure 5.6C**). Two of these groups were characterized by well-established mutational processes. Group 2 (signature B mutations) resembles signatures associated with activity of the AID/APOBEC family of cytidine deaminases[27,28]. This mutational process is common in human cancer including ALL[26] and was postulated to occur in short bursts initiated by retrotransposon mobility[29]. The mutational burden in this group was relatively low (99-156 acquired mutations) compared with the other three groups (group 1; 220-1,210, group 3; 860-1,218, group 4; 104-710). Group 3 cases have high contribution of signature C, which clusters with mismatch repair (MMR)-associated signatures, with highest similarity to SBS15 (**Figure 5.6B**). Indeed, the three relapses in this group all had biallelic mutations in one of the MMR genes (see **Supplementary Data**) and had high levels of single-base insertions or deletions in simple repeats (**Figure 5.6D**), a feature of MMR deficiency[30]. Genetic alterations in the MMR pathway have been associated with resistance to drugs such as thiopurines in ALL[19], suggesting that this mechanism of hypermutation directly contributed to relapse in these cases. Two signatures could not be assigned to known mutational processes. Signature D (group 4) lacked a strong bias for a particular trinucleotide context and showed similarity to multiple mutational signatures. Signature A (group 1) resembled clock-like signature SBS1, which is a known consequence of a slow but progressive accumulation of C to T transitions at CpG sites owed to spontaneous deamination of methylated cytosines and is more apparent in cancers diagnosed at older age[30]. Because the patients in our cohort are young, this process appears to be accelerated by an acquired imbalance between damage and repair. This was not caused by alterations in genes encoding regulators of DNA deamination. Interestingly, the signature A mutations in SJETV010 were subclonal (MAF $< 0.5\%$) in the first relapse, but showed a much wider spread of allele frequencies in the second relapse, suggesting an ongoing endogenous mutational process initiated in a minor clone at first relapse (**Figure 5.6E**). Recently, an SBS1-like signature (SBS74) has been reported that appears to be associated with MMR deficiency[31]. Indeed, all hypermutated ALL relapses with MMR deficiency show signature A mutations (**Figure 5.6C** and **Supplementary Data**).

To further compare the characteristics of signature A with the clock-like signature SBS1, we performed WGS of relapse and remission samples of SJETV010, SJHYPER022, and SJHYPO117, followed by somatic SNV calling and mutational signature extraction. In line with our findings in the coding regions, we identified highly concordant mutation densities, as well as composition and relative contribution of the four mutational signatures (**Figures S5.10C-S5.10F**). Signature SBS1 mutations have been described to occur throughout the genome and do not show strand asymmetry in transcribed regions[26,32]. We confirmed these observations using three recently published colon organoid samples[33], which are characterized by high prevalence of signature SBS1 mutations, mainly outside gene bodies. In contrast, we found that signature A mutation density was highest in gene bodies and showed strong transcription strand bias, particularly in genes with high expression in the respective samples (**Figures 5.6F**, **5.6G** and **S5.10G**). We did not observe strand asymmetry associated with DNA replication for either signature A or signature SBS1 (**Figure S5.10H**). Transcriptional strand asymmetry can be the result of more efficient repair of the transcribed strand, or increased damage on the single-stranded, nontranscribed strand, two mechanisms that show opposite correlations with expression[34]. Because signature A mutations are enriched in highly expressed genes, they may originate from transcription-coupled damage at the single-stranded, nontranscribed strand, as has been described for liver cancer[34].


A
Relative contribution
Signature A
SBS1 = 0.91
Signature B
SBS2 = 0.78
SBS13 = 0.67
Signature C
SBS15 = 0.84
Signature D
SBS5 = 0.90
0.2
0.1
0.0
C>A
C>G
C>T
T>A
T>C
T>G
B
A
B
C
D
13 2 10 18 4 24 29 3 8 5 16 9 19 23 7 11 30 6 15 1 14 20 27 22 25 21 12 26 17 28
SBS signature
0.00 0.25 0.50 0.75 1.00
Cosine similarity
C
Group 1
Group 2
Group 3
Group 4
Absolute contribution (no. mutations)
R1 R1 R2 R1 R1 R2 R1 R2 R1 R1 R1 R2 R1 R2 R1 R2 R1
SJBALL113
SJETV010
SJHYPER022
SJHYPO117
SJTALL023
SJBALL022432
SJETV048
SJTALL091
SJHYPER127
SJTALL023
SJTALL057
SJBALL113
SJHYPO126
SJHYPO146
SJTALL139
Signature A B C D
D
Group 1
Group 2
Group 3
Group 4
Average number of indels per case
Size indel (bp)
Repeat length (number of repetitive subunits)
E
Number of mutations
SJETV010 D
SJETV010 R1
SJETV010 R2
<5 5-10 10-15 15-20 20-25 25-30 30-35 35-40 40-45 >45
MAF
F
Mutation density (C>T in CpGs) per mb
SJETV010 R2
SJHYPER022 R1
Colon (average)
Inside gene bodies
Outside gene bodies
G
SJETV010 R2
SJHYPER022 R1
Colon organoids (average)
Absolute contribution
Mutation density (C>T in CpGs) per mb
Expression level
C>T in CpGs in transcribed strand
C>T in CpGs in untranscribed strand

◄**Figure 5.6.** Mutational signature analysis of hypermutated relapses identifies multiple distinct mutational processes in hypermutation. **(A)** Four mutational signatures identified in hypermutated ALL. Relative contribution of the different mutation types in their trinucleotide context, and cosine similarity values to reported COSMIC signatures are shown. **(B)** Cosine similarity heatmap showing the hierarchical clustering of *de novo* signatures identified in this study with 30 known SBS signatures, including those associated with AID/APOBEC (orange bar), spontaneous deamination of methylated cytosines (red bar), and mismatch repair (blue bar). **(C)** Absolute contribution of each of the four signatures to the acquired mutations in 17 hypermutated relapse samples from 13 patients. Samples are grouped on the basis of the most prominent contributing signature. **(D)** Average number and size of acquired indels in samples assigned to each group (top) and the number of repetitive subunits surrounding an inserted or deleted subunit (bottom). A value of 0 indicates that the indel is not located within a simple repeat. **(E)** Total number of mutations (acquired and preserved) assigned with >95% confidence to signature A in the tumors of SJETV010, binned based on the mutation allele frequency (MAF). **(F)** Density of C>T transitions in CpGs inside and outside gene bodies of two hypermutated relapses (SJETV010R2 and SJHYPER022R1) with high contribution of signature A mutations (top and middle) and healthy colon organoids with high contribution of SBS1 mutations (average of 3 organoids; bottom). **(G)** Bar plots showing number of C>T transitions in CpGs on the transcribed and nontranscribed strand in relation to gene expression (top) and density of C>T transition in CpGs (bottom) in genes with no, low (<median) and high (≥median) expression (*, $P < 0.05$).

The high prevalence of hypermutation in relapsed ALL suggested that this may result in increased generation of expressed neoepitopes that may be exploited by immunotherapeutic approaches to enhance antitumor responses of autologous T cells. We used WGS and RNA sequencing (RNA-seq) data to infer HLA class I types from each sample[35], predicted the binding affinities of all unique 8-12 amino acid peptides corresponding to SNVs and fusion proteins[36], and developed unweighted (UPAS) and weighted putative antigenicity scores (WPAS), the latter of which incorporates predicted sample-specific peptide: MHC binding with variant-specific expression. Although we observed variation in the number of fusion-encoded, predicted MHC-binding peptides across individual fusions (0-20, median = 1), fusion-encoded neoepitopes remained unchanged over time (**Table S5.21**). In contrast, we observed that the number of predicted HLA-binding mutant peptides[37] (≤500 nmol/L) per tumor rose with disease progression as a function of increased mutation burden, and thus particularly in hypermutated samples ($P < 0.001$, **Table S5.25** and **Figure S5.11A**). In addition, the number of predicted MHC-binding peptides per tumor was significantly correlated with disease (B- and T-ALL), disease progression (D, R1, and R2), and signatures of hypermutation (**Figure S5.11B**). An expression-weighted antigenicity analysis comprising the subset of missense SNVs with expression data showed a significant effect of disease progression, with median WPAS lowest in R1 and highest in R2 variants (**Figure S5.11C**), and was particularly marked for known cancer genes (**Figure S5.11D**). These patterns may correspond to variations in immunologic pressure owed to, for instance, the distinct etiologies underlying B- and T-ALL and the successive use of immunosuppressive agents in treatment, respectively.

DISCUSSION

Using multiple orthogonal approaches, we have described the patterns, dynamics, and drivers of clonal evolution in a large cohort of childhood relapsed ALL. These results have implications for the development of new approaches to monitor and treat ALL more effectively.

The scope of our study allowed us to identify relapsed enriched driver mutations more comprehensively than in prior studies and included newly identified targets of mutation as well as recurrent mutations in genes previously identified at relapse[11,38,39]. We were able to identify distinct patterns of temporal acquisition across relapse-enriched targets of mutation, with mutations in genes such as *CREBBP* preserved from or acquired after diagnosis, and others in genes such as *NT5C2* and *USH2A* only observed after initial therapy; importantly, these findings suggest a role for therapy in the induction of mutation, and/or a deleterious effect on initial leukemic fitness of these mutations[38]. Mutations observed in different gene regulation pathways showed different frequencies of relapse-enriched genes between B- and T-ALL, indicating that distinct biological mechanisms drive the genetic alterations of the disease progression.

Tumors that evolved from, or retained multiple clones all had a short time to relapse, supporting a model in which early relapses are associated with dynamic clonal evolution and late relapses to a more inert pattern[40]. These patterns of evolution, and description of the targets of mutation, have important implications for anticipation of relapse and modulation of therapy. Over half of relapses arise from a minor clone that commonly harbors, or acquires, relapse-enriched mutations that drive drug resistance. Analysis of this large cohort also enabled demonstration that relapse enriched drivers such as alterations of *CREBBP*, *IKZF1*, and *NT5C2* are rarely lost if present at diagnosis or subsequently acquired, indicating that early detection may predict an increased likelihood of treatment failure. Moreover, in parallel studies using a subset of the xenografts described here, we have shown relapse-fated minor clones already exhibiting resistance to therapy are present at the time of diagnosis[23,24]. Thus, it is imperative that mutational profiling must now strive to achieve MRD levels of mutation detection at diagnosis or early in therapy, and we have shown the feasibility of this approach using ddPCR. An alternative approach is capture-based deep sequencing of regions of sequence and structural variation in ALL. Early detection will facilitate close monitoring and consideration of alternative treatment approaches such as intensification, immunotherapy, or novel drug approaches for drug-specific resistance (e.g., *NT5C2* and thiopurines and *CREBBP* and corticosteroid resistance).

We show hypermutation is common at relapse, and driven by distinct mechanisms of mutation, including tumor-intrinsic processes, such as cytidine deaminase DNA editing activity (AID/APOBEC) as previously observed in diagnosis samples[41] and experimental models[42], or the acquisition of mutations that cause MMR deficiency, which may represent a mechanism of MMR-induced resistance to thiopurines in ALL[19]. In addition, we identified a new SBS1-like signature (Signature A), characterized by transcriptional strand asymmetry and enrichment in expressed genes, caused by an unknown mutational mechanism that is acquired during leukemia progression. Importantly, the identification of hypermutation as a common phenomenon in relapsed ALL suggests that immunotherapeutic approaches intended to restore autoreactivity against neoantigen expression, such as checkpoint blockade, should be formally explored. Although long assumed to be poor targets for immunotherapy due to the relatively low mutation burden in comparison with other tumors[43], recent data have demonstrated that pediatric hematologic malignancies promote the generation of abundant and functional immune responses to tumor-specific neoepitopes despite the apparent inability of the immune system to control those tumors[44]. In conjunction with those findings, our results suggest that not only does hypermutation drive the generation of HLA-restricted neoepitopes, but that these are expressed in an immunologically tolerized milieu that may be exploited with strategies to augment T-cell antitumor reactivity. It will now be of great interest to formally document the presence of autoreactive T-cell clones directed at neoepitopes induced by hypermutation, as we have described at diagnosis in ALL[44], and to formally test, in experimental models, whether immunomodulatory approaches can augment or restore antitumor reactivity in hypermutated ALL.

METHODS

Subjects and Samples

The cohort included 92 children (31 female, 61 male) with relapsed B-progenitor (B-ALL, *n* = 67) or T-lineage [T-ALL, *n* = 25; including 7 early T-cell precursor ALL (ETP-ALL)] ALL patients treated on St. Jude Total Therapy Studies XI-XVI[45-47] (median age at diagnosis 7.8 years; range 1 month-18.7 years; **Table S5.1**). The median time from diagnosis to first relapse was 2.7 years (range 3 months-9.9 years). Sixteen cases developed a second relapse (range 3 months-7 years). Relapse was very early (<1 year) in 17 patients, early (1-2 years) in 21 patients, and late (>5 years) in 14 patients. Nine cases developed a second tumor of different lineage, including 2 basal cell carcinomas, 1 B-cell lymphoma, 1 CML, and 5 AMLs. A total of 46 patients received bone marrow transplants at a median age of 12 years (range, 7 months-22 years), and all except one were allogeneic. Tumor samples with a blast percentage of less than 80% were flow sorted for the tumor population. Written informed consent was obtained from the patient and/or parent. The study was conducted in accordance with the Declaration of Helsinki, and was approved by the St. Jude Children's Research Hospital Institutional Review Board.

Genomic Analyses

DNA copy number aberrations were determined using SNP 6.0 microarrays (Affymetrix) in 307 samples from 92 patients (92 diagnosis, 84 relapse, 14 second relapse, 5 second tumor, and 91 germline samples; **Table S5.1**). Data were analyzed using reference normalization[48] and circular binary segmentation[49].

We performed whole exome sequencing (WES) on 276 samples from 92 cases (**Table S5.26**) and WGS on 99 samples from 36 cases (**Table S5.27**). Exomes were captured using the TruSeq Exome Library Prep Kit (67 Mb, 1 µg DNA input) or the Nextera Rapid Capture Expanded Exome Kit (62 Mb, 50 ng DNA input; Illumina). Paired-end sequencing was performed with the HiSeq 2500 sequencer (Illumina). The data were mapped to human reference genome hg19 and variant calling was performed using Bambino[50-52]. All somatic SNVs and indels identified at diagnosis or relapse were validated using NimbleGen SeqCap Target Enrichment according to manufacturer's instructions (Roche NimbleGen) and resequenced using a HiSeq 2500 genome sequencer to a mean coverage >350 × (250-500 ng DNA input; **Table S5.28**). We performed transcriptome sequencing (RNA-seq) for TRIzol extracted RNA for 115 samples obtained from 66 cases (**Table S5.29**). We used 1 µg RNA for library preparation with the TruSeq RNA Library Prep Kit v2 (Illumina) and 2 × 100 bp paired-end sequencing was performed on a HiSeq 2500 sequencer (Illumina). RNA-seq data were mapped to human reference genome hg19 using StrongArm and fusions were identified using CICERO[53] and FusionCatcher[54,55].

B-ALL Subtyping Based on Gene Expression Profile from RNA-seq

Read counts for annotated genes (Ensembl Homo sapiens GRCh37 v75) were called by HTSeq[56] (version 0.6.0) and processed by DESeq2 R package[57] to normalize gene expression into regularized log2 values (rlog). A subtype predication model was trained by Prediction Analysis of Microarrays based on a cohort of 309 samples from our previous studies[16,53,58], which consists of 8 B-ALL subtypes: *IGH-DUX4* (*n* = 42), *TCF3-PBX1* (*n* = 41), *ETV6-RUNX1* (*n* = 42), hyperdiploid (*n* = 46), *MEF2D*-rearranged (*n* = 21), *KMT2A* (*MLL*)-rearranged (*n* = 44), *BCR-ABL1* (*n* = 44), and *ZNF384*-rearranged (*n* = 27).

The trained model was applied with 100 evenly divided thresholds (control selected feature genes from 5,000 to 50) and the probability score was averaged to predict subtypes for the enrolled RNA-seq samples.

Mutation Analysis and Clonal Modeling

Variants with a total coverage of <5 reads (combining WGS/WES and validation) were excluded. Variants with ≤2 variant reads were considered wild type; those with 3 to 8 variant reads only considered mutant when both WES/WGS and capture validation techniques identified the variant or the variant was called with higher coverage in other (tumor) samples of the same patient. A MAF of <30% was considered subclonal, 30% to 75% heterozygous, and MAF ≥ 75% homozygous.

MutSigCV[59] and the Genomic Random Interval Model[17] (GRIN) analyses were performed to identify potential driver lesions (**Tables S5.11** and **S5.12**). MutSigCV is limited to analysis of sequence mutations, whereas GRIN incorporates multimodal genomic data including sequence and structural variants with robust adjustment for background mutation rate to identify significantly altered genes/regions, and unlike CNA-specific tools such as GISTIC[60], is not influenced by full chromosomal aneuploidies.

Two-dimensional MAF plots were used to visualize the relationship between sequential samples[4]. We used sciClone[14] and manual curation incorporating xenograft data to assign variants to clones. Clonal evolution was visualized using ClonEvol[61] and fishplot[13]. Noncoding variants were also considered to resolve the nature of clonal evolution in cases with presumed evolution from a major clone, which resulted in required reclassification from evolution from a major to a minor clone in 5 of 27 cases. Retention of multiple clones was deemed polyclonal evolution, and relapse from an ancestral precursor where D/R tumors share only the founding translocation and up to two SNV/indels.

Xenografting

Diagnosis and/or relapse tumors of 8 cases [3 *ETV6-RUNX1*, 1 *ETV6-RUNX1*-like, 2 *KMT2A* (*MLL*)-rearranged, 1 *DUX4* and 1 case without subtype data; **Table S5.1**] were transplanted at limiting dilution from 250,000 cells to 10 cells into the femur of 8 to 12 week-old sublethally irradiated (225 cGy) female NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG) mice. Engrafted tumor cells were harvested from bone marrow, spleen, and the central nervous system when mice displayed evidence of disease or at 30 weeks posttransplantation. Cells from the bone marrow and spleen were purified using the

Miltenyi Mouse Cell Depletion Kit (Miltenyi Biotec; samples with >20% engraftment) or by cell sorting. All animal experiments were done in accordance with institutional guidelines approved by the University Health Network (Toronto, Canada) Animal Resource Centre (AUP#1117.37).

ddPCR

Seven relapse-associated mutations were genotyped by ddPCR technique (RainDance Technologies) using custom (*NT5C2* p.P414A, *CREBBP* p.R1446C, *NRAS* p.G12R, and *WHCS1* p.E1099K) or available (*NT5C2* p.R39Q, *KRAS* p.G12S, and *KRAS* p.G12C) primers and probes (**Table S5.30**). An average of 7 million droplets were generated by the RainDrop Source instrument, and emulsion PCR was performed using the C1000 Thermal Cycler (Bio-Rad, Hercules, CA, USA). Droplet fluorescence of the amplified product was detected by the RainDrop Sense instrument, and data analysis was carried out using the RainDrop Analyst II Software. Detection limits were determined by testing serial dilutions of flow purified-mutant leukemia cells in wild-type REH cells. A frequency of >0.005% (>7 copies) could be consistently detected (**Figure S5.9** and **Table S5.24**). ddPCR MAF correlated well with the MAFs called from WES (r = 0.971) or CapVal (r = 0.9964; **Figure S5.9** and **Table S5.23**).

IGH and TCR Rearrangement Sequencing

IGH and *TCRB* loci were genotyped by ImmunoSeq (Adaptive Biotechnologies) to analyze clonal relationships of putative second tumors lacking shared genomic alteration (34 samples from 15 cases; **Table S5.19**).

Germline Analyses

Germline copy number variants were filtered by the Database of Genomic Variants[62] and in-house databases. To prioritize germline SNV/indel variants, we filtered for rare variants (<0.01% population frequency in ExAC, dbSNP, GoNL, ESP, Wellderly, Kaviar, and Complete Genomics' 60 genomes databases) that were predicted to be deleterious (nonsense, frameshift, or canonical splice site variants, and missense variants with PhyloP > 3) and were present in genes known to be associated with leukemia susceptibility and pediatric cancer syndromes (**Tables S5.31** and **S5.32**). Second, we used St. Jude's Medal Ceremony algorithm to identify Gold Medal variants[63] (truncating mutations in tumor suppressors, matches to hotspots or truncating mutations in somatic mutation databases, and matches to locus-specific databases).

Mutational Signature Analysis

We defined hypermutation as samples containing >85 SNV/indels (~1.3 mutations/Mb) based on the density histogram of the number of variants per sample in our cohort (**Figure S5.4**), which is more conservative than the cutoff determined by Segmented[25] (>56 SNV/indels, 0.8 mutations/Mb).

For *de novo* extraction of somatic mutational signatures, we selected 22 hypermutated ALL samples from this cohort, and WES and WGS from patients with hypermutated B-ALL (*n* = 38) and WGS from B-cell lymphoma samples (*n* = 10, ICGC). Mutational signatures were extracted using NMF[64,65] and MutationalPatterns[66]. Similarity between two signatures, A and B, defined as non-negative vectors with *n* mutation types, was calculated using cosine similarity. We defined two signatures to be the same if the cosine similarity is ≥0.95 (range 0-1). We calculated the cosine similarity between the mutational profiles of the samples and the mutational signatures and included the COSMIC mutational signatures and the *de novo* extracted mutational signatures in this analysis.

Finally, we studied the strand asymmetry of signature A mutations in the context of transcription and replication. This analysis requires that individual mutations are assigned to a single signature. Using the number of trinucleotide changes and the signature probabilities per sample, we calculated the relative contributions of each signature for each trinucleotide context in a sample. Only trinucleotide changes with a relative contribution of ≥95% to one of the four signatures were assigned to that particular signature.

Replication timing in leukemia samples was determined using Repli-seq data obtained from five lymphoblastoid cell lines in the ENCODE project[67] (GM06990, GM12801, GM12812, GM12813, GM12878), using median values per 1-kb bin[33]. For intestinal organoids, we used Repli-seq data described previously[68]. Predefined signature A mutations were assigned to early, intermediate, and late replicating regions, as described previously[33]. Replication asymmetry analysis was done using the MutationalPatterns R package[66]. On the basis of the distribution of rlog values from RNA-seq, genes were stratified into three groups for each sample: not expressed (genes with no or few supporting reads), genes with low expression (below the median level), and genes with high expression (median expression level or higher). We confirmed presence of SBS1 mutations by comparing their mutational profile with signature SBS1 for each sample (cosine similarity = 0.98; **Figure S5.10**). Gene expression stratification for healthy colon organoids was performed as described for the ALL samples. Testing of strand asymmetries within three groups was done using Poisson test for strand asymmetry significance testing.

Neoepitope Analyses

To characterize the antigenic potential of missense variants, we developed putative antigenicity scores that consider predicted patient-specific peptide: MHC binding variant-specific expression. WGS and RNA-seq data were used to infer class I HLA alleles to four-digit resolution for each patient using OptiType[35]. For each missense SNV and patient HLA allele, we then used NetMHCcons 1.1[36] to predict the binding affinities of all unique, mutated peptides of lengths 8-12 amino acids, excluding peptides that could be found elsewhere in the human proteome. Predicted binding affinities are often categorized as presumptive binders ($\leq$500 nmol/L) and nonbinders, which can be useful for narrowing epitope targets[44], but to estimate total antigenic potential, we also conceptualized binding affinities as correlated with probabilities of peptide: MHC binding to consider all predicted binding affinities additively. The UPAS was calculated as the natural logarithm of the summation of the inverse of all putative binding affinities; for the subset of SNVs for which expression data were available, this value was then weighted by adding the natural logarithm of 0.01 + the fraction of expressed alternate-to-total bases to generate a WPAS specifically for those variants. Each of these scores is best considered in comparison across variants, with increasingly positive scores indicative of increasing putative antigenicity.

To investigate potential correlates of antigenic potential (e.g., disease, disease progression, hypermutator status, and interactions thereof) while controlling for the nonindependence of the data owed to multiple variants across patients, we used the lme4 R package[69] to fit linear mixed models with patient as a random effect. The car R package[70] was used to assess the statistical significance of the fixed effects, and residuals of models with significant effects were verified as unbiased and homoscedastic via visual inspection.

SUPPLEMENTARY DATA

Supplementary Note

Rise in mutation burden with disease progression

All somatic sequence variants detected in the exomes of diagnosis and relapse samples (*n* = 16,615) were subjected to confirmatory capture-based sequencing to provide a more accurate estimate of mutant allele frequency (MAF) and facilitate detection of subclonal variants. Overall, there was excellent concordance in mutation calling between discovery and verification sequencing (89.7% of mutations), with discordance in defining subclonal (MAF <30%), heterozygous (MAF, 30%~75%) and homozygous (MAF >75%) status in 2% of the mutations (**Figures S5.1A**-**S5.1C**). Importantly, 3.4% of the mutations initially considered relapse-specific from analysis of WES/WGS data were shown to be present at diagnosis, including *NOTCH1*, *NRAS*, *KRAS*, *TP53*, and *CREBBP* (**Figure S5.1D**). In line, our deep sequencing efforts revealed tumor specific mutations in the complete remission samples (**Table S5.3**). A total of 936 tumor specific mutations (6.7% of the mutations) were detected in the complete remission samples of 72 patients (median of 4 mutations per sample, range 1-195 mutations). The median MAF of tumor specific mutations in the complete remission samples was 1.43% (range 0.1% - 29.8%), which increased to 43.2% (range 1.2% - 99.5%) in the corresponding tumor samples. The median increase in MAF from complete remission to tumor sample was 95.4% (range 19.5% - 99.9%). This indicates various degrees of discernable tumor contamination in the complete remission samples. To test whether the residual of somatic mutations in complete remission is an indicator of relapse, we identified 12 diagnosis-relapse cases (drawn from this study) and 20 ALL cases that did not relapse with remission samples obtained between 27-49 days (median 44 days) after diagnosis. High depth exome sequencing was performed for all germline samples (depth 377-515X, median 423X) and the somatic mutations identified in tumor samples were examined in the remission samples (**Table S5.2**). In contrast to prior data in AML, there was no association between the presence of residual somatic mutations in the germline samples and the risk of relapse (5 out of 12 relapses ALL cases have residual somatic mutations in remission samples compared to 12 out of 20 non-relapse ALL cases; Fisher exact test, *P* = 0.467).

The most frequent mutation types identified in the cohort were missense mutations (coding, silent or located in the UTRs) and single copy number deletions (**Figure S5.2**). The burden of mutations increased with disease progress from a median of 23 (range 1-164) single nucleotide variants (SNVs) and short insertions/deletions (indels) at diagnosis to 41 (range 4-1,139) at relapse and 117 (range 26-1,699) at second relapse

(**Figure S5.3**). This increase was observed for coding and noncoding variants, and for clonal and subclonal variants, suggesting ongoing mutational processes. No differences in mutation burden were observed according to disease lineage, sex, outcome (cure or death due to disease), or by age at diagnosis (with the exception of cases of infant/*KMT2A*-rearranged leukemia that have a low mutation burden; **Figure S5.3** and **Tables S5.3**-**S5.5**).

A modest increase in the number of structural genetic alterations was also observed, with a median of 9, 10 and 23 DNA copy number alterations (CNAs) at diagnosis, first and second relapse respectively (**Figure S5.3**, and **Tables S5.6**-**S5.7)**. In exceptional cases, mutations compromising DNA repair were associated with genomic instability. For example, case SJBALL022431 had somatic bi-allelic alterations of *ATM* from deletion of 11q and a nonsense mutation, and over 100 small losses and gains involving chromosomes 11, 15, and 21 reminiscent of a chromothripsis event (D: n = 109 affecting 368.5 Mb and R: n = 113 affecting 467.8 Mb; **Table S5.7**). This is in contrast to other cases with relatively high CNA burden but without DNA repair defects (e.g., SJETV043 and SJTALL023 with over 50 CNAs per sample), which contained predominantly small deletions but no amplifications (**Figure S5.2B**).

Early relapses

Next to *CREBBP* mutations, we found that mutations in *ERG* (n = 3) and *ARID2* (n = 4) associate with relatively late relapses (mean 7.0 vs 2.9 years, P = 0.011; and mean 6.6 vs 2.9 years, P = 0.013, respectively). Mutations in *ERG* and *ARID2* occurred exclusively in leukemias of the *DUX4* subtype. Several genes were correlated to early relapse, but the number of mutations was low (**Table S5.9**). We identified 8 cases who relapsed disproportionately early (>1.5 inter-quartile range) within their ALL subtype. Aberrations that were present exclusively in this group were i(9)(q10) with homozygous *PAX5* p.P80R mutation (SJBALL192) and *NR3C1* frameshift mutations (SJETV010 and SJBALL199). Other relapse-associated aberrations in this group were (combinations of) hypermutation (SJBALL022428 and SJETV010), SNV/indels in *NT5C2* (SJBALL192 and SJHYPER098), SNV/indels in *TP53* (SJBALL022428 and SJETV010) and SNV/indels and CNAs in *IKZF1* [SJBALL0212924 (p.G158S and del7-8), SJHYPER098 (del1-4), SJPHALL020 (del1-8), and SJTALL093 (p.E76*)], CNAs in *TBL1XR1* (SJETV010), CNAs in *ETV6* (SJBALL199 and SJETV010) and homozygous loss of *CDKN2A/B* (SJBALL192, SJETV010, SJHYPER098, and SJTALL093). These data indicate that time to relapse is influenced by the combined action of unique aberrations within ALL subtypes.

Frequently mutated genes and pathways

A total of 4,509 genes harbored nonsilent SNV/indels mutations, of which 1,063 were mutated in at least two cases (**Table S5.3**). Two-hundred and twenty-one were known targets of mutation in cancer according to the COSMIC Cancer Gene Census databse[71] or leukemia studies (**Figure S5.7A** and **Table S5.16**). Among the 293 genes recurrently (n ≥ 2) and exclusively targeted in the major clones at relapse (i.e., not targeted at diagnosis) by acquired SNV/indels or CNAs (**Table S5.16**), 82 were targeted by nonsilent variants, of which 6 were CGC genes (reported in the COSMIC Cancer Gene Census or in recent leukemia studies; *NCOR2, NT5C2, FOXA1, MED12, PABC1,* and *PPMID*)[72]. Twelve genes harbored recurrent clonal and acquired nonsilent mutations uniquely at second relapse, including *AASDH*, *AGBL3*, *ARGHAP29*, *ATP11C*, *CSF3R*, *DCK*, *LRP6*, *SCRIB*, *SDK2*, *TENM3*, *TNIP1*, and *ZNF436* (**Table S5.17**), suggesting a possible role in treatment resistance or evasion.

In general, B- and T-lineage ALL show distinct genetic alteration profiles, e.g., genes involved in B cell development are enriched in B-ALL (Fisher exact test, $P = 0.002$), while mutations in transcription factor *NOTCH1* involved in maturation of both CD4+ and CD8+ cells in the thymus were exclusively observed in T-ALL (**Figure S5.6**). Fifteen genes were associated with disease lineage, e.g., *VPREB1* and *CREBBP* with B-ALL, and *NOTCH1*, *WT1*, *PTEN*, *STIL*, and *NT5C2* with T-ALL (Fisher exact test $P < 0.05$). The most common alteration perturbing B cell development involved the lymphoid transcription factor gene *IKZF1*, which is commonly mutated in *BCR-ABL1*[73], Ph-like[39] and *DUX4*-rearranged ALL[16]. Consistent with the known association of *IKZF1* alterations with high-risk B-ALL[39,73] *IKZF1* alterations were strongly enriched at relapse, with preservation (n = 21) or acquisition (n = 5) in 26 cases (**Table S5.15**). Subtype-associated gene mutations were also observed such as *CREBBP* in hyperdiploid (9 of 14) and low hypodiploid (2 of 3) ALL; *ARID2* mutations in *DUX4*-rearranged (4 of 7) and *ARID5B* in hyperdiploid (4 of 14) ALL.

Overall, 76% B-ALL and 80% T-ALL harbored at least one signaling pathway mutation (median 3, range 1-8). The most frequently were the RAS (B-ALL 47.8%, T-ALL 36%), JAK-STAT (B-ALL 19.4%, T-ALL 28%) and PI3K-AKT (B-ALL 13.4%, T-ALL 28%; **Figures 5.1** and **S5.5A**). RAS signaling pathway mutations were observed in 41 cases, and 26 (63.4%) of them were preserved from diagnosis to relapse, with 24.4% (n = 10) acquired and 12.2% (n = 5) lost in disease progression. The data also provide new insight by showing common multiclonality of RAS pathway mutations at diagnosis and frequent convergence to a dominant mutation at relapse. Of 61 cases with signaling pathway mutations, 31 harbored at least one RAS pathway mutation at diagnosis, with

11 cases having multiple, commonly subclonal RAS mutations (**Table S5.14**). Seven cases showed convergence to a single or two clonal RAS pathway mutations. For the JAK-STAT pathway, similar enrichment in preserved mutations was observed (*n* = 10, 50.0% preserved; *n* = 5, 25.0% acquired; and *n* = 5, 25.0% lost). However, most of the PI3K-AKT pathway mutations in B-ALL were acquired in relapse (8 out of 10), whereas in T-ALL, most of the mutations identified in diagnosis were lost at relapse (11 out of 13; **Tables S5.13**-**S5.14**).

Genes encoding regulators of cell cycle and tumor suppression were commonly mutated in both B-ALL (*n* = 42, 62.7%) and T-ALL (*n* = 21, 84%), with gene-specific patterns of enrichment. Ten *TP53* alterations (8 cases) were preserved from diagnosis or acquired as new alterations at relapse. In contrast, *CDKN2A* mutations were lost (*n* = 6), preserved (*n* = 34) or acquired (*n* = 15) in disease progression. Epigenetic modifier/regulator mutations were observed in 67.2% of B-ALL (*n* = 45) and 72% of T-ALL (*n* = 18) (**Figure S5.5B**), most of which were acquired (*n* = 85, 56.7%) or preserved (*n* = 48, 32.0%) from diagnosis to relapse. *CREBBP* mutations (22 mutations in 17 B-ALL and 1 T-ALL case) showed preservation or acquisition in all but one case.

Second primary leukemia

To verify clonal relatedness in a subset of patients, we examined *IGH* (11 B-ALL patients) or *TCRB* (3 T-ALL patients) gene rearrangements by targeted next generation sequencing, which revealed that the relapses were clonally related to major (*n* = 5) or minor (down to 0.005%; *n* = 7) clones at diagnosis (**Table S5.19**). Three cases (SJBALL006, SJTALL049 and SJTALL142) revealed fully discordant tumors for all genetic alterations (SV, CNA, SNV, Indel and IgH/TCRβ rearrangements), raising the possibility of a distinct second primary leukemia. SJBALL006 is described in the main text. This patient exhibited intellectual impairment, short stature, and thrombocytopenia. SJTALL049 is a boy and described in the main text. He suffered a grand mal seizure at age 4 and has a mild nystagmus. Analysis of the germline WES data revealed two indels on one allele in *TSC2,* leading to an in-frame deletion of 3 amino acids. Mutations in *TSC2* are associated with tuberous sclerosis (OMIM #613254) and can cause brain and kidney tumors, but leukemia has not been reported. SJTALL142 developed T-ALL at 15 years of age and *KMT2A-MLLT10* rearranged MDS at 18. Additional WGS analysis revealed no shared SV or SNVs (**Table S5.20**). The boy had neck surgery at 6 months of age due to cysts, and he has a history of renal insufficiency. The family history was inconclusive for cancer occurrence and the boy did not harbor any constitutional variants in a gene on our gene list.

