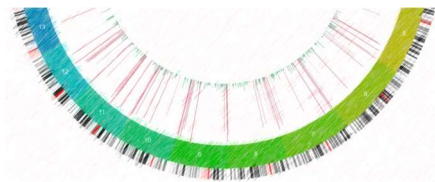


Decoding Germline and Somatic Genomes of Children with Cancer through Optical Genome Mapping and Next-Generation Sequencing



Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Danielle Brandes

aus Aachen

Düsseldorf, September 2025

aus dem Institut für
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Arndt Borkhardt
2. Prof. Dr. Sebastian Wesselborg

Tag der mündlichen Prüfung: 23.01.2026

EIDESSTATTLICHE ERKLÄRUNG

Ich versichere an Eides Statt, dass die Dissertation „Decoding Germline and Somatic Genomes of Children with Cancer Through Optical Genome Mapping and Next-Generation Sequencing“ von mir selbständig und ohne unzuverlässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Die vorliegende Dissertation wurde in dieser oder ähnlicher Weise noch an keiner anderen Fakultät vorgelegt und es wurde kein vorheriger Promotionsversuch unternommen.



Danielle Brandes

Düsseldorf, September 2025

TABLE OF CONTENTS

Eidesstattliche Erklärung	3
Table of contents	4
List of Abbreviations	6
Summary	8
Zusammenfassung	10
1. Introduction	12
1.1 The human genome	12
1.1.2 Human reference genomes	12
1.1.3 Human genetic variation and its role in disease.....	13
1.1.3.1 Germline and somatic genetic variation	15
1.1.4 Methods to resolve genomic structural variation	15
1.1.4.1 Optical genome mapping	17
1.1.4.2 Single molecule real-time (SMRT) long-read sequencing.....	19
1.1.5 Interpretation of genomic variation	21
1.2 Pediatric cancer	22
1.2.1 Etiology of pediatric cancer	23
1.2.2 Genetic cancer predisposition in children	24
1.2.2.1 Cancer predisposition syndromes	24
1.2.2.2 Germline genetic variation in children with cancer	26
1.3 Leukemia.....	26
1.3.1 Acute lymphoblastic leukemia.....	27
1.3.2 B-cell acute lymphoblastic leukemia	28
1.3.2.1 Hyperdiploid BCP-ALL	30
1.3.2.1.1 Clinical features	30
1.3.2.1.2 Etiology and genetic characteristics	31
1.3.2.2 <i>ETV6::RUNX1</i> BCP-ALL	32
1.3.2.2.1 Clinical features	32
1.3.2.2.2 Etiology and genetic characteristics	32
1.3.3 Diagnosis and treatment of childhood acute lymphoblastic leukemia	33
1.3.3.1 Diagnostic work-up of acute lymphoblastic leukemia	34
1.3.3.2 Treatment of childhood acute lymphoblastic leukemia.....	35
2. Aims of the thesis.....	36
3. Original Publications	37
3.1 Summary of Original Publication I	37
3.2 Original Publication I.....	39

3.2.1 Original Publication I: Supplementary Methods and Figures	47
3.2.2 Original Publication I: Supplementary data	51
3.3 Summary of Original Publication II	53
3.4 Original Publication II	55
3.4.1 Original Publication II: Supplementary Figures	69
3.4.2 Original Publication II: Supplementary data	79
4. Discussion	81
4.1 Utility of OGM in clinical routine testing	81
4.2 Limitations of OGM in clinical routine testing	82
4.3 OGM and NGS-based approaches in constitutional disease	83
4.3.1 Clinical utility of OGM in constitutional disorders	83
4.3.2 Advantages of long-read sequencing	84
4.3.3 Current limitations of advanced technologies in constitutional diagnostics	85
4.4. Genetic germline variation in pediatric cancer predisposition	86
4.4.1 Genotyping approaches in pediatric cancer patients	86
4.4.2 Contribution of structural variation and multi-omics profiling	87
4.4.3 Adult-onset CPGs and pediatric cancer	88
4.4.4 Candidate cancer predisposition genes and functional characterization	89
4.5. OGM and NGS-based approaches in hematological malignancies	90
4.5.1 Clinical utility of OGM in hematological malignancies	90
4.5.2 Strategies of OGM data integration and analysis	92
4.5.3 Long-read sequencing in hematological malignancies and other cancers	94
4.6. Contribution of structural variation in somatic landscapes of pediatric ALL	94
4.6.1 Recurrent alterations and mechanistic insights	95
4.6.2 Relapse-associated alterations	96
4.6.3 Non-coding variation and functional characterization	97
5. Conclusion and outlook	99
References	102
List of figures	115
List of tables	115
APPENDIX	i
PUBLICATIONS	i
SCIENTIFIC PRESENTATIONS	ii
FUNDING & AWARDS	iii
ACKNOWLEDGMENTS	iv

LIST OF ABBREVIATIONS

ACMG	American college of medical genetics and genomics
AI	Artificial intelligence
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
BCP-ALL	B-cell precursor acute lymphoblastic leukemia
BCR	B-cell receptor
CCS	Circular consensus sequencing
CLL	Chronic lymphoid leukemia
CLR	Continuous long-read
CML	Chronic myeloid leukemia
CBA	Chromosomal banding analysis
CN	Copy number
CNS	Central nervous system
CNV	Copy number variation
CPG	Cancer predisposition gene
CPS	Cancer predisposition syndrome
CRE	Cis-regulatory element
DN	De novo assembly
dNTP	Deoxynucleotide triphosphates
E::R	<i>ETV6::RUNX1</i>
FISH	Fluorescence in situ hybridization
GWAS	Genome wide association studies
HBOC	Hereditary breast and ovarian cancer
HD	Hyperdiploid
HGP	Human Genome Project
HiFi	High fidelity
HMW	High molecular weight
HR	High risk
InDel	Insertion and deletion
IR	Intermediate risk
Kbp	kilobase pairs
LOH	Loss of heterozygosity
Mb	Megabases
MAR	Minimal altered region
MDS	Myelodysplastic syndrome
MLPA	Multiplex ligation-dependent probe amplification
NB	Neuroblastoma
NGS	Next-generation-sequencing
OGM	Optical genome mapping
ONT	Oxford nanopore technology
PacBio	Pacific Biosciences
RVP	Rare variant pipeline
SMRT	Single molecule real-time
SNP	Single nucleotide polymorphism

SNV	Single nucleotide variants
SR	Standard risk
SV	Structural variant
TAD	Topologically associated domain
T-ALL	T-cell acute lymphoblastic leukemia
TCR	T-cell receptor
T2T	telomere-to-telomere
VAF	Variant allele frequency
WBC	White blood cell count
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WHO	World Health organization
ZMW	Zero-mode waveguides

SUMMARY

Genomic variation plays a significant role in pediatric cancers, contributing to both oncogenic somatic driver events and inherited cancer predisposition. While pathogenic single nucleotide variants have been extensively described and are well characterized, the contribution of larger structural variants (SVs) in this context are less understood. The full impact of these alterations still needs to be elucidated. Traditional cytogenetic methods and short-read based next-generation sequencing (NGS) approaches have limitations in comprehensively detecting the full spectrum of such larger genomic lesions. Optical genome mapping (OGM) and long-read sequencing are emerging new technologies that are capable of reliably detecting SVs in constitutional and cancer genomes. This thesis is based on two original publications that explore the application of advanced genomic technologies and NGS-based integration in distinct contexts of pediatric cancer diagnosis and research.

The first original publication focuses on patients with a potential underlying cancer predisposition syndrome but non-conclusive results from whole-exome sequencing (WES). OGM was applied to perform an unbiased screening of germline structural variation in 34 pediatric cancer patients, including available parental samples. The study identified a median of 49 rare SVs per patient and three *de novo* alterations based on 18 child-parent trios. Notably, two significant germline alterations were highlighted that were discovered in two respective leukemia patients: (i) a likely pathogenic deletion of exon three in the known cancer predisposition gene *BRCA2*, segregating within a family with a history of breast cancer; and (ii) a *de novo* duplication affecting a potentially new cancer predisposition gene candidate *RPA1*. Functional testing of patient-derived cells compared to non-affected cells showed that *RPA1* duplication might increase the sensitivity to genotoxic agents.

The second original publication applied OGM with multi-omics integration in a cohort of 60 pediatric BCP-ALL cases focusing on the two most common genetic subtypes, *ETV6::RUNX1* (E::R) translocated and classical hyperdiploid (HD) BCP-ALL. OGM was combined with conventional cyto- and molecular genetic methods, WES data, and targeted validation through long-read and RNA sequencing to comprehensively characterize the genetic somatic landscape and novel recurrent genetic events. OGM demonstrated a 95% concordance with SVs in diagnostic samples previously reported by molecular cytogenetics and SNP-arrays. As frontline method, OGM detected 677 additional SVs that were missed by regular diagnostic testing, including prognostically relevant focal deletions affecting genes such as *IKZF1*. The study

further demonstrated that SV profiles of E::R and HD BCP-ALL are profoundly different, with E::R cases harboring 2.7-fold more SVs, particularly focal deletions, compared to HD BCP-ALL. Integration of OGM and WES data identified 52 minimal altered regions (MARs) in the analyzed pediatric BCP-ALL cohort, including 19 recurrent regions that reveal potential novel leukemic drivers such as *FOCAD*, *IKBKB*, *MIB1* as well as double-hit targets *STAG2* and *MANBA*. Additionally, nine translocations were detected that may lead to novel in-frame fusion genes. In conclusion, this thesis contributes to the development of a genomic reference platform for the identification and interpretation of structural variation in both germline and somatic genomes of children with cancer. Both presented studies demonstrate the advantage of integrating genomic technologies such as OGM with NGS-based methods. Together, these findings reveal a more comprehensive spectrum of genomic alterations, offering novel insights into pediatric cancer predisposition and mechanisms underlying leukemogenesis of pediatric BCP-ALL. This work provides a foundation for future functional studies and clinical investigations aimed at improving diagnosis, risk stratification and treatment in pediatric oncology.

ZUSAMMENFASSUNG

Genomische Variation spielt eine bedeutende Rolle bei Krebserkrankungen des Kindesalters und trägt sowohl zu genetischer Prädisposition in der Keimbahn als auch zu somatischen onkogenen Treibermutationen bei. Während pathogene Einzelbasenpaar-Veränderungen umfangreich beschrieben wurden, ist der Beitrag von strukturellen Varianten (SV) in diesem Zusammenhang weniger erforscht, und die Auswirkungen dieser Veränderungen müssen noch aufgeklärt werden. Traditionelle zytogenetische Methoden und „*short-read*“ Sequenzierungsmethoden stoßen an ihre Grenzen das gesamte Spektrum größerer Veränderungen umfassend zu detektieren. Die optische Genomkartierung (OGM) und „*long-read*“ Sequenzierung haben sich als neue Technologien etabliert, strukturelle Variation im Kontext von konstitutionellen Erkrankungen und Krebs zuverlässig zu erkennen. Diese Dissertation basiert auf zwei Originalpublikationen, die den Einsatz dieser Technologien sowie deren Integration mit sequenzierungsbasierten Methoden in verschiedenen Kontexten der Krebsdiagnostik und -forschung von kindlichen Tumoren behandeln.

Die erste Studie konzentriert sich auf Kinder mit Krebserkrankungen und einem klinischen Verdacht auf ein zugrundeliegendes Krebsprädispositionssyndrom und einem unklaren Befund aus der Exom-Sequenzierung (WES). OGM wurde für ein Keimbahn-Screening der strukturellen Variationen bei 34 Kindern mit Krebs sowie deren Eltern eingesetzt. Die Studie identifizierte einen Median von 49 seltenen SV pro Fall sowie drei *de novo*-Veränderungen in 18 Kind-Eltern-Trio-Analysen. Darüber hinaus hebt die Publikation zwei signifikante Keimbahnveränderungen hervor, die bei zwei Leukämiefällen entdeckt wurden: (i) eine pathogene Deletion von Exon 3 im bekannten Krebsprädispositionsgen *BRCA2*, die innerhalb einer Familie mit Brustkrebs auftrat; und (ii) eine *de novo* Duplikation, die ein potenziell neues Krebsprädispositionsgen *RPA1* betrifft und in den Patientenzellen verglichen zu nicht betroffenen Patientenzellen eine Sensitivität gegenüber genotoxischen Reagenzien verursachen könnte.

Die zweite Studie setzte OGM und Multi-Omics-Integration in einer Kohorte (n=60) von den zwei häufigsten genetischen BCP-ALL Subtypen ein, die eine *ETV6::RUNX1* (E::R) Translokation oder eine klassische Hyperdiploidie aufwiesen. OGM wurde mit konventionellen zytogenetischen und molekulargenetischen Methoden sowie WES-Daten kombiniert und durch Validierung mittels „*long-read*“ und RNA-Sequenzierung ergänzt, um die somatische genetische Landschaft umfassend zu charakterisieren und neue wiederkehrende genetische Events zu identifizieren. OGM zeigte eine 95%ige Übereinstimmung der SVs in diagnostischen

Proben im Vergleich zu zuvor mittels molekularer Zytogenetik und SNP-Array berichteten Befunden. Als primäre Methode erkannte OGM zusätzliche 677 SV, die durch die Standardtests übersehen wurden, darunter auch prognostisch relevante Deletionen, die beispielsweise das *IKZF1* Gen betrafen. Die Studie zeigte darüber hinaus, dass sich die SV-Landschaften von E::R und HD BCP-ALL deutlich unterscheiden, wobei E::R-Fälle die 2.7-fache Anzahl an SVs, insbesondere fokale Deletionen, aufzeigten. Zudem wurden 9 Translokationen entdeckt, die potentiell neuartige Fusionsgene erzeugen. Die Kombination aus OGM- und WES-Daten identifizierte 52 minimal veränderte Regionen (MAR) in den untersuchten BCP-ALL Fällen, darunter 19 wiederkehrende Regionen, die potenziell neue Gene in der Leukämieentstehung bedeuten könnten, unter anderem *FOCAD*, *IKBKB*, *MIB1* und *STAG2*, *MANBA* die durch Doppeltreffer verändert wurden.

Zusammenfassend trägt diese Dissertation zur Entwicklung einer genomischen Referenzplattform zur Identifikation und Interpretation von strukturellen Variationen in Keimbahn- und Tumorgenomen krebskranker Kinder bei. Beide originalen Publikationen zeigten den Vorteil der Integration genomischer Technologien wie der optischen Genomkartierung mit NGS-basierten Methoden. Die dargelegten Ergebnisse offenbaren ein umfassendes Spektrum genomischer Veränderungen in kindlichen Krebserkrankungen und liefern neue Erkenntnisse über genetische Prädisposition und die Leukämogenese. Diese Arbeit bildet eine Grundlage für anschließende funktionelle und klinische Untersuchungen, die auf eine Verbesserung von Diagnostik, Risikostratifizierung und Therapieansätzen bei Kindern mit Krebs abzielen.

1. INTRODUCTION

1.1 The human genome

The human genome encodes approximately three billion base pairs of DNA, organized in 23 pairs of chromosomes. Each individual inherits one chromosome of each pair from each parent. A single chromosome consists of two identical sister chromatids joined at a constricted region known as centromere which consist of highly complex DNA compartments [1]. The centromere divides the chromosome into a short p- and a long q-arm. It includes the kinetochore region which regulates chromosome attachment to the spindle apparatus ensuring accurate chromosomal segregation during cell division [1]. Each chromatid starts and ends with a telomere region with highly repetitive DNA sequences that protect the chromosomes from degradation [2]. DNA is packaged around histone proteins into a complex known as chromatin which condenses to form visible chromosomes during cell division. Depending on the compression degree, chromatin can be subdivided into accessible euchromatin and inaccessible heterochromatin [3]. The largest proportion of the genome consists of non-coding DNA, which plays regulatory, structural or yet unknown roles, while only about 1-2% is protein-coding [4]. These regulatory elements include transcription start sites of a gene which is composed of a promotor region, typically spanning <1 kilobase pairs (kbp) [5]. Distal regulatory elements comprise e.g. enhancer or silencer regions that can regulate gene transcription from distances up to 1 Megabases (Mb) [5]. Topologically associated domains (TADs) are structural domains that organize the 3D genome by grouping genes and their regulatory elements into insulated functional units [6]. DNA and its regulation are essential for development, function and health of the human body building the basis for protein synthesis and cellular processes.