Our cohort contained five patients who developed a myeloid tumor after T-ALL. Two of these tumors were identified as clonally unrelated second primary leukemias. However, three AMLs were clonally related relapses originating from a minor clone (SJTALL124: T-ALL; 20 mutations shared, e.g., *TP53* p.R273H and *CDKN2A* deletion), a major clone (SJTALL164: ETP-ALL; 20 mutations shared, e.g., *NF1* deletion), or polyclonal (SJTALL008: ETP-ALL; 20 mutations shared, e.g., *NRAS* p.Q61H), indicating that lymphoid and myeloid tumors can share a common origin. This is in line with lineage plasticity and multipotent common ancestor cells identified in mixed phenotype acute leukemia[20] and histiocytic tumors[74].

Constitutional variants and ALL susceptibility

We analyzed our complete cohort for constitutional variants in genes known to be associated with leukemia prone syndromes or with leukemia development (**Tables S5.31** and **S5.32**). We found two truncating variants in *TP53* (p.Y163Tfs*7 and p.R196*), which are causative for Li Fraumeni syndrome in patients SJHYPO126 and SJTALL164, respectively. The tumors of both patients showed loss of the wild type allele through somatic deletion of the short arm of chromosome 17. SJHYPO126 developed hypodiploid leukemia and has an extensive family history for cancer (breast cancer, liposarcoma, lung cancer, colon cancer and leukemia). SJTALL164 developed ETP-ALL with a clonally related but phenotypically myeloid relapse. In addition to the truncating *TP53* variant, SJTALL164 harbored constitutional truncating variants in *NOTCH1* (p.Q2440*) and in *DNM2* (p.T78Nfs*18), both genes that are involved in leukemia development[75,76]. Clinical data was not available for this patient.

In patient SJBALL022480 we found a constitutional truncating mutation in *PRDM1* (p.R25Sfs*13), which is a DNA-binding transcriptional repressor involved in B-cell differentiation. The probability of intolerance for this gene is very high (pLI = 0.98), suggesting a severe effect of functional loss. Finally, SJBALL013 harbored a truncating mutation (p.I1085Hfs*60) in *PTCH1,* which is associated with basal cell nevus syndrome or Gorlin syndrome (OMIM #109400). The patient presented with low set ears, macrocephaly and intellectual disability. In addition, the boy developed a basal cell carcinoma at age 5, concurring the diagnosis of Gorlin syndrome.

We identified 13 heterozygous variants (13 patients) in genes associated with autosomal recessively inherited syndromes (e.g., Fanconi Anemia, CMMRD). Likely, these latter variants are not directly causative for the leukemia in these patients. None of the variants, except for *PMS* (p.G108R) in SJHYPER127, obtained a second hit in the tumors of the patients. We did not identify homozygous or compound heterozygous variants (**Table S5.32**). In SJHYPER021925, we found a missense variant in *PHF6* (p.R163H), a gene associated to the X-linked recessive Borjeson-Forssman-Lehman syndrome (OMIM #301900) and somatically affected in ALL[77,78]. The patient was female with no phenotypical features of Borjeson-Forssman-Lehman syndrome reported. A large cohort study is needed to reveal whether heterozygous *PHF6* variants in females lead to a susceptibility to leukemia development.

We identified 12 variants of unknown significance (in 11 patients) in genes associated with autosomal dominantly inherited cancer prone disorders. Three variants were found in genes associated with Noonan syndrome (OMIM #616564, #610733, and #616559) or Noonan-like syndrome (OMIM #613563). *LZTR1* p.G711R (SJHYPO126) is located in the BTB/POZ domain; *SOS1* p.N622S (SJHYPER009) is located in the RasGEFN domain, and *SOS2* p.E422K (SJTALL098) and *CBL* p.P684S (SJERG023) were not located in a defined domain. None of the genes harbored second hits in the tumors of the patients. The phenotype of SJHYPO126 and SJHYPER009 did not resemble Noonan syndrome. SJHYPER009 was diagnosed with Duane syndrome (OMIM #126800) and developmental delay with speech deficits. SJERG023 presented with skin moles on the back, but no Noonan-like syndrome phenotype. Phenotypic data was not available for SJTALL098.

We included *SAMD9* and *SAMD9L* in our gene list because of recent reports associating missense mutations in these genes to MIRAGE syndrome (OMIM #617053)[79], myelodysplasia and leukemia syndrome with monosomy 7 (*MLSM7*; OMIM #252270)[80] and ataxia-pancytopenia syndrome (*ATXPC*; OMIM #159550)[81]. We identified variants in *SAMD9* in the germlines of SJBALL022479 (p.I1549Sfs*10) and SJPHALL029 (p.C283Y). The frameshift variant is not likely to be associated with leukemia development in this patient as *SAMD9* is completely tolerant to loss of function mutations (ExAC pLI = 0). For both these patients' phenotypic data was not available.

Finally, we found variants of unknown significance in the genes *PAX5* (associated with leukemia susceptibility, OMIM #615545), *TYK2* (associated with leukemia susceptibility) [82], *NF1* (associated with neurofibromatosis, OMIM #16220), and *CDH1* (associated with familial gastric cancer, OMIM #137215). We identified a 7 amino acid in-frame deletion p.P51del7aa in *PAX5* in SJHYPER021999. The variant p.G183S in the octapeptide domain of *PAX5* was associated with susceptibility to ALL development in three families[83,84], but further studies should reveal whether a deletion of 7 amino acids in the paired box domain of this gene can drive leukemia development. No second hits in this gene were found in the tumor of the patient. We identified a p.P120L variant in *TYK2* in patient SJTALL055. Variants in the DPG motif of *TYK2* have been associated with leukemia susceptibility[82], but no data is available on whether *TYK2* is functionally affected by the p.P120L variant in the FERM domain. There was no second hit in the tumor, and phenotypical data for this patient was not available. The *NF1* variant p.K1385E is located in the RasGAP Neurofibromin domain and was found in SJTALL092. For this patient no phenotype was reported, but the maternal branch of the family did harbor some cases of cancer (lymphoma in a maternal uncle, breast cancer in a maternal aunt and colorectal cancer in the maternal grandfather). A second hit in this gene was not present in the tumors of this patient, which suggests that this variant is not causative for the leukemia in SJTALL092. SJPHALL018 carried a p.P373L variant in *CDH1*, a gene associated with familial gastric cancer[85], though no cancer was reported in the family. A link with leukemia predisposition has not yet been reported for *CDH1*.

Hypermutation and MMR mutations

Patients with hypermutated relapses with the MMR associated mutational profile (group 3) all harbored biallelic mutations in one of the MMR genes. These included a germline *PMS2* p.G108R mutation and a somatic deletion of *PMS2* with retention of the mutant allele in SJHYPER127-R1; somatic homozygous deletion of *MSH2* and *MSH6* in SJTALL023-R1; somatic deletion and homozygous p.R680* mutation of *MSH2* plus a hemizygous deletion of *MSH6* in SJTALL023-R2; and somatic homozygous mutation in *MSH2* (p.R340Nfs*17) in SJTALL057-R1. In group 1, SJETV10 harbored biallelic copy number deletions in *MLH1* in both relapses, while an acquired subclonal missense mutation in *MSH6* was identified in the relapse of SJHYPER022. SJHYPO177 had losses of all chromosomes where MMR genes are located (chromosomes 2, 3, and 7), but no additional mutations in these genes.

Supplementary Tables

Supplementary Tables are available online using the following link: https://doi.org/10.1158/0008-5472.BCD-19-0041.

Supplementary Figures

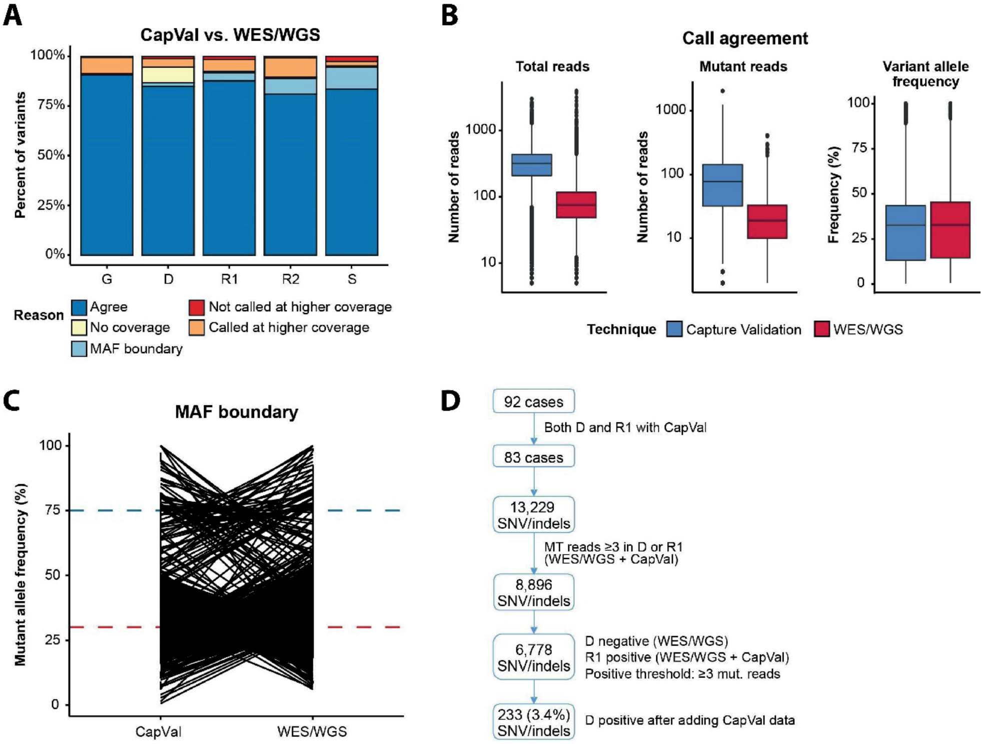


Figure S5.1. Capture Validation and WES/WGS calling are highly concordant. **(A)** For most of the variants (90%), capture validation (CapVal) and WES/WGS agree on the variant calling. D, diagnosis; R1, first relapse; R2, second relapse; S, second tumor. **(B)** Sequencing metrics between the two techniques for the agreed variants calls (dark blue in A). The total reads number and the mutant reads per variant are higher for the CapVal, but the mutant allele frequency (MAF) for both techniques is comparable for the variants. The box and whiskers plots depict the median (black line), 25th and 75th percentiles (box borders) and the highest/lowest value that is within 1.5*IQR (inter-quartile range; whiskers). Data plotted as individual points beyond the end of the whiskers are outliers. **(C)** For a total of 1,466 variants CapVal and WES/WGS techniques differed on the calling category based on MAF. Most somatic mutations were called with MAF around 30% (indicated in red dashed line), which was used as a cut off for subclonal vs heterozygous (clonal) state; a few hover around the hetero-homozygous cut off of 75% (indicated in blue dashed line). **(D)** Somatic mutations in diagnosis samples missed in WES/WGS and rescued by CapVal. The analysis was done in cases with both diagnosis and first relapse samples sequenced by CapVal. With a minimum number (≥3) of mutant (mut.) reads support for a mutation, 3.4% of the somatic mutations were recovered by CapVal in the diagnosis samples.

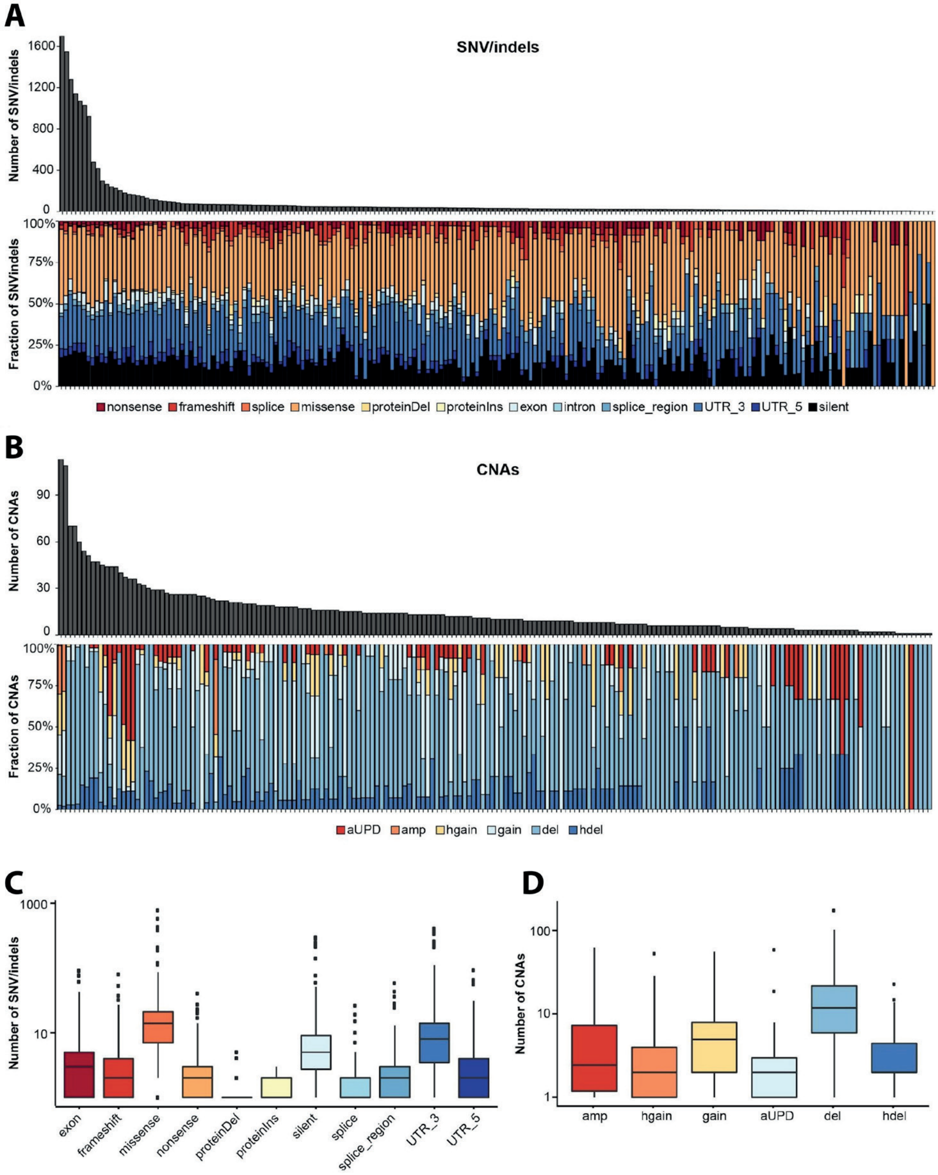
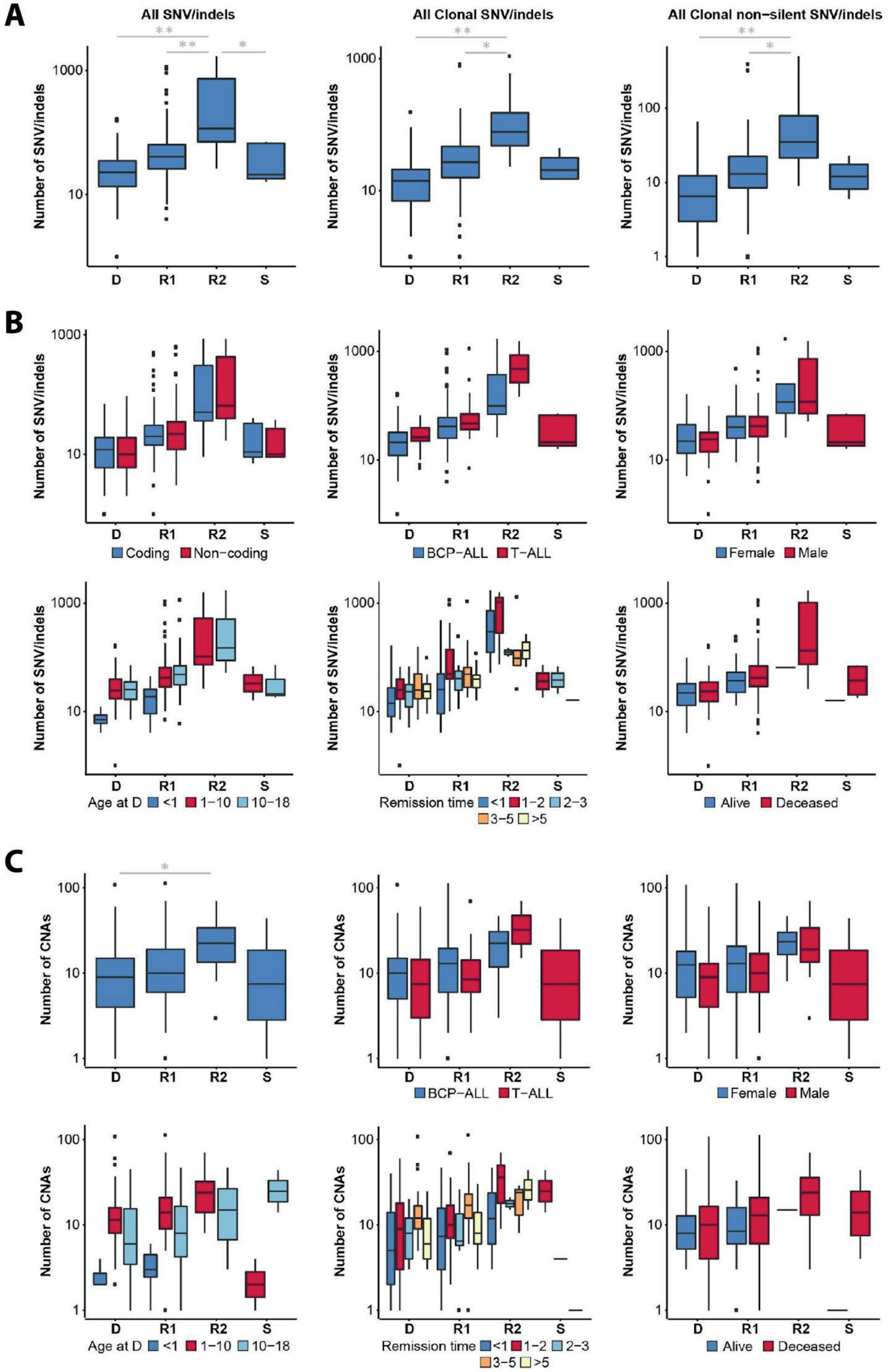
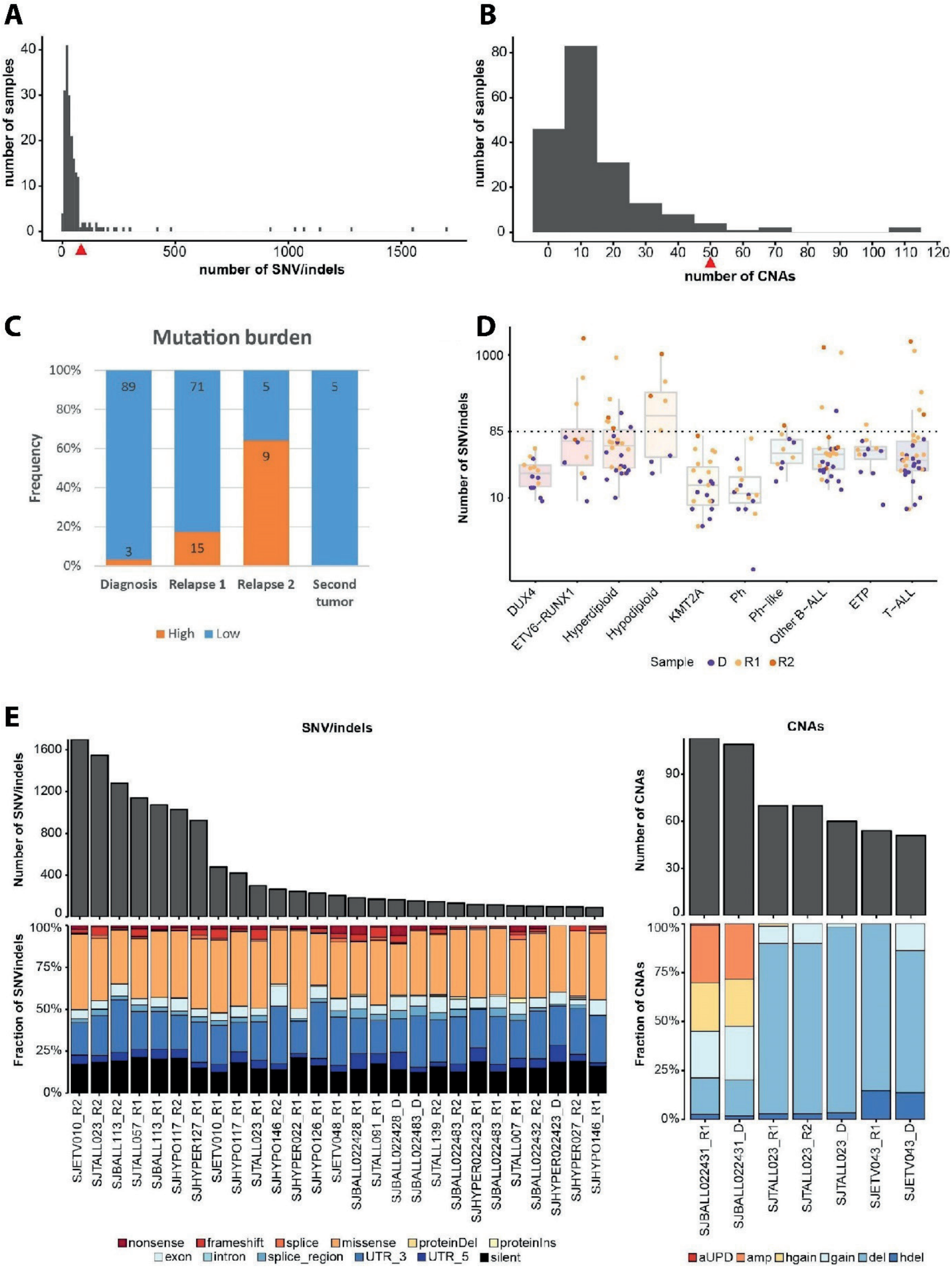


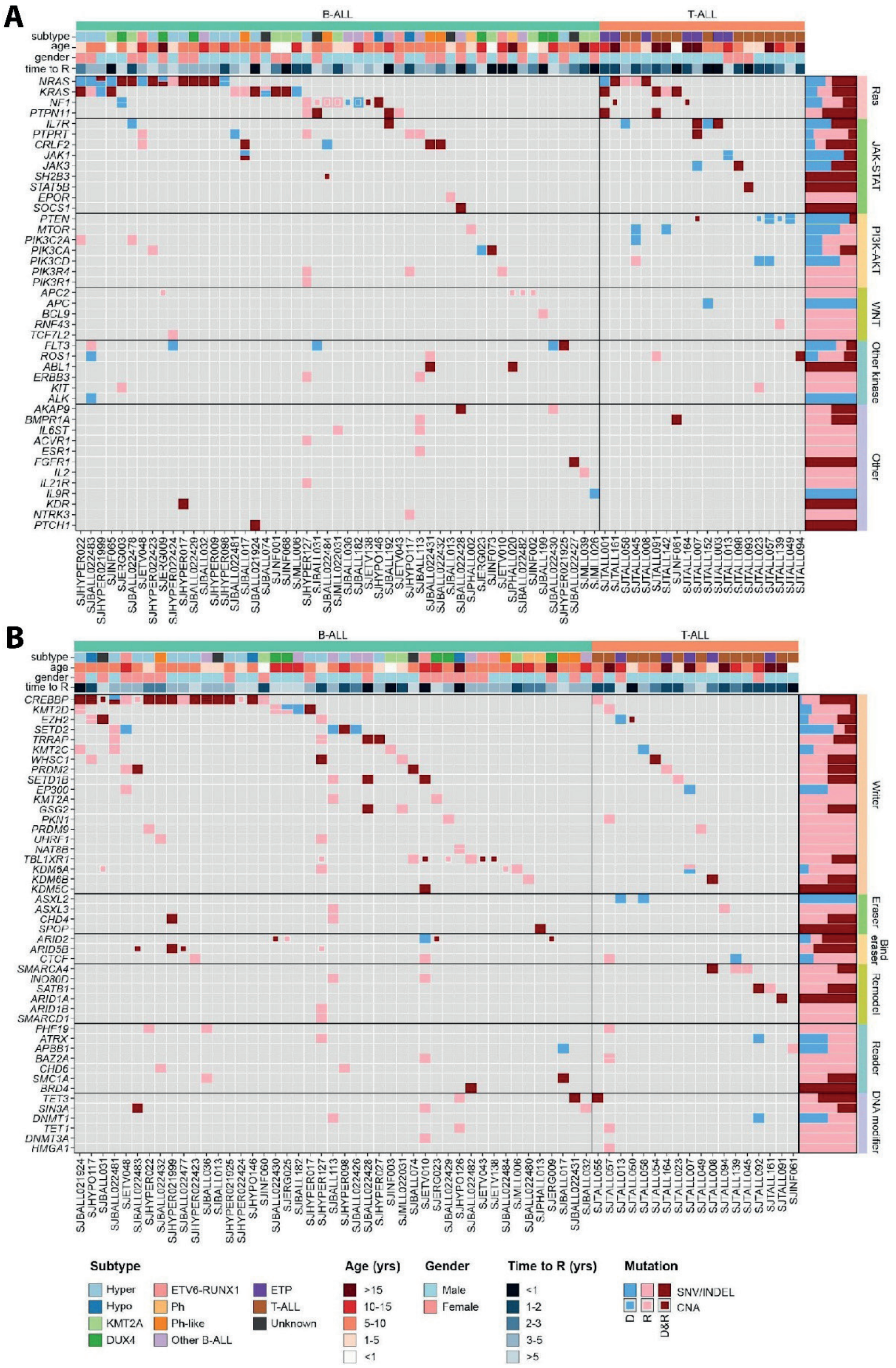
Figure S5.2. SNV/indel and CNA mutation burden. **(A)** Histogram of the number of SNV/indels (top) and the relative contribution of the mutation types (bottom) in each sample sorted from left to right by the total number of SNV/indels. The coding variants are indicated by shades of red, whereas the noncoding variants are indicated in shades of blue. **(B)** Histogram of the number of CNAs (top) and the relative contribution of the mutation types (bottom) in each sample sorted from left to right by the total number of CNAs. **(C)** Missense, silent and 3-prime UTR mutations are most frequent SNV/indels found in this study. **(D)** Deletions are the most common CNA in this study. Amp: amplification - gain of >2 copies, hgain: 2 copy gain, gain: 1 copy gain, aUPD: acquired uniparental disomy - copy number neutral loss of heterozygosity, del: 1 copy deletion, hdel: 2 copy deletion. The box and whiskers plots depict the median (black line), 25th and 75th percentiles (box borders) and the highest/lowest value that is within 1.5*IQR (inter-quartile range; whiskers). Data plotted as individual points beyond the end of the whiskers are outliers.


A
All SNV/indels
All Clonal SNV/indels
All Clonal non−silent SNV/indels
Number of SNV/indels
1000
100
10
1
D
R1
R2
S
**
*
B
Coding
Non−coding
BCP−ALL
T−ALL
Female
Male
Age at D
<1
1−10
10−18
Remission time
<1
1−2
2−3
3−5
>5
Alive
Deceased
C
Number of CNAs

◄**Figure S5.3.** Mutation burden increases with disease progression. **(A)** The number of all SNV/indels (left), of clonal (MAF>30%; middle), and clonal nonsilent SNV/indels (right) increase from diagnosis to the subsequent relapses. This effect remains when the outliers are removed (data not shown). ** $P < 0.0001$, * $P < 0.001$. **(B)** The SNV/indel mutation burden is not influenced by the type of mutation, tumor lineage, gender, age at diagnosis, remission time and vital status of the patient. **(C)** The number of CNAs increase in second relapse (* $P < 0.05$). This effect is not influenced by tumor lineage, gender, age at diagnosis, remission time and vital status of the patient. The box and whiskers plots depict the median (black line), 25th and 75th percentiles (box borders) and the highest/lowest value that is within 1.5*IQR (inter-quartile range; whiskers). Data plotted as individual points beyond the end of the whiskers are outliers. D, diagnosis; R1, first relapse; R2, second relapse; S, second tumor.


A
number of samples
number of SNV/indels
B
number of samples
number of CNAs
C
Mutation burden
Frequency
Diagnosis
Relapse 1
Relapse 2
Second tumor
High
Low
89
3
71
15
5
9
5
D
Number of SNV/indels
1000
85
10
DUX4
ETV6-RUNX1
Hyperdiploid
Hypodiploid
KMT2A
Ph
Ph-like
Other B-ALL
ETP
T-ALL
Sample
D
R1
R2
E
SNV/indels
Number of SNV/indels
Fraction of SNV/indels
CNAs
Number of CNAs
Fraction of CNAs
SJETV010_R2
SJTALL023_R2
SJBALL113_R2
SJTALL057_R1
SJBALL113_R1
SJHYPO117_R2
SJHYPER127_R1
SJETV010_R1
SJHYPO117_R1
SJTALL023_R1
SJHYPO146_R2
SJHYPER022_R1
SJHYPO126_R1
SJETV048_R1
SJBALL022428_R1
SJTALL091_R1
SJBALL022428_D
SJBALL022483_D
SJTALL139_R2
SJBALL022483_R2
SJHYPER022423_R1
SJBALL022483_R1
SJTALL007_R1
SJBALL022432_R2
SJHYPER022423_D
SJHYPER027_R2
SJHYPO146_R1
nonsense
frameshift
splice
missense
proteinDel
proteinIns
exon
intron
splice_region
UTR_3
UTR_5
silent
SJBALL022431_R1
SJBALL022431_D
SJTALL023_R1
SJTALL023_R2
SJTALL023_D
SJETV043_R1
SJETV043_D
aUPD
amp
hgain
gain
del
hdel

◂**Figure S5.4.** Cut-off determination for hypermutated samples. **(A)** Histogram of the number of SNV/indels in all samples. **(B)** Histogram of the number of CNAs in all samples. Red arrows indicate the cut-off number of SNV/indels (*n* = 85) and CNA (*n* = 50) used for defining hypermutators. **(C)** Hypermutator distribution in different disease stages. The ratio of hypermutated samples was relatively high in the second relapses. **(D)** The number of SNV/indels detected in each sample categorized by subtype of ALL at diagnosis and colored by disease stage, showing variable number of mutations across subtypes. The horizontal dotted line indicates the cut-off of >85 SNV/indels. Box and Whisker plots in the background indicate the median (thick line), 25th and 75th percentiles (box borders) and the highest/lowest value that is within 1.5*IQR (inter-quartile range; whiskers) number of SNV/indels. The ALL cases are grouped into well-defined disease subtypes, which include *DUX4*-rearranged (DUX4), *ETV6-RUNX1*, hyperdiploid, hypodiploid, *KMT2A* (*MLL*)-rearranged, *BCR-ABL1* (Ph), Ph-like, other B-ALL subtypes, early T-cell precursor ALL (ETP) and T-lineage ALL non-ETP (T-ALL). **(E)** Histogram of the number of SNV/indels (top) and the relative contribution of the mutation types (bottom) in each hypermutator (>85 SNV/indels) sorted from left to right by the total number of SNV/indels. The coding variants are indicated by shades of red, whereas the noncoding variants are indicated in shades of blue. **(F)** Histogram of the number of CNAs (top) and the relative contribution of the mutation types (bottom) in each CNA hypermutator (>50 CNAs) sorted from left to right by the total number of CNAs.


A
B-ALL
T-ALL
subtype
age
gender
time to R
NRAS
KRAS
NF1
PTPN11
IL7R
PTPRT
CRLF2
JAK1
JAK3
SH2B3
STAT5B
EPOR
SOCS1
PTEN
MTOR
PIK3C2A
PIK3CA
PIK3CD
PIK3R4
PIK3R1
APC2
APC
BCL9
RNF43
TCF7L2
FLT3
ROS1
ABL1
ERBB3
KIT
ALK
AKAP9
BMPR1A
IL6ST
ACVR1
ESR1
FGFR1
IL2
IL21R
IL9R
KDR
NTRK3
PTCH1
Ras
JAK-STAT
PI3K-AKT
WNT
Other kinase
Other
B
CREBBP
KMT2D
EZH2
SETD2
TRRAP
KMT2C
WHSC1
PRDM2
SETD1B
EP300
KMT2A
GSG2
PKN1
PRDM9
UHRF1
NAT8B
TBL1XR1
KDM6A
KDM6B
KDM5C
ASXL2
ASXL3
CHD4
SPOP
ARID2
ARID5B
CTCF
SMARCA4
INO80D
SATB1
ARID1A
ARID1B
SMARCD1
PHF19
ATRX
APBB1
BAZ2A
CHD6
SMC1A
BRD4
TET3
SIN3A
DNMT1
TET1
DNMT3A
HMGA1
Writer
Eraser
Bind eraser
Remodel
Reader
DNA modifier
Subtype
Hyper
Hypo
KMT2A
DUX4
ETV6-RUNX1
Ph
Ph-like
Other B-ALL
ETP
T-ALL
Unknown
Age (yrs)
>15
10-15
5-10
1-5
<1
Gender
Male
Female
Time to R (yrs)
<1
1-2
2-3
3-5
>5
Mutation
SNV/INDEL
CNA
D
R
D&R

◀**Figure S5.5.** Signaling pathway and epigenomic mutations in ALL at diagnosis and relapse. **(A)** Genes involved in signaling pathways with somatic mutations identified in diagnosis (D) and/or first available relapse (R) sample per case. **(B)** Epigenetic regulators with somatic mutations identified in diagnosis (D) and/or first available relapse (R) sample per case are grouped into 6 epigenetic categories[86]. Same mutation annotation schema is applied as **Figure 1A**.

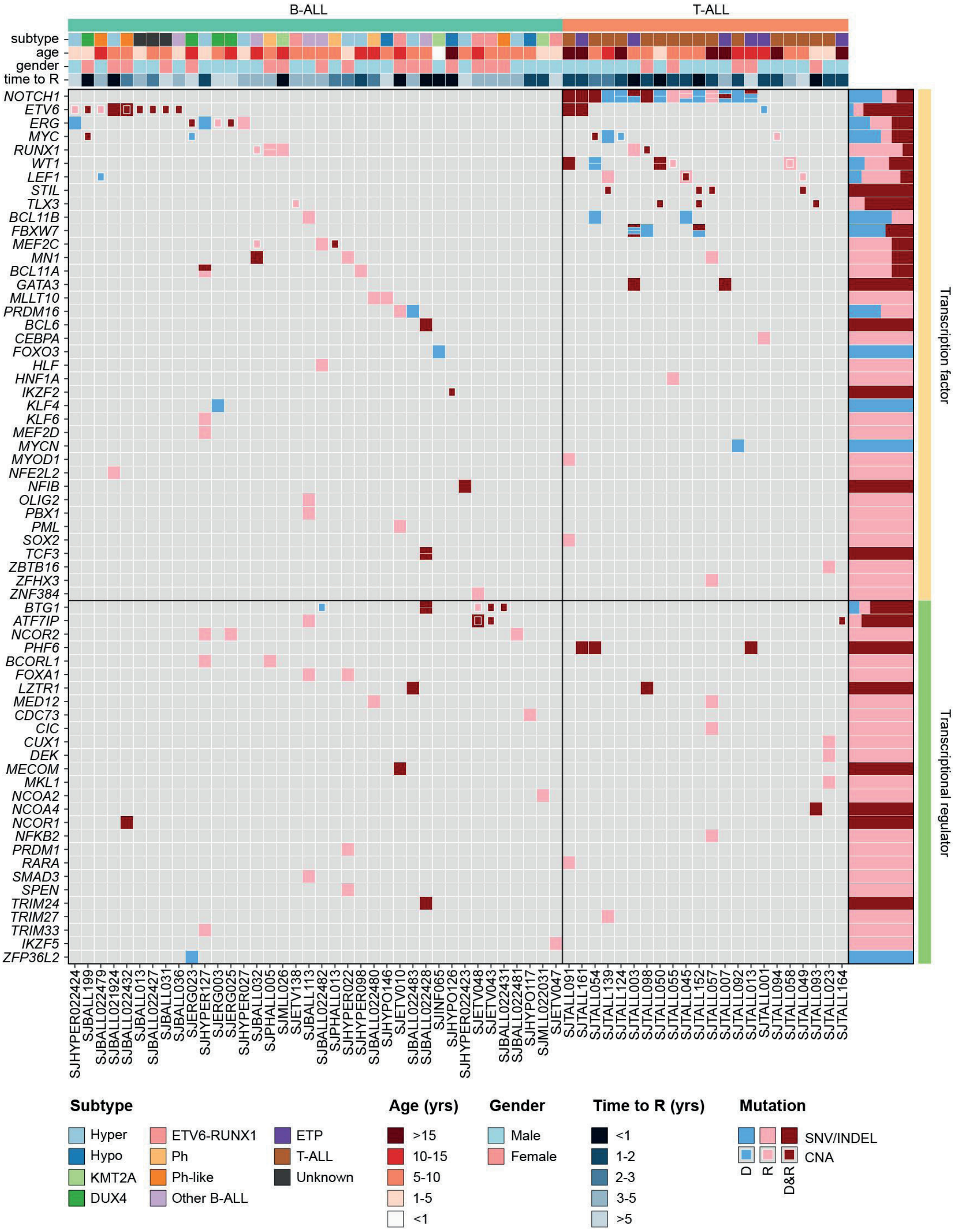


Figure S5.6. Transcription factor/regulator mutations in ALL at diagnosis and relapse. Genes with somatic mutations identified in diagnosis (D) and/or first available relapse (R) sample per case are grouped into transcription factors and transcriptional regulators. Same mutation annotation schema is applied as **Figure 1A**.

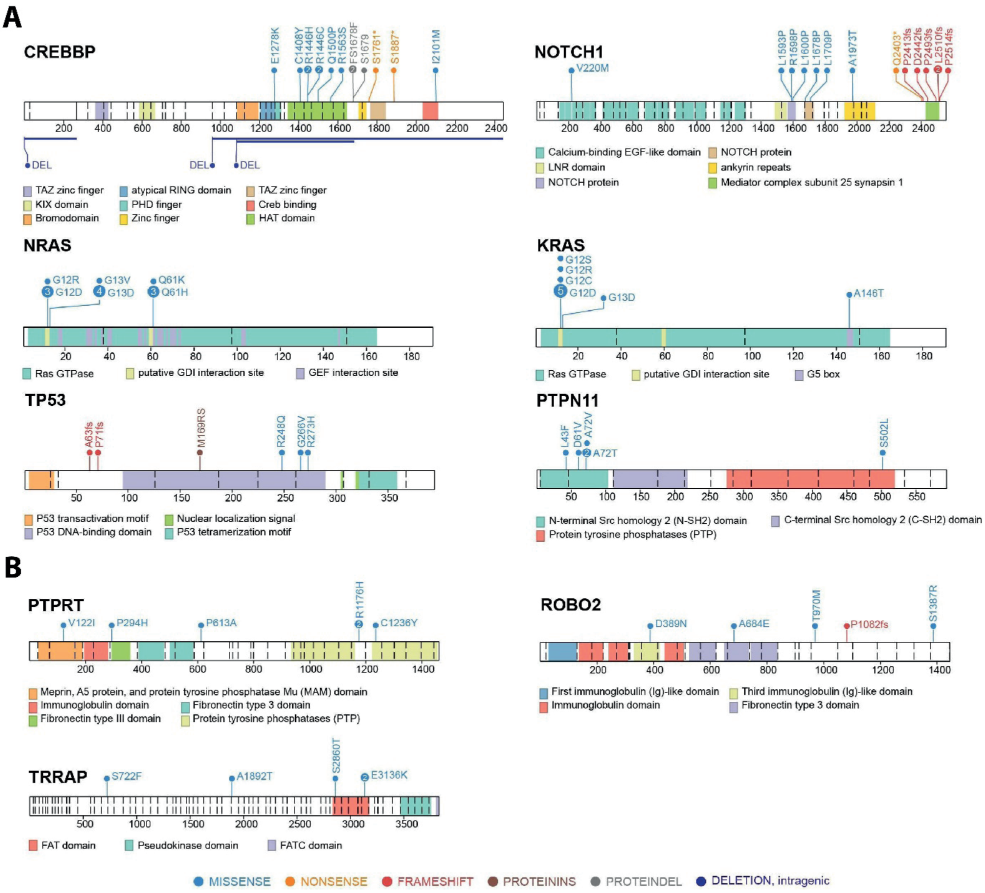
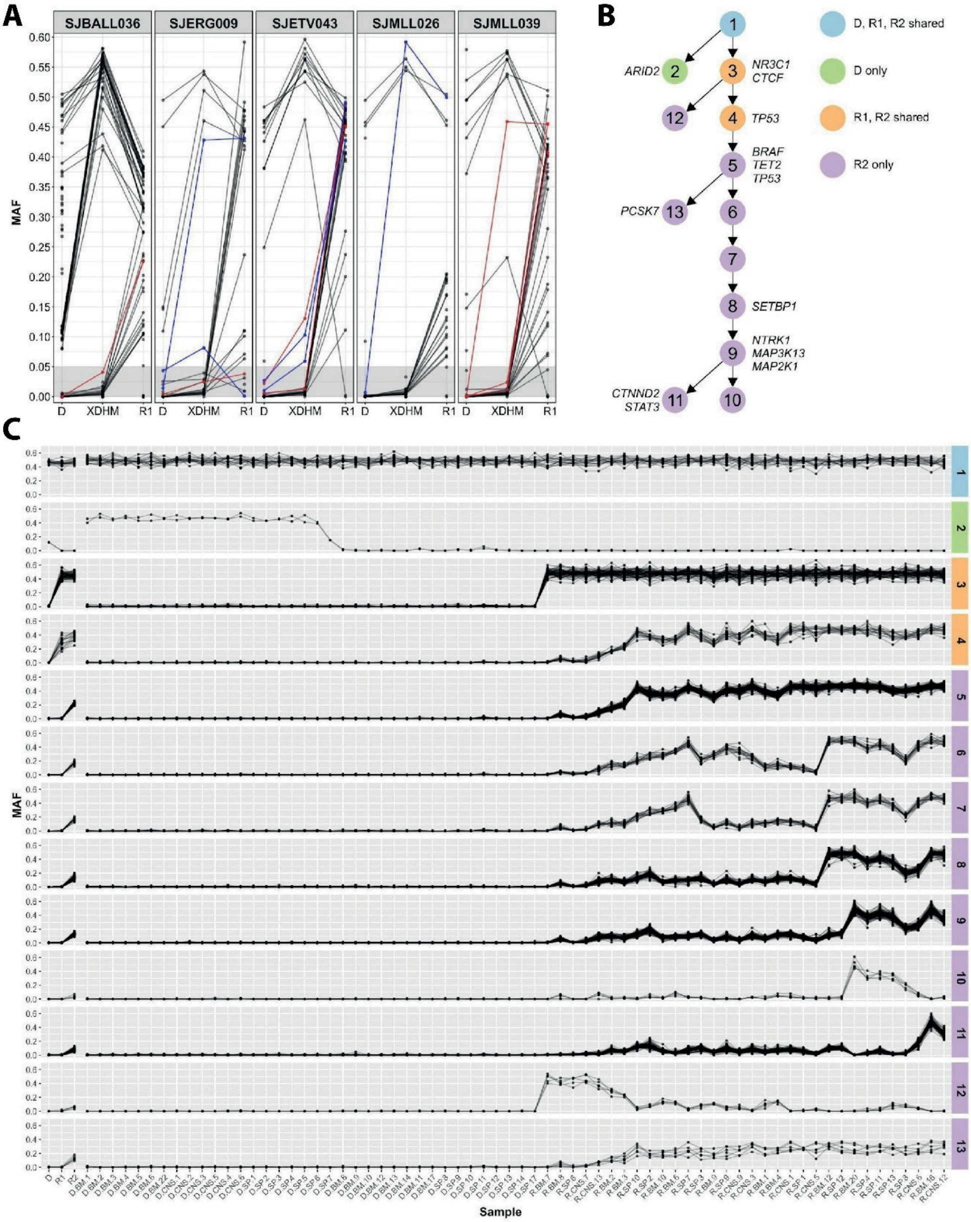
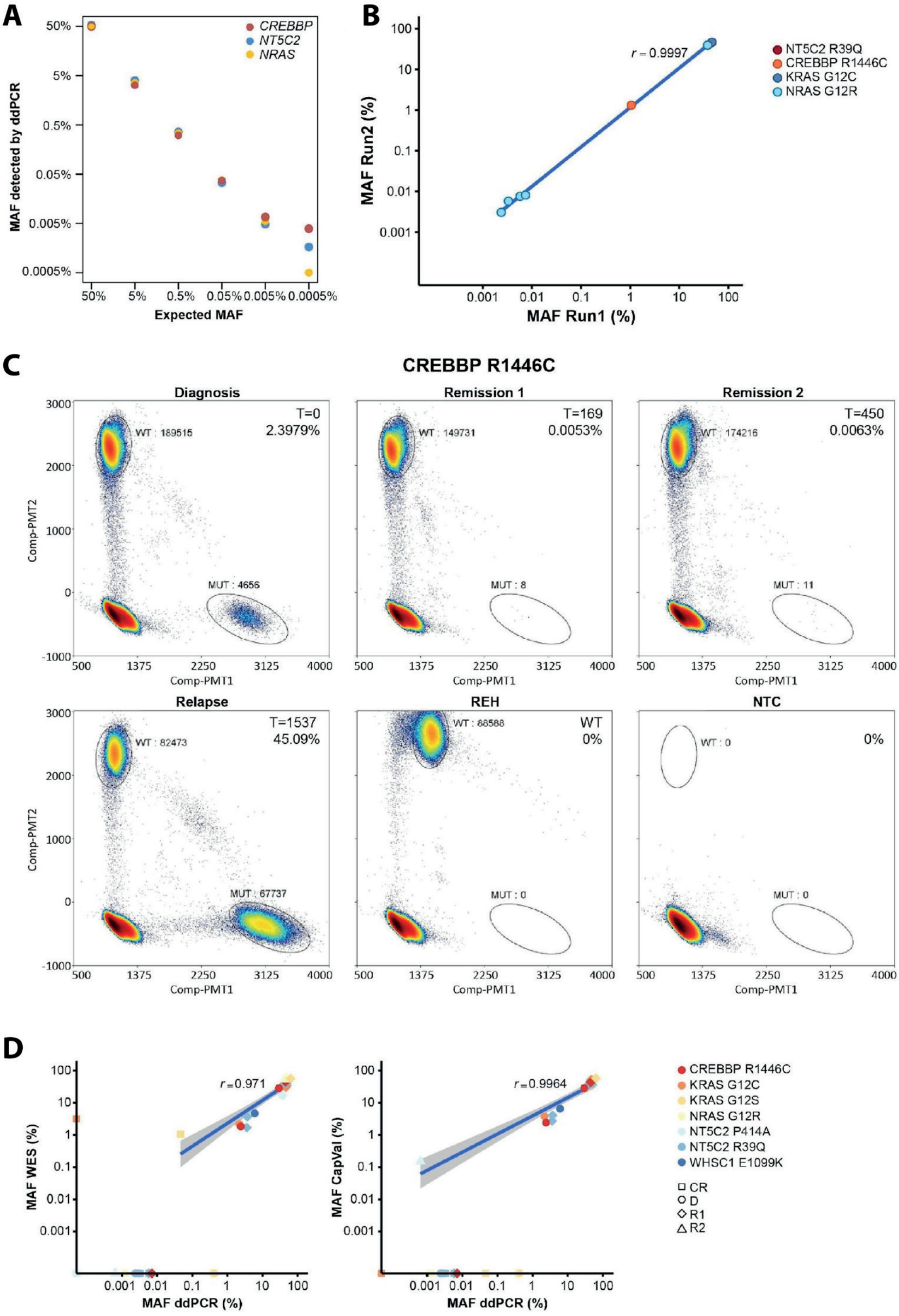


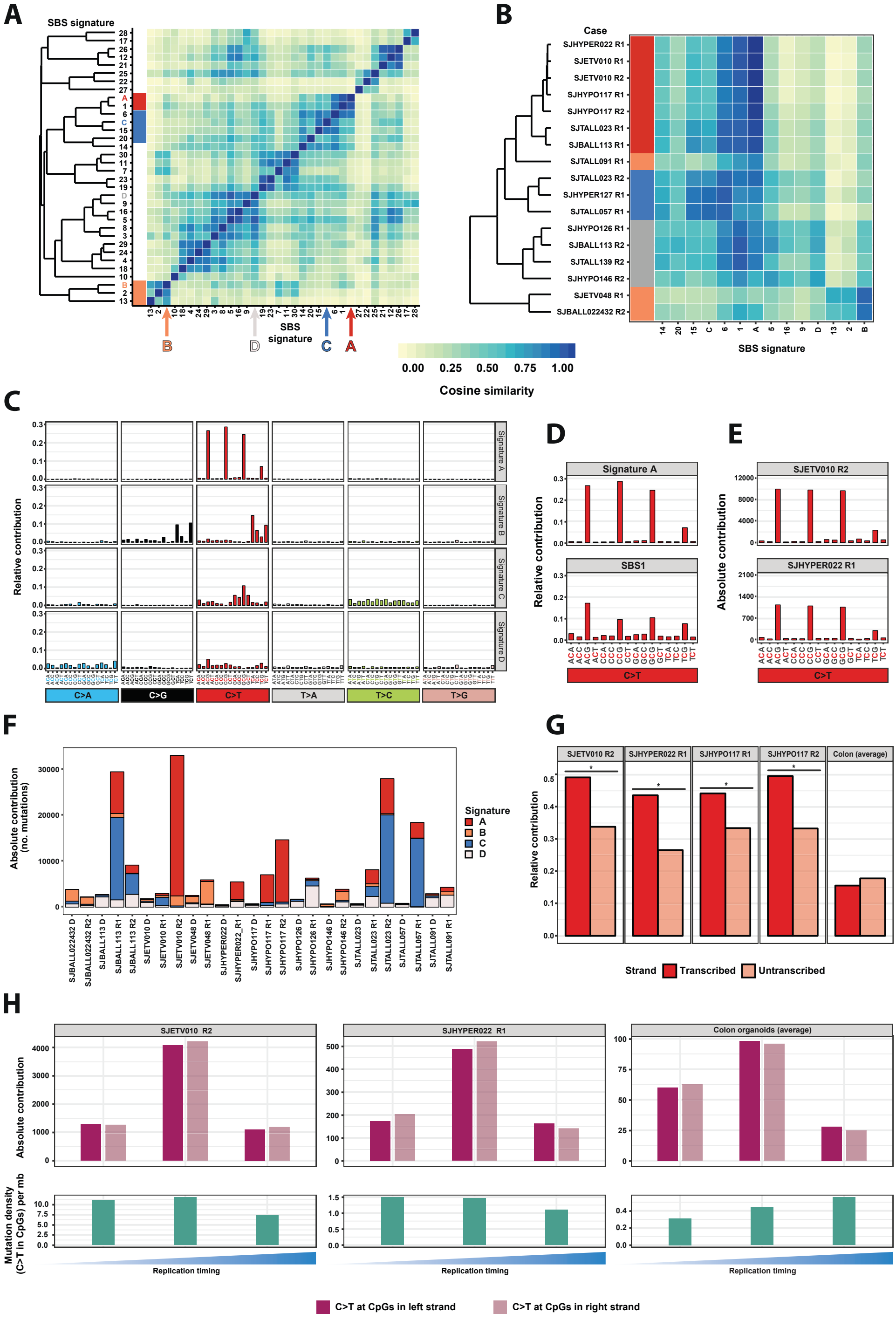
Figure S5.7. Frequently mutated genes in relapsed ALL. **(A)** ProteinPaint visualizations of the most frequently clonally mutated genes by nonsilent SNV/indels. **(B)** ProteinPaint visualizations of three genes that were exclusively affected by clonal nonsilent SNV/indels in the tumors of more than 5 patients in our cohort.


A
SJBALL036
SJERG009
SJETV043
SJMLL026
SJMLL039
MAF
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
D
XDHM
R1
B
1
ARID2 2
3 NR3C1 CTCF
12
4 TP53
5 BRAF TET2 TP53
PCSK7 13
6
7
8 SETBP1
9 NTRK1 MAP3K13 MAP2K1
CTNND2 STAT3 11
10
D, R1, R2 shared
D only
R1, R2 shared
R2 only
C
MAF
1
2
3
4
5
6
7
8
9
10
11
12
13
Sample

◀**Figure S5.8.** Mutational landscape of xenografts resolves clonal structure. **(A)** Elevated somatic mutation detection sensitivity in xenografted leukemia samples. To indicate the sensitivity of mutation calling, the highest MAF in the xenografts generated from the diagnosis sample (XDHM) was picked for each somatic mutation. Mutations showing increased MAF from D to XDHM are linked by lines. The mutations showing significant increment ($MAF<0.05$ in D and $\geqslant 0.05$ in XDHM) are highlighted in blue and the mutations missed in D (<3 mutant reads) but rescued (⩾3 mutant reads) in XDHM sample are highlighted in red. D: diagnosis patient sample, R1: first relapse patient sample. **(B)** Clonal evolution model of SJETV010 inferred from MAF of somatic mutations detected from primary patient samples and xenografted samples. In this case, only diagnosis and second relapse (R2) samples were available for transplantation. Nonsilent mutations in cancer genes (COSMIC Cancer Gene Census) in each clone are listed. The colors indicate whether mutations detected in bulk samples could be denoted as a parental clone (shared by D, R1 and R2), a D unique clone, an R2 unique clone and two relapse shared clones. **(C)** Distribution of somatic mutations' MAF in D, R1, R2 and xenografted leukemia samples from SJETV010. The color scheme for different clones is the same as in B. Without xenografts' information, clones 3 and 4 and the clones 5-13 would be indistinguishable. D.*.#, samples originating from D; R.*.#, originating from R2, *.BM.#, samples collected from bone marrow, *.CNS.#, collected from central nervous system; *.SP.#, collected from spleen.


A
MAF detected by ddPCR
50%
5%
0.5%
0.05%
0.005%
0.0005%
Expected MAF
CREBBP
NT5C2
NRAS
B
MAF Run2 (%)
MAF Run1 (%)
r = 0.9997
100
10
1
0.1
0.01
0.001
NT5C2 R39Q
CREBBP R1446C
KRAS G12C
NRAS G12R
C
CREBBP R1446C
Diagnosis
T=0
2.3979%
WT : 189515
MUT : 4656
Remission 1
T=169
0.0053%
WT : 149731
MUT : 8
Remission 2
T=450
0.0063%
WT : 174216
MUT : 11
Relapse
T=1537
45.09%
WT : 82473
MUT : 67737
REH
WT
0%
WT : 88588
MUT : 0
NTC
0%
WT : 0
MUT : 0
Comp-PMT2
Comp-PMT1
3000
2000
1000
0
-1000
500
1375
2250
3125
4000
D
MAF WES (%)
MAF CapVal (%)
MAF ddPCR (%)
r = 0.971
r = 0.9964
CREBBP R1446C
KRAS G12C
KRAS G12S
NRAS G12R
NT5C2 P414A
NT5C2 R39Q
WHSC1 E1099K
CR
D
R1
R2

◂**Figure S5.9.** Validation of digital droplet PCR (ddPCR) assay. **(A)** Test sensitivity for the *NT5C2* p.R39Q, *CREBBP* p.R1446C and *NRAS* p.G12R assays was determined using serial dilutions of patient derived leukemic cells containing the mutation with REH cells that were wild type for all mutations tested. The R-squared coefficient was 0.9994, 0.9987 and 0.9993 respectively. **(B)** A total of 9 samples were measured twice with high correlation (*r* = 0.9997). The four different assays are indicated by colors. **(C)** Representative images of the ddPCR output. Depicted are the results of the *CREBBP* p.R1446C assay for 2 tumor samples and 2 remission samples from patient SJBALL013, plus the REH wild type control and the no template control (NTC). Wild type droplets are labeled with the VIC fluorescent reporter dye and droplets containing mutant PCR products are labeled with FAM fluorescent dye. The gates used for counting the wild type (WT) and mutant (MUT) droplets are indicated, as well as the number of droplets within these gates. T indicates the time point of the sample in number of days after diagnosis. **(D)** The MAF as measured by WES (left) or capture validation (CapVal, right) correlates well with the values measured by ddPCR. A total of 11 samples were found negative (MAF = 0%) on the listed mutations in WES and/or CapVal, but mutant droplets were detected with ddPCR. One sample was positive in WES (SJTALL001 *KRAS* p.G12C MAF = 3.2%), but measured 0% in both CapVal and ddPCR (**Table S5.23**).


A
SBS signature
28
17
26
12
21
25
22
27
A
1
6
C
15
20
14
30
11
7
23
19
D
9
16
5
8
3
29
24
4
18
10
B
2
13
SBS signature
B
D
C
A
0.00
0.25
0.50
0.75
1.00
Cosine similarity
B
Case
SJHYPER022 R1
SJETV010 R1
SJETV010 R2
SJHYPO117 R1
SJHYPO117 R2
SJTALL023 R1
SJBALL113 R1
SJTALL091 R1
SJTALL023 R2
SJHYPER127 R1
SJTALL057 R1
SJHYPO126 R1
SJBALL113 R2
SJTALL139 R2
SJHYPO146 R2
SJETV048 R1
SJBALL022432 R2
14
20
15
C
6
1
A
5
16
9
D
13
2
B
SBS signature
C
Relative contribution
Signature A
Signature B
Signature C
Signature D
C>A
C>G
C>T
T>A
T>C
T>G
D
Signature A
SBS1
Relative contribution
C>T
E
SJETV010 R2
SJHYPER022 R1
Absolute contribution
C>T
F
Absolute contribution (no. mutations)
30000
20000
10000
0
Signature
A
B
C
D
SJBALL022432 D
SJBALL022432 R2
SJBALL113 D
SJBALL113 R1
SJBALL113 R2
SJETV010 D
SJETV010 R1
SJETV010 R2
SJETV048 D
SJETV048 R1
SJHYPER022 D
SJHYPER022_R1
SJHYPO117 D
SJHYPO117 R1
SJHYPO117 R2
SJHYPO126 D
SJHYPO126 R1
SJHYPO146 D
SJHYPO146 R2
SJTALL023 D
SJTALL023 R1
SJTALL023 R2
SJTALL057 D
SJTALL057 R1
SJTALL091 D
SJTALL091 R1
G
SJETV010 R2
SJHYPER022 R1
SJHYPO117 R1
SJHYPO117 R2
Colon (average)
Relative contribution
Strand
Transcribed
Untranscribed
H
SJETV010 R2
SJHYPER022 R1
Colon organoids (average)
Absolute contribution
Mutation density (C>T in CpGs) per mb
Replication timing
Replication timing
Replication timing
C>T at CpGs in left strand
C>T at CpGs in right strand

◀**Figure S5.10.** Mutational signatures. **(A)** Comparison between the *de novo* extracted signatures and known COSMIC signatures. Signature A resembles signature SBS1, characterized by spontaneous deamination of methylated cytosines in CpG context (red bar). Signature B closely resembles signatures SBS2 and SBS13, both AID/APOBEC associated signatures (orange bar). Signature C clusters close to signatures SBS6, SBS15, and SBS20, all signatures associated with mismatch repair (MMR; blue bar). Signature D shows highest resemblance to a group of signatures with less distinct characteristics. **(B)** Cosine similarity heatmap of the complete set of acquired mutations per sample relative to involved COSMIC and *de novo* extracted signatures (based on clusters identified in panel A) shows clustering of samples in the four groups presented in **Figure 5.6.** Clusters are indicated by colored bars as in panel A. Samples marked by a grey bar all have a major contribution of the less distinctive mutational profile D. **(C)** *De novo* extracted mutational signatures using WGS data. Signature profiles are highly concordant with signatures extracted from WES data. **(D)** *De novo* extracted somatic mutational signature A using WGS data (top panel) and signature SBS1 (bottom panel) according to the 96-substitution classification. Signature A is hallmarked by a higher relative number of ApCpG, CpCpG and GpCpG compared to TpCpG. **(E)** Total mutation profiles (containing all substitutions) of SJETV10-R2 and SJHYPER022-R1 based on WGS data. **(F)** Absolute contribution of each of the four signatures (WGS data) to each of 26 diagnosis and relapse samples. Composition of the four signatures in individual samples was highly concordant with the composition obtained from WES data. **(G)** Relative contribution of C>T transitions at CpGs in transcribed and untranscribed strand of four samples with a prominent signature A contribution. **(H)** Absolute contribution of C>T transitions at CpGs in the leading and lagging strands (top panels) and CpG>TpG mutation density in early, intermediate and late replicated regions of two hypermutated relapses and three healthy colon organoids.

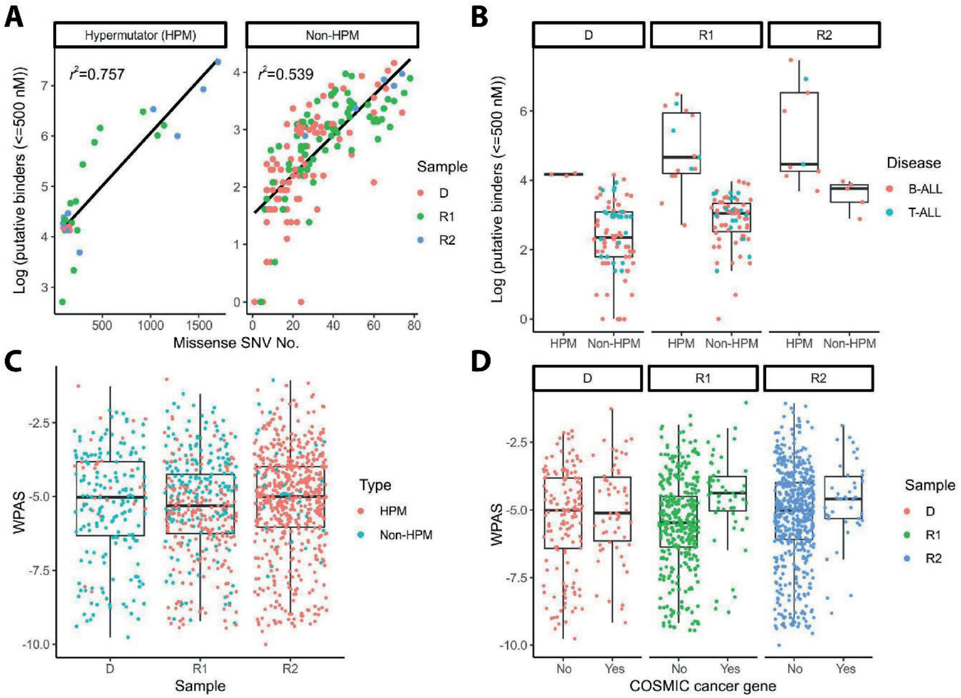


Figure S5.11. Neoepitope analysis of somatic missense SNVs. **(A)** Number of predicted HLA-binding mutant peptides (<=500nM) per tumor correlates significantly with missense mutation burden ($P < 0.001$). Correlation efficiency (r^2) of the fitted functions are shown in hypermutators (HPM) and non-HPM samples, respectively. Different disease stages, including diagnosis (D), first relapse (R1), and second relapse (R2), are shown in different colors. The number of predicted binders also varies significantly as an effect of HPM status ($P < 0.001$), with significant interactions between HPM status and mutation burden ($P < 0.001$) and between disease progression and mutation burden ($P = 0.0452$). **(B)** The number of predicted HLA-binding mutant peptides (<=500nM) per tumor varies as a function of disease stage (D-R1-R2; $P < 0.001$) and HPM status ($P < 0.001$), with more putative binders evident in hypermutated tumors and as disease progresses. Differences in the number of binders per tumor owed to effects of disease (B- and T-ALL), though visually apparent, did not reach the threshold for statistical significance ($P = 0.0511$). No significant interaction between disease progression and HPM status was identified. **(C)** Distribution of WPAS (Weighted Putative Antigenicity Score) scores across disease progression using a subset of SNVs with available expression data. WPAS varies as an effect of disease progression ($P = 0.015$), with median WPAS highest at the R2 stage (suggesting more antigenic variants in tumors from prolonged disease progression). **(D)** Distribution of WPAS scores categorized by COSMIC cancer gene status in different disease stages. WPAS again varies as an effect of disease progression (as in C) but also according to COSMIC cancer gene status categories ($P = 0.005$), with COSMIC genes showing overall higher putative antigenicity. All P values were calculated by the ANOVA test.

FUNDING

This work was supported in part by the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital, a St. Baldrick's Foundation Scholar Award, the Henry Schueler 41 & 9 Foundation, and a Robert J. Arceci Innovation Award (to CGM), the NCI grants P30 CA021765 (St. Jude Cancer Center Support Grant), P50 GM115279 (to CGM, MVR, WEE, GN, JZ, and JJY), Outstanding Investigator Award R35 CA197695 (to CGM), the Dutch Cancer Society (KUN2012-5366, to EW), the American Society of Hematology Scholar Award (to ZG), the Leukemia & Lymphoma Society's Career Development Program Special Fellow (to ZG), the NCI of the NIH under award number K99CA241297 (to ZG), the KiKa Foundation (KiKa150 to RPK), the China Scholarship Council (CSC 201304910347, to JY), and the Royal Netherlands Academy of Arts and Sciences Ter Meulen Grant (to JY). Work to JED was supported by funds from: the Princess Margaret Cancer Centre Foundation, Ontario Institute for Cancer Research with funding from the Province of Ontario, Canadian Institutes for Health Research, Canadian Cancer Society Research Institute, Terry Fox Research Institute, Genome Canada through the Ontario Genomics Institute, and a Canada Research Chair.

DATA AVAILABILITY

WES, WGS, transcriptome sequencing, and SNP array data are available at the European Genome-Phenome Archive (accession no. EGAS00001003975). The genomic landscape reported in this study can be explored at the St. Jude PeCan Data Portal https://stjuderesearch.org/site/data/relapsed-all.

CONFLICTS OF INTEREST

MVR reports receiving a commercial research grant from Servier Pharmaceuticals. PGT has received speakers bureau honoraria from Illumina and PACT Pharma and has ownership interest (including patents) in PCT/US2016/064735. JED is a Scientific Advisory Board member at Trillium Therapeutics, reports receiving a commercial research grant from Celgene, has ownership interest in a patent licensed to University Health Network. CGM reports receiving commercial research grants from Pfizer, AbbVie, and Loxo Oncology, has received speakers bureau honoraria from Amgen and Pfizer, and is a consultant/advisory board member for Illumina. No potential conflicts of interest were disclosed by the other authors.

CONTRIBUTIONS

EW: Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing-original draft, project administration, writing-review, and editing. ZG: Resources, data curation, software, formal analysis, validation, investigation, visualization, methodology, writing-original draft, writing-review, and editing. SMD: Resources, data curation, software, formal analysis, investigation, visualization, methodology, writing-original draft, writing-review, and editing. ŽA: Software, formal analysis, investigation, methodology, writing-original draft, writing-review and, editing. JCC: Resources, data curation, software, formal analysis, investigation, methodology, writing-original draft, writing-review, and editing. XM: Resources, data curation, software, formal analysis, investigation, methodology, writing-original draft, writing-review, and editing. MNE: Data curation, software, formal analysis, and methodology. DPT: Data curation, software, formal analysis, investigation, and methodology. MvdV: Data curation and investigation. MCJJ: Data curation. IM: Resources, data curation, and visualization. XZ: Resources and visualization. JW: Resources and visualization. LS: Resources, software, formal analysis, investigation, visualization, and methodology. SP: Software, formal analysis, investigation, and methodology. DP: Software, formal analysis, investigation, and methodology. CC: Formal analysis. GS: Formal analysis. YF: Formal analysis. YS: Formal analysis, validation, and investigation. MR: Resources, validation, and investigation. KM: Resources and investigation. JY: Validation and investigation. RvB: Formal analysis, validation, and investigation. FB: Formal analysis. II: Resources, data curation, formal analysis, validation, investigation, and methodology. KGR: Resources, data curation, validation, investigation, methodology, and project administration. JW: Resources, formal analysis, investigation, and project administration. GW: Formal analysis and investigation. JM: Formal analysis and investigation. JE: Formal analysis, validation, and investigation. GN: Validation and investigation. SRO: Investigation. KEN: Resources and investigation. C-HP: Resources. JZ: Resources, software, and formal analysis. WEE: Resources, software, and formal analysis. MVR: Resources. JJY: Resources. PGT: Conceptualization, resources, data curation, software, formal analysis, and investigation. JED: Conceptualization, resources, data curation, software, formal analysis, supervision, funding acquisition, investigation, methodology, writing-original draft, project administration, writing-review, and editing. RPK: Conceptualization, resources, data curation, software, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing-original draft, project administration, writing-review, and editing. CGM: Conceptualization, resources, data curation, software, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing-original draft, project administration, writing-review, and editing.

ACKNOWLEDGMENTS

We thank the Biorepository, Genome Sequencing Facility of the St. Jude Hartwell Center for Bioinformatics and Biotechnology, and the Flow Cytometry and Cell Sorting Core Facility of St. Jude Children's Research Hospital. We acknowledge WM Nillesen and AR Mensenkamp for expert advice on germline mutation interpretation and NB Waanders for bioinformatic assistance.