1.1.2 Human reference genomes

As a first milestone, the Human Genome Project (HGP) was established with the goal of sequencing the human genome. The consortium created a reference genome as representative example of the genetic code to compare human DNA sequences and facilitate annotation of genomic variation [7, 8]. This reference genome is a consensus sequence as a single representation of multiple genomes [7]. The first major genome assembly, GRCh35 (hg17), was released by the HGP in 2004 [9], followed by more accurate and improved versions (GRCh37, hg19 and GRCh38, hg38) up to the first complete human reference genome from the telomere-

to-telomere (T2T) consortium from one haplotype-resolved individual (T2T-CHM13) in 2022 [10]. The T2T reference filled the 8% of remaining gaps of the previous genome references covering the entire length of every chromosome [10]. These advances were achieved through ultra- and high-fidelity long-read sequencing, optical genome mapping and other methods (refer to section 1.1.4). In contrast to linear genome references, the concept of pangenomes has emerged in recent years. Pangenomes capture the genetic diversity from many individuals and ethnicities within the same species and circumvents a reference bias by comparing multiple alternate alleles offering a more complex representation of genomic variation [11].

While the sequence of the human genome is largely complete, ongoing research continues to explore gene functions, non-coding elements, genetic variation and their implications for health and disease.

1.1.3 Human genetic variation and its role in disease

Human genetic variation arises from four main classes of genetic alterations: single nucleotide variants (SNVs), small insertions and deletions (indels, <50bp) and structural variants (SVs, >50bp) (Figure 1). Larger chromosomal abnormalities are often referred to as copy number variation (CNV) including whole chromosome gains or losses. CNVs can result in aneuploid genomes containing more or less chromosomes than the normal human karyotype. SVs are further subdivided into deletions (loss of genetic material), duplications (gain of genetic material), inversions (copy number neutral rearrangements) and translocations. Translocations involve the exchange of material between two or more chromosomes (interchromosomal) or within one chromosome (intrachromosomal) and can be balanced (copy number neutral) or unbalanced resulting in loss or gain of genetic material. Another subclass of SVs includes insertions, which represents the addition of genetic material into a genomic region that is not present in the reference sequence. SNVs and SVs arise through different mutagenic processes. SNVs typically result from errors in DNA replication or repair of single-strand damage, whereas SVs are generated by errors in DNA double-strand break repair, mitotic or meiotic recombination or chromosomal missegregation leading to numerical chromosomal abnormalities [12]. SVs are typically formed through mechanisms such as non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ) upon DNA double-strand break or non-allelic homologous recombination (NAHR) during meiotic crossing-over [13].

While SNVs outnumber other types of variants with >4,000,000 variations per genome, SVs make up more than 10Mb of the genome affecting approximately 81% of the genome [14]. SVs

largely contribute to individual diversity with an average of 2.500 SVs per individual genome and are 50-fold more likely to alter gene expression compared to SNVs [15]. Germline SVs are therefore widely contributing to differences in gene expression in humans [15]. Their pathogenic potential has been implicated in a wide range of genetic disorders and disease processes [16]. Notably, more than 98% of the human genome may be altered by genetic variation in the non-coding fraction like intronic or intergenic regions. Non-coding lesions can affect cis-regulatory elements (CREs), that regulate gene expression during developmental processes and cellular differentiation and are broadly classified as promoters, enhancers and insulators [17]. Intergenic regions are further relevant for organizing the genome into compartmentalized domains with preferential chromatin interactions [18]. Disruption of TAD boundaries can lead to *de novo* contacts between CREs and gene dysregulation [18]. Regulatory variation by influencing gene expression and genomic architecture can thereby contribute to inherited disease [19] and cancer biology [20, 21].

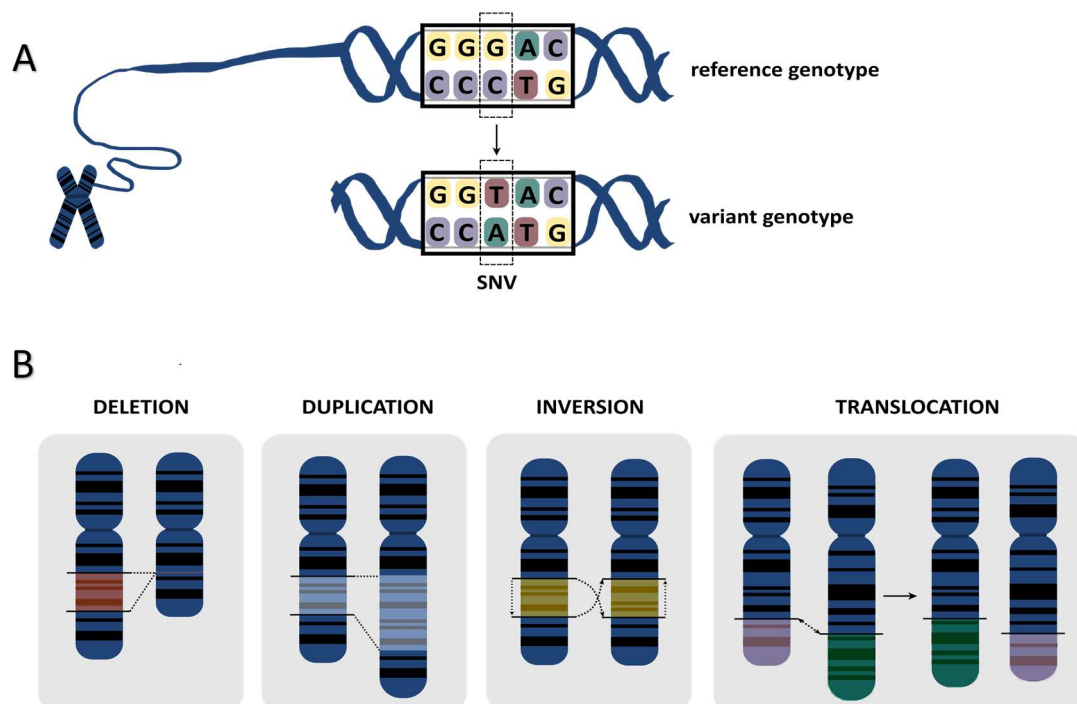


Figure 1: Types of human genetic variation.

(A) Chromosomes consist of DNA that is wrapped around histone proteins with heterochromatin being more condensed and with inactive gene transcription; and euchromatin being more open and with active gene transcription (depicted as dark and light blue chromosomal bands). The genetic code consists of DNA sequences of five distinct nucleotides: adenine (green), thymine (red), guanine (yellow) and cytosine (purple). Single nucleotide variants are single base pair changes in comparison to human reference genome. **(B)** Structural variants are genomic alterations >50bp and can be subdivided into deletions leading to loss of material (red), duplications leading to gain of material (blue), inversions (yellow) and translocations. An interchromosomal translocation is depicted showing the exchange of material of two chromosomes (purple and green). Two DNA double-strand breaks (black lines) are needed within one or two chromosomes for the structural alteration to occur. Abbreviations: single nucleotide variant (SNV).

1.1.3.1 Germline and somatic genetic variation

Germline genetic variants are inherited or arise *de novo* in parental germ cells or during very early embryonic development. An allele is a variable version of a DNA sequence at a specific genomic locus. Each individual inherits one allele from each parent with identical alleles being homozygous and different alleles being heterozygous. A set of alleles that are physically grouped together on the same chromosome segment tend to be inherited together and are referred to as haplotype [22]. As constitutional alterations, they are present in every cell of the body. Germline mutations accumulate over time and vary among individuals, reflecting dynamic genetic signatures across populations and generations that are shaped by evolutionary factors [12]. In contrast, somatic mutations accumulate within cells from one individual from the beginning of early development and continuing throughout life. These mutations are driven by endogenous and exogenous influences, resulting in selection pressure on cells [23]. A mixture of germline and somatic variation, including SNVs and SVs, shape human phenotypic diversity and can both contribute to risk factors for a wide range of human disease [12], including cancer [24]. Genetic cancer predisposition can be caused by pathogenic germline variants that lead to earlier onset of disease compared to the general population, contributing to the occurrence of cancer in children and adolescents [25-31] (refer to section 1.2.2). A comprehensive characterization of both SNVs and SVs, across coding and non-coding regions, is crucial to fully understand the spectrum of genetic diversity, identify pathogenic mutations in affected individuals, and elucidate their role in disease progression.

1.1.4 Methods to resolve genomic structural variation

Structural alterations can be identified using conventional cytogenetic and molecular methods such as karyotyping via chromosomal banding analysis (CBA), fluorescence *in situ* hybridization (FISH) and array-based methods including comparative genomic hybridization (CGH) or single nucleotide polymorphism (SNP)-array (Figure 2). These techniques are widely implemented in clinical diagnostic workflows to detect structural aberrations in both constitutional disease and cancer. CBA is the oldest method that stains chromosomal bands to assemble karyotypes from metaphase chromosomes identifying genomic alterations >5Mb. FISH enables the detection of specific target DNA sequences by hybridization of fluorescently labeled probes, while array-based methods analyze the samples by hybridizing patient DNA to microarrays containing either sequence-specific probes (CGH) or polymorphic loci (SNP-array). Variations in fluorescence intensity is used to computationally infer copy number variations. Due to several

limitations of these conventional methods such as low resolution, required prior knowledge of SVs or inability to detect balanced aberrations, the complete spectrum of structural aberrations remains to be explored [32]. Since second generation next-generation-sequencing (NGS) methods emerged, SNVs and short indels can be characterized by high-throughput short-read sequencing like whole-exome sequencing (WES) or whole-genome sequencing (WGS) (from Illumina) with a high accuracy [33]. However, short-read based approaches like WGS have limited ability to characterize larger SVs, mainly due to their short-read length (usually 100-150bp) and challenges of mapping reads in complex and repetitive regions of the genome [34]. A critical step is the correct alignment of sequences to the reference genome that is difficult in repetitive or low-complexity regions and for split reads that align to different regions of the reference. Notably, many SVs are thought to arise in such difficult repetitive regions [35], which make up two thirds of the human genome, particularly consisting of transposable elements and segmental duplications [7]. This further highlights the need for other technologies to fully characterize the spectrum of genomic variation.

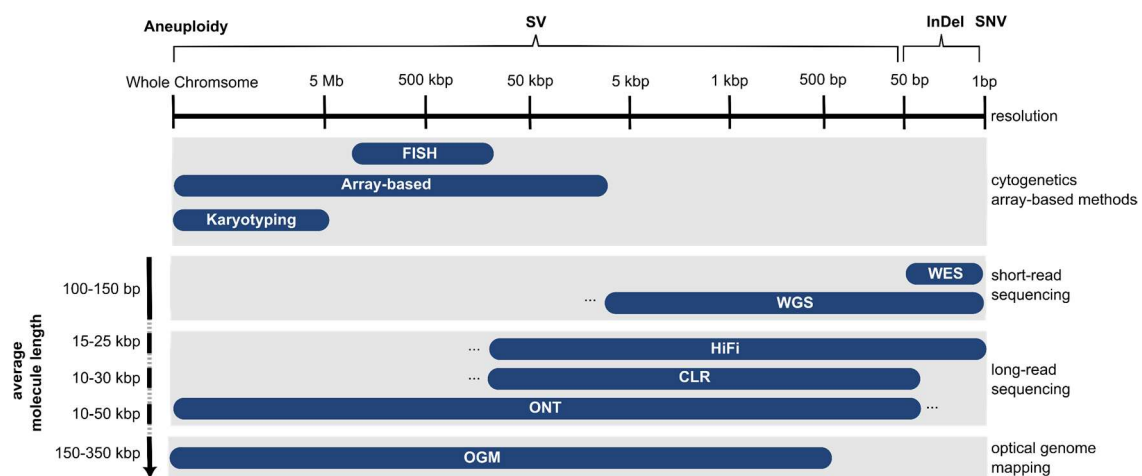


Figure 2: Technical approaches to detect genomic variation: from SNV to aneuploidy.

Overview of genomic variant detection methods by variant size and molecular resolution (x-axis). The graph summarizes the capability of variant detection using various cytogenetic, array-based and sequencing methods across the spectrum of genomic variation, ranging from SNVs, indels, SVs up to whole chromosome aneuploidy. The effective detection range for each technique is depicted in blue bars. For sequencing approaches and OGM, the average molecule length is indicated on the y-axis. Traditional cytogenetic approaches such as FISH and karyotyping are effective for large-scale variants, while NGS approaches as short-read WES/WGS and long-read platforms from PacBio (CLR + HiFi) and ONT extend the detection of smaller variants. HiFi sequencing has a higher accuracy compared to CLR sequencing which enables detection of SNVs. OGM covers a broad spectrum of large SVs with the highest average molecule length. Abbreviations: single nucleotide variant (SNVs), insertion/deletion (InDel), structural variant (SV), fluorescence in situ hybridization (FISH), whole-exome sequencing (WES), whole-genome sequencing (WGS), high-fidelity (HiFi), continuous long-read (CLR), Oxford nanopore technology (ONT), optical genome mapping (OGM), next-generation-sequencing (NGS). Figure partly adapted from Bionano Genomics.

In recent years, third generation long-read sequencing methods and optical mapping platforms have been developed, offering a more robust approach to resolve larger and complex SVs. Optical genome mapping (OGM), is a novel mapping approach that analyzes ultra-long DNA molecules and is able to detect SVs >500bp (refer to section 1.1.4.1). The two leading long-read sequencing technologies include single molecule real-time (SMRT) sequencing by Pacific Biosciences (PacBio, refer to section 1.1.4.2) and the Oxford nanopore sequencing technology (ONT) [36]. Unlike short-read sequencing, these technologies generate long sequencing reads (>10kbp in length) from native DNA without the need for PCR amplification. They further have demonstrated improved performance in detecting SVs, largely contributing to the detection of alterations in low mappability and repetitive regions [37]. A comprehensive cross-platform approach applied by the Human Genome Structural Variation Consortium resulted in a seven-fold increase in the detection of the number of SVs in human reference samples compared to previously identified variation based on short-read sequencing, with PacBio long-read data contributing to the majority of novel variation [38]. This study showed that OGM outperformed all other methods in detecting large deletions and duplications ranging from 10kbp to 10Mb. Further, whole-genome sequencing or mapping approaches have the advantage to detect variation in the non-coding genome and thus provide a more comprehensive view on genetic variation compared to WES and targeted approaches.

1.1.4.1 Optical genome mapping

Over the last years, DNA-based optical genome mapping has been used to identify SVs and aneuploidy in constitutional and cancer disease as next-generation cytogenetic tool [39-47]. Most studies applied the OGM technology that was developed by Bionano Genomics and advocate OGM as next-generation cytogenetics tool due to its feasible implementation in clinical routine testing [48-51]. Especially in the diagnostic set-up of hematological malignancies, OGM showed a high concordance with conventional testing methods and the detection of clinically relevant SVs that have significant implications for risk stratification and treatment decisions [40, 42, 45, 49, 51-54]. OGM is based on imaging of fluorescently tagged DNA molecules, converting them into *in silico* maps that are used for down-stream analysis and SV detection [41, 53] (Figure 3).