REFERENCES

1. Hunger SP, Mullighan CG. Acute lymphoblastic leukemia in children. N Engl J Med. 2015;373(16):1541-1552.
2. Nguyen K, Devidas M, Cheng SC, La M, Raetz EA, Carroll WL, et al. Factors influencing survival after relapse from acute lymphoblastic leukemia: a Children's Oncology Group study. Leukemia. 2008;22(12):2142-2150.
3. Iacobucci I, Mullighan CG. Genetic basis of acute lymphoblastic leukemia. J Clin Oncol. 2017;35(9):975-983.
4. Ma X, Edmonson M, Yergeau D, Muzny DM, Hampton OA, Rusch M, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun. 2015;6(1):6604.
5. Schwartzman O, Savino AM, Gombert M, Palmi C, Cario G, Schrappe M, et al. Suppressors and activators of JAK-STAT signaling at diagnosis and relapse of acute lymphoblastic leukemia in Down syndrome. Proc Natl Acad Sci U S A. 2017;114(20):E4030-E4039.
6. Meyer JA, Wang J, Hogan LE, Yang JJ, Dandekar S, Patel JP, et al. Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. Nat Genet. 2013;45(3):290-294.
7. Tzoneva G, Perez-Garcia A, Carpenter Z, Khiabanian H, Tosello V, Allegretta M, et al. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. Nat Med. 2013;19(3):368-371.
8. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science. 2008;322(5906):1377-1380.
9. Mar BG, Bullinger LB, McLean KM, Grauman PV, Harris MH, Stevenson K, et al. Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. Nat Commun. 2014;5(1):3469.
10. Oshima K, Khiabanian H, da Silva-Almeida AC, Tzoneva G, Abate F, Ambesi-Impiombato A, et al. Mutational landscape, clonal evolution patterns, and role of RAS mutations in relapsed acute lymphoblastic leukemia. Proc Natl Acad Sci U S A. 2016;113(40):11306-11311.
11. Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011;471(7337):235-239.
12. Irving J, Matheson E, Minto L, Blair H, Case M, Halsey C, et al. Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. Blood. 2014;124(23):3420-3430.
13. Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, et al. Visualizing tumor evolution with the fishplot package for R. BMC Genom. 2016;17(1):880.
14. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. PLoS Comput Biol. 2014;10(8):e1003665.
15. Klco JM, Miller CA, Griffith M, Petti A, Spencer DH, Ketkar-Kulkarni S, et al. Association between mutation clearance after induction therapy and outcomes in acute myeloid leukemia. JAMA. 2015;314(8):811-822.
16. Zhang J, McCastlain K, Yoshihara H, Xu B, Chang Y, Churchman ML, et al. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. Nat Genet. 2016;48(12):1481-1489.

17. Pounds S, Cheng C, Li S, Liu Z, Zhang J, Mullighan C. A genomic random interval model for statistical analysis of genomic lesion data. Bioinformatics. 2013;29(17):2088-2095.
18. Andries V, Vandepoele K, Staes K, Berx G, Bogaert P, Van Isterdael G, et al. NBPF1, a tumor suppressor candidate in neuroblastoma, exerts growth inhibitory effects by inducing a G1 cell cycle arrest. BMC Cancer. 2015;15:391.
19. Evensen NA, Madhusoodhan PP, Meyer J, Saliba J, Chowdhury A, Araten DJ, et al. MSH6 haploinsufficiency at relapse contributes to the development of thiopurine resistance in pediatric B-lymphoblastic leukemia. Haematologica. 2018;103(5):830-839.
20. Alexander TB, Gu Z, Iacobucci I, Dickerson K, Choi JK, Xu B, et al. The genetic basis and cell of origin of mixed phenotype acute leukaemia. Nature. 2018;562(7727):373-379.
21. Notta F, Mullighan CG, Wang JCY, Poeppl A, Doulatov S, Phillips LA, et al. Evolution of human BCR–ABL1 lymphoblastic leukaemia-initiating cells. Nature. 2011;469(7330):362-367.
22. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature. 2011;469(7330):356-361.
23. Garcia Prat L, Dobson SM, Chan-Seng-Yue M, Vanner R, Murison A, Wintersinger J, et al. Relapse-initiating clones preexisting at diagnosis in B- cell acute lymphoblastic leukemia help predict molecular pathways of relapse. Blood. 2018;132(Supplement 1):915-915.
24. Dobson SM, García-Prat L, Vanner RJ, Wintersinger J, Waanders E, Gu Z, et al. Relapse-fated latent diagnosis subclones in acute B lineage leukemia are drug tolerant and possess distinct metabolic programs. Cancer Discov. 2020;10(4):568-587.
25. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive analysis of hypermutation in human cancer. Cell. 2017;171(5):1042-1056.e1010.
26. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-421.
27. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. 2012;149(5):994-1007.
28. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012;149(5):979-993.
29. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. Cell. 2019;176(6):1282-1294.e1220.
30. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47(12):1402-1407.
31. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101.
32. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-specific mutation accumulation in human adult stem cells during life. Nature. 2016;538(7624):260-264.
33. Roerink SF, Sasaki N, Lee-Six H, Young MD, Alexandrov LB, Behjati S, et al. Intra-tumour diversification in colorectal cancer at the single-cell level. Nature. 2018;556(7702):457-462.
34. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, et al. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. Cell. 2016;164(3):538-549.
35. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics. 2014;30(23):3310-3316.

36. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. Immunogenetics. 2012;64(3):177-186.

37. Eisen HN, Hou XH, Shen C, Wang K, Tanguturi VK, Smith C, et al. Promiscuous binding of extracellular peptides to cell surface class I MHC protein. Proc Natl Acad Sci U S A. 2012;109(12):4580.

38. Tzoneva G, Dieck CL, Oshima K, Ambesi-Impiombato A, Sánchez-Martín M, Madubata CJ, et al. Clonal evolution mechanisms in NT5C2 mutant-relapsed acute lymphoblastic leukaemia. Nature. 2018;553(7689):511-514.

39. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009;360(5):470-480.

40. Spinella JF, Richer C, Cassart P, Ouimet M, Healy J, Sinnett D. Mutational dynamics of early and late relapsed childhood ALL: rapid clonal expansion and long-term dormancy. Blood Adv. 2018;2(3):177-188.

41. Gawad C, Koh W, Quake SR. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. Proc Natl Acad Sci U S A. 2014;111(50):17947-17952.

42. Swaminathan S, Klemm L, Park E, Papaemmanuil E, Ford A, Kweon SM, et al. Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. Nat Immunol. 2015;16(7):766-774.

43. Zamora AE, Crawford JC, Thomas PG. Hitting the target: how T cells detect and eliminate tumors. J Immunol. 2018;200(2):392-399.

44. Zamora AE, Crawford JC, Allen EK, Guo XJ, Bakke J, Carter RA, et al. Pediatric patients with acute lymphoblastic leukemia generate abundant and functional neoantigen-specific CD8(+) T cell responses. Sci Transl Med. 2019;11(498):eaat8549.

45. Pui CH, Pei D, Sandlund JT, Ribeiro RC, Rubnitz JE, Raimondi SC, et al. Long-term results of St Jude total therapy studies 11, 12, 13A, 13B, and 14 for childhood acute lymphoblastic leukemia. Leukemia. 2010;24(2):371-382.

46. Pui CH, Sandlund JT, Pei D, Campana D, Rivera GK, Ribeiro RC, et al. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. Blood. 2004;104(9):2690-2696.

47. Pui C-H, Campana D, Pei D, Bowman WP, Sandlund JT, Kaste SC, et al. Treating childhood acute lymphoblastic leukemia without cranial irradiation. N Engl J Med. 2009;360(26):2730-2741.

48. Pounds S, Cheng C, Mullighan C, Raimondi SC, Shurtleff S, Downing JR. Reference alignment of SNP microarray signals for copy number analysis of tumors. Bioinformatics. 2009;25(3):315-321.

49. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics. 2007;23(6):657-663.

50. Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nat Genet. 2013;45(3):242-252.

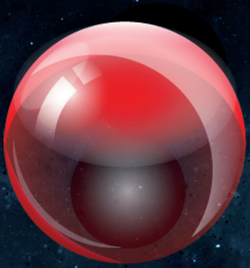
51. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012;481(7380):157-163.

52. Edmonson MN, Zhang J, Yan C, Finney RP, Meerzaman DM, Buetow KH. Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. Bioinformatics. 2011;27(6):865-866.

53. Roberts KG, Li Y, Payne-Turner D, Harvey RC, Yang Y-L, Pei D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. N Engl J Med. 2014;371(11):1005-1015.

54. Nicorici D, Satalan M, Edgren H, Kangaspeska S, Murumagi A, Kallioniemi O, et al. FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv; 2014:011650.
55. Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, et al. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. Genome Biol. 2011;12(1):R6.
56. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-169.
57. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
58. Gu Z, Churchman M, Roberts K, Li Y, Liu Y, Harvey RC, et al. Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. Nat Commun. 2016;7:13331.
59. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214-218.
60. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12(4):R41.
61. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, et al. ClonEvol: clonal ordering and visualization in cancer sequencing. Ann Oncol. 2017;28(12):3076-3082.
62. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The database of genomic variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 2014;42(D1):D986-D992.
63. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, et al. Germline mutations in predisposition genes in pediatric cancer. N Engl J Med. 2015;373(24):2336-2346.
64. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010;11(1):367.
65. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013;3(1):246-259.
66. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10(1):33.
67. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.
68. Drost J, van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science. 2017;358(6360):234-238.
69. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models Using lme4. J Stat Softw. 2015;67(1):1-48.
70. Fox J, Weisberg, Sanford. An R Companion to Applied Regression: Thousand Oaks: Sage; 2011.
71. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43(D1):D805-D811.
72. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45(D1):D777-D783.

73. Mullighan CG, Miller CB, Radtke I, Phillips LA, Dalton J, Ma J, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature. 2008;453(7191):110-114.
74. Waanders E, Hebeda KM, Kamping EJ, Groenen PJ, Simons A, Hoischen A, et al. Independent development of lymphoid and histiocytic malignancies from a shared early precursor. Leukemia. 2016;30(4):955-958.
75. Weng AP, Ferrando AA, Lee W, Morris JPt, Silverman LB, Sanchez-Irizarry C, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science. 2004;306(5694):269-271.
76. Ge Z, Li M, Zhao G, Xiao L, Gu Y, Zhou X, et al. Novel dynamin 2 mutations in adult T-cell acute lymphoblastic leukemia. Oncol Lett. 2016;12(4):2746-2751.
77. Chao MM, Todd MA, Kontny U, Neas K, Sullivan MJ, Hunter AG, et al. T-cell acute lymphoblastic leukemia in association with Borjeson-Forssman-Lehmann syndrome due to a mutation in PHF6. Pediatr Blood Cancer. 2010;55(4):722-724.
78. Van Vlierberghe P, Palomero T, Khiabanian H, Van der Meulen J, Castillo M, Van Roy N, et al. PHF6 mutations in T-cell acute lymphoblastic leukemia. Nat Genet. 2010;42(4):338-342.
79. Narumi S, Amano N, Ishii T, Katsumata N, Muroya K, Adachi M, et al. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. Nat Genet. 2016;48(7):792-797.
80. Schwartz JR, Wang S, Ma J, Lamprecht T, Walsh M, Song G, et al. Germline SAMD9 mutation in siblings with monosomy 7 and myelodysplastic syndrome. Leukemia. 2017;31(8):1827-1830.
81. Chen DH, Below JE, Shimamura A, Keel SB, Matsushita M, Wolff J, et al. Ataxia-pancytopenia syndrome is caused by missense mutations in SAMD9L. Am J Hum Genet. 2016;98(6):1146-1158.
82. Waanders E, Scheijen B, Jongmans MC, Venselaar H, van Reijmersdal SV, van Dijk AH, et al. Germline activating TYK2 mutations in pediatric patients with two primary acute lymphoblastic leukemia occurrences. Leukemia. 2017;31(4):821-828.
83. Shah S, Schrader KA, Waanders E, Timms AE, Vijai J, Miething C, et al. A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. Nat Genet. 2013;45(10):1226-1231.
84. Auer F, Ruschendorf F, Gombert M, Husemann P, Ginzel S, Izraeli S, et al. Inherited susceptibility to pre B-ALL caused by germline transmission of PAX5 c.547G>A. Leukemia. 2014;28(5):1136-1138.
85. Russell LJ, Capasso M, Vater I, Akasaka T, Bernard OA, Calasanz MJ, et al. Deregulated expression of cytokine receptor gene, CRLF2, is involved in lymphoid transformation in B-cell precursor acute lymphoblastic leukemia. Blood. 2009;114(13):2688-2698.
86. Huether R, Dong L, Chen X, Wu G, Parker M, Wei L, et al. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. Nat Commun. 2014;5:3630.



CHAPTER 6

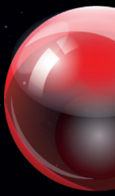
UNRAVELLING THE SEQUENTIAL INTERPLAY OF MUTATIONAL MECHANISMS DURING CLONAL EVOLUTION IN RELAPSED PEDIATRIC ACUTE LYMPHOBLASTIC LEUKEMIA

Željko Antić[1], Stefan H. Lelieveld[1*], Cédric G. van der Ham[1], Edwin Sonneveld[1,2], Peter M. Hoogerbrugge[1,2] and Roland P. Kuiper[1,3]*

[*] These authors contributed equally to this study as lead authors

[1] Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands
[2] Dutch Childhood Oncology Group, Utrecht, The Netherlands
[3] Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

Genes (Basel); February 2021; 12(2):214



ABSTRACT

Pediatric acute lymphoblastic leukemia (ALL) is the most common pediatric malignancy and is characterized by clonal heterogeneity. Genomic mutations can increase proliferative potential of leukemic cells and cause treatment resistance. However, mechanisms driving mutagenesis and clonal diversification in ALL are not fully understood. In this proof-of-principle study, we performed whole genome sequencing of two cases with multiple relapses in order to investigate whether groups of mutations separated in time show distinct mutational signatures. Based on mutation allele frequencies at diagnosis and subsequent relapses, we clustered mutations into groups and performed cluster-specific mutational profile analysis and *de novo* signature extraction. In patient 1, who experienced two relapses, the analysis unraveled a continuous interplay of aberrant activation induced cytidine deaminase (AID)/apolipoprotein B editing complex (APOBEC) activity. The associated signatures SBS2 and SBS13 were present already at diagnosis, and although emerging mutations were lost in later relapses, the process remained active throughout disease evolution. Patient 2 had three relapses. We identified episodic mutational processes at diagnosis and first relapse leading to mutations resembling ultraviolet light-driven DNA damage, and thiopurine-associated damage at first relapse. In conclusion, our data shows that investigation of mutational processes in clusters separated in time may aid in understanding the mutational mechanisms and discovery of underlying causes.

INTRODUCTION

Pediatric acute lymphoblastic leukemia (ALL) represents the most common pediatric malignancy[1-4]. Despite improvements in treatment, around 10-15% of the children do not achieve long-term remission and outcome among them remains poor[5,6]. Previous next-generation sequencing (NGS) studies revealed unprecedented diversity in the genomic alteration of ALL[7-12]. However, only a small subset of these alterations occurs in known cancer-driver genes, which have the potential to initiate and propel disease progression. Furthermore, these studies revealed genetic alterations which are not essential for cancer development but may drive treatment resistance and eventually give rise to a relapse[7,8,13-20]. Although the majority of genetic alterations in ALL represent passenger mutations and do not confer with selective advantage of leukemic cells, they still contain valuable information about tumor evolution, clonal dynamics, and mechanisms driving mutagenesis[7,8,11,12,21-24].

In contrast to adult cancers, where external factors like ultraviolet (UV) light exposure, tobacco smoke, and alcohol consumption contribute to the incidence of cancer in the population, pediatric cancers are more likely the result of dysregulated intrinsic processes which directly impact normal development[8,11,12,21,23-25]. In some cases, a mutational process can become active in a (pre)malignant cell that may accelerate mutation accumulation and disease progression. Examples of these mutational processes are aberrant activity of the activation induced cytidine deaminase (AID)/ apolipoprotein B editing complex (APOBEC) class cytidine deaminases and mismatch repair deficiency (MMR), processes that also have been reported in ALL[8,11,12,21,23,24,26]. Each of these mutagenic processes shows biases represented by specific changes which occur in distinct genomic contexts[8,21,23]. These changes are recognized as footprints of underlying biological processes, also known as mutational signatures[21,23,24]. For example, mutations driven by aberrant AID/APOBEC activity typically present as C>T/G substitutions in TpCpN context. In addition, certain mutational processes may be accompanied by additional features, e.g., small indels in simple repeats in MMR deficiency, while others exhibit stronger mutational patterns in specific genomic regions, e.g., introns in AID/APOBEC-driven mutational signature[21,23,24]. So far, 72 single-base substitution (SBS) signatures have been described in the Catalogue of Somatic Mutations in Cancer (COSMIC), many of which are of unknown etiology[21]. Tumor samples taken at defined disease stages may reveal the footprints of several of these processes, but these mutational processes may have been active at different times and in different cells during tumor development. For example, compared to tumors at primary diagnosis, relapses may exhibit specific mutational signatures that are a direct consequence of

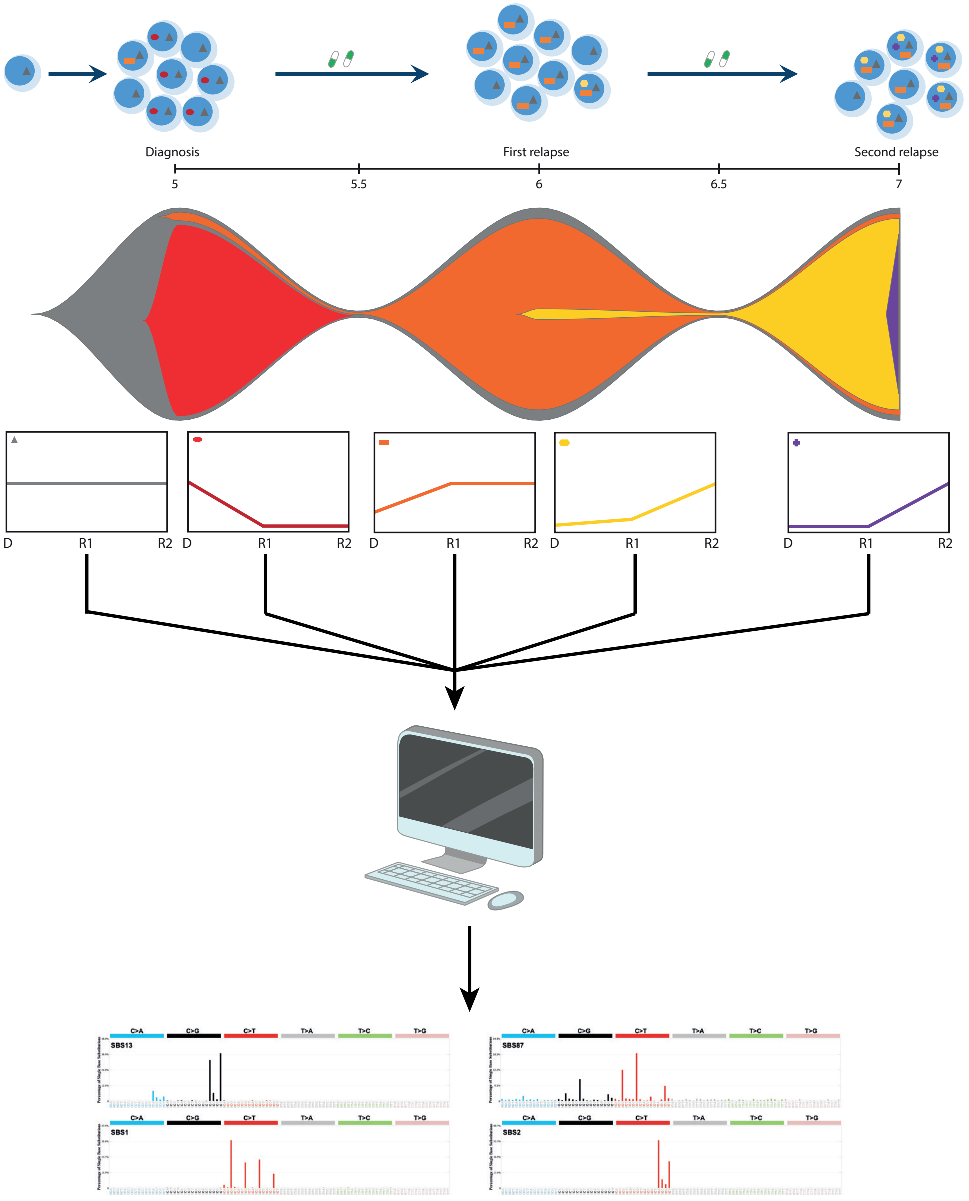

Diagnosis
First relapse
Second relapse
5
5.5
6
6.5
7
D
R1
R2
D
R1
R2
D
R1
R2
D
R1
R2
D
R1
R2
C>A
C>G
C>T
T>A
T>C
T>G
SBS13
SBS87
SBS1
SBS2

Figure 6.1. Schematic representation of the approach. Mutational profiles are analyzed in clusters of mutations that follow similar dynamics over different time points. Leukemic cells are thought to collect mutations sequentially, indicated as different colored symbols in each cell (top panel). New clones evolving from the ancestral clone, depicted in different colors in the fish plot, can be extracted based on the dynamics of the variant allele frequency (VAF) between different time points, which are then used to investigate active mutational mechanisms by analyzing their mutational profiles and performing *de novo* signature extraction in a clone-specific manner.

prior treatment with, e.g., DNA damaging agents as part of multidrug chemotherapy regimen[23,27]. Unraveling mutational patterns of processes driving mutagenesis in a spatial and temporal manner may give insights into the intrinsic and external influences and impact of therapy on disease evolution.

We and others have previously demonstrated that whole genome sequencing (WGS) of leukemia at diagnosis and relapse may aid in deciphering unique evolutionary trajectories of individual clones emerging from each other due to different quantities in which these clones occur at each time point[7,8,22,27]. This provides the opportunity to study mutational mechanisms that, for example, occurred before and after therapy, or can even be taken a step further by sequencing tumors from multiple tumor sites. Sequencing of multiple samples with different spatio-temporal origin can aid in discriminating clusters of mutations with similar clonal dynamics, which opens the possibility to allocate mutational mechanisms to individual clones with different behavior in time and space.

Based on our previous findings, we hypothesize that analyzing clusters of mutations in a clone-specific manner will improve the identification of mutational processes that are active in individual leukemic clones and different stages of disease progression (**Figure 6.1**). Here, we show a proof-of-principle for this approach by performing whole genome sequencing of 10 samples from two pediatric ALL patients with multiple relapses, revealing the sequential action of multiple mutational mechanisms in each of the two cases.

MATERIALS AND METHODS

Patient samples

In order to follow dynamics of a maximal number of individual somatic mutations at multiple time points during leukemia progression, we selected two patients with multiple relapses and performed WGS. Individual clones within a leukemia may have different dynamics during disease progression, which can be disentangled with greater resolution when more time points are available.

Patient 1 was a girl with Down syndrome who developed *CRLF2*-positive B-cell leukemia at the age of 7, who received treatment according to the Dutch Childhood Oncology Group (DCOG) ALL9 non-high risk protocol. After achieving complete remission, the patient experienced a first relapse 4 years after initial diagnosis and a second relapse 6.9 years after diagnosis (**Figure 6.2A**). She achieved complete remission and received an allogeneic stem cell transplant (SCT) but relapsed one year later, from which she passed away. No material was available from the third relapse.

Patient 2 is a boy who developed B-cell leukemia at the age of 2 and was treated according to the DCOG-ALL10 standard risk treatment protocol. A preserved *DDX3X-MLLT10* gene fusion was detected at all time points, but no subtype-specific abnormalities were identified. First relapse occurred 5.7 years after initial diagnosis, followed by a second relapse at 7.9 years and a third relapse at 8.3 years after initial diagnosis. As part of the treatment of the first relapse, this patient received an allogenic SCT (**Figure 6.3A**).

DNA was isolated from mononuclear cells derived from bone marrow or peripheral blood. The percentage of blast cells in tumor samples was high for most of the samples (>80% for 6 samples) (**Table S6.1**). 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).

Whole Genome Sequencing

Whole genome sequencing for patient 1 and patient 2 was performed at Novagene (Hong Kong, China) and the Hartwig Medical Foundation (Amsterdam, The Netherlands), respectively. The library was constructed using NEBNext DNA Library Preparation Kit, following sequencing on an Illumina NovaSeq 6000 platform using 150 base-pair paired-end reads. The sample-specific overview of achieved sequencing depth can be found in **Table S6.1**. All samples were aligned to the HG38/GRCh38 of the human reference genome by using the Burrows-Wheeler aligner (BWA)[28]. Duplicate reads were marked using Picard. Tumor purity estimations were performed based on the allele frequencies of high-quality somatic variants detected in the WGS data. Manta (version 1.6.0) was used for structural variant detection on the WGS data[29].

Somatic variant calling, annotation, and filtering

Somatic variants were called by the MUTECT2 software of GATK package version 4.1.1.0 followed by the FilterMutectCalls function as recommended by the authors. We removed the somatic variant that did not have a “PASS” filter status to ensure the highest likelihood of true somatic variants. Next, we annotated the filtered somatic variants with the Variant Effector Prediction (VEP) version 92[30]. Somatic variants were annotated with: (i) frequencies of the 76,156 whole genome sequenced samples from the GnomAD release 3.0[31], and (ii) frequencies from 498 whole genome sequenced samples of the Genome of the Netherlands[31,32].

Somatic variants were filtered on the following criteria: (i) allele Frequency in the GnomAD database and the Genome of the Netherlands below 0.01, (ii) a minimal overall coverage of at least 20X in all samples and a minimum of at least 3 reads containing the variant, (iii) a minimal variant frequency of 25% in at least one of the samples from a patient, (iv) location outside the centromere locations (as defined in the UCSC genome browser), and (v) no variant reads in the control samples.

Mutation clustering

We applied *k*-mean clustering, integrated in R package stats (version 4.0.2), of high-confidence somatic variants to define clusters of variants with distinctive patterns of variant allele frequency (VAF) change between time points. These distinct clusters represent the same evolutionary trajectories as the cells from which they originate and, therefore, these clusters can be used to track dynamics of individual clones. For Patients 1 and 2, we created 10 and 15 clusters, respectively (**Figures S6.1** and **S6.2**). Next, to increase the number of mutations per cluster, we merged clusters with the same evolutionary trajectory together based on manual inspection. Finally, we cleaned the merged clusters of outlier mutations and divided the cleaned clusters to create biological relevant clusters that we subjected to mutational signature analysis (**Figures S6.1** and **S6.2**).

Mutational profile analysis

For each cluster of somatic substitutions, the 96-trinucleotide count matrices and mutation profiles were computed and visualized by the R package MutationalPatterns (version 2.0.0)[33]. For each cluster of mutations, we computed the cosine similarity of the 96-trinucleotide profile to the repertoire of 72 known single-base substitution (SBS) signatures reported in the Catalogue of Somatic Mutations in Cancer (COSMIC v3, URL: https://cancer.sanger.ac.uk/cosmic/signatures/SBS/index.tt). The cosine similarity is a measure ranging from 0 to 1 and is used to compute differences between two mutational profiles, where a value of 1 indicates an identical profile. Cosine similarity scores between the profiles of the clusters and the COSMIC signatures aids in the identification of recurrent active mutational mechanisms.

De novo mutational signature extraction

Mutational signatures were *de novo* extracted from the 96-trinucleotide mutation count matrix using non-negative matrix factorization (NMF). We used the R packages

MutationalPatterns (version 2.0.0)[33] and NMF (version 0.23.0)[34] to perform *de novo* signature extraction on the profiles of the somatic mutation clusters. To increase the power to perform *de novo* signature extraction, somatic variants of 214 whole genome sequenced pediatric ALL patients from a recent pan-cancer study were included[11]. The annotation and filtering of these somatic mutations were performed with the same pipelines and settings as the seven in-house sequenced tumor samples. We combined the 96-trinucleotide count matrices of the seven tumors of our two patients and the 214 and performed the *de novo* extraction of signatures on the combined set of 221 samples. The relative and absolute contribution of the *de novo* extracted signatures in the mutational profiles of the somatic mutation clusters were computed using the MutationalPatterns R package. Reconstruction of mutational profiles using *de novo* extracted mutational signatures was performed using the MutationalPatterns package in R[33].

RESULTS

In order to unravel somatic mutations at each time point during tumor evolution, we performed WGS of tumor samples at diagnosis, remission, and all subsequent relapses of two patients diagnosed with B-cell precursor acute lymphoblastic leukemia (BCP-ALL). We specifically selected cases with multiple relapses, because they offer much greater resolution to distinguish individual clones, compared to non-relapsed cases. Since patient 2 received an allogeneic SCT before his second relapse, we also sequenced a remission sample taken after SCT in order to filter out donor-derived variants identified in the second and third relapse. In total, we detected 8,922 and 8,759 single-base substitutions and 686 and 646 indels in patients 1 and 2, respectively (**Tables S6.2**-**S6.4**). Since the association between mutational signatures and underlying processes is best understood for single-base substitutions[21], we focused on this type of mutations in the present study.

Using *k*-mean clustering of the mutation allele frequencies observed at each time point, we extracted multiple clusters of single-base substitutions that followed similar dynamics over the different time points for both patients. These clusters were subsequently manually curated based on biological criteria, resulting in six and five clusters of sufficient size or biological relevance for patients 1 and 2, respectively (**Figures S6.1** and **S6.2**). Interestingly, the mutational profiles of the defined clusters in each of the two patients were highly different, even between clusters that co-occurred at one or more time points (**Table S6.5**). Furthermore, the mutation profiles of some of these clusters showed high similarity with known COSMIC signatures (**Table S6.6**), suggesting that indeed clone-specific mutational processes could be revealed in this manner. Below, we will describe the clonal trajectories with underlying mutational processes in more detail.

(Re)activation of aberrant AID/APOBEC expression follows evolution of individual clones

In patient 1, we identified a total of six clusters of somatic single-base substitutions that presented with different dynamics between diagnosis, first relapse, and second relapse (**Figures 6.2B** and **S6.1**). Cluster 1 was composed of 745 mutations that were preserved between all time-points, suggesting their origin from a (pre)leukemic ancestral clone. In addition, we observed a large second cluster of mutations (5,562 mutations) representing a falling clone that was dominant at diagnosis but disappeared at first relapse. The mutations in clusters 3 and 4 appeared at time of first relapse, but whereas the mutations in cluster 3 (*n* = 562) were preserved in the second relapse, the cluster 4 mutations (*n* = 65) were lost at relapse 2 (R2). This suggests that the dominant clone in the first relapse contained mutations from both of these clusters but was eradicated during treatment. In contrast, a minor preceding clone, which did not carry the cluster 4 mutations, evolved into a second relapse. This second relapse also carried new mutations, of which about half (cluster 5; *n* = 1,094) were already detectable in low amounts in relapse 1, whereas 887 mutations (cluster 6) appeared to be newly acquired. Therefore, the six mutational clusters represent distinct mutational episodes during progression of the disease in this patient.

Next, we examined the mutational profiles of each of these clusters, which revealed that four of them showed high similarity with known COSMIC signatures (**Figure 6.2C**, **Tables S6.5** and **S6.6**). Cluster 1, which carried the preserved ancestral mutations, showed high similarity with the clock-like mutational signature SBS1 (cosine similarity = 0.89). These mutations are associated with spontaneous deamination of methylated cytosines at CpGs and may have occurred during the pre-malignant phase of the leukemia-initiating cell. Three clusters showed high similarity with the COSMIC signatures SBS2 and SBS13, which are both attributed to AID/APOBEC mutagenesis. Cosine similarity of the merged SBS2/13 signature with clusters 2, 5, and 6 was 0.98, 0.99, and 0.95, respectively (**Figure S6.3**). Clusters 3 and 4 appeared as mixed signatures, which were likely at least in part composed of SBS1 and SBS2/13. To confirm that this was indeed the case, we performed a *de novo* signature extraction in which we included all individual clusters as well as 214 publicly available ALL samples (see Methods, **Tables S6.7** and **S6.8**). The *de novo* signature analysis revealed eight signatures with high cosine similarities ranging to COSMIC signatures that resembled a variety of mutational processes (**Figure S6.4** and **Tables S6.7-S6.9**). Four extracted signatures were prominently found in at least one of the clusters from the two patients (**Figures 6.2D** and **6.3D**).

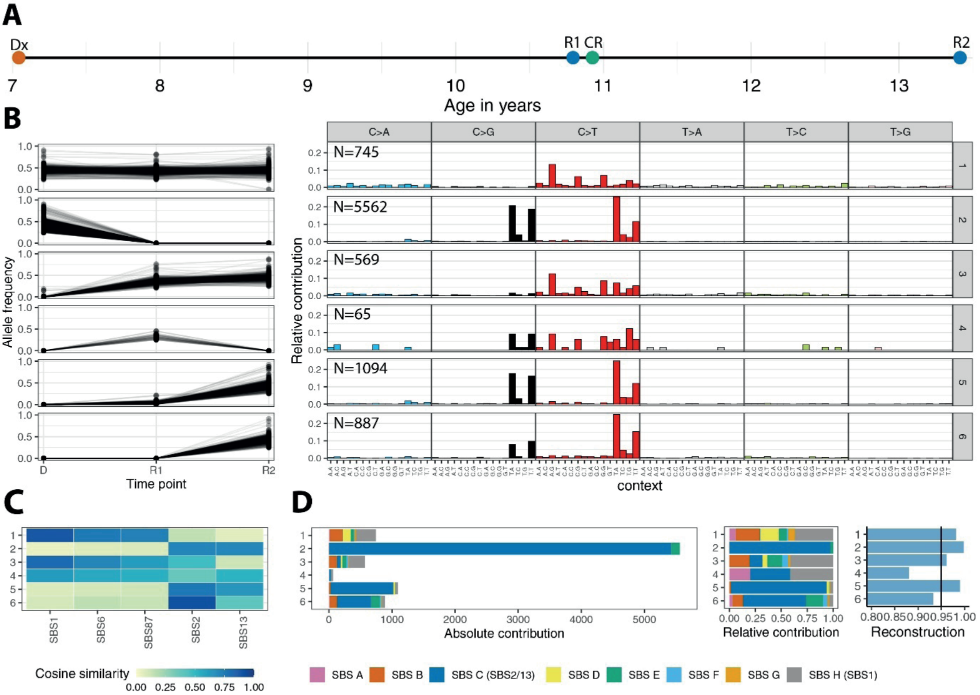


Figure 6.2. Analysis of mutational clusters in patient 1. **(A)** Schematic timeline of the analyzed samples with age at diagnosis (Dx) or relapse (R). A complete remission sample (CR) after first relapse was used as a normal control. **(B)** Clustering of high-confident somatic single-base substitutions based on allele frequencies at time of diagnosis and the two relapses (left) and the associated 96 trinucleotide mutation profiles (right). The clusters represent a (pre)leukemic ancestral clone (cluster 1), falling (cluster 2) and two rising clones (clusters 3 and 4) at first relapse, and rising clones at second relapse (clusters 5 and 6). The number of somatic mutations per cluster is indicated in the top right corner of each panel. **(C)** Heatmap showing the COSMIC signatures with a cosine similarity higher than 0.7 to the mutational profiles of at least one of the six clusters. For the complete overview of the six clusters to all 72 COSMIC v3 signatures, see **Figure S6.3**. **(D)** The absolute and relative contribution and reconstruction of eight extracted *de novo* signatures (**Figure S6.4**) for each of the clusters. The right panel depicts cosine similarities between the actual mutational profiles and the reconstructed profile based on the identified *de novo* extracted signatures, with a reliability threshold set at 90%.

In patient 1, this includes SBS H (cosine similarity of 0.98 to COSMIC signature SBS1), which represents 41% of the mutations in both clusters 1 and 3 (**Figure 6.2D**). Signature SBS C (cosine similarity of 0.98 to the combined COSMIC signatures SBS2 and SBS13), which is indeed the prominent signature in clusters 2, 5, and 6, contributed

92%, 97%, and 61% of all somatic substitutions (**Figure 6.2D**). Also, in cluster 4, which only contains 65 mutations, resulting in lower accuracy, AID/APOBEC activity appears to be responsible for 38% of the mutations. Furthermore, we noted that clusters 5 and 6 differ in their relative contributions between SBS2 and SBS13 (cluster 5: 0.75 and 0.65 vs. cluster 6: 0.91 and 0.40, respectively; **Figure 6.2C**), suggesting that SBS13 mutations were more prominent in the cluster 5 mutations, which were already present subclonally at R1 compared to the cluster 6 mutations. In conclusion, the six clusters of mutations demonstrate that AID/APOBEC-mediated mutagenesis appeared to be active in emerging clones at different stages of leukemia development in patient 1, particularly at diagnosis and relapse 2. Strikingly, these SBS2/13 mutations were hardly preserved but the underlying cause remained present in the different stages of disease.

Episodic activity of mutational mechanisms

In patient 2, we identified a cluster of preserved mutations (cluster 1), a falling clone at diagnosis (cluster 2), and rising clones at the first, second, and third relapse (clusters 3, 4, and 5; **Figures 6.3B** and **S6.2**). Therefore, similar to patient 1, patient 2 presented with different dominant clones at diagnosis and first relapse, which shared the cluster 1 mutations combined with the mutations in either cluster 2 or cluster 3, respectively. Strikingly, however, clusters 1, 2, and 3 were highly similar in their mutational profile and strongly resembled a known COSMIC mutational signature associated with UV-associated DNA damage (SBS7a, cosine similarities of clusters 1, 2, and 3 were 0.93, 0.97, and 0.89, respectively; **Figure 6.3C** and **Table S6.6**). Thus, this mutational mechanism strongly contributed to the majority of mutations in both diagnosis and first relapse. The mutations in clusters 4 and 5 were clearly different, suggesting that after first relapse, this mutational mechanism was no longer active. To gain more insight into the mutational mechanisms active at all time points, *de novo* signature extraction on the individual clusters was performed, which confirmed the presence of the mutational signature SBS7a, in clusters 1, 2, and 3 (**Figure 6.3C**, **Tables S6.7** and **S6.8**). Signature SBS7a has been associated with UV light exposure and is commonly found in head and neck cancers and melanoma. Previous studies reported this mutational signature also in several pediatric ALL patients[11,27], but the etiology is still unknown. However, despite the fact that the number of mutations in clusters 4 and 5 was too low to reliably assign signatures, SBS7a seemed to be absent, suggesting that this mutational mechanism was no longer present after first relapse. Furthermore, SBS7a represented only 89% of the mutations acquired in relapse 1-specific cluster 3, and it may be possible that these mutations were already present at low levels at time of diagnosis. Additionally, relapse

1-specific cluster 3 showed a strong presence of a different, therapy-related signature, as 39% of the mutations in cluster 3 were assigned to *de novo* extracted signature SBS A that have a cosine similarity of 0.97 to COSMIC signature SBS87. SBS87 was recently identified in relapsed BCP-ALL and was associated with thiopurine treatment[27]. The latter study reported pathogenic mutations in the cytosolic 5'-nucleotidase II gene *NT5C2* in five patients with SBS87-associated mutations at relapse, but this mutation was not identified in patient 2, and no other pathogenic mutations were found that could explain the presence of thiopurine-related damage (**Tables S6.8** and **S6.9**).

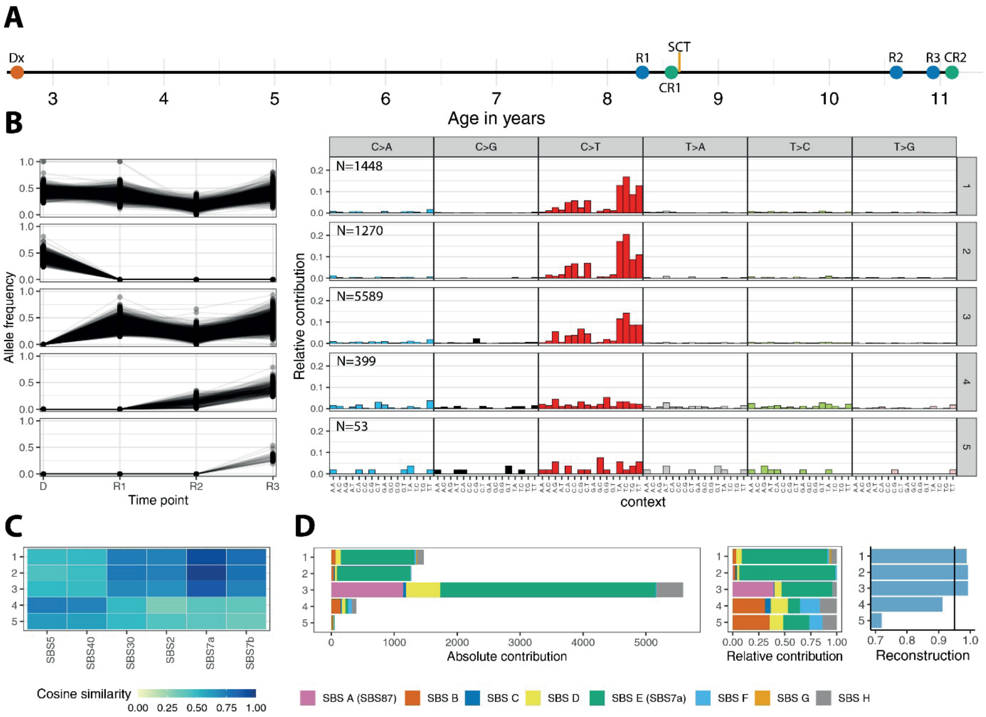


Figure 6.3. Analysis of mutational clusters in patient 2. **(A)** Schematic timeline of the analyzed samples with the age at diagnosis (Dx) or relapse (R). The patient received an allogenic stem cell transplantation (SCT) after his first relapse. Complete remission before (CR1) and after (CR2) were sequenced to correct for patient- and donor-derived normal variation. **(B)** Clustering of all high-confidence somatic single-base substitutions based on allele frequencies at diagnosis and three relapses (left) and the associated 96-trinucleotide mutation profiles corresponding to these clusters (right). The clusters represent a (pre)leukemic ancestral clone (cluster 1), falling (cluster 2) and rising clone (cluster 3) at relapse 1, and rising clones at relapse 2 (cluster 4) and relapse 3 (cluster 5). Total number of somatic mutations in each cluster is indicated in the top right corner of each profile. **(C)** Heatmap showing the COSMIC signatures with a cosine similarity higher than 0.7 to the mutational profiles of at least one of the five clusters. For the complete overview of all five clusters to the 72 COSMIC signatures, see **Figure S6.3**. **(D)** The absolute and relative contribution of eight extracted *de novo* signatures (**Figure S6.4**) for each of the clusters. The contribution of SBS E (with a cosine similarity of 0.98 to COSMIC signature SBS7a) in clusters 1, 2, and 3 is 93%, 97%, and 89%, respectively. The contribution of SBS A (with a cosine similarity of 0.97 to COSMIC signature SBS87) in cluster 3 is estimated to be 39%. The right panel depicts cosine similarities between the actual mutational profiles and the reconstructed profile based on the identified *de novo* extracted signatures, with a reliability threshold set at 90%.

DISCUSSION

Pediatric ALL is a heterogeneous disease characterized by the presence of multiple leukemic clones[7-9]. Clonal diversification at different stages during disease development may be a ramification of mutational mechanisms that are active during episodes of disease progression or throughout the entire disease course[8,26,27]. These processes drive clonal heterogeneity and may lead to selection of therapy-resistant clones[27]. Various mutational mechanisms have been identified in (relapsed) ALL cases, but the spatial and temporal activity of these mechanisms is still far from understood. In this whole genome sequencing study on two BCP-ALL patients with multiple relapses, we demonstrated that by clustering mutations based on the clonal dynamics during disease progression, distinct mutational processes can be unraveled, and their timing can be specified. This proof-of-principle can be applied to past and future whole genome sequencing studies involving multiple samples of the same tumor, separated in time or space in order to unravel the sequential interplay of mutational mechanisms during cancer progression.

Both cases revealed complex evolutionary patterns, with multiple clones rising and falling during disease progression. The clusters of mutations underlying these individual clones revealed distinct mutational mechanisms in these patients. In patient 1, the outgrowth of multiple relapses appeared to be derived from subclones that acquired new mutations by a sustained AID/APOBEC-driven mutational mechanism. This mechanism is responsible for high mutational burden in a subset of ALL[8,23,27], and can be recognized by the presence of two distinct signatures, SBS2 and SBS13. These signatures are thought to arise from the same mutational mechanism (cytosines to uracil deamination at TpC dinucleotides), but whereas SBS2 directly results from the replication of a U:A mismatch, SBS13 appears to be caused by error-prone polymerases that fill in the excised uracil[23]. SBS2 and SBS13 often co-occur in similar amounts, which explains why the two patterns can be recognized as a single mutational signature[8]. Interestingly, however, we observed a difference in the relative contribution of SBS13 between clusters 5 and 6 in patient 1. Since cluster 5 contains relapse 2 mutations that were already present subclonally at the first relapse, this observation may be suggestive of a relatively higher level of uracil excision in the subclonal stage, for example because of lower replication rates. Inclusion of more ALL cases with underlying AID/APOBEC mutagenesis in future studies may further clarify this aspect.

In patient 2, two mutational mechanisms were identified which, in contrast to patient 1, were active only temporarily. The mutations detected at diagnosis could be linked to mutational signature SBS7a, which has been associated with DNA damage caused by UV light exposure[11,21,23,24,27]. SBS7a-associated mutations have been observed occasionally in ALL[11,27], but the underlying cause of these mutations in the bone marrow is still unknown. Acquired mutations in the first relapse also contained a substantial number of SBS7a-associated mutations, but in the later relapses, this mutational mechanism was no longer active. In fact, we cannot rule out the possibility that relapse 1 emerged from a subclone that was already present at diagnosis, and that the SBS7a mutations all arose before first diagnosis. Whereas SBS7a appears to be caused by an intrinsic process, a substantial number of mutations in the first relapse could be linked to thiopurine treatment (SBS87)[27]. Presence of these thiopurine-related scars may indicate that this therapy was (partially) ineffective, which may have contributed to the development of relapse. Strikingly, however, also, this mutational mechanism was no longer present in later relapses, despite the use of thiopurines in the ALL-R3-based therapy he received after his first relapse. The absence of thiopurine scars may suggest that the later relapses were either completely resistant or completely sensitive to thiopurine-induced DNA damage.

CONCLUSIONS

In summary, we showed that the use of samples taken at multiple time points during tumor evolution may improve separation of distinct cell populations that evolve independently, thereby unravelling clone-specific mutational mechanisms. Clonal inference can be further improved by deep targeted resequencing of somatic mutations, as well as utilization of limited dilution xenograft models, as previously demonstrated[8,35]. Furthermore, we showed that independent examination of mutational clusters reveals distinct mutational profiles which correspond to previously reported mutational signatures. We therefore conclude that analyzing mutational signatures in the clusters of mutations, in a clone-specific manner, can improve our understanding of mutational processes occurring in a single parental cell.

SUPPLEMENTARY DATA

Supplementary Tables

Supplementary Tables are available online using the following link: https://doi.org/10.3390/genes12020214.

Supplementary Figures

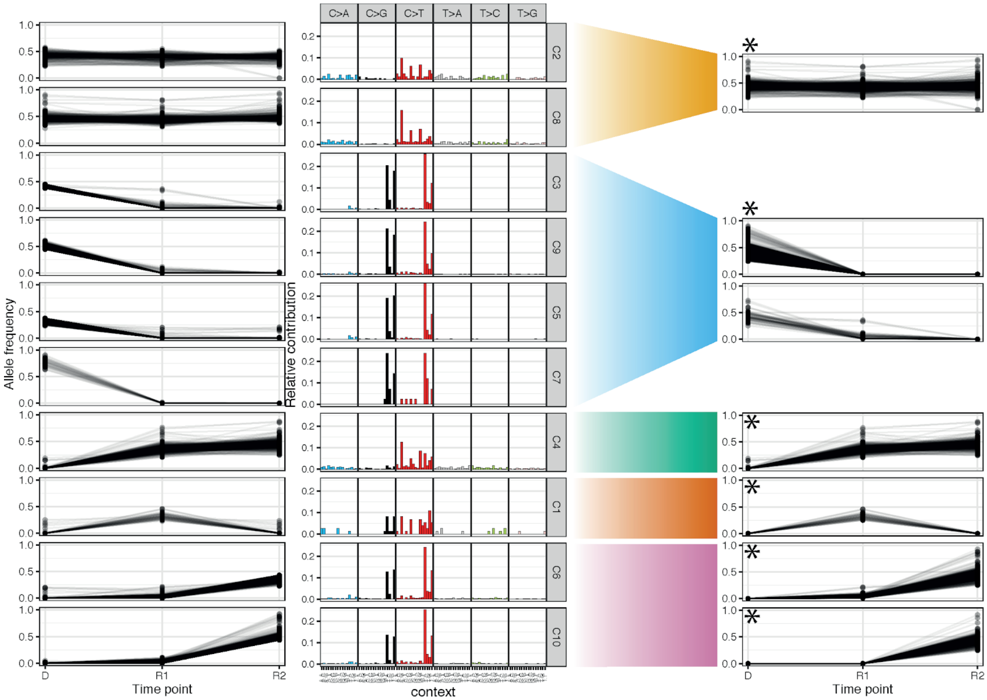


Figure S6.1. Mutation clustering in patient 1. The left panel shows the 10 clusters (*k*-means clustering; *k* = 10) of high-confidence filtered somatic mutations identified in the whole-genome sequence experiments in patient 1. The middle panel shows the 96 trinucleotide mutation profiles corresponding to 10 clusters on the left. We manually merged clusters with the same evolutionary trajectory (left panel) and similar mutational profiles together. The clusters represent a (pre)leukemic ancestral clone (yellow), falling (blue) and two rising clones (green and red) at first relapse, and rising clones at second relapse (pink). Finally, we cleaned the merged clusters of outlier mutations and divided the cleaned clusters to create biological relevant clusters, indicated with an asterisk (*), that were subjected to mutational signature analysis.

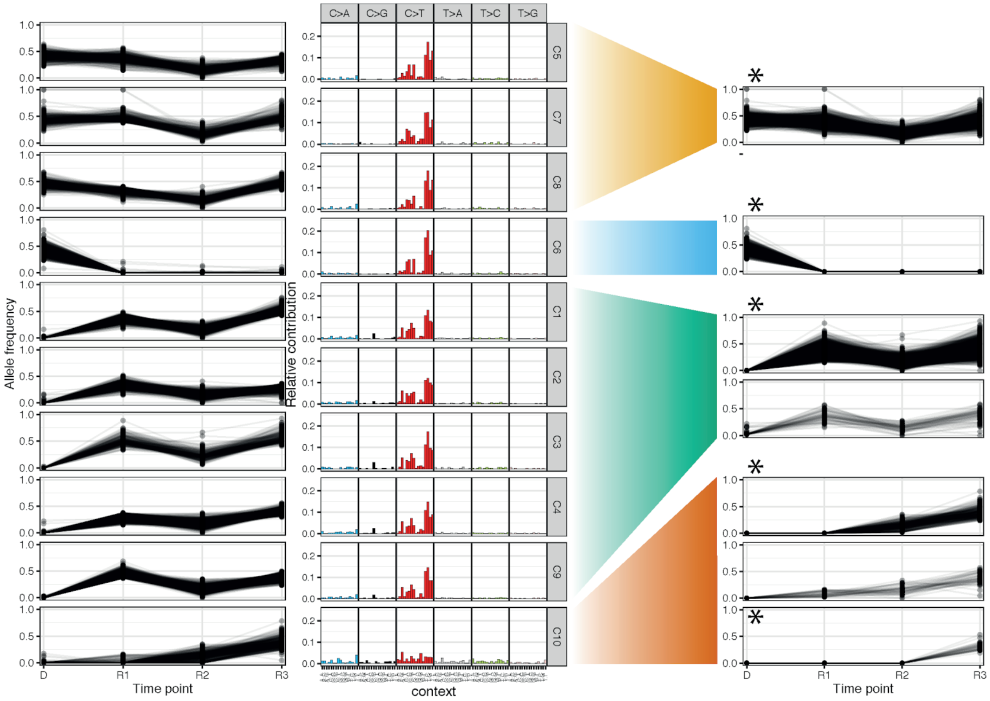


Figure S6.2. Mutation clustering in patient 2. The left panel shows the 10 clusters (*k*-means clustering; *k* = 10) of high-confidence filtered somatic mutations identified in the whole-genome sequence experiments in patient 2. The middle panel shows the 96 trinucleotide mutation profiles corresponding to 10 clusters on the left. We manually merged clusters with the same evolutionary trajectory (left panel) and similar mutational profiles together. The clusters represent a (pre)leukemic ancestral clone (yellow), falling (blue) and rising clone (green) at relapse 1, and rising clones at relapse 2 (red; top cluster) and relapse 3 (red; bottom cluster). Finally, we cleaned the merged clusters of outlier mutations and divided the cleaned clusters to create biological relevant clusters, indicated with an asterisk (*), that were subjected to mutational signature analysis.

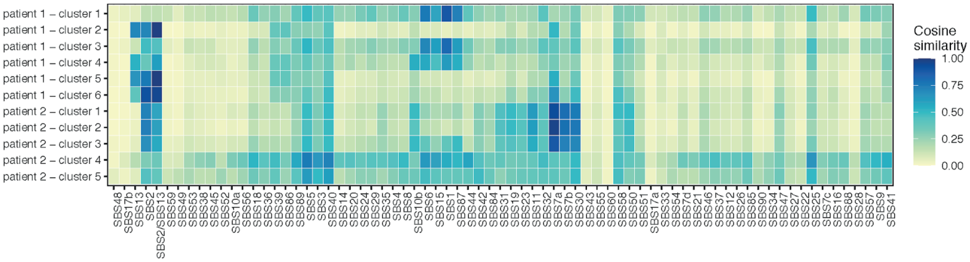


Figure S6.3. Cosine similarity heatmap of the mutational profiles. Heatmap showing the cosine similarity of the profiles of the 6 clusters extracted from patient 1 and the 5 clusters extracted from patient 2 versus the 72 single-base signatures present in the version 3.1 of the COSMIC database.

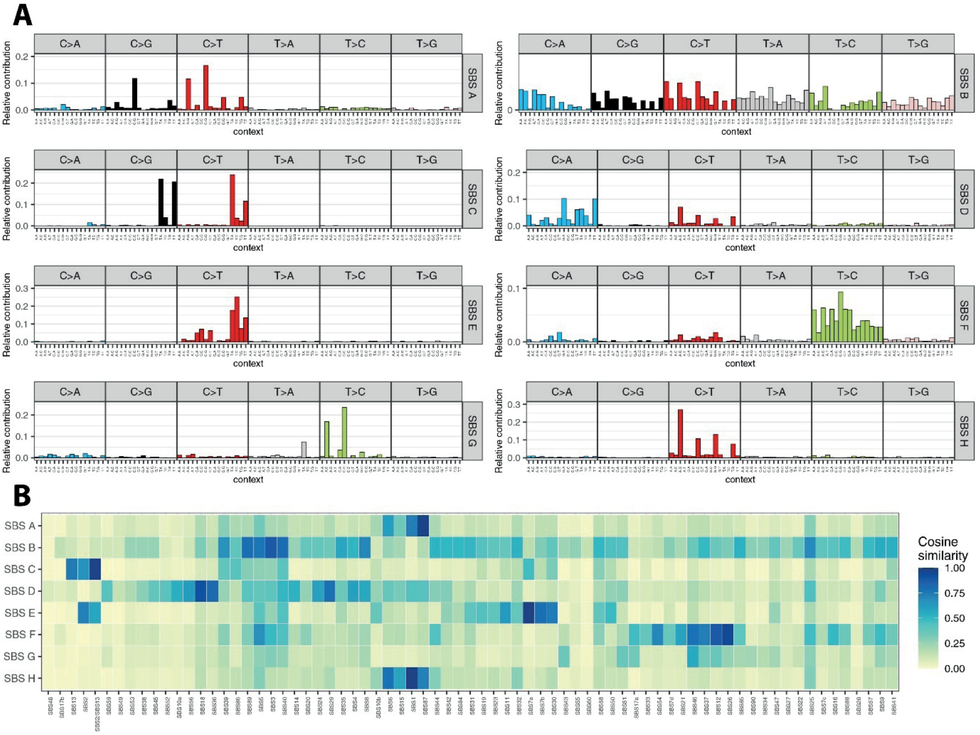


Figure S6.4. *De novo* mutational signature extraction. **(A)** *De novo* mutational signature extraction yielded 8 single-base substitution signatures that resemble known COSMIC mutational signatures. The four extracted signatures that were found to contribute significantly to the two patients in this study were SBS A (cosine similarity of 0.97 to COSMIC signature SBS87), SBS C (cosine similarity of 0.98 to the merged COSMC signatures SBS2 and SBS13), SBS E (cosine similarity of 0.98 to COSMIC signature SBS7a, and SBS H (cosine similarity of 0.98 to COSMIC signature SBS1).The complete overview of the cosine similarity comparison to the COSMIC signatures can be found in **Tables S6.7**, **S6.8** and **S6.9**. **(B)** Cosine similarity heatmap of all eight extracted signatures to the known COSMIC signatures.

FUNDING

This work was supported by grants from Stichting Kinderen Kankervrij and Dutch Cancer Society (KWF-12482 to RPK).

DATA AVAILABILITY STATEMENT

The data presented in this study have been deposited in the European Genome-phenome Archive (https://ega-archive.org); accession number EGAS00001005001.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

CONTRIBUTIONS

RPK and ŽA conceived the study. RPK, ŽA and SHL designed the study. ŽA, and CGvdH performed experiments and analyzed the data. SHL performed bioinformatic analyses. ŽA, SHL and RPK wrote the manuscript and created the figures and tables. ES and PMH provided samples and clinical information. All authors critically reviewed the manuscript and approved the final submitted manuscript.

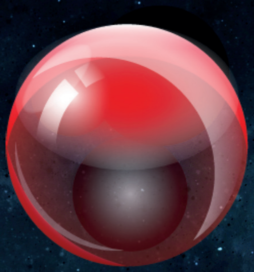
REFERENCES

1. Barr RD, Ferrari A, Ries L, Whelan J, Bleyer WA. Cancer in adolescents and young adults: a narrative review of the current status and a view of the future. JAMA Pediatr. 2016;170(5):495-501.
2. Katanoda K, Shibata A, Matsuda T, Hori M, Nakata K, Narita Y, et al. Childhood, adolescent and young adult cancer incidence in Japan in 2009-2011. Jpn J Clin Oncol. 2017;47(8):762-771.
3. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. CA Cancer J Clin. 2014;64(2):83-103.
4. Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al. International incidence of childhood cancer, 2001-10: a population-based registry study. Lancet Oncol. 2017;18(6):719-731.
5. Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the Children's Oncology Group. J Clin Oncol. 2012;30(14):1663-1669.
6. Pieters R, Groot-Kruseman Hd, Velden VVd, Fiocco M, Berg Hvd, Bont Ed, et al. Successful therapy reduction and intensification for childhood acute lymphoblastic leukemia based on minimal residual disease monitoring: study ALL10 from the Dutch Childhood Oncology Group. J Clin Oncol. 2016;34(22):2591-2601.
7. Ma X, Edmonson M, Yergeau D, Muzny DM, Hampton OA, Rusch M, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun. 2015;6(1):6604.
8. Waanders E, Gu Z, Dobson SM, Antić Ž, Crawford JC, Ma X, et al. Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. Blood Cancer Discov. 2020;1(1):96-111.
9. Antić Ž, Yu J, Van Reijmersdal SV, Van Dijk A, Dekker L, Segerink WH, et al. Multiclonal complexity of pediatric acute lymphoblastic leukemia and the prognostic relevance of subclonal mutations. Haematologica. 2021;106(12):3046-3055.
10. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature. 2011;469(7330):356-361.
11. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature. 2018;555(7696):371-376.
12. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. Nature. 2018;555(7696):321-327.
13. Malinowska-Ozdowy K, Frech C, Schönegger A, Eckert C, Cazzaniga G, Stanulla M, et al. KRAS and CREBBP mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia. Leukemia. 2015;29(8):1656-1667.
14. Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011;471(7337):235-239.
15. Irving J, Matheson E, Minto L, Blair H, Case M, Halsey C, et al. Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. Blood. 2014;124(23):3420-3430.

16. Tzoneva G, Perez-Garcia A, Carpenter Z, Khiabanian H, Tosello V, Allegretta M, et al. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. Nat Med. 2013;19(3):368-371.
17. Jerchel IS, Hoogkamer AQ, Ariës IM, Steeghs EMP, Boer JM, Besselink NJM, et al. RAS pathway mutations as a predictive biomarker for treatment adaptation in pediatric B-cell precursor acute lymphoblastic leukemia. Leukemia. 2018;32(4):931-940.
18. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009;360(5):470-480.
19. Kuiper RP, Waanders E, van der Velden VH, van Reijmersdal SV, Venkatachalam R, Scheijen B, et al. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. Leukemia. 2010;24(7):1258-1264.
20. Jaffe JD, Wang Y, Chan HM, Zhang J, Huether R, Kryukov GV, et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nat Genet. 2013;45(11):1386-1391.
21. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101.
22. Spinella JF, Richer C, Cassart P, Ouimet M, Healy J, Sinnett D. Mutational dynamics of early and late relapsed childhood ALL: rapid clonal expansion and long-term dormancy. Blood Adv. 2018;2(3):177-188.
23. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet. 2014;15(9):585-598.
24. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-421.
25. Campbell PJ, Getz G, Korbel JO, Stuart JM, Jennings JL, Stein LD, et al. Pan-cancer analysis of whole genomes. Nature. 2020;578(7793):82-93.
26. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. Cell. 2019;176(6):1282-1294.e1220.
27. Li B, Brady SW, Ma X, Shen S, Zhang Y, Li Y, et al. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. Blood. 2020;135(1):41-55.
28. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754-1760.
29. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32(8):1220-1222.
30. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl variant effect predictor. Genome Biol. 2016;17(1):122.
31. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-443.
32. Francioli LC, Menelaou A, Pulit SL, van Dijk F, Palamara PF, Elbers CC, et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet. 2014;46(8):818-825.
33. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10(1):33.

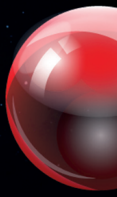
34. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010;11(1):367.

35. Dobson SM, García-Prat L, Vanner RJ, Wintersinger J, Waanders E, Gu Z, et al. Relapse-fated latent diagnosis subclones in acute B lineage leukemia are drug tolerant and possess distinct metabolic programs. Cancer Discov. 2020;10(4):568-587.



CHAPTER 7

GENERAL DISCUSSION



Acute lymphoblastic leukemia (ALL) represents the most common childhood malignancy. Despite the fact that cure rates approach 90% in developed countries[1-4], relapse and the associated treatment resistance still represent major clinical and scientific challenges. Among cases that relapse, outcome is much poorer[5] and prevention of relapse may aid further improvement in the outcome. Although numerous studies have vastly improved insight into the genomic landscape of ALL, overall survival rates have not improved dramatically in the contemporary clinical trials[2]. Furthermore, it is still unclear how different factors contribute to treatment failure. In this thesis we aimed at performing comprehensive characterization of relapsed pediatric ALL in order to investigate the mechanisms driving relapse and treatment failure.

In general, there are several possible causes for patients to develop a relapse (**Figure 7.1**). For example, relapses may be driven by undertreatment of leukemia due to treatment modifications, omissions or interruptions caused by drug toxicity and infections. In other cases, relapses may be driven by outgrowth of the clones residing in distinct body compartments (niches), in which they are protected from chemotherapy, e.g., eye and testis[6,7]. Furthermore, dormancy of leukemic cells may decrease their vulnerability against commonly used chemotherapeutics. These dormant, non-proliferative, leukemic cells can exit their dormant state at the end of treatment leading to development of relapse[8,9]. Similarly, the presence of genetic alterations associated with treatment resistance in leukemic clones may enable their survival, outgrowth and relapse development[10-19]. These genetic alterations may be present already at the time of diagnosis, often in a minor clone and undetectable by routine diagnostic methods, or can be newly acquired during the treatment[10,13,17,18], and sometimes even driven by the treatment[20,21]. Despite differences in the etiology of relapse, lack of treatment pressure against a subpopulation of leukemic cells is the main force driving clonal selection and clonal outgrowth. Therefore, better understanding of how these different mechanisms contribute to relapse development is essential in order to prevent relapse.

In this thesis we have examined the genetic basis of relapse, and how different genomic alterations may contribute to the development of treatment resistance, clonal outgrowth and relapse. Furthermore, we investigated mutagenic processes that drive acquisition of these genetic alterations and their activity during disease progression. Finally, we examined the predictive role of relapse-associated genomic alterations for relapse development in the context of contemporary Dutch ALL treatment protocols.

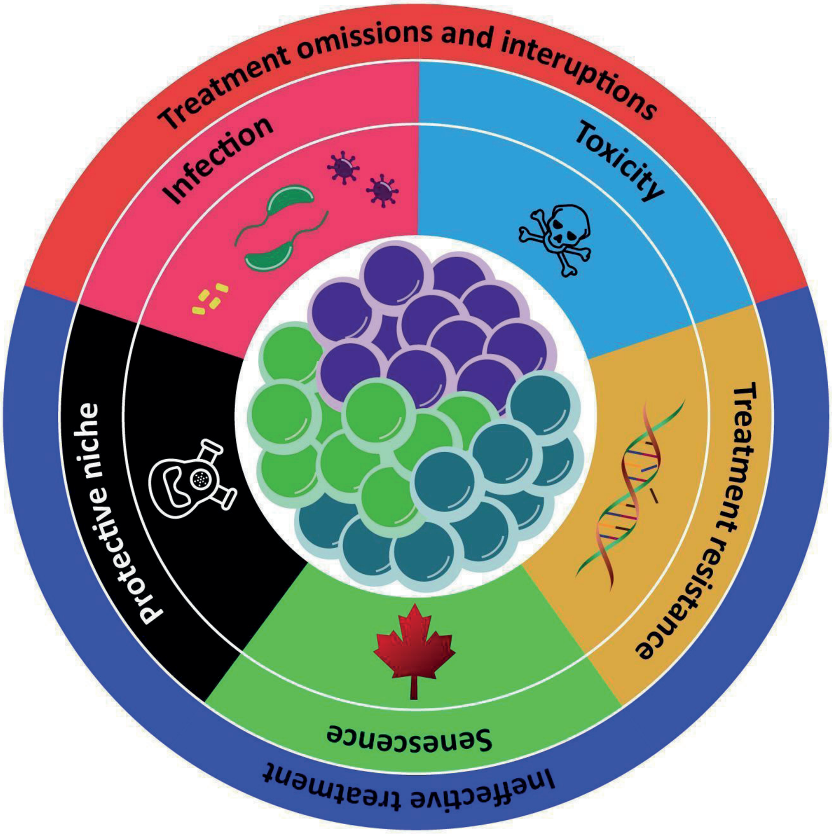

Treatment omissions and interuptions
Infection
Toxicity
Protective niche
Treatment resistance
Senescence
Ineffective treatment

Figure 7.1. Schematic representation of the mechanisms driving relapse in pediatric ALL. Relapses may occur due to ineffective treatment (blue outer circle), which can be driven by drug resistance (orange), cellular senescence (green) or inability of a drug to reach therapeutic concentrations in a niche in which leukemic cells reside (black). Remaining relapses occur due to treatment omissions and interruptions (red outer circle), which can be caused by drug toxicity (blue) and infections (magenta).

Clonal heterogeneity and evolution of ALL

In his landmark article from 1976, Peter Nowell proposed a model explaining the development of a malignant tumor in a multi-step process from a single, initially normal cell, very similar to Darwin's model on the evolution of species[22]. In his model, a single normal cell acquires an alteration that provides an evolutionary advantage over other normal cells. This advantage can be metabolic or proliferative, but on its own it is not enough to cause malignant transformation of the cell. During a variable period of time this cell, or its progeny, can acquire additional alterations with variable effects on their fitness. Cells with alterations that give evolutionary advantage continue to proliferate and acquire new alterations until a single precancerous cell becomes malignantly transformed. Up to this point, tumor initiation and evolution from a single ancestral cell (monoclonal tumors) is linear. Although Nowell left open the possibility that multiple cells enter initiation processes, each of them would need to acquire alterations necessary to reach complete malignant transformation before emergence of tumors with different ancestry (polyclonal tumors). After malignant transformation is complete, the transformed cell and its progeny gain the ability to proliferate uncontrollably, while at the same time acquiring new genomic alterations. At this point, cancer cells harboring genomic alterations with competitive advantage over the first malignantly transformed progenitor may arise, thus resulting in branching of propagating tumors in multiple clones[22]. Remarkably, Nowell's model of cancer evolution has survived decades of cancer research, and even the large number of comprehensive genomics studies, including the ones presented in this thesis, have confirmed its validity. For instance, deep sequencing studies revealed a plethora of mutations unique for specific time points, with variable allele frequencies, consistent with Nowell's model of tumor cell branching into multiple clones.

Studies in twins with pediatric ALL have confirmed that the presence of strong drivers, such as a *ETV6-RUNX1* translocation, are essential for leukemia development, but these lesions on their own are not sufficient for malignant transformation of the lymphocytes[23,24]. Once new mutations occur in a single cell, they may lead to relative outgrowth of the new clone. Similarly, deep sequencing studies in adults have also identified many crucial, leukemia-driving lesions in non-leukemic cells in the hematopoietic stem cell (HSC) compartment. Interestingly, non-malignant HSCs harboring these genetic lesions showed higher proliferative potential over HSCs that do not harbor these lesions, leading to formation of healthy HSC subclones, a phenomenon known as clonal hematopoiesis[25,26]. Although these alterations are not enough for disease development, their gradual, stepwise accumulation in a non-transformed HSC may eventually lead to acquisition of alterations that induce HSC transformation and development of leukemia, similar to what was observed in solid malignancies, and resembling a linear model of tumor evolution[27]. In addition, as we demonstrated in **chapter 5**, in a small subset of cases the presence of leukemia-driving alterations in preleukemic HSCs can lead to branching of a new clone and development of second leukemia, seemingly as a relapse event. This finding suggests that branching of (pre)leukemic cells can occur at any point during disease development, even before transformation of HSC into malignant cells.