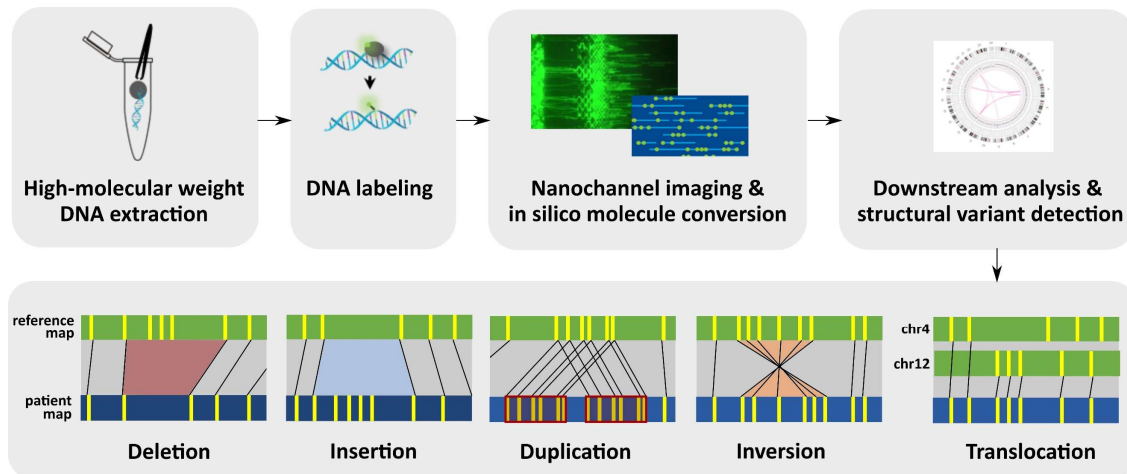


Figure 3: Technical workflow of optical genome mapping.

A simplified workflow for generation of OGM data is shown. The OGM workflow begins with the extraction of HMW DNA with molecules >150kbp extending to several megabases in size. Following DNA precipitation, HMW DNA is bound to nanobind disks which are used to wash and elute the DNA to minimize shearing and to preserve the integrity of long DNA fragments. The purified DNA is subjected to a labeling reaction where label patterns are introduced along the DNA molecules through a direct label enzyme (DLE-1). The DLE-1 globally recognizes a –CTTAAG– motif that is present on average every 5kbp in the genome and covalently connects a DL-Green fluorophore marker at this specific target site. The labeled DNA is loaded onto a chip where each DNA molecule is linearized within nanochannels and is subsequently imaged by a nanochannel scanner from the Bionano instrument. The captured fluorescent DNA sequence images are computationally processed and converted into digital in silico representations of the DNA molecules. These digital maps reflect the spatial arrangement of labels across the genome and are used for further down-stream analysis and SV detection according to the motif. Based on differences in label spacing or deviations of label patterns from the reference genome map distinct types of SVs such as deletions, insertions, duplications and translocation can be identified. The optical genome maps with specific label patterns (yellow) of the investigated DNA sample (e.g., patient sample, blue) are compared to a reference genome map (green) to detect deletions (red), insertions (light blue), duplications (red square), inversions (orange) and translocations. This analysis is visualized and annotated by the automated Bionano Access Software. Figure partly adapted from Bionano Genomics. Abbreviations= high molecular weight (HMW); Optical genome mapping (OGM), structural variant (SV). Figure partly adapted from Bionano Genomics.

Following data generation, the automated down-stream analysis software Bionano Access with Bionano Solve is used for SV detection. OGM data can be processed using three different bioinformatic pipelines: i) the *de novo* assembly (DN); ii) the rare variant pipeline (RVP) and iii) the CN pipeline (Figure 4). The DN assembly is efficient to detect haplotype-resolved germline variation and has been primarily applied in the context of constitutional genetic disorders [39, 46, 55]. The RVP is optimized for somatic variation to efficiently detect SVs with a variant allele frequency (VAF) >5% [41]. This approach has been frequently applied for hematological malignancies, where OGM effectively identified somatic and subclonal genomic alterations [40-43, 50, 52, 54]. The CN pipeline allows for the detection of larger chromosomal events and whole chromosome alterations. The CN approach has important diagnostic applications, such as identifying e.g. constitutional trisomy [39] or aneuploidy in cancer cells [41].

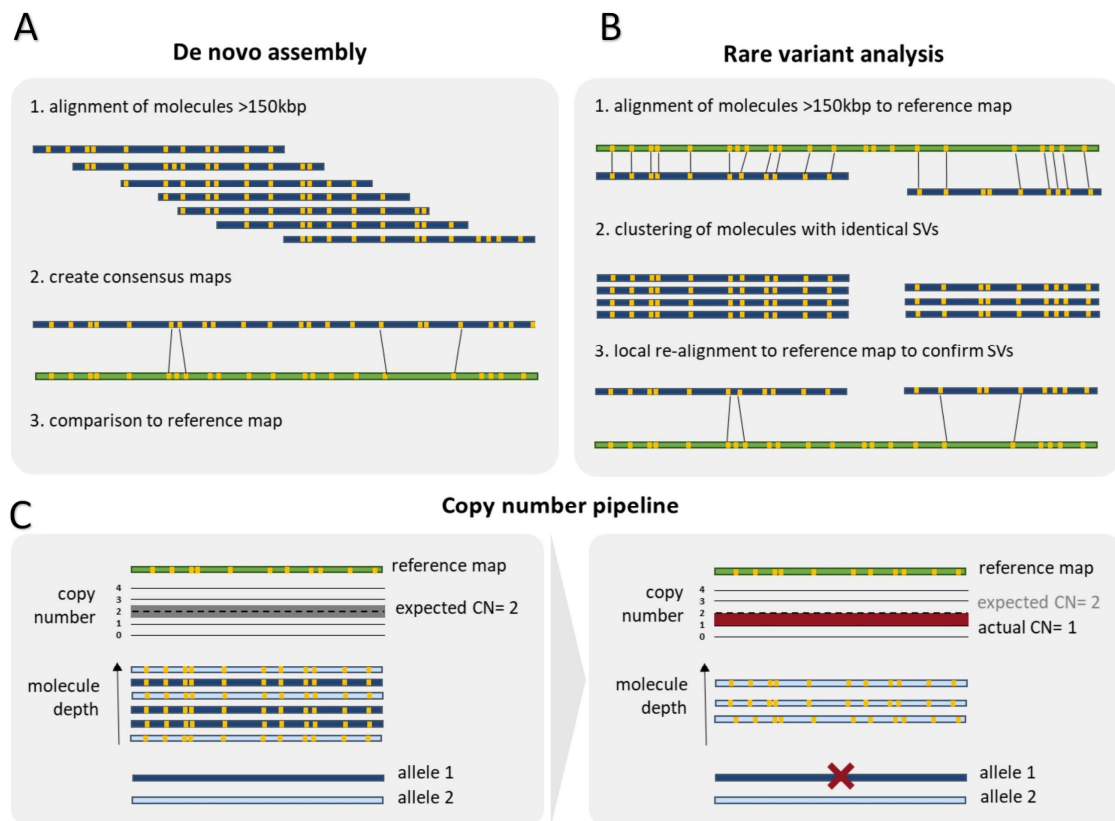


Figure 4: Optical genome mapping pipelines for detection of structural variants and copy number alterations.

(A) The DN analysis is applied for approximately 420Gb of raw OGM data yielding around 100x effective genome coverage. This approach enables the detection of SVs >500bp with haplotype resolution. The de novo assembly pipeline starts with the initial alignment of in silico molecules (blue) >150kbp to each other, based on similar labeling (yellow) distances, in order to assemble them into long genome consensus maps in a second step. These consensus maps are then aligned to a reference genome map (green) to identify SVs. **(B)** The rare variant analysis is designed to process 1300Gb of raw OGM data, generating >300x effective whole-genome coverage. The pipeline directly starts with the alignment of molecules >150kbp to the reference genome map to call potential SVs. All molecules showing identical SVs are then clustered and locally re-aligned to the reference genome map to confirm the SVs that are >5kbp. **(C)** The CN pipeline is a coverage-based approach that is automatically integrated into the workflows of both SV (DN+RVP) pipelines, enabling the detection of deletions and duplications >500kbp up to whole chromosome alterations. The overall molecule depth along the reference genome map is calculated and normalized to establish a baseline for the expected copy number of e.g. a normal human diploid genome expecting two alleles. At genomic loci where the molecule coverage deviates from the baseline, the CN caller reports a corresponding reduction or increase in molecule depth and estimates the actual copy number at that location. Variations of molecule depth are used to call copy number changes at a specific region of the reference genome map, indicating loss (as illustrated, red highlight) or gain of one allele. Abbreviations: structural variants (SVs); copy number (CN), de novo assembly (DN), rare variant pipeline (RVP).

1.1.4.2 Single molecule real-time (SMRT) long-read sequencing

SMRT sequencing developed by PacBio is a long-read sequencing technology that operates on single DNA molecules without the need for prior DNA amplification (Figure 5). Depending on the analyzed DNA template size, ranging from one to several megabases in length [36, 56], two sequencing modes are available: (i) continuous long-reads (CLR); or (ii) circular consensus (CCS)

reads. CLR sequencing typically achieves a base call accuracy ranging from 85 to 92% [36], whereas CCS sequencing achieves high-fidelity (HiFi) consensus reads with an overall base call accuracy up to 99.9% [36].

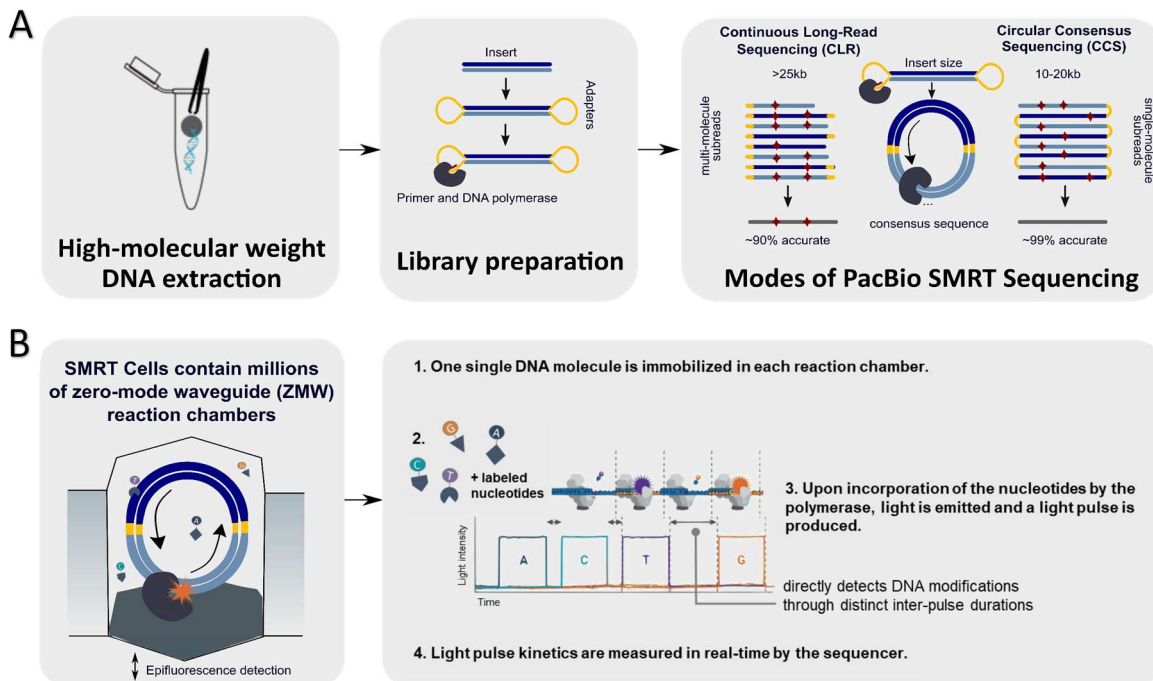


Figure 5: Technical workflow to generate SMRT long-read sequencing data.

Simplified workflow to generate whole-genome long-read sequencing data using the PacBio SMRT technology. **(A)** After HMW DNA extraction and DNA size selection, the fragmented DNA is submitted for library preparation. Library preparation involves ligation of hairpin adapters to both ends of double-stranded HMW DNA, thereby creating circular DNA constructs called “SMRTbell” templates. Before loading on the flow cell and sequencing, these templates are primed with a polymerase. Based on the DNA insert size, two modes of PacBio sequencing are available: (i) CLR with DNA fragments in size > 25kbp leading to multi-molecule subreads, and a consensus sequence with ~90% accuracy; and (ii) CCS with DNA fragments 10-20kbp leading to single-molecule subreads and consensus sequence with >99% accuracy. In case of the CLR sequencing approach, the larger insert size limits the number of full passes of the polymerase to approximately one to five, depending on the duration of the data collection time and size of the template. CCS is based on smaller DNA insert sizes, which enables the DNA polymerase to progress multiple complete passes over the circular DNA template for approximately seven to 12 times or more. This process generates multiple subreads that are combined to highly accurate consensus sequence, called HiFi read. **(B)** The DNA-templates are introduced into a SMRT cell, which contains up to eight million ZMWs. Within each single ZMW, one DNA-polymerase complex is immobilized at the bottom, allowing for DNA synthesis at a single-molecule level. For each nucleotide, distinct fluorescently labeled dNTPs are introduced into the sequencing chambers. The polymerase starts synthesizing a new DNA strand by incorporating each nucleotide along the circular DNA template. While the polymerase holds and cleaves a fluorescently labeled nucleotide, a specific light pulse for each nucleotide is produced through excitation at the bottom of each reaction chamber before incorporation of the next dNTP. This process is called sequencing by synthesis. The emitted fluorescence intensity is recorded over time by the camera of the sequencer after each nucleotide incorporation and cleavage, generating distinct fluorescence patterns that reflect the underlying DNA sequences. SMRT sequencing additionally enables the simultaneous detection of DNA modifications, such as 5-methylcytosine, through distinct inter-pulse durations generated during nucleotide incorporation. The light pulse kinetics are measured in real-time by the PacBio sequencing instrument. Abbreviations: high-molecular weight (HMW); single molecule real-time (SMRT); continuous long-read sequencing (CLR); circular consensus sequencing (CCS); zero-mode waveguide (ZMW), deoxynucleotide triphosphates (dNTPs), high-fidelity (HiFi). Figure partly adapted from PacBio.

Although long-read sequencing technologies, including the SMRT technology, are increasingly utilized in research settings, their integration into clinical routine diagnostics remains limited [57]. Long-read platforms provide several clinically relevant advantages, including enhanced sensitivity to detect SVs and complex genomic alterations, the ability to generate high quality alignments at breakpoint sites, accurate haplotype-reconstruction and simultaneous detection of DNA methylation [34, 57-60]. The implementation in clinical practice is currently still hindered by the high cost of instrumentation, the large volume of generated data, the need for automated workflows and advanced bioinformatic expertise.

1.1.5 Interpretation of genomic variation

The interpretation of genomic variation in health and disease relies on a multifactorial strategy that includes annotation and assessment of genomic variation and evaluating their potential impact. This process integrates clinical evidence, population allele frequencies, computational prediction tools and functional studies to classify the significance of the observed genetic changes. The American College of Medical Genetics and Genomics (ACMG) has established standardized guidelines for variant interpretation [25, 61] and further adaption for the assessment of somatic variation [62, 63]. Genetic variants are classified based on a five-tier system where assigning points to categories of evidence are summarized to identify their association with the underlying disease. Genetic variation that is not considered disease-causing is classified as (i) benign or (ii) likely benign. Variants with insufficient or conflicting evidence of the impact on the disease are classified as (iii) variants of unknown significance (VUS). Variants that are definitive or potentially disease-causing are classified as (iv) likely pathogenic or (v) pathogenic. The classification is determined by weighing multiple criteria to assess the impact of the genetic alteration. These include the type and location of the alteration (e.g., null variants such as frameshift or stop gain mutations or those impacting functional protein domains), *de novo* occurrence or familial segregation, prevalence (e.g., case-control studies, population databases from healthy individuals), computational prediction tools assessing deleterious effects (e.g., conservation, splicing impact) [61]. In addition, the focus of the interpretation of somatic variation is the determination of oncogenicity and clinical actionability including the prediction of therapy response or resistance [62].