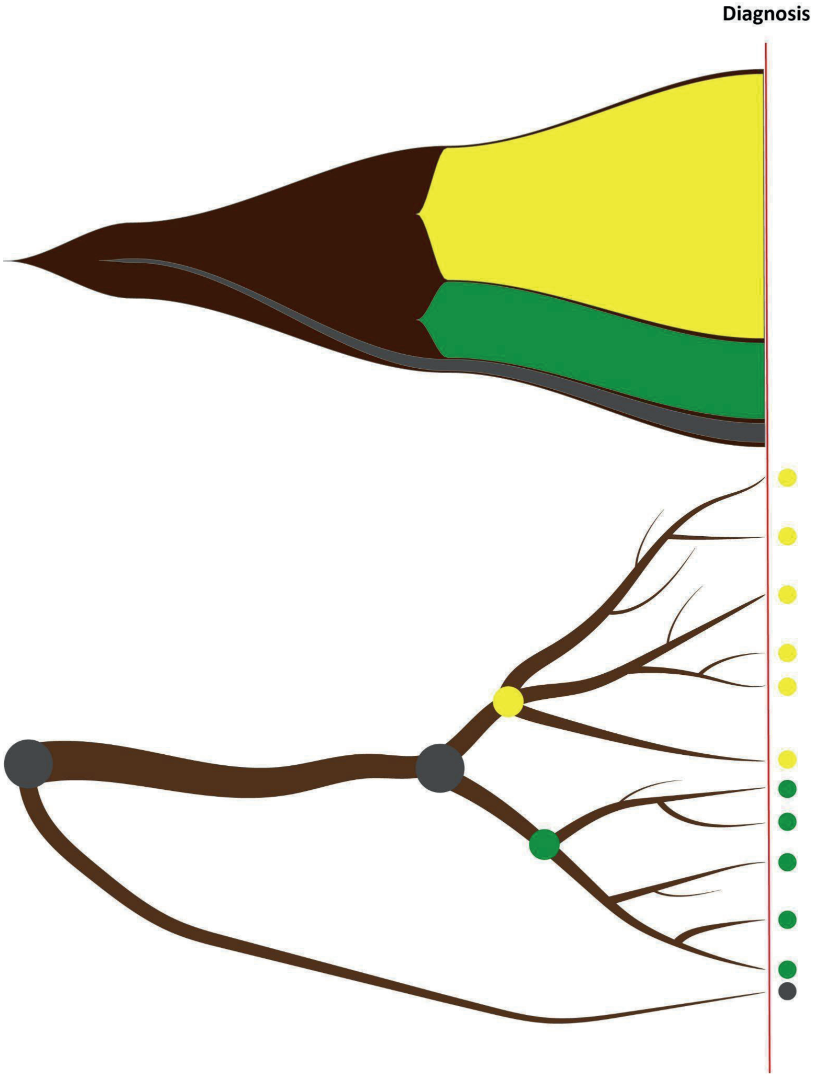


Figure 7.2. Schematic representation of leukemia development and the emergence of clonal heterogeneity. Through acquisition of genetic alterations, leukemia branches into smaller subpopulations and gives rise to new subclones. The majority of the mutations will stay below the detection limit of current sequencing assays, making the observed heterogeneity in ALL much less complex. Higher sequencing depth may enable more of these subclones to reach the detection threshold of the assay. However, only real disease-driving alterations will enable progeny of this new clone to grow out and become dominant in the tumor (depicted in yellow, green and gray).

Through spontaneous mutations that occur with every cell division, as well as by the influence of various mutagenic mechanisms, new clones arise that lead to the development of heterogeneity in ALL (**Figure 7.2**). A minor subset of these mutations increases proliferative potential and cellular fitness, leading to relative expansion of the new clones. Despite the fact that each leukemic cell in diagnosis has unique mutations, making it a new subclone, only clones with the potential to rise above the detection threshold of currently used techniques will be detected. Therefore, the remarkable heterogeneity of ALL observed in current studies represents only the tip of the iceberg of true clonal diversity, with many unique subclones still remaining hidden.

Clonal dynamics in relapsed ALL

During development of ALL, clonal evolution is driven by natural selection and competition between healthy HSC and heterogeneous populations of (pre)leukemic clones. Therefore, at diagnosis, the genetic makeup of the most prominent clones will be represented by those alterations that increase the natural fitness of leukemic cells. Initiation of treatment represents a game-changer in the evolution of ALL, since treatment-challenged leukemic clones now need to compete in different conditions compared to the ones in which they initially evolved. Selective pressure introduced by the treatment forces leukemic cells to adapt to the new conditions in order to survive, break through the constraints introduced by the treatment, and give rise to a relapse. Using evolutionary modeling, we demonstrated that more than half of the relapses evolved from a minor subclone at diagnosis (**chapter 5**), which suggests that indeed selection has taken place. In the remaining cases, relapse-driving clones originated from the major clone observed at diagnosis (27%), or had a polyclonal origin (18%). In addition, in **chapter 4**, where we studied very early, on-treatment relapses, in almost all of these cases relapses originated from a minor subclone in diagnosis, suggesting that these minor subclones may harbor lesions driving relapse. The dynamic clonal evolution we observed in these studies, with clones rising and falling from diagnosis to relapse in all cases that relapsed during treatment, indicates that these cases undergo strong clonal selection due to selective pressure of the treatment, and supports a branched model of clonal evolution. Furthermore, based on the clonal dynamics observed between major clones in diagnosis and release, we were able to discern four different patterns leading to the development and outgrowth of relapse clones. First, we observed relapses from the major clone, which are rare and can be driven by aberrations associated with very poor prognosis, like *TCF3-HLF*-positive leukemias (**Figure 7.3A**). Most other scenarios do involve development of relapse as a consequence of continuous clonal evolution, thus from a subclone in diagnosis. For example, a second model was observed in cases where a relapse-driving alteration is present in a subclone already at diagnosis, and where a treatment-driven selective pressure may lead to outgrowth of this subclone early during remission (**Figure 7.3B**). Examples of alterations associated with this type of evolution model are those in *IKZF1*, which can be found preserved from a major or minor clone, or those in *WHSC1* and *CREBBP*, which typically arise from a minor clone at diagnosis. Third, compared to a model where relapse-prone clones are present already in diagnosis, intensive treatment of ALL may lead to acquisition of alterations that drive relapse-development, resulting in formation of a new relapse clone at a later time point during remission (**Figure 7.3C**). *NT5C2* and *TP53* mutations are examples that typically (but not exclusively) follow this model of evolution. Finally, a premalignant clone, or one of its early branches, may still survive the treatment and give rise to a new leukemia presenting as a very late relapse (**Figure 7.3D**). In **chapter 5** we have identified several examples of this latter model, including two cases with *BCR-ABL1* ALL that did not share mutations between diagnosis and relapse, raising the suspicion that these patients actually developed a second ALL rather than a relapse. Subsequent whole genome sequencing (WGS) for one of these cases, however, revealed identical breakpoints of the *BCR-ABL1* fusion between diagnosis and relapse, as well as 65 shared somatic variants, demonstrating that this was indeed relapse from a premalignant common ancestral clone.

The presence of multiple, genetically heterogeneous clones creates a pool of diverse cell populations that may drive mutual competition, natural selection and clonal selection induced by the treatment. After treatment has been initiated, the evolutionary bottleneck created by the therapy becomes the major factor influencing the evolution of tumors and mutational landscape of relapsed ALL. Although previous studies have shown that alterations in these genes confer treatment resistance in ALL *in vivo* and *in vitro*, only alterations affecting *IKZF1* have been identified as a strong predictor for relapse development in different treatment protocols[11,12,16,28-31], and have been demonstrated to cause resistance to several drugs *in vitro* and *in vivo*[32-35]. Lack of such a strong association for other genes, and often discordant findings between different clinical studies, suggests that despite the clear connection with treatment resistance *in vivo* and *in vitro*, the presence of additional factors may be needed for the development of relapse. Furthermore, despite growing knowledge about genetic heterogeneity, it remains unclear how functional heterogeneity, e.g., different responses to medicines in genetically similar leukemic clones, may influence relapse development. Therefore, examining the functional heterogeneity of these tumors and the evolution of clones which harbor alterations in genes frequently mutated in relapse, during treatment itself, may further improve our understanding of factors contributing to relapse development.

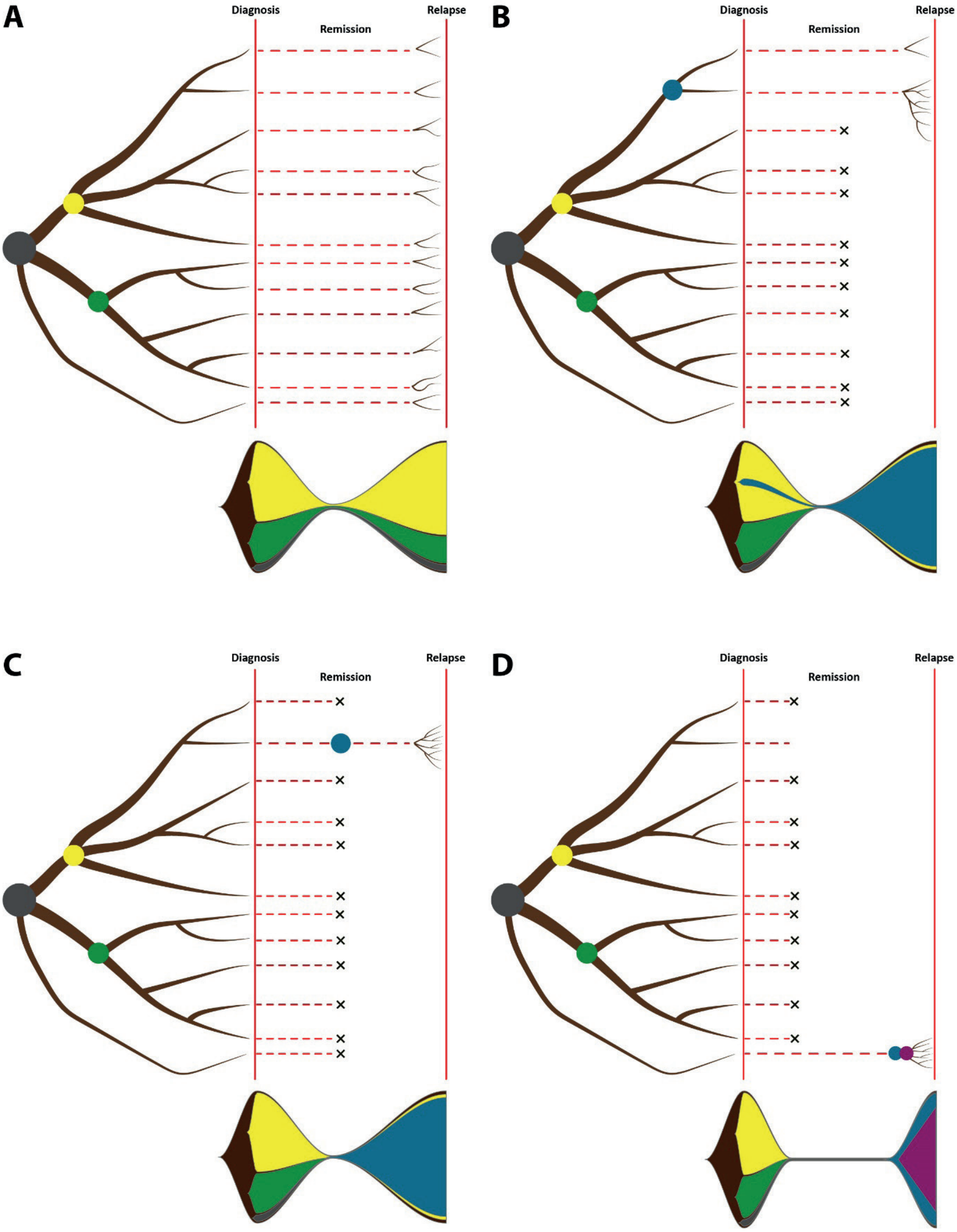

A
Diagnosis
Remission
Relapse
B
Diagnosis
Remission
Relapse
C
Diagnosis
Remission
Relapse
D
Diagnosis
Remission
Relapse

Figure 7.3. Models of clonal evolution from diagnosis to relapse. In children with ALL, harboring genomic lesions associated with a poor prognosis, relapse can develop from all branches of the major clone in diagnosis, and the subclonal heterogeneity present at diagnosis may be preserved at relapse **(A)**. In other cases, relapse evolves from a subclone already present in diagnosis **(B)** or one that may evolve during treatment **(C)**. Relapse can also evolve from a premalignant clone that branched off early during leukemia development and survived the therapy in which the malignant cells were eradicated **(D)**. Dots with different colors represent clone-founding driver mutations.

Although many of these genes are drivers associated with relapse, it is unclear how big their impact is to drive relapse by themselves. As such, genes involved in relapse formation may not always be prognostically relevant.

Methods to study clonal evolution and heterogeneity in ALL

Whole exome and genome sequencing (WES, WGS) approaches, which we used in **chapters 4**, **5** and **6**, dramatically improved the resolution in which the process of clonal evolution can be studied. However, although the costs of sequencing are dropping, a choice has to be made in these studies between sample size, which increases the power of a study, and sequencing depth, where the sensitivity to detect subclones is higher. Because of the heterogeneity that is observed between individual patients with ALL, inclusion of more samples usually has the preference. As a result, limited sequencing depth can cause subclonal mutations that are below the detection limit to be missed, which hampers the ability to fully capture the diversity of clonal composition in ALL. Although increasing sequencing depth may aid the discovery of mutations occurring in smaller subclones, it may also increase the number of sequencing artefacts. These sequencing artefacts, which usually present with low alternate read counts, may be hard to distinguish from real subclonal mutations, leading to unreliable results. We have tried to tackle this problem by performing targeted deep-sequencing of all detected mutations in samples from a single individual, using an independent targeted next generation sequencing (NGS) approach (**chapters 4** and **5**). In this way, true positive subclonal mutations can be effectively identified, but also a more accurate estimate of the subclonal allele frequency can be obtained.

Another strategy that we used to identify relevant clones and subclonal mutations is to study their evolutionary trajectories over time. When the subclone detected at diagnosis expands and becomes dominant at relapse, the effect of mutations in this subclone on relapse development becomes noticeable. With every subsequent time point, chances for divergence of evolutionary trajectories of individual clones are increasing, therefore the more time points available for sequencing, the better evolutionary modeling works in individual samples. Although the majority of mutations in these (sub)clones are non-pathogenic, their identification helps in defining distinctive clones and clonal dynamics. This approach was proven to be particularly beneficial in cases where genomic alterations were found to coexist in independent clones on a similar frequency in individual samples (**chapters 4**, **5** and **6**). In situations when samples taken at subsequent time points are not available, the use of xenograft models can aid in further uncovering clonal composition and clonal heterogeneity of ALL. Besides a characterization of the genetic landscape of individual clones, xenograft models can provide information about their preferences for different body compartments and response to treatment[36-39]. Therefore xenograft models provide a powerful tool to identify different aspects of clonal heterogeneity, including the presence of relapse initiating subclones at diagnosis[36].

Finally, the emerging single cell sequencing (SCS) approaches offer unprecedented resolution to study the mutational landscape of individual cells and heterogeneity of total tumor. These SCS-based techniques offer the possibility to study even the rarest population of cells, and the unique genetic and epigenetic contexts in which relapse-prone clones evolve[40-48]. Furthermore, SCS techniques give unprecedented resolution for minimal residual disease (MRD) monitoring while the patient is undergoing treatment. However, SCS-based techniques are still under development, which, in addition to the high costs and the need for viable cells, limits their broad application.

Detection of subclonal alterations and their prognostic relevance

The observation that relapsed ALL frequently originates from a minor clone at diagnosis suggests that these minor subclones may have had a selective advantage during therapy. If these subclones harbor alterations in specific genes associated with treatment resistance and relapse, the early detection of these alterations could have prognostic relevance. However, to reveal a possible correlation with relapse, an unbiased approach is needed to detect all subclonal mutations in such genes in a large cohort of diagnosis samples. The studies dealing with clinical relevance of subclonal genetic alterations in different cancer types, including ALL, are sparse. A good method to reliably detect subclonal alterations is digital droplet PCR (ddPCR), which offers much higher sensitivity compared to conventional NGS techniques. However, ddPCR is mutation specific, which makes it unsuitable to perform high-throughput detection of subclonal mutations in an entire gene. Furthermore, single cell technologies, although promising for studying subclonality and clonal heterogeneity in ALL, are not suitable for high-throughput detection of subclonal mutations.

We tackled this problem in **chapter 2**, where we demonstrated that the use of random molecular tagging and deep sequencing can be used to reliably detect low-level mosaic mutations in a large cohort of patients in a predefined set of genes. The scalability of this approach offers the possibility to add at any point more molecular targets and perform deep sequencing while correcting for possible amplification biases. We subsequently used this approach to investigate the clinical relevance of subclonal alterations in relapse-associated genes by screening a large cohort of patients treated according to two Dutch treatment protocols: DCOG-ALL9 and DCOG-ALL10 (**chapter 3**). We selected the genes *KRAS*, *NRAS*, *PTPN11*, *CREBBP*, *WHSC1*, *TP53*, and *NT5C2*, for which it was previously shown that they are frequently altered in relapsed ALL. Subclonal mutations were found to be common, particularly in RAS pathway genes (*KRAS*, *NRAS*, *PTPN11*). Despite their abundance, alterations in examined genes were not predictive of relapse development, neither when present in the major, nor in the minor clones. This finding suggests that the presence of alterations in relapse-associated genes may not be sufficient to give rise to a relapse, and may require that leukemic cells acquire additional genetic alterations. An example is presented in **chapter 4**, where we described a patient with a subclonal mutation in *TP53*. During treatment, this subclone lost the p-arm of chromosome 17, resulting in a complete loss of the wild type *TP53* gene, which subsequently expanded into a new major clone in relapse. Therefore, alterations in relapse-associated genes may give slight proliferative advantage at diagnosis, resulting in detectability of this subclone, due to relative outgrowth of the leukemic cells harboring these alterations. However, these alterations also may predispose subclones to develop into more aggressive phenotypes after acquiring additional alterations, which is why they are frequently found in major clones at relapse.

Deletions affecting *IKZF1* were previously shown to be a strong predictor of relapse in different treatment protocols[11,12,16,28-31,49]. Therefore, subclonal *IKZF1* deletion may show the same strong association with relapse development. To examine this hypothesis, we included *IKZF1* exon 4-7 deletions (**chapter 3**), which represent around 25% of all *IKZF1* deletions[50]. These deletions harbor tightly clustered breakpoints[51], thereby enabling their detection using breakpoint-spanning semi-quantitative PCR. Surprisingly, although we observed a strong association of *IKZF1* exon 4-7 deletions in major clones with relapse, this was not the case with subclonal deletions. Furthermore, failure to find back any of these *IKZF1* 4-7 deletions in matched relapse samples, even on a subclonal level, suggests that they may have been eradicated during treatment, which further underscores their lack of association with relapse development. The reason behind this striking difference remains unclear. Possibly, the functional impact of full-clonal *IKZF1* deletions, which arise early during leukemia development, is different from those that occur in later stages when the leukemia has already fully expanded. For example, it was suggested that *IKZF1* acts as a metabolic gatekeeper in the developing B-cell by repressing glucose uptake and pentose phosphatase activity in cells harboring oncogenic lesions, leading to energy stress and cell death[52,53]. *IKZF1* haploinsufficiency may release the B-cell from the limitations in energy supply, allowing pathological proliferation of these cells. Subclonal *IKZF1* deletions arise during proliferation, where the cell may have overcome this metabolic restriction in another way (e.g., by a deletion of *PAX5*). On the other hand, the presence of an *IKZF1* deletion in a leukemic cell may increase cell survival under therapy, but the majority of children reaches complete remission, indicating that most *IKZF1*-deleted cells are killed effectively. Therefore, at the cellular level, therapy resistance and relapse development will also have a stochastic component, which may minimize the selective advantage of minor *IKZF1*-deleted subclones. Although in this study we did not assess the prognostic relevance of whole gene and rare intragenic *IKZF1* deletions, a previous study has shown that these deletions have similar unfavorable outcomes as *IKZF1* 4-7 deletions[50]. Whole gene *IKZF1* deletions, as well as deletions encompassing exon 2, which contains the ATG translation start site[54], are considered to cause haploinsufficiency, thus likely having similar functional consequences and prognostic value. In contrast, loss of four zinc finger DNA binding domains in *IKZF1* 4-7 deletions results in expression of a dominant-negative isoform (isoform 6). This dominant-negative isoform lacks DNA binding activity, but retains the ability to interact with wild type IKAROS, thereby effectively altering its function. Studies on *IKZF1* knockout mouse models have demonstrated that while haploinsufficiency leads to the development of B- and T-cell tumors, dominant-negative mutants experience widespread failure of hematopoiesis and shorter latency in tumorigenesis compared to wildtype mice[55]. Furthermore, a previous study has demonstrated different outcomes in *BCR-ABL1*-negative adult ALL patients, depending on whether *IKZF1* deletion resulted in haploinsufficiency or expression of a dominant-negative isoform[56]. Taken together, these findings indicate that *IKZF1* deletions with expression of a dominant-negative isoform, such as the case with exon 4-7 deletions, have different mechanisms in which they can contribute to leukemia development and treatment failure, compared to the whole gene and deletions involving exon 2. Thus, we cannot completely rule out that the other *IKZF1* deletions may have different effects than the ones observed in subclonal *IKZF1* exon 4-7 deletions.

Previous studies, including the ones presented in this thesis, observed that *NT5C2* mutations are almost exclusively found in relapse[17,18,57], and when they do occur at time of diagnosis, they are present in a minor subclone. Previously, it was demonstrated that upfront treatment influences the genetic composition of relapsed ALL through selective pressure of the drugs that represent the backbone of different treatment protocols[58]. This strong influence of individual drugs used in frontline ALL treatment is particularly important for alterations affecting *NT5C2*. NT5C2 is a cytosolic 5' nucleotidase responsible for dephosphorylation of 6-hydroxypurine monophosphates, such as inosine monophosphate (IMP) and guanosine monophosphate (GMP). Furthermore, in patients receiving thiopurines, e.g., 6-mercaptopurine, NT5C2 catalyzes dephosphorylation of their thiolated metabolites, e.g., 6-thioinositol monophosphate and 6-thioguanosine monophosphate[17,59-62]. Dephosphorylated endogenous and thiopurine-originating 6-hydroxypurine products are then exported from the cell, while endogenous 6-hydroxypurine products can also be used as substrates in the degradation or purine salvage pathways[59,60,62]. Under conditions present in untreated leukemia, activating mutations in *NT5C2* result in an increased export of nucleosides from the cell, thus decreasing cell fitness and proliferative potential, which may ultimately lead to collapse of the clone. However, under the pressure of thiopurine treatment, clones harboring these mutations not only survive, but even have a higher proliferative potential compared to clones without these mutations[17,57,59]. This striking effect of activating *NT5C2* mutations during treatment with thiopurines can be explained by the higher substrate affinity of mutated NT5C2 for thiolated 6-hydroxypurine monophosphates[60]. Therefore, mutated NT5C2 increases dephosphorylation and transport of 6-MP metabolites, while their lower substrate affinity for endogenous purine nucleotides effectively increases the pool of available purines for DNA synthesis. Exclusive enrichment of *NT5C2* mutations in relapses (**chapters 4** and **5**), and the lack of *NT5C2* mutations, even subclonal ones, in diagnosis (**chapter 3**) can therefore be explained by conditions favoring expansion of the *NT5C2* mutated clones during treatment, as well as by a limited detection threshold of the NGS assays in diagnosis.

Mutational processes in relapsed ALL

Genomic alterations in ALL have the potential to drive disease development, clonal expansion, treatment resistance and relapse. However, our knowledge about the mechanisms that drive the acquisition of these genomic alterations is still limited. Although mutagenic processes do not directly cause disease development, they may introduce leukemia-driving and treatment resistance-associated lesions, thus leading to the emergence of disease and facilitating adaptation of leukemic cells to treatment. Therefore, better understanding of the mutational processes that are active in ALL may result in improved risk-stratification and even better strategies for treatment.

Previous studies have discovered a diverse range of different mutational mechanisms active in ALL, some of which may cause extensive lesions such as chromothripsis in dicentric rob(15;21)c-associated iAMP21[63,64], a breakage-fusion-bridge mechanism often followed by chromothripsis in sporadic iAMP21[64,65], or focal deletions due to cryptic activity of the RAG family recombinases[66-68]. In this thesis we focused on the processes driving development of single-base substitutions in relapsed ALL. When studied in a trinucleotide context, groups of single-base substitutions (SBS) can be linked to underlying mutational mechanisms, which represent a fingerprint known as a mutational signature[69-71]. Until now, 72 SBS signatures have been described in the Catalogue of Somatic Mutations in Cancer (COSMIC), of which 41 have a known etiology.

Mutational signatures can be *de novo* extracted from a well-curated set of somatic mutations using the dimensionality reduction method of non-negative matrix factorization (NMF)[72,73]. Although powerful, this method is limited by the number of samples with distinct mutational patterns. One solution to overcome limitations related to the sample size is to extend the dataset by procuring data derived from samples of similar origin, as we applied in **chapter 5**. Another approach in mutational signature analysis is based on the fitting of mutational profiles of individual samples to known COSMIC SBS signatures in order to quantify their contributions to the overall mutational spectrum[72]. This approach is independent of other samples included in the same analysis, which makes it suitable in situations where the number of samples is limited. However, the biggest drawback of this method remains the inability to identify novel mutational mechanisms due to its dependency from the previously defined SBS signatures.

The initiation and duration of mutation accumulation appears to be different between the various mutational processes that are found to be active in ALL. Whereas some processes are regularly found to be active already at diagnosis, others arise later as a direct consequence of therapy or because of an acquired pathogenic mutation (**Figure 7.4**).

In **chapter 5**, we analyzed WES data of 92 pediatric ALL cases obtained from diagnosis and relapse. At diagnosis we observed a high mutational load in a limited set of samples (3%), but the mutational load increased with every subsequent relapse, suggesting that hypermutation is associated with cell proliferation and/or the impact of therapy. Therefore, in **chapters 4**, **5**, and **6** we examined the phenomenon of hypermutation and analyzed mutational mechanisms driving mutagenesis. Our analysis unraveled that the hypermutation may occur at each time point during leukemia development and can be driven by intrinsic processes such as AID/APOBEC mutagenesis, or acquired mechanisms, like mismatch repair deficiency (MMR). Alternatively, it can be caused by treatment, such as the case with mutational signature SBS87, which is associated with thiopurine treatment.

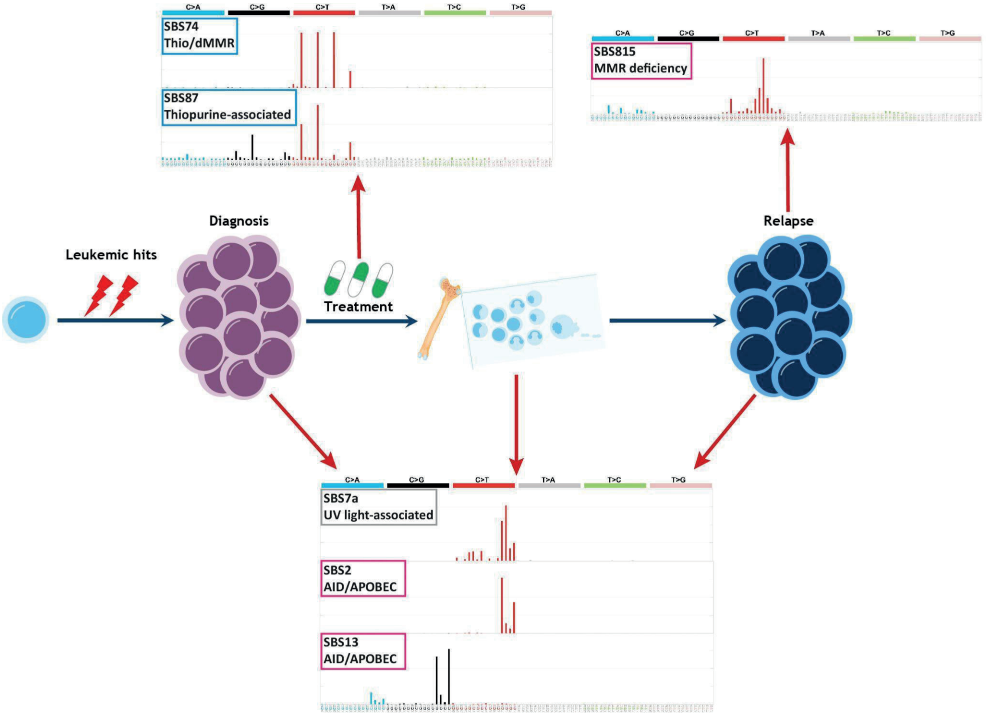

C>A
C>G
C>T
T>A
T>C
T>G
SBS74
Thio/dMMR
SBS87
Thiopurine-associated
SBS815
MMR deficiency
Diagnosis
Relapse
Leukemic hits
Treatment
SBS7a
UV light-associated
SBS2
AID/APOBEC
SBS13
AID/APOBEC

Figure 7.4. Schematic representation depicting stages of ALL development and evolution during which specific mutational processes may become active. Intrinsic mutational processes are outlined in magenta, external are in blue, while the processes without apparent etiology in ALL are in grey. Although signs of activity of distinct mutational processes may be found at each stage during ALL evolution, certain processes are more likely to be more active at specific time points. For example, mutations caused by aberrant AID/APOBEC activity and UV light exposure-associated DNA damage are often found at diagnosis, while MMR deficiency and thiopurine-associated mutations are often found in relapses.

In this thesis we identified the presence of several known mutational signatures, including SBS15, associated with the mismatch repair deficiency (MMR), SBS2 and SBS13, both associated with aberrant AID/APOBEC mutagenesis, and, surprisingly, SBS7a associated with UV light exposure. Although we were not able to unravel the exact etiology of SBS7a, UV light exposure in ALL seems unlikely, and further analysis may aid our understanding of this mutational process. For example, defects of one of the DNA repair pathways may make these leukemic cells more susceptible for UV light-like DNA damage. Furthermore, we and others have demonstrated that SBS7a was often found at diagnosis, but never in the second relapse, suggesting that the impact of this mutational process is temporary and may even be limited only to diagnosis[20,74,75]. In this scenario, the acquired SBS7a mutations at first relapse may in fact represent rising clones that were present at a subclonal level already at diagnosis. We also identified a novel signature strongly resembling SBS signature 1, a clock-like signature previously associated with spontaneous deamination of methylated cytosines in TpCpG context. The high load of newly acquired mutations in relapse samples of these patients, comparable to the ones observed in cases with MMR deficiency, already suggests that this process might have been accelerated by an acquired imbalance between damage and repair. In addition, in independent cases we observed a wide spread of mutation allele frequencies (MAF) that could be assigned to the SBS1-like mutational signature. This finding was particularly prominent in case SJETV010, where most of the SBS1-like mutations occurred on a low subclonal level (MAF < 0.5%) in the first relapse. In the second relapse these mutations were much wider spread, indicating that the SBS1-like mutations arose in an ongoing mutational process, which was initiated in the subclone at first relapse.

To further unravel the basis of the SBS1-like mutational signature, we performed WGS and RNA sequencing in a subset of cases that showed a high contribution of this mutational process, and in addition used a previously published dataset of colon organoids for which it was shown to have mutations driven by COSMIC SBS signature 1. In contrast to previously published data on SBS1, our analysis revealed strand asymmetry for the transcribed regions in cases with the SBS1-like signature, complemented by high C>T mutational densities in CpGs. Furthermore, the mutational densities increased and strand asymmetry widened with higher gene expression levels in respective samples, a finding that could not be recapitulated in colon organoid samples with COSMIC SBS1 mutations. High prevalence of C>T substitutions in CpGs in gene bodies and gene expression-coupled strand asymmetry suggests that this mutational process may be caused by deficient transcription-coupled repair mechanisms, or increased damage on a single stranded, nontranscribed DNA strand. Indeed, a recent study involving children with relapsed ALL from the United States and China recapitulated our findings on SBS1-like mutational signatures and further unraveled the connection of this mutational signature with thiopurine treatment in leukemia with MMR deficiency (here referred to as the thio-dMMR signature)[21]. SBS1-like/thio-dMMR signature is very similar to the COSMIC SBS74 mutational signature, for which association with MMR deficiency was already suggested[70] (**Figure 7.4**). Moreover, incorporation of an active thiopurine metabolite, 6-thioguanine, causes stalling of RNA polymerase and activations of transcription-coupled nucleotide excision repair, which can explain our observation of high mutational densities in the transcribed DNA strand of expressed genes[76]. Nevertheless, it remains unclear why thiopurine treatment in the context of MMR deficiency leads to different mutational signatures compared to previously observed SBS87. Although SBS74 was very prominent in relapsed ALL, we did not observe this mutational signature in the leukemia of patients that experienced relapse after treatment with one of the DCOG treatment protocols. This leaves the possibility that SBS74 is not caused only by thiopurine treatment in MMR deficient leukemia, but that the specific regimen of drug administration may also play a role. These findings emphasize the need to further examine the etiology of this SBS1-like/thio-dMMR signature in the context of different treatment conditions.

The finding that individual clones may harbor mutations driven by mutational processes unique for these clones, triggered the hypothesis that studying mutational signatures in a clone-specific manner may improve their identification and our understanding of clonal heterogeneity and clonal expansion at different stages of disease development. To test this hypothesis, we performed WGS of two cases which experienced multiple ALL relapses (**chapter 6**). Based on the MAF between diagnosis and subsequent relapses in these two cases, we clustered mutations in groups following unique evolutionary trajectories and performed mutational signature analysis. Mutations within the same cluster are likely to occur in the same cell which potentially creates a clearer picture of mutational mechanisms that occurred at a specific time, as well as changes in the activity of these mutational processes over time. In this study, we demonstrated that analysis of mutational processes in a clone-specific manner can unravel unique mutational processes active in these clones. To our surprise, the observation that multiple mutational processes may be present in a single tumor (**chapter 5**) could be recapitulated even in individual clones. This finding suggests that distinctive mutational processes may affect multiple cell populations during tumor evolution, rather than a single clone. Furthermore, variable contribution of individual mutational processes across different clones in a single tumor suggests that their activity may vary during tumor development (temporal dynamics). In this scenario multiple clones, harboring variable contributions of different mutational mechanisms would evolve in stepwise processes, which is in line with previous observations showing sudden bursts of mutations specific to certain mutational processes at a single time point, e.g., AID/APOBEC mutagenesis[10,77,78]. However, there may be additional factors that can contribute to the change of activity of these processes in a single time point, e.g., interaction with the microenvironment or conditions in the leukemic niche (spatial dynamics). In this case, these clones would evolve in parallel, with occasional clonal branching, and differences in the contribution of individual mutational processes between clones would not solely depend on the episodic character of certain mutational mechanisms, such as

AID/APOBEC activity or thiopurine treatment, but also how much these mechanisms were restricted by other factors in the environment in which leukemic cells were residing. Furthermore, these mutational processes may drive alterations necessary for another mutational mechanism to become active, or create a predisposition for other mutational processes to occur. Indeed, a previous study has shown that the activity of BFB mechanism in the formation of sporadic iAMP21 creates dicentric chromosomes which dramatically increases the risk of chromothripsis, a phenomenon commonly observed in sporadic iAMP21[63,64]. Finally, the prognostic relevance of the mutational mechanisms active in leukemic cells remains unclear.

Future perspectives

Despite highly effective treatment strategies, relapse is still the most common cause of death in children with ALL[5]. This thesis gives directions for future research that may lead to our improved understanding of relapsed ALL and mechanisms leading to relapse development.

In recent years, clonal heterogeneity has become one of the frequently investigated aspects of ALL and cancer in general. In this thesis we provided an additional layer to already complex models of heterogeneity in ALL, by showing that individual clones may also differ in the mutational mechanisms that are active. It remains to be elucidated whether activity of mutational processes has prognostic relevance in current treatment protocols, which mechanisms drive (de)activation of individual mutational mechanisms, as well as how these mutational processes may contribute to clonal heterogeneity and clonal outgrowth. The advent of single cell technologies may further deepen our understanding of tumor heterogeneity and the complex mechanisms driving clonal diversification and mutagenesis. Single cell sequencing may be particularly relevant in further identifying prognostic relevance of genetic lesions in the rare, subclonal, population of leukemic cells. Furthermore, additional layers of epigenetic information have already proved invaluable in improving our understanding of ALL, and may reveal novel aspects of how individual clones evolve from diagnosis to relapse. Finally, functional characterization of individual tumor clones *in vivo* and in *vitro* may aid our understanding of therapy resistance and context in which different (epi)genetic lesions may contribute to treatment failure.

The analysis of mutational processes and our understanding of how they operate in human cancers has dramatically improved since 2013, when the first comprehensive catalogue of mutational processes in human cancers was published. Since then, mutational signature analysis has become an integral part of every comprehensive NGS study of human cancer. Next to being a useful tool in cancer characterization in research settings, analysis of mutational signatures may find its way in routine cancer diagnostics, once the associated etiology and risk factors are fully unraveled. For example, identification of the thiopurine and thio-dMMR mutational signatures, as well as the presence of thiopurine-associated mutations in genes previously linked with relapse development[20], suggests that these patients may benefit from further treatment adaptation.

In this thesis, we have also shown that the presence of subclonal alterations in a set of relapse-associated genes are not of prognostic relevance. However, their higher frequency in cases that relapse, compared to the ones that do not, requires further study. For example, these subclonal alterations may still provide susceptibility for relapse development in a specific genomic context, by acquiring additional driving alterations, or subclonal alteration in other relapse-associated genes. In addition, subclonal alterations in other relapse-associated genes, not studied in this thesis, may prove to be associated with relapse and unfavorable outcome. Finally, it is important to be aware of the fact that the introduction of new treatment strategies may change the landscape of genetic alterations associated with relapse and, therefore, the predictive value of these (sub)clonal alterations at diagnosis.

During the last decade our knowledge on genomic lesions driving treatment resistance has rapidly increased and with it our treatment strategies have improved, leading to better survival rates[2,5,79]. Although the majority of novel therapies are being assessed or introduced for salvage treatment in cases with relapsed or refractory disease, the ones with most exciting results are rapidly moving to the frontline of treatment[80,81]. For example, tyrosine kinase inhibitors (TKI) are already part of the frontline treatment of *BCR-ABL1* positive ALL, with second generation of TKI inhibitors, e.g., dasatinib and nilotinib, already showing improved efficacy profile against leukemia with mutation in *BCR-ABL1*, compared to the first generation TKI inhibitor imatinib. Third generation TKI inhibitors are undergoing clinical testing and offer promising results in *BCR-ABL1*-positive ALL cases with resistance to the second generation of TKI inhibitors[80,81]. Detection of B-cell specific surface markers, e.g., CD19, CD20 and CD22, has led to development of different immunotherapeutic approaches, such as monoclonal bodies, antibody-drug conjugates, bispecific T-cell engaging antibodies, and chimeric antigen receptor (CAR) T-cells[80-82]. Particularly promising results were achieved with blinatumomab[83], a bispecific antibody engaging both CD19+ B-cells and CD3+ cytotoxic T-cells, which leads to T-cells activation and apoptosis of CD19+ B-cells.

One of the most important observations that came from the studies comparing the genetic landscape of diagnosis and relapse is that the treatment-induced selective pressure represents the strongest individual factor influencing genetic composition of relapse. Although lesions such as *IKZF1* may not be a strong predictor of unfavorable outcome as before, novel medicines and treatment strategies will undoubtedly lead to development of new ways in which leukemic cells survive treatment. Therefore, treatment adaptations and introduction of new therapeutics may have as a result that the cases which relapse today remain in remission tomorrow. For example, in the clinical trials using blinatumomab in the frontline ALL treatment, loss of CD19 antigen already emerged as a reason for treatment resistance[82], and was also observed in other CD19 directed therapies, e.g., CAR T-cells[82,84]. Loss of CD19 antigen was frequently connected with genomic alterations affecting CD19[82,84], however, in a subset of cases the exact molecular mechanism is still unknown[82], or could be explained by disrupted CD19 export from the endoplasmic reticulum[85]. The latter finding represents a completely novel mechanism of treatment resistance in ALL and illustrates new challenges that need to be tackled from both clinical and research perspective in order to overcome treatment resistance. Therefore, all the cases relapsing after treatment with novel drugs, should be carefully evaluated in order to understand mechanisms driving treatment failure.

REFERENCES

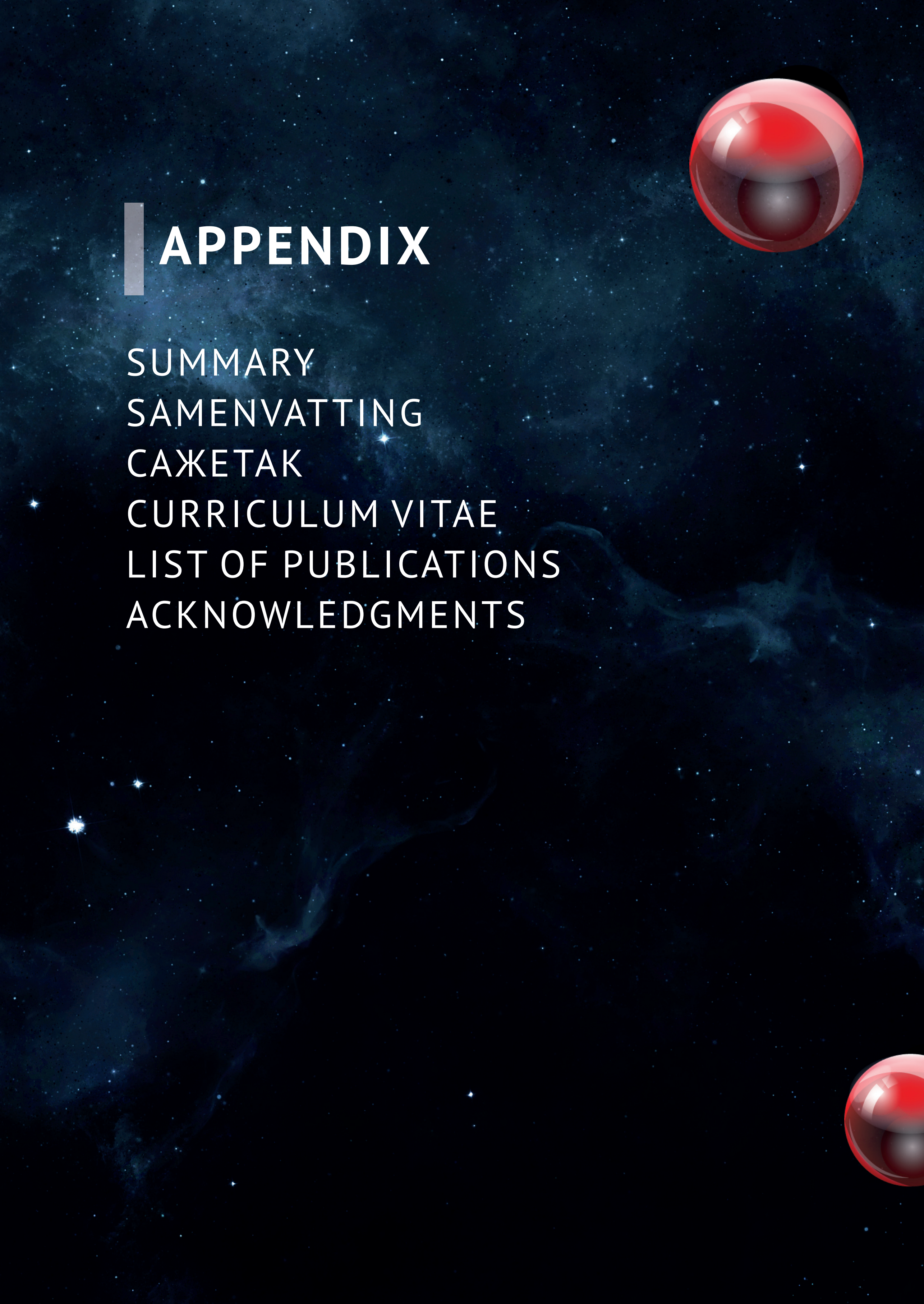
1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30.
2. Inaba H, Mullighan CG. Pediatric acute lymphoblastic leukemia. Haematologica. 2020;105(11):2524-2539.
3. Pieters R, Groot-Kruseman Hd, Velden VVd, Fiocco M, Berg Hvd, Bont Ed, et al. Successful therapy reduction and intensification for childhood acute lymphoblastic leukemia based on minimal residual disease monitoring: study ALL10 from the Dutch Childhood Oncology Group. J Clin Oncol. 2016;34(22):2591-2601.
4. Jeha S, Pei D, Choi J, Cheng C, Sandlund JT, Coustan-Smith E, et al. Improved CNS control of childhood acute lymphoblastic leukemia without cranial irradiation: St Jude total therapy study 16. J Clin Oncol. 2019;37(35):3377-3391.
5. Cooper SL, Brown PA. Treatment of pediatric acute lymphoblastic leukemia. Pediatr Clin North Am. 2015;62(1):61-73.
6. Gaudichon J, Jakobczyk H, Debaize L, Cousin E, Galibert M-D, Troadec M-B, et al. Mechanisms of extramedullary relapse in acute lymphoblastic leukemia: Reconciling biological concepts and clinical issues. Blood Rev. 2019;36:40-56.
7. Scharff BFSS, Modvig S, Marquart HV, Christensen C. Integrin-mediated adhesion and chemoresistance of acute lymphoblastic leukemia cells residing in the bone marrow or the central nervous system. Front Oncol. 2020;10:775.
8. Spinella JF, Richer C, Cassart P, Ouimet M, Healy J, Sinnett D. Mutational dynamics of early and late relapsed childhood ALL: rapid clonal expansion and long-term dormancy. Blood Adv. 2018;2(3):177-188.
9. Cahu X, Calvo J, Poglio S, Prade N, Colsch B, Arcangeli ML, et al. Bone marrow sites differently imprint dormancy and chemoresistance to T-cell acute lymphoblastic leukemia. Blood Adv. 2017;1(20):1760-1772.
10. Waanders E, Gu Z, Dobson SM, Antić Ž, Crawford JC, Ma X, et al. Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. Blood Cancer Discov. 2020;1(1):96-111.
11. Kuiper RP, Waanders E, van der Velden VH, van Reijmersdal SV, Venkatachalam R, Scheijen B, et al. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. Leukemia. 2010;24(7):1258-1264.
12. van der Veer A, Waanders E, Pieters R, Willemse ME, Van Reijmersdal SV, Russell LJ, et al. Independent prognostic value of BCR-ABL1-like signature and IKZF1 deletion, but not high CRLF2 expression, in children with B-cell precursor ALL. Blood. 2013;122(15):2622-2629.
13. Ma X, Edmonson M, Yergeau D, Muzny DM, Hampton OA, Rusch M, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun. 2015;6(1):6604.
14. Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011;471(7337):235-239.
15. Jaffe JD, Wang Y, Chan HM, Zhang J, Huether R, Kryukov GV, et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nat Genet. 2013;45(11):1386-1391.
16. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009;360(5):470-480.

17. Tzoneva G, Perez-Garcia A, Carpenter Z, Khiabanian H, Tosello V, Allegretta M, et al. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. Nat Med. 2013;19(3):368-371.
18. Meyer JA, Wang J, Hogan LE, Yang JJ, Dandekar S, Patel JP, et al. Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. Nat Genet. 2013;45(3):290-294.
19. Antić Ž, Yu J, Van Reijmersdal SV, Van Dijk A, Dekker L, Segerink WH, et al. Multiclonal complexity of pediatric acute lymphoblastic leukemia and the prognostic relevance of subclonal mutations. Haematologica. 2021;106(12):3046-3055.
20. Li B, Brady SW, Ma X, Shen S, Zhang Y, Li Y, et al. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. Blood. 2020;135(1):41-55.
21. Yang F, Brady SW, Tang C, Sun H, Du L, Barz MJ, et al. Chemotherapy and mismatch repair deficiency cooperate to fuel TP53 mutagenesis and ALL relapse. Nat Cancer. 2021;2(8):819-834.
22. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976;194(4260):23-28.
23. Ford AM, Bennett CA, Price CM, Bruin MC, Van Wering ER, Greaves M. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. Proc Natl Acad Sci U S A. 1998;95(8):4584-4588.
24. Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G, et al. Prenatal origin of acute lymphoblastic leukaemia in children. Lancet. 1999;354(9189):1499-1503.
25. Acuna-Hidalgo R, Sengul H, Steehouwer M, van de Vorst M, Vermeulen SH, Kiemeney L, et al. Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. Am J Hum Genet. 2017;101(1):50-64.
26. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. Science. 2019;366(6465).
27. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61(5):759-767.
28. Hinze L, Möricke A, Zimmermann M, Junk S, Cario G, Dagdan E, et al. Prognostic impact of IKZF1 deletions in association with vincristine–dexamethasone pulses during maintenance treatment of childhood acute lymphoblastic leukemia on trial ALL-BFM 95. Leukemia. 2017;31(8):1840-1842.
29. Clappier E, Grardel N, Bakkus M, Rapion J, De Moerloose B, Kastner P, et al. IKZF1 deletion is an independent prognostic marker in childhood B-cell precursor acute lymphoblastic leukemia, and distinguishes patients benefiting from pulses during maintenance therapy: results of the EORTC Children's Leukemia Group study 58951. Leukemia. 2015;29(11):2154-2161.
30. Dörge P, Meissner B, Zimmermann M, Möricke A, Schrauder A, Bouquin JP, et al. IKZF1 deletion is an independent predictor of outcome in pediatric acute lymphoblastic leukemia treated according to the ALL-BFM 2000 protocol. Haematologica. 2013;98(3):428-432.
31. Asai D, Imamura T, Suenobu S, Saito A, Hasegawa D, Deguchi T, et al. IKZF1 deletion is associated with a poor outcome in pediatric B-cell precursor acute lymphoblastic leukemia in Japan. Cancer Med. 2013;2(3):412-419.
32. Marke R, Havinga J, Cloos J, Demkes M, Poelmans G, Yuniati L, et al. Tumor suppressor IKZF1 mediates glucocorticoid resistance in B-cell precursor acute lymphoblastic leukemia. Leukemia. 2016;30(7):1599-1603.
33. Churchman ML, Low J, Qu C, Paietta EM, Kasper LH, Chang Y, et al. Efficacy of retinoids in IKZF1-mutated BCR-ABL1 acute lymphoblastic leukemia. Cancer Cell. 2015;28(3):343-356.
34. van der Veer A, Zaliova M, Mottadelli F, De Lorenzo P, Te Kronnie G, Harrison CJ, et al. IKZF1 status as a prognostic feature in BCR-ABL1-positive childhood ALL. Blood. 2014;123(11):1691-1698.

35. Steeghs EMP, Boer JM, Hoogkamer AQ, Boeree A, de Haas V, de Groot-Kruseman HA, et al. Copy number alterations in B-cell development genes, drug resistance, and clinical outcome in pediatric B-cell precursor acute lymphoblastic leukemia. Sci Rep. 2019;9(1):4634.
36. Dobson SM, García-Prat L, Vanner RJ, Wintersinger J, Waanders E, Gu Z, et al. Relapse-fated latent diagnosis subclones in acute B lineage leukemia are drug tolerant and possess distinct metabolic programs. Cancer Discov. 2020;10(4):568-587.
37. Clappier E, Gerby B, Sigaux F, Delord M, Touzri F, Hernandez L, et al. Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. J Exp Med. 2011;208(4):653-661.
38. Ebinger S, Özdemir EZ, Ziegenhain C, Tiedt S, Castro Alves C, Grunert M, et al. Characterization of rare, dormant, and therapy-resistant cells in acute lymphoblastic leukemia. Cancer Cell. 2016;30(6):849-862.
39. Paul BS, Helen HB, Sarra LR, Lars B, Joanna C, Jake C, et al. Dynamic clonal progression in xenografts of acute lymphoblastic leukemia with intrachromosomal amplification of chromosome 21. Haematologica. 2018;103(4):634-644.
40. Candelli T, Schneider P, Garrido Castro P, Jones LA, Bodewes E, Rockx-Brouwer D, et al. Identification and characterization of relapse-initiating cells in MLL-rearranged infant ALL by single-cell transcriptomics. Leukemia. 2022;36(1):58-67.
41. Gawad C, Koh W, Quake SR. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. Proc Natl Acad Sci U S A. 2014;111(50):17947-17952.
42. Albertí-Servera L, Demeyer S, Govaerts I, Swings T, De Bie J, Gielen O, et al. Single-cell DNA amplicon sequencing reveals clonal heterogeneity and evolution in T-cell acute lymphoblastic leukemia. Blood. 2021;137(6):801-811.
43. Wu L, Jiang M, Yu P, Li J, Ouyang W, Feng C, et al. Single-cell transcriptome analysis identifies ligand – receptor pairs associated with BCP-ALL prognosis. Front Oncol. 2021;11(322):639013.
44. Anand P, Guillaumet-Adkins A, Dimitrova V, Yun H, Drier Y, Sotudeh N, et al. Single-cell RNA-seq reveals developmental plasticity with coexisting oncogenic states and immune evasion programs in ETP-ALL. Blood. 2021;137(18):2463-2480.
45. Caron M, St-Onge P, Sontag T, Wang YC, Richer C, Ragoussis I, et al. Single-cell analysis of childhood leukemia reveals a link between developmental states and ribosomal protein expression as a source of intra-individual heterogeneity. Sci Rep. 2020;10(1):8079.
46. De Bie J, Demeyer S, Alberti-Servera L, Geerdens E, Segers H, Broux M, et al. Single-cell sequencing reveals the origin and the order of mutation acquisition in T-cell acute lymphoblastic leukemia. Leukemia. 2018;32(6):1358-1369.
47. Mehtonen J, Teppo S, Lahnalampi M, Kokko A, Kaukonen R, Oksa L, et al. Single cell characterization of B-lymphoid differentiation and leukemic cell states during chemotherapy in ETV6-RUNX1-positive pediatric leukemia identifies drug-targetable transcription factor activities. Genome Med. 2020;12(1):99.
48. Witkowski MT, Dolgalev I, Evensen NA, Ma C, Chambers T, Roberts KG, et al. Extensive remodeling of the immune microenvironment in B cell acute lymphoblastic leukemia. Cancer Cell. 2020;37(6):867-882.e812.
49. Ofverholm I, Tran AN, Heyman M, Zachariadis V, Nordenskjöld M, Nordgren A, et al. Impact of IKZF1 deletions and PAX5 amplifications in pediatric B-cell precursor ALL treated according to NOPHO protocols. Leukemia. 2013;27(9):1936-1939.

50. Boer JM, van der Veer A, Rizopoulos D, Fiocco M, Sonneveld E, de Groot-Kruseman HA, et al. Prognostic value of rare IKZF1 deletion in childhood B-cell precursor acute lymphoblastic leukemia: an international collaborative study. Leukemia. 2016;30(1):32-38.
51. Caye A, Beldjord K, Mass-Malo K, Drunat S, Soulier J, Gandemer V, et al. Breakpoint-specific multiplex polymerase chain reaction allows the detection of IKZF1 intragenic deletions and minimal residual disease monitoring in B-cell precursor acute lymphoblastic leukemia. Haematologica. 2013;98(4):597-601.
52. Chan LN, Chen Z, Braas D, Lee J-W, Xiao G, Geng H, et al. Metabolic gatekeeper function of B-lymphoid transcription factors. Nature. 2017;542(7642):479-483.
53. Xiao G, Chan LN, Klemm L, Braas D, Chen Z, Geng H, et al. B-cell-specific diversion of glucose carbon utilization reveals a unique vulnerability in B cell malignancies. Cell. 2018;173(2):470-484.e418.
54. Kaufmann C, Yoshida T, Perotti EA, Landhuis E, Wu P, Georgopoulos K. A complex network of regulatory elements in Ikaros and their activity during hemo-lymphopoiesis. EMBO J. 2003;22(9):2211-2223.
55. Marke R, van Leeuwen FN, Scheijen B. The many faces of IKZF1 in B-cell precursor acute lymphoblastic leukemia. Haematologica. 2018;103(4):565-574.
56. Kobitzsch B, Gökbuget N, Schwartz S, Reinhardt R, Brüggemann M, Viardot A, et al. Loss-of-function but not dominant-negative intragenic IKZF1 deletions are associated with an adverse prognosis in adult BCR-ABL-negative acute lymphoblastic leukemia. Haematologica. 2017;102(10):1739-1747.
57. Tzoneva G, Dieck CL, Oshima K, Ambesi-Impiombato A, Sánchez-Martín M, Madubata CJ, et al. Clonal evolution mechanisms in NT5C2 mutant-relapsed acute lymphoblastic leukaemia. Nature. 2018;553(7689):511-514.
58. Yu J, Waanders E, van Reijmersdal SV, Antić Ž, van Bosbeek CM, Sonneveld E, et al. Upfront treatment influences the composition of genetic alterations in relapsed pediatric B-cell precursor acute lymphoblastic leukemia. Hemasphere. 2020;4(1):e318.
59. Dieck CL, Ferrando A. Genetics and mechanisms of NT5C2-driven chemotherapy resistance in relapsed ALL. Blood. 2019;133(21):2263-2268.
60. Moriyama T, Liu S, Li J, Meyer J, Zhao X, Yang W, et al. Mechanisms of NT5C2-mediated thiopurine resistance in acute lymphoblastic leukemia. Mol Cancer Ther. 2019;18(10):1887-1895.
61. Brouwer C, Vogels-Mentink TM, Keizer-Garritsen JJ, Trijbels FJ, Bökkerink JP, Hoogerbrugge PM, et al. Role of 5'-nucleotidase in thiopurine metabolism: enzyme kinetic profile and association with thio-GMP levels in patients with acute lymphoblastic leukemia during 6-mercaptopurine treatment. Clin Chim Acta. 2005;361(1-2):95-103.
62. Welin M, Nordlund P. Understanding specificity in metabolic pathways--structural biology of human nucleotide metabolism. Biochem Biophys Res Commun. 2010;396(1):157-163.
63. Harrison CJ. Blood spotlight on iAMP21 acute lymphoblastic leukemia (ALL), a high-risk pediatric disease. Blood. 2015;125(9):1383-1386.
64. Li Y, Schwab C, Ryan SL, Papaemmanuil E, Robinson HM, Jacobs P, et al. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. Nature. 2014;508(7494):98-102.
65. Robinson HM, Harrison CJ, Moorman AV, Chudoba I, Strefford JC. Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage-fusion-bridge cycle. Genes Chromosomes Cancer. 2007;46(4):318-326.

66. Waanders E, Scheijen B, van der Meer LT, van Reijmersdal SV, van Emst L, Kroeze Y, et al. The origin and nature of tightly clustered BTG1 deletions in precursor B-cell acute lymphoblastic leukemia support a model of multiclonal evolution. PLoS Genet. 2012;8(2):e1002533.
67. Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet. 2014;46(2):116-125.
68. Kuiper RP, Waanders E. A RAG driver on the road to pediatric ALL. Nat Genet. 2014;46(2):96-98.
69. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-421.
70. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101.
71. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet. 2014;15(9):585-598.
72. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10(1):33.
73. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010;11(1):367.
74. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature. 2018;555(7696):371-376.
75. Newman S, Nakitandwe J, Kesserwan CA, Azzato EM, Wheeler DA, Rusch M, et al. Genomes for kids: the scope of pathogenic mutations in pediatric cancer revealed by comprehensive DNA and RNA sequencing. Cancer Discov. 2021;11(12):3008-3027.
76. You C, Dai X, Yuan B, Wang Y. Effects of 6-thioguanine and S6-methylthioguanine on transcription in vitro and in human cells. J Biol Chem. 2012;287(49):40915-40923.
77. Antić Ž, Lelieveld SH, van der Ham CG, Sonneveld E, Hoogerbrugge PM, Kuiper RP. Unravelling the sequential interplay of mutational mechanisms during clonal evolution in relapsed pediatric acute lymphoblastic leukemia. Genes (Basel). 2021;12(2):214.
78. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. Cell. 2019;176(6):1282-1294.e1220.
79. Pieters R. Acute lymphoblastic leukaemia in children and adolescents: chance of cure now higher than 80%. Ned Tijdschr Geneeskd. 2010;154:A1577.
80. Phelan KW, Advani AS. Novel therapies in acute lymphoblastic leukemia. Curr Hematol Malig Rep. 2018;13(4):289-299.
81. Inaba H, Pui C-H. Advances in the diagnosis and treatment of pediatric acute lymphoblastic leukemia. J Clin Med. 2021;10(9):1926.
82. Schultz L, Gardner R. Mechanisms of and approaches to overcoming resistance to immunotherapy. Hematology Am Soc Hematol Educ Program. 2019;2019(1):226-232.
83. Queudeville M, Ebinger M. Blinatumomab in pediatric acute lymphoblastic leukemia-from salvage to first line therapy (a systematic review). J Clin Med. 2021;10(12):2544.
84. Shah NN, Fry TJ. Mechanisms of resistance to CAR T cell therapy. Nat Rev Clin Oncol. 2019;16(6):372-385.
85. Braig F, Brandt A, Goebeler M, Tony HP, Kurze AK, Nollau P, et al. Resistance to anti-CD19/CD3 BiTE in acute lymphoblastic leukemia may be mediated by disrupted CD19 membrane trafficking. Blood. 2017;129(1):100-104.



APPENDIX

SUMMARY
SAMENVATTING
САЖЕТАК
CURRICULUM VITAE
LIST OF PUBLICATIONS
ACKNOWLEDGMENTS

SUMMARY

In the last seven decades, the outcome of pediatric acute lymphoblastic leukemia (ALL) and survival rates have improved dramatically, reaching 90% in the contemporary treatment protocols. However, around 10% of the children experience relapse and among them the outcome remains poor. Therefore, relapse remains a major clinical and scientific issue in pediatric ALL, and understanding etiology and mechanisms driving treatment failure may improve outcome and prevent relapse development. Because cancer is predominantly a genetic disease, and genetic alterations have a major role in driving development of ALL and treatment failure, we focused in this thesis on examining the genetic basis for relapse development and the competition between leukemic clones during treatment. In **chapter 1** we give a broad overview of the current knowledge of the genetics of ALL, mutagenesis, relapse development, treatment resistance and clonal evolution. In the subsequent chapters we use specific approaches to investigate these topics in the context of relapsed ALL and explain mechanisms in which leukemic clones return after treatment has been successfully initiated.

Previous studies comparing differences between samples from diagnosis and relapse have shown that the relapse driving mutations can be found already at diagnosis, often on a subclonal level and undetectable by routine diagnostic methods. However, the clinical relevance of these relapse-associated alterations in subclones has so far not been investigated. In order to accurately and sensitively detect mutations present even in a small fraction of cells, we developed an assay using single molecule Molecular Inversion Probe (smMIP), which we describe in the **chapter 2**. The smMIP approach utilizes random molecular tagging for each captured locus which enables correction for amplification and sequencing artefacts and their distinction from true somatic mutations. Using smMIP-based NGS sequencing we were able to reliably detect mutations with allele frequencies as low as 0.4%. We subsequently utilized this approach in **chapter 3**, to perform screening for mutations in the relapse-associated genes *CREBBP*, *KRAS*, *NRAS*, *PTPN11*, *TP53* and *WHSC1*, in a cohort of 503 pediatric ALL patients enrolled in DCOG-ALL9 and DCOG-ALL10 studies. In addition, we developed a semi-quantitative PCR assay to study an intragenic exon 4-7 *IKZF1* deletion, which occurs in 25% of the cases and represents the second most common *IKZF1* deletion. In total we identified 660 alterations in 285 patients of which 495 (75%) were subclonal. Despite a high number of subclonal alterations, we did not find an association of either subclonal or major clone alterations with relapse development. The only exceptions were major clone *IKZF1* exon 4-7 deletions, in line with previous studies done on all *IKZF1* deletions, suggesting their strong association with relapse development. In addition, we

traced the alterations detected at diagnosis in the relapse samples of the same patients and found that only 7% of subclonal alterations were preserved from diagnosis to major clones in relapse, compared to 56% of major clone alterations. Results were even more striking for *IKZF1* exon 4-7 deletions where all detected deletions in major clones were preserved in available relapse samples, while all subclonal deletions from diagnosis were lost at time of relapse. These data suggest that for the investigated genes there is no basis to use subclonal alterations as a prognostic marker.

Around one quarter of all the relapses occur very early (<18 months from diagnosis), while the treatment is still ongoing, and prognosis among these cases is poor compared to cases that relapse after treatment has ended. In **chapter 4** we aimed to unravel the genomic basis of very early relapse. We performed whole exome sequencing (WES) of 12 diagnosis-relapse pairs of children with ALL who relapsed very early, followed by a deep-sequencing validation of all identified mutations. In addition, we included one case with a good initial treatment response and on-treatment relapse at the end of upfront therapy. Despite short remission time, we observed dynamic clonal evolution, with relapse originating from a subclone in almost all the cases. In addition, we observed several driver mutations that may have influenced relapse development and outgrowth of treatment-resistant clones. For example, in the MRD-standard risk patient with *ETV6-RUNX1*-positive ALL we observed outgrowth of the *TP53* mutated clone after loss of the wildtype allele, resulting in relapse development. Furthermore, two patients with *TCF3-PBX1*-positive leukemia, a subtype with favorable prognosis in the contemporary treatment protocols, harbored *WHSC1* p.E1099K mutation in diagnosis. Since cases with *TCF3-PBX1*-positive ALL rarely relapse, and prognosis is poor in case of relapse, we collected 16 additional *TCF3-PBX1*-positive relapsed cases from eight collaborating centers, including six with very early relapse, and performed screening for *WHSC1* mutations in these cases. Our analysis revealed one additional case with a p.E1099K mutation who experienced very early relapse, suggesting that *WHSC1* mutations may not be a strong predictor of relapse in *TCF3-PBX1*-positive ALL by itself, but in combination with additional alterations may contribute to relapse development in these cases. In addition to alterations in known relapse drivers, we identified two cases with truncating mutations in cohesin complex gene *RAD21*, which was previously not associated with relapsed ALL.

In **chapter 5** we performed a study on a large cohort of relapsed ALL patients in collaboration with colleagues from St. Jude Children’s research hospital and the University of Toronto. In this study we show that relapses may develop from ancestral, major and minor clones at diagnosis. Furthermore, we demonstrate that hypermutation is a common phenomenon in relapsed ALL and that different mechanisms, often in combination, can drive this hypermutation phenotype. In addition to mutational signatures caused by mismatch repair deficiency and aberrant AID/APOBEC activity, we also describe a new mutational mechanism in this chapter, which is very similar to the already known age-related mutational signature SBS1. This SBS1-like signature is characterized by C>T substitutions occurring at CpGs but, in contrast to the known COSMIC SBS1 signature, shows a strong transcription bias in CpGs which correlates with gene expression. After our finding, another study revealed that this signature developed due to thiopurine-associated DNA damage in leukemias with mismatch repair (MMR) deficiency, therefore named thio-dMMR. This finding suggests that this thio-dMMR signature may be driven by a mechanism causing damage on a single stranded DNA during transcription, or lack of transcription coupled repair mechanism. Finally, we show that the hypermutation causes formation of neoepitopes which may be used as targets for immunotherapy.

The observation that multiple processes can drive mutagenesis in a single tumor, led us to examine whether analysis of mutational signatures in a clone-specific manner may aid uncovering specific mutational processes active at particular time points during tumor development. Therefore, in **chapter 6** we selected two cases with multiple relapses and performed whole genome sequencing (WGS) of samples taken at diagnosis and every subsequent relapse. We subsequently performed clustering of detected somatic mutations based on their allele frequencies and performed *de novo* mutational signature extraction in order to examine whether groups of mutations separated in time harbor distinct mutational processes. We demonstrated that examining mutational signatures in a clone-specific manner can improve our understanding of mechanisms active in a single parental cell. Furthermore, we demonstrated that multiple mutational processes can be active in a single clone, and that some processes are active at specific timepoints during tumor evolution, e.g., signatures resembling ultraviolet light-driven DNA damage and thiopurine-associated damage, while others show continuous interplay of aberrant activation, e.g., AID/APOBEC.

In **chapter 7** we discuss the clinical relevance of subclonal mutations in relapse-associated genes, clonal heterogeneity, tumor evolution and mechanisms in which individual clones and genomic alterations may break through treatment-induced selective pressure and cause relapse. Furthermore, we discuss mechanisms driving mutagenesis in diagnosis and relapse and how our understanding of these processes may improve future ALL treatment. Finally, we elaborate on the future perspectives on examining clonal heterogeneity, mechanisms of relapse, treatment and prevention of ALL.

SAMENVATTING

Sinds de jaren '50 van de vorige eeuw is de overleving van kinderen met acute lymfatische leukemie (ALL) drastisch verbeterd, tot 90% in de huidige behandelingsprotocollen. Echter, ongeveer 10% van de kinderen ontwikkelt een recidief en bij hen blijft de uitkomst slecht. Het voorkomen van recidieven blijft daarom een belangrijk klinisch en wetenschappelijk aandachtspunt, waarbij een beter begrip van de oorzaak van een falende behandeling uiteindelijk moet leiden tot nog betere genezing van kinderen met ALL. In de basis is kanker een genetische ziekte en genetische afwijkingen spelen een belangrijke rol bij de ontwikkeling van ALL. Sommige van deze genetische afwijkingen kunnen ook de reden zijn dat de behandeling niet aanslaat, en in dit proefschrift hebben we de genetische basis voor het ontwikkelen van een recidief en de competitie tussen leukemische klonen tijdens de uitgroei en behandeling onderzocht. In **hoofdstuk 1** geven we een uitgebreid overzicht van de huidige kennis van de genetica van ALL bij kinderen, het ontstaan van genetische afwijkingen, klonale evolutie, resistentie tegen behandeling en de ontwikkeling van een recidief. In de daaropvolgende hoofdstukken onderzoeken we deze onderwerpen in de context van recidief ALL met specifieke benaderingen en beschrijven we verschillende mechanismen die de terugkeer van leukemische klonen na een behandeling kunnen veroorzaken.