More recently, ACMG recommendations and adaptations for assessment of copy number alterations were established including multigenic or single-gene copy number variations with distinct scoring metrics for duplications and deletions [64, 65]. Exemplarily, for a deletion to be

considered as potentially pathogenic, it should affect a gene coding region or functionally important element or an established haploinsufficient region. To assess CNV population frequencies the curated database of genomic variation (DGV) is widely used, which however has limitations based on breakpoint uncertainty due to usage of distinct technologies [64]. A broad clinical adoption of NGS-based methods will lead to an increased detection of genetic alterations including SVs and variation in non-coding regions, underscoring the need of laboratory internal and extended population databases with sequence level variation based on advanced techniques. The ongoing development of comprehensive guidelines for all types of DNA alterations aims to improve the accuracy and consistency of variant interpretation.

1.2 Pediatric cancer

Each year, around 400 000 children and young adolescents are diagnosed with cancer worldwide, still making it the leading cause of disease-associated death for children from birth to 19 years of age [66]. Pediatric cancer encompasses heterogeneous malignancies that can be broadly subdivided into hematological malignancies, central nervous system (CNS) tumors, and non-CNS solid tumors. These main categories include numerous subtypes that are defined by distinct clinical, genetic and prognostic characteristics (Figure 6).

According to the German childhood cancer registry (1980-2018; [67]), the most prevalent pediatric cancers are leukemia, CNS tumors and lymphomas, occurring in around 30%, 24% and 15% of children with cancer, respectively. Incidence rates vary by tumor type and demographic factors. Lymphoid leukemia e.g., shows a higher incidence in males compared to females (4.5 vs 3.5 cases). The highest 5-year survival rate of ~95% are observed in Hodgkin lymphoma and retinoblastoma. In contrast, certain cancers such as embryonal tumors and rhabdomyosarcoma have a dismal disease with 5-year survival rates of 20% to 30%.

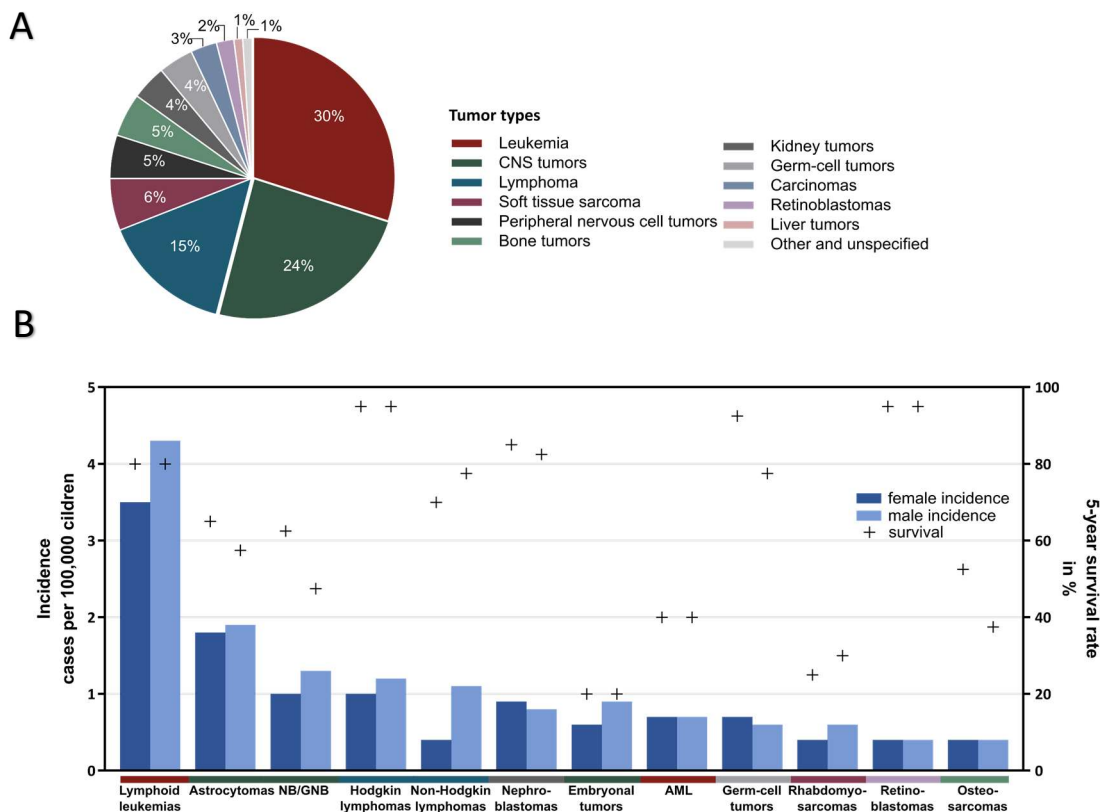


Figure 6: Frequency, incidence and survival of pediatric tumor diagnoses in Germany.

(A) Frequency of tumor entities diagnosed in children with cancer. Different tumor types are indicated by distinct colors. (B) Incidence per 100,000 children of different tumor types and the corresponding survival rate (in %) is shown for male and female sex. Data was extracted from the annual report 2019 of the German childhood cancer registry [67]. Abbreviations: AML=acute myeloid leukemia; GNB= ganglioneuroblastoma;

1.2.1 Etiology of pediatric cancer

The etiology of childhood cancer remains to be fully understood. While life-style risk factors such as UV-light exposure, infectious diseases, tobacco and alcohol consumption are contributing to adult cancer risk, these factors not adequately explain the cause of cancer disease in children and young adolescence, mainly due to their young age. In contrast to adult tumors, pediatric cancers tend to arise from stem and progenitor cells, exhibit a low mutational burden, and are frequently characterized by oncogenic gene fusions or epigenetic dysregulation [27]. To date, a small proportion of childhood cancers can partially be attributed to pre- and postnatal exposure to environmental factors [68], inherited genetic risk factors [69], or cancer predisposition syndromes [70]. Recent advances in sequencing technologies and comprehensive genomic studies provided deeper insights into the role of germline variation in modulating cancer susceptibility in children [26, 27, 29, 31, 71] (refer to section 1.2.2).

1.2.2 Genetic cancer predisposition in children

1.2.2.1 Cancer predisposition syndromes

Patients with an inherited cancer predisposition syndrome (CPS) harbor an increased risk of developing cancer. Typically, the causal link to a CPS is inherited or *de novo* mutation affecting either autosomal recessive or dominant cancer predisposition genes (CPG) [29]. Germline mutations frequently affect CPGs involved in DNA damage repair and/or cell cycle regulation. These alterations increase rates of somatic mutation, cause genomic instability and the likelihood of cancer development [72-78]. This pathogenic mechanism is observed in a variety of well-characterized CPS and the corresponding CPG, including Ataxia telangiectasia (*ATM*, [72]); Bloom Syndrome (*BLM*, [73, 74]), Fanconi anemia (e.g. *BRCA2*, *PALB2* etc. [75]), Nijmegen breakage syndrome (*NBN*, [76]) and constitutional mismatch repair (CMMR) deficiency (e.g. *MSH2*, *MSH6* [77, 78]).

To facilitate systematic clinical screening of pediatric cancer patients, diagnostic criteria for identifying a suspected CPS have been established by Jongmans *et al.* [79] and Ripperberger *et al.* [80]. These criteria include a prominent family history of cancer, specific cancer entities known to be associated with a CPS, the presence of multiple primary malignancies, congenital anomalies and excessive treatment-related toxicity, among other indicators [79, 80]. Notably, severe toxicity during therapy may result from pre-existing immunodeficiency or dysfunctional DNA repair mechanisms [81] or from mutations in genes involved in drug-metabolism [82]. These factors can lead to an impaired tissue regeneration or adverse reactions during chemo- and/or radiotherapy. The identification of cancer susceptibility in pediatric patients can inform clinical-decision making, lead to modified treatment protocols [83], improve prognostic assessment [84], identify strategies for tumor surveillance strategies [85], and ensure early detection of secondary malignancies [86]. Trio sequencing from child-parent samples enables the detection of *de novo* mutations and is further helpful, as the identification of germline predisposition in a child with cancer can have direct implications for parents and siblings [87-89]. Notably, secondary malignancies occur in around 3-4% of pediatric cancer survivors within 20-25 years after initial diagnosis [90]. These secondary events can be associated with the therapy of the initial diagnosis or with underlying germline mutations in genes such as *TP53*, *RB1* and *NF1* [91]. CPS often remain undetected when relying solely on clinical phenotypes due to e.g., heterogeneity of clinical presentations or the absence of a clear family cancer history [30, 79, 87], highlighting the importance of comprehensive genetic testing. Due to

increased clinical screening and genetic testing, the number of recognized CPS and associated CPGs are increasingly expanding (Table 1).

Table 1: Cancer predisposition genes in pediatric malignancies.

Summary of selected cancer predisposition genes, the corresponding syndrome and the frequently associated cancer type. Abbreviations: Atypical teratoid rhabdoid tumor (AT/RT); Acute lymphoblastic leukemia (ALL); Acute myeloid leukemia (AML); Carcinoma (CA); Embryonal tumor with multilayered rosettes (ETMR); Gastro-intestinal stromal tumor (GIST); high-grade (H-G); Juvenile myelomonocytic leukemia (JMML); Medulloblastoma (MB); Myelodysplastic syndrome (MDS); Meningioma (MG); Malignant peripheral nerve sheath tumor (MPNST); Neuroblastoma (NB); Osteosarcoma (OS); Pilocytic astrocytoma (PAC); Pineal blastoma (PB); Pheochromocytoma/Paraganglioma (PMC/PG); Pleuropulmonary blastoma (PPB); Retinoblastoma (RB); Rhabdomyosarcoma (RMS); Sertoli-Leydig cell ovarian TU (SLCOT); Tumors (TU). Table was adapted from [83] and created in <https://Biorender.de>.

Gene(s)	Associated syndrome(s)	Associated pediatric cancer(s)				
		ABDOMINAL TU	BRAIN TU	CERVICAL & THORACIC TU	LEUKAEMIA & LYMPHOMA	SARCOMA
<i>APC</i>	-, APC-associated polyposis		MB	Thyroid CA		Desmoid TU
<i>ALK</i>	ALK syndrome	NB				
<i>ATM</i>	Ataxia telangiectasia				ALL, Lymphoma	
<i>BLM</i>	Bloom syndrome	Wilms TU			ALL	OS
<i>BRCA2, PALB2</i>	Fanconi anaemia		MB			
<i>DICER1</i>	DICER1 syndrome	Wilms TU	PB, ETMR	Thyroid CA, PPB, SLCOT		RMS
<i>ETV6</i>	ETV6 deficiency				ALL	
<i>GATA2</i>	GATA2 deficiency				AML/MDS	
<i>GPR161</i>	GPR161 deficiency		MB			
<i>HRAS</i>	Costello syndrome	NB				
<i>IKZF1</i>	IKZF1 deficiency				ALL	
<i>NRAS, KRAS</i>	Rasopathies				JMML	
<i>MSH2/6, MLH1, PMS2</i>	Constitutional MMR deficiency		MB, H-G glioma		ALL, AML/MSD, Lymphoma	
<i>NBN</i>	Nijmegen breakage syndrome				ALL	
<i>NBS1</i>	Nijmegen breakage syndrome				Lymphoma	RMS
<i>NF1</i>	Type 1 Neurofibromatosis, Rasopathies	NB	PAC	PMC/PG	JMML	RMS, MPNST, GIST
<i>NF2</i>	Type 2 Neurofibromatosis		MG, Spinal ependymoma			
<i>PAX5</i>	PAX5 deficiency				ALL	
<i>POLE</i>	Polymerase exonuclease deficiency		MB, H-G glioma			
<i>PTCH1, SUFU</i>	Gorlin syndrome		MB, MG			
<i>PTEN</i>	Cowden syndrome, -		MG	Thyroid CA		
<i>PTPN11</i>	Noonan syndrome, Rasopathies	NB			ALL, JMML	RMS
<i>RB1</i>	RB predisposition		RB			
<i>RET</i>	Multiple endocrine neoplasia			Thyroid CA, PMC/PG		
<i>RUNX1</i>	Robertsonian t15;21, Ring chr21, RUNX1 deficiency, Trisomy 21				ALL, AML/MDS	
<i>SAMD9/L</i>	Ataxia Pancytopenia syndrome				AML/MDS	
<i>SMARCA4</i>	Rhabdoid TU predisposition syndrome	SCCOHT	AT/RT			Rhabdoid TU
<i>SMARCB1</i>	Rhabdoid TU predisposition syndrome, Familial MG		AT/RT, MG			Rhabdoid TU
<i>TP53</i>	Li Fraumeni syndrome	NB, Adrenocortical CA	MB, H-G glioma, Choroid Plexus CA		ALL, AML/MDS	RMS, OS
<i>TRIM28</i>	Familial Wilms TU	Wilms TU				
<i>WT1</i>	Familial Wilms TU	Wilms TU				

1.2.2.2 Germline genetic variation in children with cancer

Over the past years, advances in comprehensive genetic testing of children with cancer showed that approximately 8-10% of pediatric patients carry a presumably pathogenic germline mutation, that may contribute to inherited cancer susceptibility [25, 26, 29-31, 71]. Most of these predisposing mutations occur in children diagnosed with solid tumors (16.7%), followed by those diagnosed with brain tumors (8.6%) and leukemia (4.4%) [71]. The most frequently altered genes across large-scale pediatric cancer cohorts include *TP53*, *APC*, *BRCA2*, *NF1*, *PMS2*, *RB1* and *RUNX1* [25, 26, 29, 31, 71]. These studies mostly applied WES or WGS to identify pathogenic SNVs or indels, typically incorporating panels of well-established CPGs. However, recent studies have increasingly integrated SV analysis, using either short- or long-read WGS approaches to check for pathogenic events in constitutional [92] or cancer disease [59]. It has been reported that e.g., around 34% of disease-causing germline alterations in an adult cancer cohort are SVs, with 1.5% of cases carrying predisposing structural alterations [59]. Pathogenic SVs commonly act as driving lesions during tumorigenesis and may contribute to cancer susceptibility by disrupting gene function, expression, or broader genomic architecture [93, 94]. Despite these advances, SVs and non-coding variation in the context of cancer predisposition remain poorly characterized, leading to significant challenges for clinical interpretation of such alterations. The currently estimated prevalence of 8-10% of detectable pathogenic germline mutations in children with cancer may represent only a fraction of the true burden and a substantial proportion may still be obscured.

1.3 Leukemia

Leukemia is a blood cancer that develops through the malignant transformation of cells during the development of the hematopoietic system (hematopoiesis) leading uncontrolled growth of abnormal cells in the bone marrow and impairment of production and function of healthy blood cells [95]. Leukemia can be subdivided into acute forms with rapid onset and progression of immature blast cells and chronic forms with slow disease progression and gradual accumulation of dysfunctional but more mature blast cells [96].