In eerdere onderzoeken naar genetische verschillen tussen ALL monsters van diagnose en recidief kon worden aangetoond dat de mutaties die het recidief veroorzaken veelal reeds aanwezig zijn bij diagnose, vaak op een subklonaal niveau en niet-detecteerbaar met standaard diagnostische testen. De klinische relevantie van deze met recidief geassocieerde afwijkingen in subklonen is tot dusver niet onderzocht. Om mutaties nauwkeurig en sensitief te detecteren, zelfs als ze slechts in een klein deel van de cellen aanwezig zijn, hebben we een test ontwikkeld met behulp van zogenaamde '*single molecule Molecular Inversion Probes*' (smMIPs), die we beschrijven in **hoofdstuk 2**. De smMIP-benadering maakt gebruik van willekeurig geformeerde moleculaire herkennings-sequenties ('*tags*') voor elke geselecteerde locus, waardoor we kunnen corrigeren voor allerlei artefacten en we de echte afwijkingen met veel hogere gevoeligheid kunnen aantonen. In onze experimenten bleek het mogelijk om met behulp van smMIP-sequencing mutaties met allelfrequenties zo laag als 0.4% op betrouwbare wijze te detecteren. In hoofdstuk 3 hebben we deze smMIP benadering gebruikt om leukemiemateriaal van 503 kinderen met ALL te screenen op mutaties in de bekende recidief-geassocieerde genen *CREBBP*, *KRAS*, *NRAS*, *PTPN11*, *TP53*, *WHSC1* en *NT5C2*. Daarnaast hebben we een semi-kwantitatieve PCR-test ontwikkeld om een exon 4-7 deletie in het gen *IKZF1* te bestuderen die in 25% van de gevallen van pediatrische ALL voorkomt. In totaal hebben we 660 mutaties bij 285 patiënten kunnen aantonen, waarvan er 495 (75%) subklonaal waren. Ondanks het grote aantal subklonale mutaties, hebben we geen verband gevonden tussen subklonale of klonale mutaties en de ontwikkeling van een recidief. Klonale deleties van exon 4-7 in *IKZF1* waren de enige uitzondering. Zoals in eerdere studies al was aangetoond, bleken deze deleties wel sterk geassocieerd te zijn met de ontwikkeling van een recidief. Door alteraties die gedetecteerd werden bij diagnose te traceren in recidiefmonsters van dezelfde patiënten ontdekten we vervolgens dat slechts 7% van de subklonale alteraties klonaal aanwezig waren in het recidiefmonster. Voor klonale alteraties was dit 56%. Voor exon 4-7 deleties in *IKZF1* waren de resultaten nog opmerkelijker. Alle deleties die klonaal gedetecteerd werden in het diagnosemonster waren behouden in de beschikbare recidiefmonsters, terwijl alle deleties die subklonaal aanwezig waren in het diagnosemonster niet gedetecteerd werden in het recidiefmonster. Deze gegevens suggereren dat er voor de onderzochte genen geen basis is om subklonale alteraties als prognostische marker te gebruiken.

Ongeveer een kwart van alle recidieven treedt zeer vroeg op (binnen 18 maanden na diagnose), terwijl de patiënt nog behandeld wordt. De prognose van deze gevallen is slechter in vergelijking met gevallen die recidiveren na de behandeling. In **hoofdstuk 4** hebben we de genomische basis van zeer vroege recidieven onderzocht. Op 12 diagnose-recidiefparen van kinderen met ALL die zeer vroeg een recidief ontwikkelden hebben we whole exome-sequencing (WES) uitgevoerd, gevolgd door een diepere sequencing om alle geïdentificeerde mutaties te valideren. Daarnaast hebben we één patiënt geïncludeerd met een goede initiële respons op de behandeling van de oorspronkelijke leukemie die binnen twee jaar (dus vlak voor het eind van de behandeling) toch een recidief ontwikkelde. Ondanks een korte remissietijd hebben we in vrijwel alle patiënten een dynamische klonale evolutie waargenomen, waarbij het recidief bijna altijd afkomstig was van een subkloon van de oorspronkelijke diagnose. Daarnaast hebben we verschillende driver-mutaties gedetecteerd die de ontwikkeling van het recidief en de uitgroei van therapieresistente klonen mogelijk hebben beïnvloed. Bij een op basis van therapie-respons laag-risico geclassificeerde patient met een *ETV6-RUNX1*-positieve ALL zagen we bijvoorbeeld uitgroei van een *TP53*-gemuteerde kloon na verlies van het wildtype allel, resulterend in de ontwikkeling van een recidief. Daarnaast vonden we bij twee patiënten met *TCF3-PBX1*-positieve leukemie, een subtype met gunstige prognose in de huidige behandelprotocollen, een *WHSC1* E1099K-mutatie bij diagnose. Aangezien gevallen met *TCF3-PBX1*-positieve ALL zelden recidiveren en de prognose bij een recidief slecht is, hebben we met acht samenwerkende centra 16 extra *TCF3-*

PBX1-positieve patiënten met een recidief verzameld, waaronder zes met een zeer vroeg recidief. Deze patiënten hebben we gescreend op *WHSC1* mutaties. Onze analyse bracht één extra patiënt aan het licht met een E1099K-mutatie en een zeer vroeg recidief. Dit suggereert dat *WHSC1*-mutaties op zichzelf geen sterke prognostische factor zijn voor recidief in *TCF3-PBX1*-positieve ALL, maar dat ze in combinatie met aanvullende afwijkingen wel kunnen bijdragen aan de ontwikkeling van een recidief in deze gevallen. Naast afwijkingen in bekende genen in de context van recidieven, identificeerden we twee gevallen met truncerende mutaties in het cohesinecomplexgen *RAD21*.

In **hoofdstuk 5** hebben we een studie uitgevoerd op een groot cohort van recidiverende ALL-patiënten in samenwerking met collega's van het St. Jude Children's Research Hospital in Memphis, de Verenigde Staten, en de Universiteit van Toronto in Canada. In deze studie laten we zien dat recidieven zich kunnen ontwikkelen uit reeds bestaande dominante klonen en subklonen bij diagnose. Verder laten we zien dat hypermutatie een veelvoorkomend fenomeen is bij recidiverende ALL en dat verschillende mechanismen, vaak in combinatie, dit hypermutatiefenotype kunnen aansturen. Naast de specifieke mutatiepatronen (mutational signatures) veroorzaakt door mismatch-repair deficiëntie en afwijkende AID/APOBEC-activiteit, beschrijven we in dit hoofdstuk ook een nieuw mutatiemechanisme dat erg lijkt op de bekende leeftijdsgerelateerde mutatiepatronen SBS1. Deze SBS1-achtige mutatiepatronen wordt gekenmerkt door C>T-substituties die optreden op CpG basevolgorden, maar vertoont, in tegenstelling tot het bekende COSMIC SBS1-patroon, een sterke transcriptie-geassocieerde onbalans in CpG's die correleert met de mate van expressie van het gen. Na onze bevinding werd in een andere studie ontdekt dat dit specifieke patroon wordt veroorzaakt door thiopurine-geassocieerde DNA-schade bij leukemieën met mismatch-repair (MMR)-deficiëntie. Daarom werd dit patroon het thio-dMMR patroon genoemd. Deze bevinding suggereert dat het thio-dMMR-patroon wordt veroorzaakt door een mechanisme dat schade toebrengt aan enkelstrengs DNA tijdens transcriptie, of door het ontbreken van een aan transcriptie gekoppeld reparatiemechanisme. Ten slotte laten we zien dat hypermutatie de vorming van neo-epitopen veroorzaakt die kunnen worden gebruikt als doelwit voor immuuntherapie.

De observatie dat mutagenese kan worden veroorzaakt door meerdere processen in één tumor bracht ons ertoe te onderzoeken of de analyse van kloonspecifieke mutatiepatronen kan helpen bij het blootleggen van mutatieprocessen die actief zijn op specifieke tijdstippen tijdens tumorontwikkeling. Daarom selecteerden we in **hoofdstuk 6** twee patiënten met meerdere recidieven en voerden we whole genome sequencing (WGS) uit op monsters afgenomen bij diagnose en bij elk volgend recidief.

Vervolgens hebben we de gedetecteerde somatische mutaties geclusterd op basis van allelfrequentie en hebben we een de novo extractie van de mutatiepatronen uitgevoerd om te onderzoeken of groepen mutaties die op verschillende momenten ontstaan zijn, door verschillende mutatieprocessen veroorzaakt worden. We hebben aangetoond dat het onderzoeken van mutatiepatronen op een kloonspecifieke manier ons begrip van de mechanismen die actief zijn in een enkele voorlopercel kan verbeteren. Verder hebben we aangetoond dat meerdere mutatieprocessen actief kunnen zijn in dezelfde kloon en dat sommige processen actief zijn op specifieke tijdstippen tijdens tumorontwikkeling, bijvoorbeeld processen die duiden op DNA schade zoals die ook kan worden veroorzaakt door blootstelling aan UV licht of op thiopurine-geassocieerde schade, terwijl andere processen, zoals b.v. AID/APOBEC activiteit, voortdurend tot nieuwe mutaties kunnen leiden.

In **hoofdstuk 7** bediscussiëren we al onze resultaten in een bredere context. We bespreken de klinische relevantie van subklonale mutaties in genen geassocieerd met recidieven en de klonale heterogeniteit van ALL in relatie tot de ontwikkeling van recidieven. Verder bespreken we de mechanismen die mutaties kunnen veroorzaken bij diagnose en recidief en hoe ons begrip van deze processen de toekomstige behandeling van ALL kan verbeteren. Tot slot gaan we dieper in op de toekomstperspectieven van het bestuderen van klonale heterogeniteit, mechanismen van recidief, behandeling en preventie van ALL.

САЖЕТАК

Током протеклих седам деценија дошло је до драматичног побољшања исхода и стопе преживљавања код оболелих од педијатријске акутне лимфобластне леукемије (АЛЛ), која у савременим терапијским протоколима достиже и до 90%. Упркос напретку у лечењу, код 10% оболелих долази до релапса болести, и код ове деце исход је лош. Стога, релапс АЛЛ-а представља важан клинички и научни проблем, и боље познавање етиологије и механизама који доводе до неуспеха у лечењу може довести до побољшања исхода и спречити развој релапса. С обзиром да је рак претежно генетска болест, а генетске алтерације имају главну улогу у покретању развоја АЛЛ-а, као и неуспеха у лечењу, у овој тези смо испитали генетске основе за развој релапса и надметања између леукемијских ћелија током лечења. У **поглављу 1** дајемо преглед тренутних сазнања о генетици АЛЛ-а, мутагенези, развоју релапса, отпорности на третман и клонској еволуцији. У осталим поглављима користимо различите методе да објаснимо механизме који доводе до експанзије леукемијских клонова након што је третман успешно започет.

Упоређујући генетску структуру тумора из узорака узетих након иницијалне дијагнозе болести и релапса, претходне студије су показале да се генетске алтерације које изазивају релапс могу детектовати већ у дијагностичким узорцима, често у ретким и малобројним популацијама ћелија (недоминантни клонови), што онемогућава њихово откривање рутинским дијагностичким методама. Међутим, клинички значај генетских алтерација повезаних са релапсом у недоминантним клоновима до сада није истраживан. Да бисмо уз високу прецизност и осетљивовост открили мутације присутне чак и у малом броју ћелија, развили смо методу засновану на single molecule Molecular Inversion Probe (smMIP), коју описујемо у **поглављу 2**. smMIP метод омогућава насумично молекуларно обележавање сваког ухваћеног локуса, што обезбеђује могућност корекције артефаката који настају током реакције ланчаног умножавања (PCR) и секвенционирања, те њихово разликовање од правих соматских мутација. Користећи smMIP технологију, и нову генерацију метода за секвенцирање ДНК, успели смо да поуздано откријемо мутације са алелском учесталошћу до чак 0,4%. Овај метод применили смо у **поглављу 3**, како би извршили скрининг на мутације у генима за повезаним са развојем релапса: *CREBBP*, *KRAS*, *NRAS*, *PTPN11*, *TP53*, *WHSC1* и *NT5C2*, у групи од 503 педијатријских АЛЛ пацијента укључених у клиничке студије DCOG-ALL9 и DCOG-ALL10. Поред тога, развили смо семи-квантитативни PCR тест како би испитали делеције егзона 4-7 *IKZF1* гена, које представљају другу најчешћу *IKZF1* делецију, и чине око 25% свих АЛЛ случајева са

IKZF1 делецијама. Укупно смо идентификовали 660 генетских алтерација код 285 пацијената, од којих је 495 (75%) било у недоминантним клоновима. Упркос великом броју детектованих генетских алтерација, нисмо пронашли повезаност између развоја релапса и генетских алтерација у доминантним, нити у недоминантним клоновима. Једини изузетак су била делеције егзона 4-7 *IKZF1* гена у доминантним клоновима, што је у складу са претходним студијама које су обухватиле све познате *IKZF1* делеције, и указује на њихову снажну повезаност са развојем релапса. Поред тога, пратили смо појединачне генетске алтерације откривене током дијагнозе, у узорцима узетим након релапса код истих пацијената. У поређењу са 56% генетских алтерација у доминантним клоновима, само 7% алтерација детектованих у недоминантним клоновима је опстало и постало доминантно током релапса. Резултати су били још упечатљивији за делеције егзона 4-7 *IKZF1* гена, где су све откривене делеције у доминантним клоновима опстале у доступним узорцима узетим након релапса, док су све делеције у недоминантним клоновима биле недетектабилне. Ови подаци сугеришу да за испитиване гене нема основа да се генетске алтерације у недоминантним клоновима користе као прогностички маркер.

Једна четвртина свих релапса јавља се веома рано (<18 месеци од постављања дијагнозе), док је лечење још увек у току, и њихова прогноза је лоша у поређењу са случајевима код којих се релапс јавља након завршеног лечења. У **4. поглављу** имали смо за циљ да откријемо генетску основу веома раног релапса. Извршили смо секвенцирање егзома узорака узетих током иницијалне дијагнозе и релапса код деце код које је дошло до веома раног релапса, након чега је уследило дубоко секвенцирање свих идентификованих мутација у циљу њихове валидације. Поред тога, укључили смо један случај са добрим терапијским одговором на почетку лечења, код кога се релапс јавио на самом крају терапије. Упркос краткој ремисији, приметили смо динамичну клонску еволуцију, са релапсом који, у скоро свим случајевима, потиче од недоминантних клонова. Такође, идентификовали смо неколико мутација које су могле да утичу на развој релапса и селекцију клонова отпорних на третман. На пример, код пацијената са *ETV6-RUNX1* леукемијом, који је на основу минималне резидуалне болести лечен у складу са терапијским протоколом за пацијенте са стандардним ризиком, дошло је до селекције и експанзију клона са мутацијом у *TP53* гену, што је након потпуног губитка другог, дивљег алела, резултирало развојем релапса. Надаље, код два пацијента са *TCF3-PBX1* леукемијом, подтипом са повољном прогнозом у савременим протоколима лечења, пронашли смо *WHSC1* p.E1099K мутацију у дијагнози. Пошто случајеви са *TCF3-PBX1*-позитивним АЛЛ ретко имају релапсе, а прогноза је лоша уколико до релапса дође, у сарадњи са осам других центара прикупили смо 16 додатних случајева са *TCF3-PBX1* леукемијом код којих је дошло до релапса, укључујући шест са веома раним релапсом, и извршили скрининг на *WHSC1* мутације. Наша анализа је открила још један случај са мутацијом p.E1099K код кога је дошло до веома раног релапса, што сугерише да иако *WHSC1* мутације нису јак показатељ ризика за развој релапса код деце са *TCF3-PBX1*-позитивним АЛЛ, у комбинацији са додатним алтерацијама могу допринети развоју раног релапса. Поред генетских алтерација у познатим узрочницима релапса, идентификовали смо и два случаја са нонсенс мутацијама у гену кохезинског комплекса *RAD21*, који раније није био повезан са релапсом у АЛЛ.

У сарадњи са колегама из дечије истраживачке болнице “St Jude” и Универзитета у Торонту, у **5. поглављу** извели смо студију на великој групи пацијената са АЛЛ, код које је дошло до релапса. У овој студији смо показали да се релапс може развити од предака леукемијских клонова, као и доминантних и недоминантних клонова приликом постављања дијагнозе. Надаље, показали смо да је хипермутабилност уобичајена појава у релапсу АЛЛ-а и да различити механизми, често у комбинацији, могу довести до хипермутабилности. Поред мутационих отисака узрокованих недостатком поправке неусклађености ДНК (MMR) и аберантном активношћу AID/APOBEC деаминаза, у овом поглављу описујемо и нови механизам мутагенезе који је веома сличан већ познатом мутационом отиску SBS1 повезаним са старењем. Овај SBS1-сличан отисак карактеришу C>T супституције које се јављају у CpG динуклеотидима, али за разлику од познатог COSMIC SBS1 отиска, значајно су чешће на транскрибованим ланцима ДНК, и корелишу са нивоом генске експресије. Након нашег истраживања, друга студија је показала да се овај мутациони механизам јавља као последица терапије тиопуринима код оболелих од леукемије са недостатком поправке неусклађености, због чега је назван thio-dMMR. Ово сугерише да thio-dMMR отисак може бити узрокован механизмом који доводи до оштећења једноланчане ДНК током транскрипције, или недостатаком механизма за поправку повезаног са транскрипцијом. Коначно, показујемо да хипермутабилност изазива формирање неоепитопа који се могу користити као мете за имунотерапију.

Запажање да више процеса може покренути мутагенезу у једном тумору, навело нас је да испитамо да ли анализа мутационих отисака на нивоу појединачних клонова може помоћи у откривању специфичних мутационих процеса активних током одређених временских периода у току развоја тумора. Стога смо у **поглављу 6** одабрали два пацијента са вишеструким релапсима и извршили секвенцирање комплетног генома узорака узетих након иницијалне дијагнозе болести и при сваком следећем релапсу. Затим смо извршили груписање откривених соматских мутација на основу њихових алелских учесталости и извршили *de novo* екстракцију мутационих отисака, како бисмо испитали да ли су групе мутација одвојене у времену узроковане различитим мутационим процесима. Показали смо да испитивање мутационих отисака на нивоу појединачних клонова може унапредити наше разумевање механизама активних у једној родитељској ћелији. Штавише, показали смо да више мутационих процеса може бити активно у једном клону, и да су поједини процеси активни у одређеним временским периодима током еволуције тумора. На пример, отисци који подсећају на оштећења ДНК изазвана ултраљубичастим зрацима, као и оштећења повезана са терапијом тиопуринима, јављају се у кратким временским периодима током развоја тумора, док други показују континуирану аберантну активност, нпр. AID/APOBEC.

У **7. поглављу** дискутујемо о клиничком значају генетских мутација, повезаних са развојем релапса, у недоминантним клоновима, разноврсности клонова, еволуцији тумора и механизмима у којима појединачни клонови и генетске алтерације могу да надвладају селективни притисак изазван применом лекова и изазову релапс. Надаље, разматрамо механизме који покрећу мутагенезу током иницијалне дијагнозе и релапса и како наше разумевање ових процеса може побољшати лечење АЛЛ-а. На крају, представљамо будуће правце за испитивање клонске разноврсности, механизама релапса, лечења и превенције акутне лимфобластне леукемије.

CURRICULUM VITAE

Željko Antić was born on the 2nd of February 1989 in Belgrade, Serbia. He grew up in the town of Smederevska Palanka, where he finished primary and secondary school. In 2008 Željko moved to Belgrade where he studied Medicine at the University of Belgrade. After graduation in 2014, Željko started his clinical internship at the University Clinical Center of Serbia in Belgrade and passed the medical license exam in 2015. In the same year Željko moved to the Netherlands, where he started his PhD on the genetics of relapsed pediatric acute lymphoblastic leukemia at the Department of Human Genetics of the Radboud University Medical Center, under supervision of Dr. Roland Kuiper. In 2016, this research group moved to the Princess Maxima Center for Pediatric Oncology in Utrecht, where Željko, continued his PhD with Prof. Frank Holstege as his promotor. Since April 2020 Željko has been working as a postdoc on the epigenetics of pediatric leukemia at the Department of Human Genetics of the Hannover Medical School (MHH).

LIST OF PUBLICATIONS

1. Antić Ž, Yu J, Bornhauser BC, Lelieveld SH, van der Ham CG, van Reijmersdal SV, Morgado L, Elitzur S, Bourquin JP, Cazzaniga G, Eckert C, Camós M, Sutton R, Cavé H, Moorman AV, Sonneveld E, Geurts van Kessel A, van Leeuwen FN, Hoogerbrugge PM, Waanders E, Kuiper RP. Clonal dynamics in pediatric B-cell precursor acute lymphoblastic leukemia with very early relapse. Pediatr Blood Cancer. 2022 Jan;69(1):e29361. doi: 10.1002/pbc.29361.

2. Antić Ž*, Yu J*, Van Reijmersdal SV, Van Dijk A, Dekker L, Segerink WH, Sonneveld E, Fiocco M, Pieters R, Hoogerbrugge PM, Van Leeuwen FN, Van Kessel AG, Waanders E, Kuiper RP. Multiclonal complexity of pediatric acute lymphoblastic leukemia and the prognostic relevance of subclonal mutations. Haematologica. 2021 Dec 1;106(12):3046-3055. doi: 10.3324/haematol.2020.259226.

3. Chouvarine P*, Antić Ž*, Lentes J, Schröder C, Alten J, Brüggemann M, Carrillo-de Santa Pau E, Illig T, Laguna T, Schewe D, Stanulla M, Tang M, Zimmermann M, Schrappe M, Schlegelberger B, Cario G, Bergmann AK. Transcriptional and Mutational Profiling of B-Other Acute Lymphoblastic Leukemia for Improved Diagnostics. Cancers (Basel). 2021 Nov 12;13(22):5653. doi: 10.3390/cancers13225653.

4. Boer JM, Valsecchi MG, Hormann FM, Antić Ž, Zaliova M, Schwab C, Cazzaniga G, Arfeuille C, Cavé H, Attarbaschi A, Strehl S, Escherich G, Imamura T, Ohki K, Grüber TA, Sutton R, Pastorczak A, Lammens T, Lambert F, Li CK, Carrillo de Santa Pau E, Hoffmann S, Möricke A, Harrison CJ, Den Boer ML, De Lorenzo P, Stam RW, Bergmann AK, Pieters R. Favorable outcome of NUTM1-rearranged infant and pediatric B cell precursor acute lymphoblastic leukemia in a collaborative international study. Leukemia. 2021 Oct;35(10):2978-2982. doi: 10.1038/s41375-021-01333-y.

5. Antić Ž*, Lelieveld SH*, van der Ham CG, Sonneveld E, Hoogerbrugge PM, Kuiper RP. Unravelling the Sequential Interplay of Mutational Mechanisms during Clonal Evolution in Relapsed Pediatric Acute Lymphoblastic Leukemia. Genes (Basel). 2021 Feb 2;12(2):214. doi: 10.3390/genes12020214.

6. Waanders E*, Gu Z*, Dobson SM*, Antić Ž*, Crawford JC, Ma X, Edmonson MN, Payne-Turner D, van de Vorst M, Jongmans MCJ, McGuire I, Zhou X, Wang J, Shi L, Pounds S, Pei D, Cheng C, Song G, Fan Y, Shao Y, Rusch M, McCastlain K, Yu J, van Boxtel R, Blokzijl F, Iacobucci I, Roberts KG, Wen J, Wu G, Ma J, Easton J, Neale G, Olsen SR, Nichols KE, Pui CH, Zhang J, Evans WE, Relling MV, Yang JJ, Thomas PG, Dick JE, Kuiper

RP, Mullighan CG. Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. Blood Cancer Discov. 2020 Jul;1(1):96-111. doi: 10.1158/0008-5472.BCD-19-0041.

7. Yu J, Waanders E, van Reijmersdal SV, Antić Ž, van Bosbeek CM, Sonneveld E, de Groot H, Fiocco M, Geurts van Kessel A, van Leeuwen FN, Pieters R, Hoogerbrugge PM, Kuiper RP. Upfront Treatment Influences the Composition of Genetic Alterations in Relapsed Pediatric B-Cell Precursor Acute Lymphoblastic Leukemia. Hemasphere. 2020 Jan 22;4(1):e318. doi: 10.1097/HS9.0000000000000318.

8. Yu J, Antić Ž, van Reijmersdal SV, Hoischen A, Sonneveld E, Waanders E, Kuiper RP. Accurate detection of low-level mosaic mutations in pediatric acute lymphoblastic leukemia using single molecule tagging and deep-sequencing. Leuk Lymphoma. 2018 Jul;59(7):1690-1699. doi: 10.1080/10428194.2017.1390232.

*Authors contributed equally as lead authors.

ACKNOWLEDGEMENTS

In June 2015 I embarked on the adventure of obtaining a PhD, which led me to move from Serbia to the faraway kingdom of the Netherlands. Looking back, I was able to grow professionally, but what I did not expect in 2015 is how the next five years in the Netherlands would influence my personal growth, as well as the way I perceive the world and the country I come from. Now, when this adventure is coming to an end, I wish to thank all the colleagues in Radboudumc, Princess Maxima Center, Hubrecht Institute, as well as friends, acquaintances and unexpected companions with whom I crossed paths and who made this adventure memorable. In particular, I would like to thank my supervisor Roland Kuiper and promotor Frank Holstege.

Frank, thank you for your help and input during the last stages of my PhD. Your comments really helped in bringing this thesis to the higher level.

Roland, thank you for believing in me and giving me the opportunity to work with you. Your calm and emphatic nature was truly inspiring and will always serve as an example in my future work. Thank you for your patience and for selflessly sharing your knowledge. Thank you for showing me how to change the perspective and helping me think outside the box.

Ad, it was a great privilege to get to know you and work with you. Thank you for your continuous support after group moved to Utrecht. Thank you for your help in the preparation of manuscripts, and your insightful comments, which really improved the quality of this thesis.

I owe special thanks to all previous and current members of the Kuiper group, **Eveline**, **Lilian**, **Simon**, **Renske**, **Sabrina**, **Marcel**, **Jiangyan**, **Richarda**, **Esmé**, **Robbert**, **Junxiao**, **Judith**, **Marjolijn**, **Reno**, **Freerk**, **Lionel**, **Rajith**, **Stefan**, **Nienke**, **Jette**, **Janna**, **Cédric**, my paranymphs **Dilys** and **Mariangela** as well as all the students and guest scientists. Thank you for your support, guidance and for making this journey special.

Dilys, you are a talented scientist with a bright future and your fun and cheerful personality made it a real pleasure to work with you. Thank you for your encouragement and all the help in the last steps of my PhD.

Mariangela, I really enjoyed our discussions about Italian language and all the snacks that you were bringing after coming back from your holidays in Italy. Your dedication to your faith, in a world where people tend to trust less and less to one another, as well

as your resourcefulness and perseverance when performing all those southern blot experiments, were truly inspiring to me. I hope one day my Italian gets good enough so I can have even longer chats with your mom. Thank you for all your encouragement and motivation during my PhD.

Stefan, also known under his pseudonym "vakantieman". It was great to have you as a colleague again, and I really enjoyed working with you during my last year in Maxima. Thank you for all your help and support with my thesis after I left Maxima.

Jiangyan, my dearest queen Bee, thank you for helping me make my transition into a new culture so smooth, for teaching me and being my best friend in the Netherlands, and not making fun of my Chinese. Thank you for choosing me as Julie's godfather. I am so glad you were able to avoid her being eaten by Minoes!

Simon, thank you for being the best party organizer in Nijmegen! I always had lots of fun when you were around, be it in Doornroosje, lab, or at the lunch table. Thank you for teaching me lab techniques and for always having a lot of patience with me.

Esmé, thank you for your guidance and valuable insights. It was a great opportunity to work with you on the Blood Cancer Discovery paper. I learned a lot from that experience.

I would also like to extend my gratitude to all the members of the van Leeuwen group. I would particularly like to acknowledge **Laurens**, **Dorette**, **René**, **Miriam** and **Trisha**. I had great time working with you and thank you for your support and help throughout my PhD. **Frank**, thank you for your advices, all the brainstorming sessions and support over the past years.

The writing of this thesis was greatly aided by **Prof. Peter Hoogerbrugge**. Thank you for providing valuable insights into treatment and bedside management of children with leukemia. Your comments were of great help in presenting the data in a way that is interesting for both basic scientists and clinicians.

I would like to thank to all "freezer room" members for making every day at work fun and enjoyable. **Lindy**, **Jordy**, **Kaylee**, **Luke**, **Tomasz**, **Susan**, **Winnie**, **Marit**, **Waleed**, **Maroussia** and **Tamara**, it was a real pleasure having you in the office and thank you for making every day special. Special thanks to all freezer room toddlers. It's great seeing you all grown up and some even back to Maxima!

Patricia, I always enjoyed our discussions and during those moments time seemed to fly. You are one of the most knowledgeable people and scientists I have ever met and I am proud to be calling you a friend.

Thomas, many thanks for your enthusiasm and always bringing good mood to the office. Thank you for being a friend and sharing a piece of your amazing floor with me, which is now exhibited on a prominent place in my living room.

Leiah, I am so glad I got to know you and I really enjoyed our coffee breaks. Your humor perfectly fits Serbian mentality and thanks to this you made me feel like I am finally at home. According to the etiquette established by George Lucas, it is only suitable to end this paragraph with an appropriate greet to Princess Leiah - May the Force be with you!

Elvin, thanks for all your pessimism. It actually made me feel that my life is not that bad! I am very glad I got to know you and I really enjoyed our political debates.

I also want to thank to all the PhD students in the PhD room of the Human Genetics Department in Radboudumc, and in particular **Francesca**, **Sarita**, **Zhaomin** and **Eugenia**. I really loved spending time with you and I really missed you after moving to Utrecht.

I would like to thank to two persons, who helped me make my first steps in science and guided me as a medical student at the University of Belgrade, **Sonja Misirlić-Denčić** and **Anđelka Isaković**.

Драга **Соња** и **Анђелка**, велико хвала на вашем труду, стрпљењу, несебичном дељењу знања, и подршци коју сте ми пружили током свих ових година.

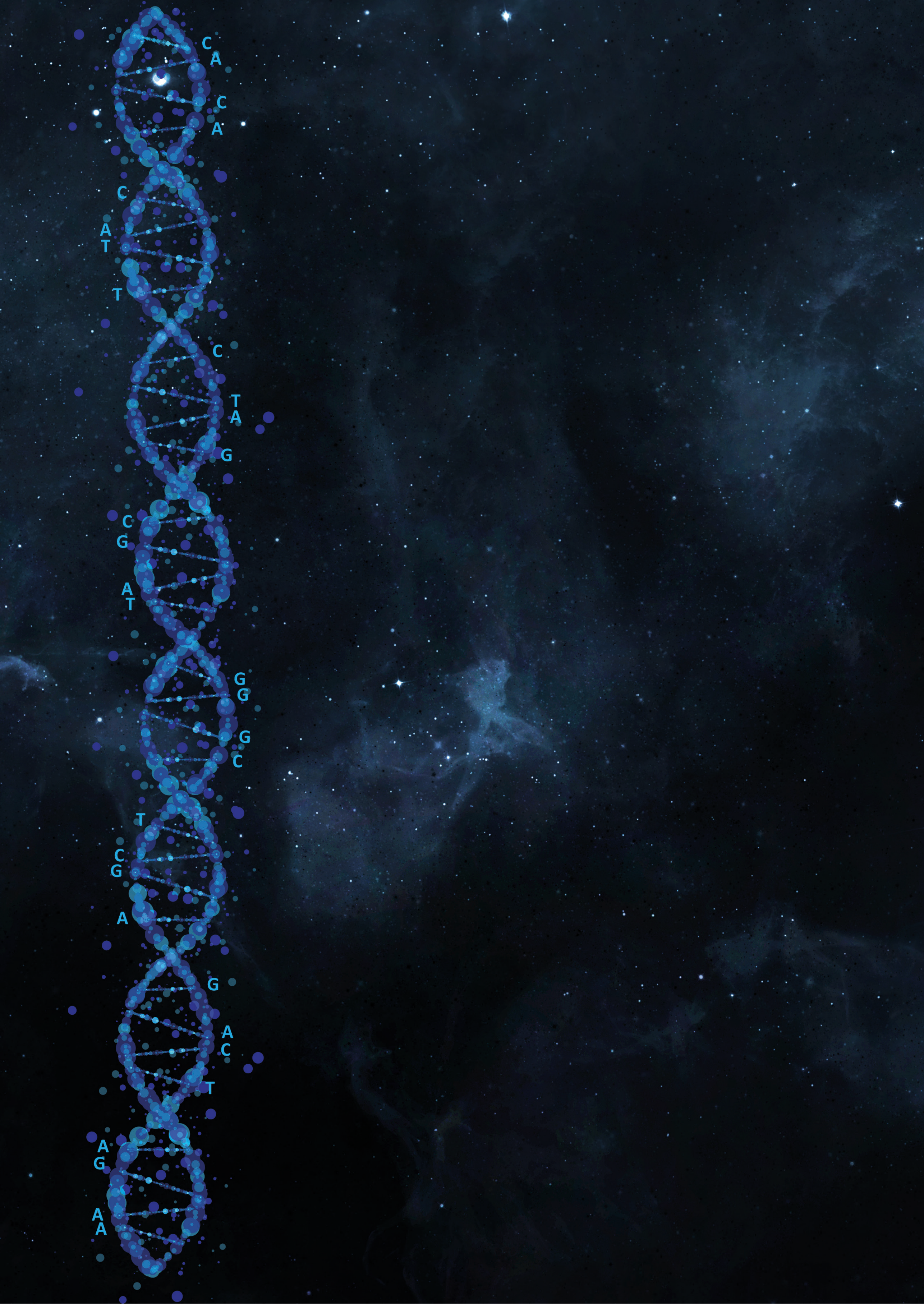
Finally, special thanks to my family and friends, who helped and encouraged me over the years.

Драга **Машо**, **Немања**, **Александра**, **Милице** и **Ведране**, велико вам хвала на подршци и свему што сте учинили за мене у протеклих 10. година. Хвала што сте остали близу иако нас је даљина раздвајала.

Драги **Мирославе**, хвала ти на охрабрењу да се отиснем у ову авантуру и на подршци коју си наставио да ми пружаш током протеклих година.

Драги **Алберто**, велико хвала на подршци у протекле 4. године, стрпљењу и помоћи у организацији одбране дисертације. Хвала на додавању недостајућих чланова, правом тревижанском тирамисуу, и венецијанској магли.

Драги мама и тата, драги бато хвала вам на свим одрицањима, подршци и стрпљењу, чак и онда када га ја нисам имао за вас. Ову тезу посвећујем вама.


C
A
C
A
C
A
T
T
C
T
A
G
C
G
A
T
G
G
G
C
T
C
G
A
G
A
C
T
A
G
A
A