Further subdivision groups leukemia based on the expanding cell characteristics such as the cell lineage, either affecting cells from the myeloid or lymphoid compartment (Figure 7). These subtypes are grouped to acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML),

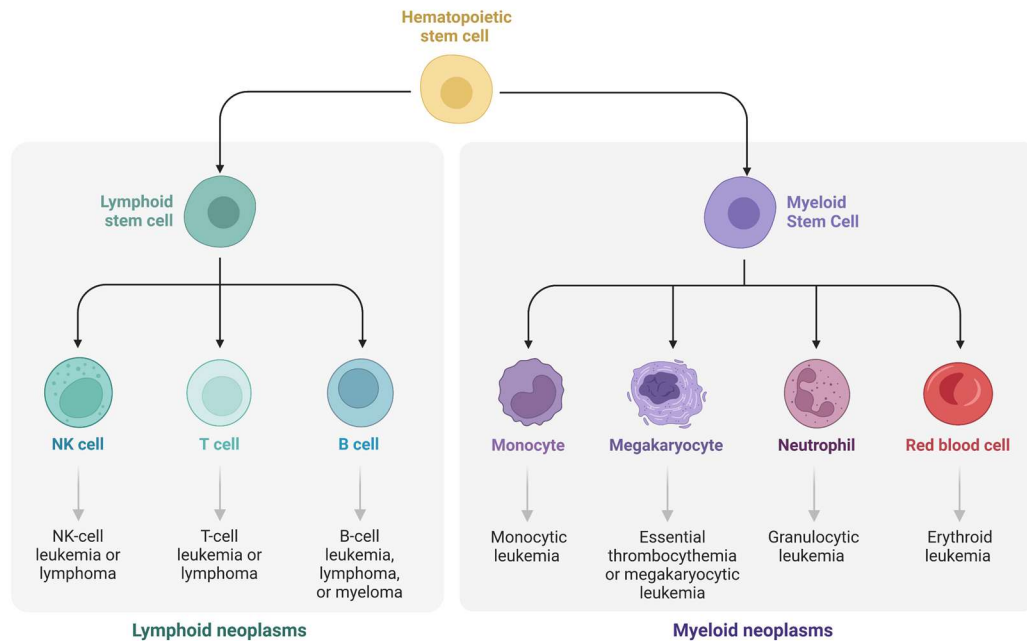


Figure 7: Leukemia subtypes resulting from distinct hematopoietic compartments.

During hematopoiesis, hematopoietic stem cells can develop into cells with lymphoid or myeloid competence going through different stem and precursor stages. The different lymphoid or myeloid neoplasms arising from different cell types are indicated. Figure template was adapted from <https://Biorender.de>.

chronic lymphoid leukemia (CLL) and chronic myeloid leukemia (CML). The ALL subtype can be further subdivided into T-cell acute lymphoblastic leukemia (T-ALL) and B-cell acute lymphoblastic leukemia (B-ALL). Based on the World Health Organization (WHO) guidelines further refinements for subtype classification include differential molecular and cytogenetic events as well as clinical and pathological features [97]. This thesis will specifically focus on the most common B-cell precursor ALL (BCP-ALL) subtypes (refer to section 1.3.2.1 and 1.3.2.2).

1.3.1 Acute lymphoblastic leukemia

ALL is the most common malignancy and leading cause of cancer-related mortality worldwide amongst children and young adolescents (<19 years) [98]. The majority of ALL cases originate from transformed precursor cells of the B-cell lineage (~83%), with a smaller proportion arising from T-cell lineage precursors (~17%) [99]. These malignant lymphoid progenitor cells, referred to as lymphoblasts, become arrested at immature cell stages during lymphoid differentiation and transform into leukemia [98]. The uncontrolled proliferation of the lymphoblasts leads to an elevated white blood cell (WBC) count and widespread infiltration of the bone marrow (>20% of leukemic blasts) and peripheral organs such as the liver, spleen and lymph nodes. The displacement of healthy hematopoietic cells leads to various cytopenia [98]. In more aggressive

cases, leukemic blasts may infiltrate the central nervous system (CNS) [100] and testis [101], thereby contributing to disease severity. Clinically, these pathological processes manifest as fatigue, neutropenic infections, fever and excessive bleeding, among other symptoms [98, 102, 103].

The etiology of ALL is multifactorial. ALL is thought to be caused by a complex interplay between environmental factors such as radiation or exposure to toxins [104, 105] and genetic variables. The majority of ALL may originate from an initiating genetic event during early hematopoiesis, referred to as “first hit”, which occurs *in utero* and gives rise to preleukemic clones [98, 106-110]. The initial genetic lesion typically either involves chromosomal translocations giving rise to oncogenic fusion genes such as *ETV6::RUNX1*, *BCR::ABL1*; *TCF3::PBX1*; *KMT2A*-rearrangements, or numerical chromosomal abnormalities like hyper- or hypodiploidy [98, 106, 107, 110]. The progression from a preleukemic state to overt full-blown leukemia usually requires additional postnatally-acquired oncogenic hits. These can be triggered by e.g., exposure to infections and dysregulated immune responses [98, 106, 111, 112]. Furthermore, inherited genetic factors may contribute to ALL risk, including GWAS-identified germline polymorphisms that may direct or indirectly affect regulators of hematopoietic and lymphoid cell development, such as *IKZF1* and *ARID5B* as well as tumor suppressor genes like *CDKN2A/B* and *TP63* [113-116]. In addition, pathogenic germline variants in CPGs including *ETV6*, *IKZF1*, *PAX5* and *TP53* [114, 116-118], have been implicated in familial or sporadic cases of ALL with variable penetrance (refer to section 1.2.2). The following sections will focus specifically on B-cell derived ALL, including details on its clinical and genetic background, as well as current treatment approaches.

1.3.2 B-cell acute lymphoblastic leukemia

B-cell precursor acute lymphoblastic leukemia (BCP-ALL) is the most common cancer in childhood usually with a good prognosis and an overall 5-year survival rate exceeding 90% in high-income countries [119]. The success of standard treatment is largely attributed to the combination of clinical, immunophenotypic and genetic characterization to define ALL subtypes enabling risk stratification and therapeutic decision-making [100, 120] (refer to section 1.3.3). The most frequent subtypes of BCP-ALL are characterized by recurrent disease-initiating genetic alterations, comprising structural abnormalities in approximately 75% of these subtypes [121] (Figure 8).

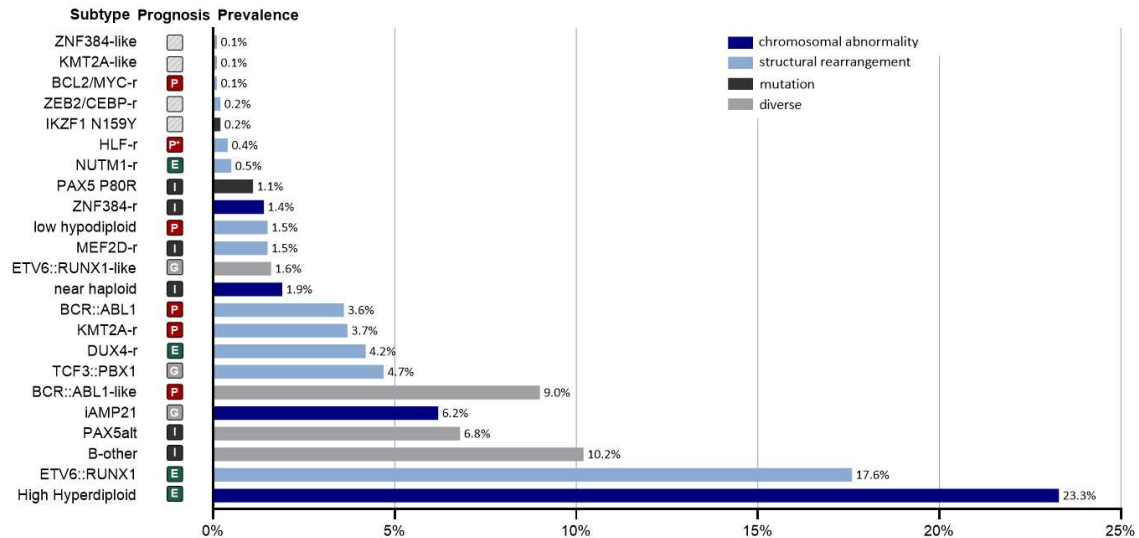


Figure 8: Prevalence and prognosis of genetic subgroups in pediatric BCP-ALL.

Prognosis and overall frequency of each genetic subgroup of B-cell derived pediatric leukemia is shown. Excellent (E), good (G), intermediate (I), poor (P) or very poor (P*) prognosis for each subtype was extracted from [122]. Relative numbers of each subtype were extracted from [99]. Coloring of the bar graph indicates the genetic alteration contributing to the primary subtype classification. Abbreviations: rearrangement (-r).

These include gross chromosomal changes such as aneuploidy and chromosomal rearrangements deregulating genes through the formation of chimeric fusion genes that alter function or activity of hematopoietic transcription factors, epigenetic modifiers, tyrosine kinases and others [122]. The underlying genetic landscape plays a central role in leukemogenesis and is closely linked to gene expression profiles, altered signaling pathways, clinical presentation, treatment response and overall prognosis [98, 123-125]. Subtype-defining genetic alterations along with cooperating secondary genetic lesions are essential for refined risk stratification, prognosis and therapeutic strategies [119, 122]. For instance, the co-occurrence of *IKZF1* deletions together with *CDKN2A/B*, *PAX5*, or *PAR1* deletions in the absence of *ERG* deletions defines a high-risk genetic profile termed *IKZF1*^{plus} in several BCP-ALL subtypes [126]. This subgroup has worse outcomes compared to patients with *IKZF1* deletions without fulfilling the *IKZF1*^{plus} criteria [126].

The two most common genetic subtypes are classical hyperdiploid BCP-ALL (HD BCP-ALL) and *ETV6::RUNX1* translocated BCP-ALL (E::R BCP-ALL) (refer to section 1.3.2.1 and 1.3.2.2), both of which are associated with a favorable prognosis and may benefit from low-intensity therapy approaches [127]. However, they constitute around 25% of relapsed cases that occur in childhood BCP-ALL [128]. Additional subtypes with structural and numerical alterations include:

(i) hypodiploid leukemia, characterized by fewer than 48 chromosomes, (ii) *TCF3::PBX1* translocated leukemia, (iii) *BCR::ABL* translocated leukemia and (iv) *KMT2A*-rearranged BCP-ALL. Among the rarest and most aggressive cases is *TCF3::HLF*-rearranged BCP-ALL, which comprises less than 1% of cases and is associated with a dismal prognosis [122]. Notably, overall survival of this rare subtype may be improved with immunotherapy such as blinatumomab [129]. The application of NGS has led to the discovery of novel subtype-defining events in BCP-ALL with clinical and therapeutic relevance, including: (i) *MEF2D*- and *ZNF384*-rearrangements, (ii) mutations in *PAX5* (P80R) or *IKZF1* (N159Y), and (iii) molecular subgroups mimicking expression profiles of known genetic subtypes referred to as *ETV6::RUNX1*-like, Ph-like, *KMT2A*-like or *ZNF384*-like BCP-ALL [122]. Notably, the third most frequent BCP-ALL subtype, referred to as “B-other”, lacks such defining alterations, making risk stratification and therapeutic intervention particularly challenging [130].

1.3.2.1 Hyperdiploid BCP-ALL

Classical HD BCP-ALL is characterized by a nonrandom gain of chromosomes, commonly involving chromosomes 4, 10, 14, 21 and X. These chromosomal gains result in an abnormal karyotype with a modal chromosome number ranging from 52 to 67 chromosomes, indicative of aneuploidy in the leukemic blasts [122, 128, 131]. The karyotype normally consists of di-, tri- and tetrasomies, where trisomies arise from the duplication of a single parental chromosome and tetrasomies involve the duplication of both parental homologs [128].

1.3.2.1.1 Clinical features

Clinically, HD BCP-ALL comprises approximately one-third of BCP-ALL diagnoses, typically presents at a median age of around 4 years and is associated with a favorable prognosis. The immune phenotype of leukemic blasts shows a low expression of CD45 and a high expression of markers such as CD9, CD20, CD22, CD58, CD66c, CD86, and CD123. CD123 is considered as immunophenotypic hallmark of classical HD leukemia [132]. Diagnosis of this subtype is based on the detection of the hyperdiploid karyotype using DNA index measurement, conventional cytogenetics, FISH or array-based methods. Further refinement of risk stratification specifically for hyperdiploid leukemia has been proposed in a study of Enshaei *et al.* [133], suggesting the existence of two distinct risk groups (good and poor risk) based on the presence or absence of copy number gains of four specific chromosomes: 17, 18, 5 and 20.

1.3.2.1.2 Etiology and genetic characteristics

HD leukemia is hypothesized to originate from a single abnormal mitosis occurring prenatally in a B-cell precursor during early embryogenesis, generating the nonrandom gains [109, 134-137]. The leukemogenic process is driven by the gene dosage effects and transcriptional dysregulation caused by the extra chromosomal copies [138, 139]. The precise mechanisms underlying the initiating chromosomal segregation defect or the corresponding genetic cause remain unresolved. There is growing evidence of inherited genetic susceptibility in the development of HD BCP-ALL. Pathogenic germline mutations in RTK/RAS pathway genes are implicated in the development of RASopathies in children with several clinical abnormalities including cancer predisposition. Particularly, children with Noonan syndrome and *PTPN11* [140, 141] or *SOS1* [141] mutations are often diseased with hyperdiploid ALL. Beyond RASopathies, pathogenic germline mutations in other genes such as *ETV6* [142], *GAB2* [143] or *LZTR1* [144] have been described to be enriched in patients with the HD BCP-ALL subtype. Genome wide association studies (GWAS), which statistically assess genotype-phenotype correlations, revealed selective risk alleles including *ARID5B* (rs7090445) [114], *CEBPE* (rs2239630) [145] and *BMI1-PIP4K2A* (rs7088318) [146] in patients with HD leukemia. Collectively, the recurrent germline mutations and risk-associated polymorphisms in HD BCP-ALL indicate genetic predisposition as contributor to the pathogenesis of this subtype.

Frequent secondary mutations in HD BCP-ALL occur in genes of the RTK/RAS pathway including *KRAS*, *NRAS*, *FLT3*, *SOS* and *PTPN11* as well as epigenetic modifiers such as *CREBBP*, *NSD2*, *SETD2* and *EZH2* [128, 131]. In early relapsed HD BCP-ALL cases, *KRAS* mutations commonly coexist with mutations in the chromatin-modifying gene *CREBBP* [147, 148], suggesting a cooperative role in disease progression and therapy resistance. Secondary structural alterations are common events and include duplication of chromosome arm 1q, deletion of chromosome arm 6q and formation of isochromosomes 17q and 7q, which involve loss of the p-arm and mirror duplication of the q-arm [134]. Focal deletions frequently target gene loci of the cell cycle regulator *CDKN2A* and the hematopoietic transcription factor *ETV6* [99].

Among HD BCP-ALL around 10-15% of patients experience relapse, that occur within the first three years after initial diagnosis. Due to the high overall incidence of HD BCP-ALL, the relapses make up 25% of all relapsed BCP-ALL cases [128]. Relapse is correlated with the acquisition of genetic hits contributing to therapy resistance and clonal evolution, including mutations in genes of the RAS pathway and the chromatin-modifier *CREBBP* [147-150]. Numerical variation such as positive selection for trisomy 8 is also associated with relapsed HD leukemia [151].

1.3.2.2 *ETV6::RUNX1* BCP-ALL

ETV6::RUNX1 translocated leukemia is defined by the recurrent chromosomal translocation t(12;21)(p13;q22) that results in the juxtaposition of the *ETV6* gene to the *RUNX1* gene locus, forming an oncogenic fusion gene *ETV6::RUNX1* (*E::R*). *ETV6* and *RUNX1* both play an essential role for embryonic and adult hematopoiesis by regulating hematopoietic stem cell differentiation into mature blood cells [152, 153]. The chimeric fusion protein acts as an *ETV6*-mediated transcriptional repressor converting *RUNX1* from a transcription activator into a repressor of its downstream target genes [154, 155], thereby causing a partial block of B-cell differentiation [156].

1.3.2.2.1 Clinical features

Clinically, around 20% of pediatric BCP-ALL cases are characterized by a favorable diagnosis [157, 158] occurring at a median age of ~4 years [103]. The leukemic blasts express B-cell lineage markers such as CD19, CD10 and TdT expression and exhibit clonal genetic characteristics such as immunoglobulin heavy chain (*IgH*) and T-cell receptor (TCR) rearrangements [159-161]. In some cases, aberrant expression of myeloid-phenotypic markers like CD13 and CD33 is observed [160]. For diagnostic confirmation of the genetic subtype, the translocation is routinely detected using reverse-transcriptase polymerase chain reaction (RT-PCR), FISH or targeted RNA-sequencing.

1.3.2.2.2 Etiology and genetic characteristics

The origin of the translocation t(12;21)(p13;q22) is frequently acquired *in utero* and can be detected in up to 5% of neonatal blood spots or cord bloods [57, 75-78]. The generation of the fusion gene is considered as first hit that arises in a pre-leukemic B-cell progenitor, conferring a genetic predisposition to leukemia. However, the presence of the *E::R* gene fusion is insufficient to drive the development to a full-blown leukemia, as postnatal secondary hits are required for leukemogenesis to occur [107-110, 162-164]. Furthermore, GWAS studies have identified germline susceptibility loci associated with increased risk of *E::R* positive leukemia. Significant associations have been reported for polymorphisms within genes of the *TP53* family, including *TP63*, as well as in the tyrosine phosphatase gene *PTPRJ* [165].

Secondary additional genetic events are essential for the progression from a preleukemic clone to overt *E::R* positive leukemia [108]. These events can largely be attributed to aberrant activity

of the recombination activating gene (RAG) recombinase activity. Mechanistically, the E::R fusion protein, along with wild-type *RUNX1*, can directly bind to enhancer and promoter regions of the *RAG1* gene, thereby upregulating its expression [166, 167]. Under physiological conditions, RAG1/2 recombinases are essential for somatic V(D)J recombination during normal lymphoid differentiation, generating antigen receptor diversity in B- and T-cells [168]. In the leukemic context, the aberrantly active RAG1/RAG2 complex promotes illegitimate recombination events near specific recombination signal sequences leading to secondary structural rearrangements [169]. Specifically, secondary copy number alterations such as loss of the second *ETV6* allele and deletion of *PAX5* and *IKZF1* are frequent genetic events in E::R positive leukemia contributing to leukemogenesis [98, 121, 169]. Other focal deletions recurrently affect *CDKN2A*, *CDKN1B*, *ATF7IP*, *BTG1*, *TBL1XR1*, *BTLA*, *CD200*, *SLX4IP* and members of the histone cluster located on chromosome 6p22.2 [99, 159]. Although less frequently, secondary mutations in genes such as *UBA2*, involved in post-translational protein modification, and *KRAS* are observed in the E::R BCP-ALL subtype [99, 170].

Among E::R translocated cases, around 20% experience relapse that typically occurs later compared to other BCP-ALL subtypes, often several years after initial diagnosis [171, 172]. Relapsed E::R cases often harbor copy number alterations such as loss of chromosome 12p13, including the second *ETV6* allele, and depletion of cell cycle regulator *CDKN1B* or *PAX5*, *IKZF1* and *CDKN2A/B* [173, 174].

1.3.3 Diagnosis and treatment of childhood acute lymphoblastic leukemia

Diagnosis and treatment strategies in pediatric ALL apply a risk stratification system that incorporates clinical factors such as age, sex, CNS or testicular involvement, whole white blood cell count, treatment response alongside with cytomorphological and cytogenetic features present in the leukemic blasts at diagnosis [119]. Based on these criteria, patients are stratified into three main risk groups: standard risk (SR), intermediate risk (IR) and high risk (HR) (Table 2). Further details on diagnosis and treatment are provided in the following sections.

Table 2: Prognostic factors influencing favorable and unfavorable outcomes in ALL.

Abbreviations: WBC= white blood cell count, CNS= central nervous system, SR= standard risk, HR= high risk. Table was created in <https://Biorender.de>.

Prognostic factors	Outcome	
	Favorable	Unfavorable
1. Clinical characteristics Age Sex Ethnicity WBC counts Lineage CNS involvement Testicular involvement	1-9 years female caucasian, asian $<50 \times 10^9$ cells/L B-ALL CNS 1 no	infant or >10 years male hispanic, african american $\geq 50 \times 10^9$ cells/L T-ALL CNS 2&3 yes
2. Genetics Down syndrome Subtypes	no hyperdiploidy ETV6::RUNX1 TCF3::PBX1 NUTM1-r	yes hypodiploidy BCR::ABL/-like TCF3::HLF KMT2A-r MEF2D-r iAMP21 BCL2/MYC-r
3. Minimal residual disease Day 33 Day 78	SR: $<0.0001\%$ -	- HR: $\geq 0.001\%$

1.3.3.1 Diagnostic work-up of acute lymphoblastic leukemia

A range of standard-of-care techniques are currently used to detect clinically relevant genetic alterations during the diagnosis of ALL, including cytogenetic methods such as karyotyping and FISH or molecular genetic techniques such as SNP-array, CHG or multiplex ligation-dependent probe amplification (MLPA) [124, 125].

Treatment response is assessed through minimal residual disease (MRD) monitored at the end of induction and consolidation therapy. MRD is a key prognostic factor and can be quantified by: (i) leukemic cell frequency with a detection threshold of one leukemic cell in 10,000 normal cells (1×10^{-4}), (ii) quantitative reverse transcription PCR (qRT-PCR) targeting leukemia-specific rearrangements (e.g., immunoglobulin or TCR genes), offering a specificity between 1×10^{-4} and 1×10^{-5} leukemic to normal cell ratios and (iii) multiparametric flow cytometry, which identifies aberrant immunophenotypes. Not yet standardized but emerging NGS-based approaches are investigated to increase MRD sensitivity and resolution during the course of therapy, targeting immunoglobulin or T-cell receptor rearrangements [175]. A critical aspect of diagnostic evaluation includes the potential leukemic involvement in CNS. Leukemic blasts can

infiltrate CNS sites which confers a worse prognosis, increased neurotoxicity and higher relapse risk where CNS is involved in approximately 30% [176].

1.3.3.2 Treatment of childhood acute lymphoblastic leukemia

In Germany, ALL is treated according to common protocols of the AIEOP-BFM-ALL and CoALL/ALLTogether consortia. These protocols are reference standards for clinical trials and routine care and are regularly updated to incorporate new risk stratification parameters such as MRD and genetics. The overall goal of ALL therapy is to maximize cure rates while minimizing treatment-related toxicities. Standard chemotherapy treatment for childhood ALL comprises four key components: (i) induction, (ii) consolidation and, (iii) maintenance therapy, as well as (iv) CNS prophylaxis [103].

Induction therapy aims to rapidly remove most of the leukemic cells to achieve remission. During induction therapy, a three-drug scheme is administered with a steroid substance (prednisone, prednisolone, or dexamethasone) together with vincristine and L-asparaginase in order to achieve complete remission. Remission is defined samples with less than 5% leukemic blasts in the bone marrow on morphological level or less than 0.01% blasts by MRD testing [177]. Consolidation phase is started after a patient achieved complete remission and intensifies therapy to maintain remission, eliminate residual leukemic cells, reduce the risk of relapse and to improve the long-term outcome [103]. Finally, maintenance therapy is lasting about two years to sustain long-term remission and prevent relapse. Composition and intensity of the chemotherapy may vary according to the ALL subtype and risk stratification, incorporating drugs such as methotrexate, 6-mercaptopurine, doxorubicin and cytarabine. Further treatment options include radiotherapy, especially for CNS prophylaxis, hematopoietic stem cell transplantation (HSCT) for high-risk or relapsed cases, immunotherapy and targeted therapy [119]. Small molecules like tyrosine kinase inhibitor Imatinib are applied to target specific molecule markers such as *BCR::ABL* positive leukemic blasts [178]. As a potential new frontline therapy, Blinatumomab, a CD19-directed T-cell engager antibody construct, has emerged as a novel transformative treatment option in pediatric BCP-ALL including relapsed and refractory disease [179].

2. AIMS OF THE THESIS

This thesis aims to decipher comprehensive genomes of children with cancer with the focus on germline and somatic structural variation by applying optical genome mapping and integration of next-generation sequencing approaches.

This thesis is based on two Original Publications with the following main objectives:

Objective I:

Identification of pathogenic structural variants and novel potential cancer predisposition genes by optical genome mapping in children with cancer (Original Publication I, [180])

The first objective was to decode germline structural variation in children with cancer that were clinically suspicious for an underlying cancer predisposition syndrome but had a negative result using whole-exome sequencing. The aim was to apply optical genome mapping to detect and evaluate germline structural variants in known or novel cancer predisposition genes. In addition, validation and functional testing of relevant candidates aimed to characterize the function of novel structural variants and their contribution to childhood cancer predisposition.

Objective II:

Deciphering the somatic structural variation landscape of pediatric *ETV6::RUNX1*+ and hyperdiploid B-cell precursor acute lymphoblastic leukemia by optical genome mapping (Original Publication II, [181])

The second objective was to evaluate the clinical application of the novel optical genome mapping technique in diagnostic B-cell precursor acute lymphoblastic leukemia (BCP-ALL) samples and to discover new potential targets relevant for BCP-ALL leukemogenesis. Technical performance and additional value of optical genome mapping was evaluated by determining the concordance with conventional cytogenetic methods and molecular array data. In addition, structural variation landscape of *ETV6::RUNX1*+ and hyperdiploid BCP-ALL was analyzed and compared using optical genome mapping. Further analyses aimed to identify recurrently altered regions and genes by integrating short-read whole-exome sequencing data with optical genome mapping and to find potential novel fusion genes. The application of long-read sequencing and/or RNA sequencing aimed to validate structural variants affecting novel recurrently altered target genes and novel fusion genes.

3. ORIGINAL PUBLICATIONS

This thesis is based on the following original publications which are referred to within the text:

3.1 Summary of Original Publication I

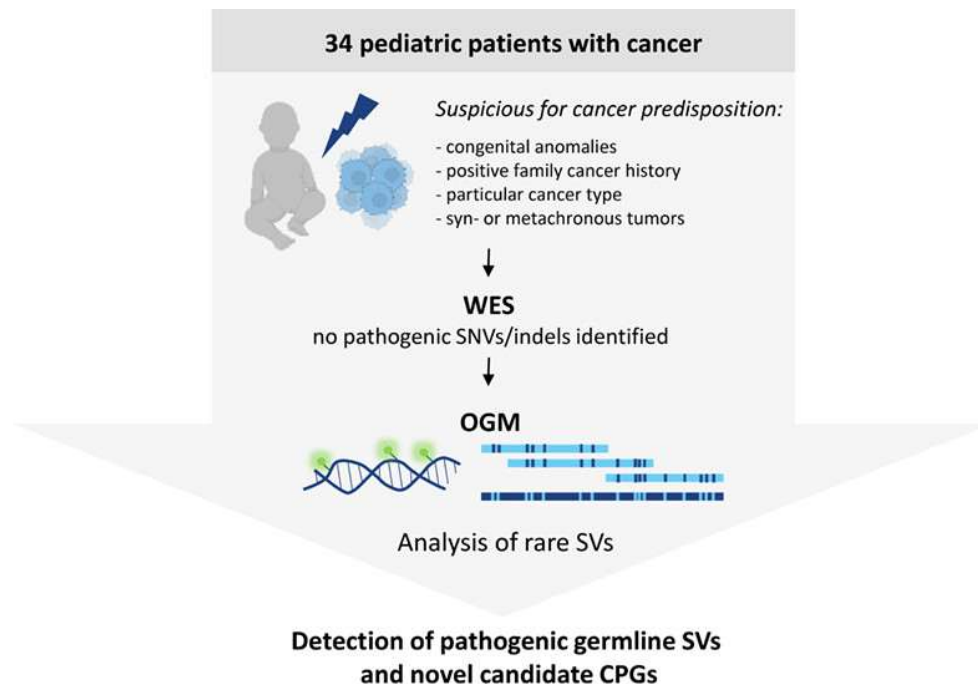
Wagener, R.*, Brandes, D.*, Jung M. Huetzen M.A., Bergmann A.K., Panier S., Picard D., Fischer U., Jachimowicz R.D., Borkhardt A., Brozou T.

Optical genome mapping identifies structural variants in potentially new cancer predisposition candidate genes in pediatric cancer patients.

International Journal of Cancer. 2024 Feb 15;154(4):607-614.

PMID: 37776287; DOI: 10.1002/ijc.34721; impact factor: 5.7, *equal contribution

Graphical abstract:



Key findings:

- OGM identified structural variants in known and novel candidate cancer predisposition genes in pediatric cancer patients clinically suspicious of an underlying cancer predisposition syndrome and a negative whole-exome sequencing result
- *De novo* structural variants can be identified by performing OGM-based parent-children trio analysis
- OGM identified an inherited pathogenic deletion affecting the known cancer predisposition gene *BRCA2* in a family with cancer segregation

- OGM identified a duplication affecting a potential new cancer predisposition gene *RPA1* in a pediatric leukemia patient
- Functional testing of patient fibroblasts with *RPA1* duplication show a higher sensitivity to DNA damage compared to age-matched controls

Contribution:

I revised and wrote the paper together with R.W. I conceptualized and designed the graphical abstract. I extracted DNA from clinical samples, performed optical genome mapping experiments as well as data analysis together with M.J. I further performed functional validation and analysis of known and candidate cancer predisposition genes including whole-exome-sequencing, reverse-transcription PCR, quantitative PCR, Sanger-sequencing and SNP-array analysis together with R.W. and M.J. Major contribution and my own experiments are represented in Table 1 & Figure 2; Supplementary Material & Methods PCR-based analysis, Supplementary Figures 1 & 4; Supplementary tables 1, 3 & 5.

3.2 Original Publication I

Received: 29 January 2023 | Revised: 1 August 2023 | Accepted: 11 August 2023
DOI: 10.1002/ijc.34721

SHORT REPORT

Innovative Tools and Methods



Optical genome mapping identifies structural variants in potentially new cancer predisposition candidate genes in pediatric cancer patients

Rabea Wagener^{1,2} | Danielle Brandes¹ | Marie Jung¹ | Maxim A. Huetzen³ | Anke K. Bergmann⁴ | Stephanie Panier^{3,5} | Daniel Picard¹ | Ute Fischer^{1,2} | Ron D. Jachimowicz^{3,6,7,8} | Arndt Borkhardt^{1,2} | Triantafyllia Brozou¹

¹Department of Pediatric Oncology, Hematology and Clinical Immunology, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

²German Cancer Consortium (DKTK), partner site Essen/Düsseldorf, Düsseldorf, Germany

³Max Planck Research Group Mechanisms of DNA Repair, Max Planck Institute for Biology of Ageing, Cologne, Germany

⁴Institute of Human Genetics, Hannover Medical School (MHH), Hannover, Germany

⁵Institute for Genome Stability in Aging and Disease, Medical Faculty, University of Cologne, Cologne, Germany

⁶Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne and Düsseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

⁷Cologne Excellence Cluster on Cellular Stress Responses in Ageing-Associated Diseases, University of Cologne, Cologne, Germany

⁸Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany

Correspondence

Arndt Borkhardt, Department of Pediatric Oncology, Hematology and Clinical Immunology, Medical Faculty, Moorenstr. 5, 40225 Düsseldorf, Germany.
Email: arndt.borkhardt@med.uni-duesseldorf.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: 495318549; Deutsche Krebshilfe; Forschungskommission of the Medical Faculty of the Heinrich Heine University Düsseldorf, Grant/Award Number: 2020-28; Forschungskommission of the Medical Faculty of the Heinrich Heine University of Düsseldorf; German José Carreras leukemia foundation, Grant/Award Numbers: DJCLS 07 R/2019, DJCLS 18R/202, DJCLS 21 R/2019; Katharina Hardt foundation; Löwenstern e.V.

Abstract

Genetic predisposition is one of the major risk factors for pediatric cancer, with ~10% of children being carriers of a predisposing germline alteration. It is likely that this is the tip of the iceberg and many children are underdiagnosed, as most of the analysis focuses on single or short nucleotide variants, not considering the full spectrum of DNA alterations. Hence, we applied optical genome mapping (OGM) to our cohort of 34 pediatric cancer patients to perform an unbiased germline screening and analyze the frequency of structural variants (SVs) and their impact on cancer predisposition. All children were clinically highly suspicious for germline alterations (concomitant conditions or congenital anomalies, positive family cancer history, particular cancer type, synchronous or metachronous tumors), but whole exome sequencing (WES) had failed to detect pathogenic variants in cancer predisposing genes. OGM detected a median of 49 rare SVs (range 27-149) per patient. By analysis of

Abbreviations: aCGH, array comparative genome hybridization; BCP-ALL, B-cell-precursor acute lymphoblastic leukemia; CPG, cancer predisposition gene; CPS, cancer predisposition syndrome; DSBs, double strand breaks; GoF, gain of function; HMW, high molecular weight; HU, hydroxyurea; indels, insertions and deletions; OGM, optical genome mapping; PBMC, peripheral blood mononuclear cells; PFBMT6, telomere-related pulmonary fibrosis and/or bone marrow failure-6 syndrome; SNVs, single nucleotide variants; STS, short telomere syndrome; SVs, structural variants; WES, whole exome sequencing.

Rabea Wagener and Danielle Brandes have contributed equally to this work and shared first authorship.

Arndt Borkhardt and Triantafyllia Brozou have contributed equally to this work and shared last authorship.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *International Journal of Cancer* published by John Wiley & Sons Ltd on behalf of UICC.

18 patient-parent trios, we identified three de novo SVs. Moreover, we discovered a likely pathogenic deletion of exon 3 in the known cancer predisposition gene *BRCA2*, and identified a duplication in *RPA1*, which might represent a new cancer predisposition gene. We conclude that optical genome mapping is a suitable tool for detecting potentially predisposing SVs in addition to WES in pediatric cancer patients.

KEYWORDS

cancer predisposition, optical genome mapping, pediatric cancer, structural variants

What's new?

Around 10% of patients with pediatric cancer have been found to harbor a cancer predisposing germline alteration. However, the entire spectrum of genetic alterations has yet to be considered. Here, the authors analyzed 34 pediatric patients with a suspected underlying cancer predisposition syndrome by using the new optical genome mapping technique. They identified de novo or inherited genetic structural variants in *BRCA2* and in the potential novel cancer predisposition gene *RPA1*. The study demonstrates that optical genome mapping is a suitable tool for detecting potentially predisposing structural variants in addition to whole exome sequencing in pediatric cancer patients.

1 | INTRODUCTION

In recent years, several studies have shown that ~10% of pediatric cancer patients harbor a cancer predisposing germline alteration.¹⁻⁵ However, most of these studies focused on single nucleotide variants (SNVs) and small insertions and deletions (indels) ≤20 bp. In all of these studies, the whole exome sequencing (WES) analysis was mainly concentrated on a predefined gene panel of known cancer predisposition genes (CPG). Therefore, there is some question as to whether the remaining ~90% of patients carry any predisposing germline alteration, in particular as not the entire spectrum of genetic alterations was considered. In line with this, no causative genetic alteration was detectable in ~80% of pediatric cancer patients with clinical signs indicative of an underlying cancer predisposition syndrome (CPS).^{4,6} Although SNVs outnumber other types of DNA sequence variants by 7:1, structural variants (SVs) are 30-fold more prone to alter the gene expression. Moreover, 34% of all disease-causing variants are larger than SNVs, with 1/3 being >20 bp in size.⁷ Based on previous studies, 1.5% of adult cancer patients harbor a cancer predisposing SV.⁸ By contrast, the frequency seems to be more variable in pediatric cancer patients, with reports of 0.4%-3.2% of patients harboring a pathogenic germline SV.^{1,5} These studies are mostly based on short-read next generation sequencing approaches that have limitations with respect to accurate alignment in challenging genomic regions, which are prone to breakpoints causing SVs.⁹ Hence, the exact prevalence of SVs in the germline of cancer patients might be substantially underestimated. Recent proof-of-principle studies have shown that optical genome mapping (OGM) is a new method of reliably detecting constitutional SVs that can be used as an alternative to karyotyping, fluorescence in situ hybridization and DNA microarrays.^{10,11}

We here report on the applicability of OGM as a complementary instrument to the existing sequencing methods for unbiased germline screening of pediatric cancer patients regarding cancer predisposing SVs.

2 | MATERIALS AND METHODS

2.1 | Patient cohort

We included 34 patients of our germline study and, if available, their parents.¹² Those patients were positive for clinical criteria suggesting the presence of an underlying CPS,^{13,14} but lacking pathogenic cancer predisposing variants in WES.⁴ Refer to [Supplementary Material and Methods](#) for a detailed description of the cohort.

2.2 | Optical genome mapping

2.2.1 | Extraction and labeling of high-molecular weight DNA

We extracted high-molecular weight (HMW) DNA peripheral blood or bone marrow mononuclear cells from remission or, if available, from fibroblasts of the children and peripheral blood of the parents. Sufficient material was available from 18 child-parent trios, six child-parent duos and 10 child-singletons. The HMW DNA (750ng) was labeled, loaded on a Saphyr G2.3 chip and molecules were imaged by the Saphyr instrument (Gen 2, Bionano Genomics). Refer to Table S1 for an overview of the run statistics and quality metrics and [Supplementary Material and Methods](#) for a detailed description.

2.2.2 | De novo assembly and variant calling

The de novo assemblies against the human reference genome GRCh38, structural variant calling and annotation were performed using Bionano Solve (version 3.5.1, Bionano Genomics). The Bionano Access software (version 1.5.2, Bionano Genomics)

was used to visualize the data. For OGM data from the patient and one or both parents, we performed “Dual” or “Trio” variant annotation, respectively, to differentiate de novo and transmitted SVs. Refer to the [Supplementary Material and Methods](#) for a detailed description of the performed analysis and SV filtering.

2.3 | Variant annotation and interpretation

The final list of SVs was manually annotated using the UCSC genome browser tracks to identify whether an SV targets a gene (NCBI RefSeq, Ensembl: MANE Project [v1.0], GENCODE V39) and whether the SV region overlaps with a common SV ($\geq 1\%$ of the population) using the NCBI dbVAR Curated Common SVs database.¹⁵ We also checked whether the SVs overlap with known regulatory regions using GeneHancer (Double Elite).¹⁶ We annotated the overlap of an SV with a CPG using a published list.⁴ Genes targeted by an SV that potentially hits a coding region were explored with regard to clinical relevance for the cancerogenesis of the respective patient.

2.4 | Validation and functional testing

We performed validation experiments using DNA microarray, PCR-based assays and/or RNA sequencing for selected SVs and performed functional testing on P28 fibroblasts and age/gender-matched controls. Refer to the [Supplementary Material and Methods](#) for a detailed description of the performed analysis.

3 | RESULTS AND DISCUSSION

3.1 | Description of patient cohort

We analyzed 34 pediatric cancer patients with a median age of 5.4 years (range 0–18.5) at initial tumor diagnosis and a male to female ratio of 1.6:1 (Table S2). All patients harbored clinical signs indicative of an underlying predisposition such as congenital anomalies, concomitant conditions, multiple tumors or excessive toxicity to the chemotherapy (Figure 1A). Fifty-five percent of the patients harbored congenital anomalies, mainly of a mild form and not corresponding to a specific syndrome. Thirty-eight percent of the patients harbored positive family histories of pediatric and/or adult cancer. Nine patients were diagnosed with types of cancer strongly associated with genetic germline predisposition, such as sonic hedgehog-activated medulloblastoma. Moreover, three patients harbored more than two tumors (synchronous or metachronous) or showed excessive toxicity after chemotherapy.

3.2 | Identification of structural variants using optical genome mapping

We generated an average of 422.5 Gb of data per sample resulting in average 106 $\times$ effective coverage of the de novo assemblies

(Table S1). SV calling yielded a median of 6455 (range 5696–6685) SVs per sample. Applying filter steps to exclude artifacts and to detect rare SVs, yielded 1831 SVs, with a median of 49 (range 27–149) SVs per patient. Deletions and insertions were the most frequent SV types (median of 30 [range 15–93] and 18 [range 10–53], respectively; Figure 1B, Table S3). 95.6% of unbalanced SVs (considering deletions, duplications and insertions that can also constitute duplications) had a length of less than 50 kbp (Figure 1C). The majority of SVs were either intergenic (55%) or were located in an intron (19%) (Table S3). However, 26% of the SVs potentially affected coding sequences of one or more genes.

Due to our unique OGM dataset of 18 child-parent trios, we were able to identify de novo SVs (two deletions and one insertion) in 17% (3/18) of the children, affecting the *DDX10* or *FLG1* gene or an intergenic locus, respectively (Figure 1D, Table S4). This is in line with estimates based on WES published by Kloosterman et al, who reported de novo SVs >20 bp or >500 bp in 16% or 8% of children, respectively.¹⁷

3.3 | Structural variants targeting known cancer predisposition genes

In our list of high-confidence SVs, we checked whether they targeted known CPGs.⁴ Twenty-one SVs overlapped with a CPG, of which five potentially affected a coding region (Table S5).

Patients P6 and P10 both carried the same 1.4 kbp deletion in 17p13.1, potentially targeting the *TP53* gene (Figure S1A,B). Further analysis revealed that the deletion maps 5.5 kbp upstream of the *TP53* gene, overlapping with a described common deletion (nssv16197110) (Figure S1C,D).

After literature mining, we classified a deletion in *BRCA2*, which occurred in P26 as likely pathogenic (Table 1, Figure 2A). A maternally transmitted 4.8 kbp deletion was detected that overlapped with the exons 2–9 of the *BRCA2* gene. Performing PCR-based Sanger sequencing, we could confirm the presence of a 4.7 kbp deletion in chr13:32318724–32 323 440 bp, leading to deletion of the third exon of the *BRCA2* gene (*BRCA2* ^{Δ ex3}, ENST00000380152) (Figure 2B). Variants leading to *BRCA2* exon 3 deletions have been reported in individuals with breast and ovarian cancer.^{18,19} Although this deletion does not alter the reading frame, a part of the highly conserved N-terminal domain is lost that is important for the interaction with PALB2²⁰ and DNA repair. Next, we performed a semiquantitative PCR to check whether both the *BRCA2* ^{Δ ex3} and *BRCA2*^{wt} transcripts were expressed. While material was available from the child and the father, unfortunately, no material was available of the mother or the initially diagnosed leukemia of the patient to be examined for down-stream analysis. Interestingly, the child exhibited a higher expression of *BRCA2* ^{Δ ex3} than *BRCA2*^{wt} compared to the father and the control samples with predominant expression of *BRCA2*^{wt} transcript (Figure 2C). There has been some debate about the pathogenic effect of exon 3 deletion, as the *BRCA2* ^{Δ ex3} transcript is constitutively expressed in tissue samples from apparently healthy individuals. However, Muller et al reported an allelic imbalance in favor of the

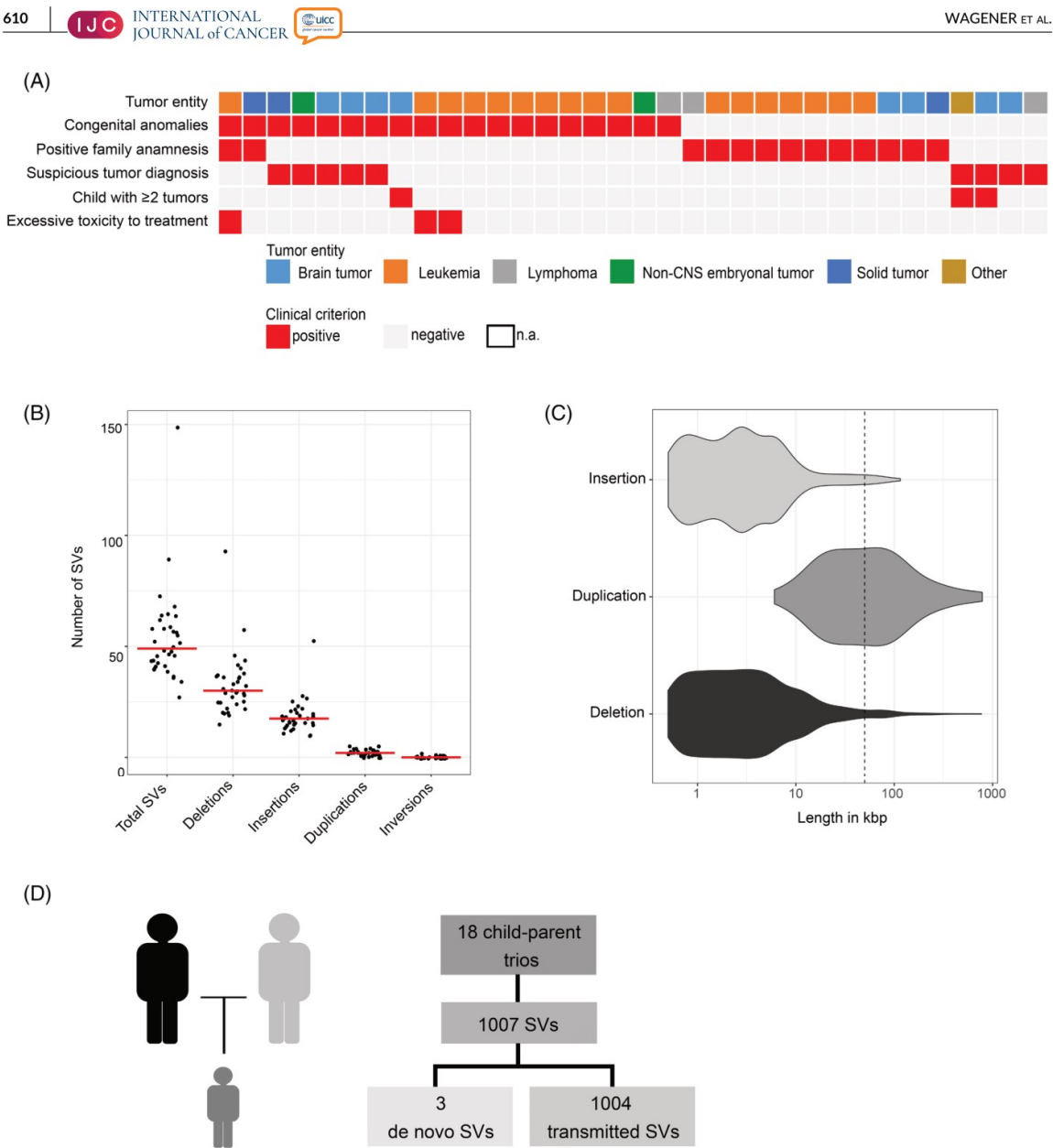


FIGURE 1 Structural variants detected by OGM in 34 pediatric patients suspicious for an underlying cancer predisposition syndrome. (A) Summary of patients' tumor diagnoses and status of clinical criteria. Each line represents a clinical feature and each column represents a patient. (B) Overview of structural variants identified using OGM. The dot plot represents the number of germline SVs per type, with every dot representing a patient. The red horizontal lines represent the median. (C) The violin plot depicts the length distribution of the identified SVs. The dashed line represents the 50 kbp threshold that is a lower limit of copy number variant length detectable by conventional DNA array-based methods. (D) Analysis of 18 child-parent trios using OGM revealed that 3/1007 detected SVs were de novo variants.

BRCA2^{Δex3} mutant allele¹⁸ and Caputo et al proposed that only variants producing transcripts lacking exon 3 should be classified as pathogenic.¹⁹ In line with this, we observed an allelic imbalance toward the *BRCA2*^{Δex3} transcript in our patient. The maternally transmitted

deletion segregated within the cancer history in the maternal lineage. The grandmother (carrier status unknown) was diagnosed with breast cancer at the age of 32, and the patient developed an *ETV6::RUNX1* rearranged BCP-ALL at the age of 12, whereas the patients' mother

TABLE 1 Structural variants and affected genes potentially playing a role in cancer predisposition of two pediatric patients.

Case	P26	P28
Tumor type	BCP-ALL	BCP-ALL
Subtype	ETV6::RUNX1	KMT2A::AFF1
Phenotype	Congenital cardiac anomaly (HP:0001627)	Arched eyebrows (HP:0002553)
Genomic region (hg38)	chr13:32 315 668-32 331 479 bp	chr17:1 747 307-1 899 562 bp
SV type	Deletion	Duplication
Size	4.8 kbp	166 kbp
Affected genes of interest	BRCA2	RPA1
Overlap common dbVar	None	None
Validation	WES, Sanger, RT-qPCR	WES, SNP-array, qPCR
Functional testing	n.a.	Telomere length, DNA damage
Interpretation	Likely pathogenic	Likely pathogenic

Note: Overlap common dbVAR: Indicating whether the region in which the SV was observed overlaps with a known copy number alteration using the common dbVar database (>1% in the population) as reference.

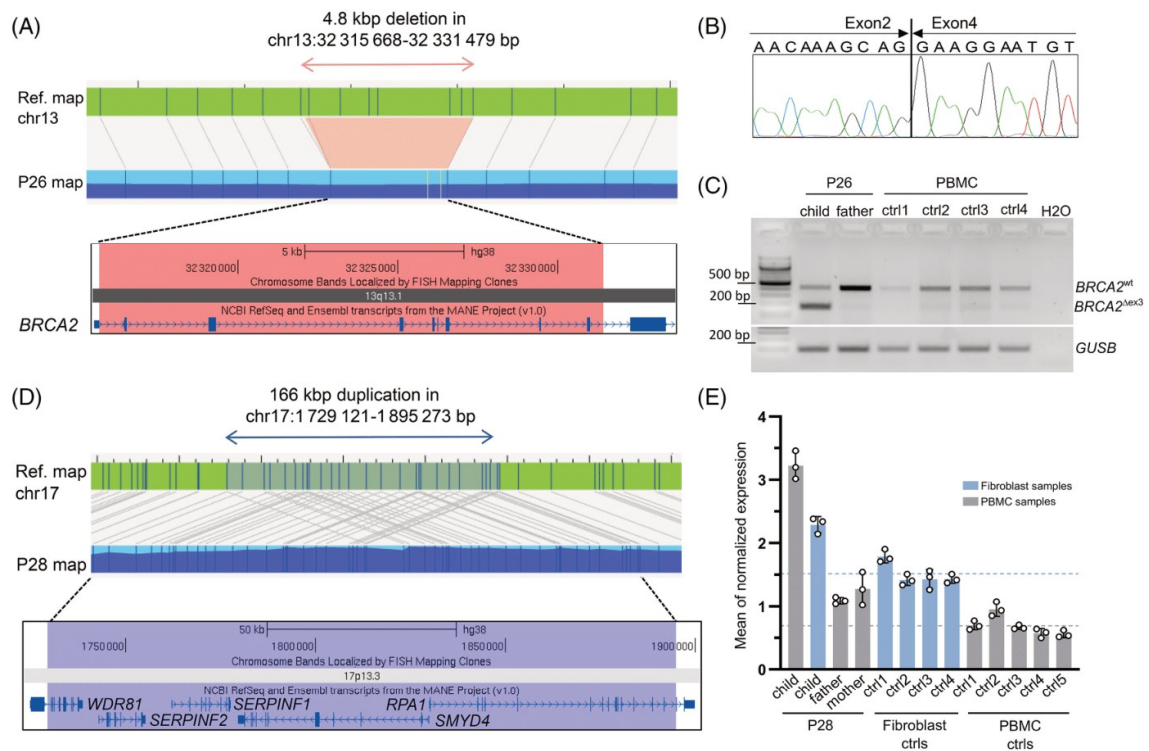


FIGURE 2 Structural variants affecting the known CPG *BRCA2* and the potential candidate CPG *RPA1* in two pediatric cancer patients. (A) Germline deletion detected in the cancer predisposition gene *BRCA2* in P26. Genome maps (green bar above reference and blue bar below patient) indicating the presence of a 4.8 kbp deletion in 13q13.1. Lower part of the figure depicts the location of the deleted region, highlighted in red, which is located within the *BRCA2* gene. (B) Sanger sequencing of the mRNA of the patient confirms the deletion of exon 3 in *BRCA2* (ENST00000380152). (C) Upper panel: Semiquantitative PCR of *BRCA2*^{wt} and *BRCA2*^{Δex3} transcript RNA extracted from the PBMCs of P26 carrying the 4.8 kbp deletion in comparison to the father (P26_father) and four control samples (PBMC ctrl 1-4). Lower panel: *GUSB* was used as control. (D) Genome map of duplicated region affecting the *RPA1* gene locus (17p13.3) in P28 (green bar represents reference map and blue bars the patient map) and the location within the genome using UCSC genome browser (duplication highlighted in blue). (E) Analysis of *RPA1* mRNA expression levels reveals higher expression in PBMCs and fibroblasts of P28 in comparison to the parents and controls. The controls were separated based on tissue of origin, including fibroblasts (blue) and PBMC (gray) control samples. The dashed lines represent the mean expression levels of the fibroblast controls (blue) and PBMC controls (gray).

10970215, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34721 by Universitat Unit Landeshbibliothek, Wiley Online Library on [09/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

had not yet developed any cancer at the age of 40. Although, *BRCA2* alterations have not been described to increase the susceptibility to ALL as yet, the internal *BRCA2* deletion might be a possible risk factor.

3.4 | Identification of novel candidate cancer predisposition genes

98.8% of all SVs identified by OGM in our study did not overlap with a known CPG. Performing a comprehensive literature search and further analysis, we identified three interesting SVs potentially targeting novel CPGs.

In patient P16, a girl who synchronously developed a medulloblastoma and an astrocytoma at the age of 3, we identified a maternally transmitted 180 kbp tandem duplication in 8p11.22, verified by SNP array and coverage-based WES data (Figure S2A-C). The duplicated region encompasses the *ADAM9* and *ADAM32* genes, which belong to the “disintegrin and metalloprotease” (ADAM) family. *ADAM9* is overexpressed in glioblastoma, where it promotes invasiveness and proliferation.^{21,22} To test the effect on *ADAM9* and *ADAM32* expression or if the duplication leads to a potential fusion transcript, we performed whole transcriptome analysis of the patient and the noncarrier father. We did not detect an *ADAM32::ADAM9* fusion transcript or altered expression of both genes in the affected child (Figure S2D). In addition, we identified compound heterozygous SVs in the *RAP1B* gene, in P10, a boy with B-ALL. The patient carried a maternally transmitted 9.3 kbp insertion and a paternally transmitted 10.8 kbp deletion, both affecting regulatory regions of *RAP1B* gene locus (Figure S3A,B). *RAP1B* encodes for a protein of the RAS family of small GTP binding proteins with mitogenic and oncogenic properties.^{23,24} Activating missense variants in *RAP1B* have been described in three patients with syndromic thrombocytopenia and multiple malformations, including microcephaly and learning difficulties.^{25,26} In line with these findings, P10 had mild global developmental delay and a growth disorder in addition to his B-ALL. However, the analysis of *RAP1B* mRNA expression in the PBMCs of the patient revealed no altered expression in comparison to the father or the controls (Figure S3C).

Conclusively, additional experiments including corresponding tumor analysis are required to clarify whether the identified germline SVs affecting *ADAM9* and *RAP1B* have a functional influence on the development of the patient's tumors.

The most interesting SV was a 166 kbp duplication in 17p13.3 in P28, an infant diagnosed with *KMT2A*-rearranged B-cell-precursor acute lymphoblastic leukemia (BCP-ALL) at the age of 4 months, who relapsed 6 months after initial diagnosis (Table 1, Figure 2D). The duplication was present in the germline and the corresponding tumor of P28 (Figure S4A). We verified the copy number gain in the child using coverage-based information from WES and could show that this gain was not present in the WES data of the parents, indicating that the SV occurred de novo (Figure S4B). Using SNP array, we could show that the duplication is located in a region of copy number neutral loss of heterozygosity in 17p13.3-p13.1 (Figure S4C). Due to the

consanguinity of the parents (first-degree cousins), 11.2% of the patient's autosomal genome is homozygous (Figure S4D). The duplication affected five genes, including *RPA1*, which has essential functions in DNA replication, repair, and recombination.²⁷ It is associated with autosomal dominant inherited telomere-related pulmonary fibrosis and/or bone marrow failure-6 syndrome (PFBMFT6, MIM:619767). The PFBMFT6 syndrome belongs to the short telomere syndromes (STS) with variable symptoms, genetic heterogeneity and variable expressivity. Analysis of the *RPA1* mRNA level in PBMCs and fibroblasts of the child revealed an increased expression in comparison to parents and controls, which points toward the duplication having a gain-of-function (GoF) effect (Figure 2E). This is in line with reports by Sharma et al, who described GoF *RPA1* missense variants in patients with short telomeres and varying clinical features of STS.²⁸ We further assessed *RPA1* protein expression in the patient's fibroblasts in comparison to two age/gender-matched fibroblast samples from children without *RPA1* alterations, which revealed similar *RPA1* expression (Figure S5A). We screened for a potential telomeropathy by performing TeloTAGGG Assay on fibroblasts of P28 carrying the *RPA1* duplication in comparison to age/gender-matched control fibroblasts (Figure S5B,C). However, we even detected a mild increase of mean telomere restriction fragment length in the *RPA1* duplicated cells compared to the controls (P28 = 13.3 kb vs Ctrl1 = 10.4 kb; Ctrl2 = 11.6 kb, respectively).

Additionally, it has been reported that *RPA1* duplication induces elevated sensitivity to DNA damage.²⁹ Therefore, we also checked for an underlying DNA repair defect in P28 by performing cell viability assays to analyze sensitivity to genotoxic agents (Figure S5D,E). We used the chemotherapeutic agent cisplatin, inducing DNA cross-linking and DNA double-strand breaks (DSBs), and the ribonucleotide reductase inhibitor hydroxyurea (HU) that depletes nucleotides and induces replication stress leading to subsequent DSBs. Fibroblasts of P28 carrying the *RPA1* duplication exhibited a significant increase in sensitivity to cisplatin und HU compared to wild-type age/gender-matched controls. Thus, in line with the published findings,²⁹ P28 fibroblasts with *RPA1* duplication showed elevated sensitivity to DNA damage. The extent to which the described genotype might be associated with the origin of the *KMT2A*+ infant ALL and if the patient, who remains very young, may develop shortened telomeres needs to be monitored in the future.

4 | CONCLUSION

By applying OGM, we comprehensively characterized the SV landscape of 34 pediatric cancer patients who were clinically highly suspicious for an underlying CPS and had a nonconclusive WES result. We identified three de novo SVs among 18 child-parent trios and two hot germline alterations affecting the known CPG *BRCA2* and the potential new CPG candidate *RPA1* in two respective leukemia patients.

Our study highlights the suitability of OGM in a clinical setting to detect potentially predisposing SVs in suspicious pediatric cases. However, there remains the necessity for applying additional

sequencing-based methods to determine the exact breakpoint boundaries when interpreting OGM reported SVs. We expect that extended germline screens, improvement of the OGM reference database and integration of the telomere-to-telomere reference genome will enhance SV detection and interpretation of disease-causing events. Further studies will likely use a combination of OGM and long-read sequencing technologies to overcome the current limitations of OGM resolution (<500 bp) to cover the full spectrum of genetic predisposition in children with cancer.

AUTHOR CONTRIBUTIONS

Rabea Wagener organized the project and wrote the paper. Rabea Wagener, Danielle Brandes, Marie Jung and Triantafyllia Brozou analyzed and interpreted the data. Danielle Brandes and Marie Jung extracted the DNA and performed the analysis. Triantafyllia Brozou recruited patients, provided clinical information, obtained informed consent and revised the manuscript. Maxim A. Huetzen, Stephanie Panier and Ron D. Jachimowicz performed functional analysis of RPA1. Daniel Picard analyzed RNA-sequencing data. Anke K. Bergmann provided cytogenetic and molecular genetic data. Ute Fischer and Arndt Borkhardt discussed the data and outlined the project. All authors contributed to and approved the final manuscript. The work reported in the paper has been performed by the authors, unless clearly specified in the text.

ACKNOWLEDGEMENTS

The authors wish to thank Mr. Daniel Scholtysik and Ms. Silke Furlan for their excellent technical assistance. Computational support and infrastructure was provided by the Centre for Information and Media Technology (ZIM) at the University of Düsseldorf (Germany). Maxim A. Huetzen is member of the Cologne Graduate School of Ageing Research. The graphical abstract was created with the help of Biorender. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This work was supported by the Katharina Hardt foundation and Löwenstern e.V. Rabea Wagener acknowledges the financial support of the Forschungskommission of the Medical Faculty of the Heinrich Heine University Düsseldorf (2020-28). Ute Fischer (DJCLS 18R/2021 and DJCLS 21 R/2019) and Arndt Borkhardt (DJCLS 07 R/2019) were funded by the German José Carreras leukemia foundation. Ute Fischer has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation—Projektnummer 495318549) and the German Cancer Aid (Priority Programme ‘Cancer Prevention—Graduate School [CPGS]’).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of the study are available from the corresponding author upon request.

ETHICS STATEMENT

The study was approved by the ethics committee of Heinrich Heine University, Düsseldorf, Germany (ethics vote number 4886 R and study registration number 2014112933).

ORCID

Rabea Wagener  <https://orcid.org/0000-0002-5394-0907>

Arndt Borkhardt  <https://orcid.org/0000-0002-6121-4737>

TWITTER

Arndt Borkhardt  [Borkhardt_Lab](#)

REFERENCES

1. Zhang J, Walsh MF, Wu G, et al. Germline mutations in predisposition genes in pediatric cancer. *N Engl J Med*. 2015;373(24):2336-2346. doi:10.1056/NEJMoa1508054
2. Diets IJ, Waanders E, Ligtenberg MJ, et al. High yield of pathogenic germline mutations causative or likely causative of the cancer phenotype in selected children with cancer. *Clin Cancer Res*. 2018;24(7):1594-1603. doi:10.1158/1078-0432.Ccr-17-1725
3. Gröbner SN, Worst BC, Weischenfeldt J, et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555(7696):321-327. doi:10.1038/nature25480
4. Wagener R, Taeubner J, Walter C, et al. Comprehensive germline-genomic and clinical profiling in 160 unselected children and adolescents with cancer. *Eur J Hum Genet*. 2021;29(8):1301-1311.
5. Wong M, Mayoh C, Lau LMS, et al. Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. *Nat Med*. 2020;26(11):1742-1753. doi:10.1038/s41591-020-1072-4
6. Byrjalsen A, Hansen TVO, Stoltze UK, et al. Nationwide germline whole genome sequencing of 198 consecutive pediatric cancer patients reveals a high incidence of cancer prone syndromes. *PLoS Genet*. 2020;16(12):e1009231. doi:10.1371/journal.pgen.1009231
7. Eichler EE. Genetic variation, comparative genomics, and the diagnosis of disease. *N Engl J Med*. 2019;381(1):64-74. doi:10.1056/NEJMr1809315
8. Whitworth J, Smith PS, Martin JE, et al. Comprehensive cancer-predisposition gene testing in an adult multiple primary tumor series shows a broad range of deleterious variants and atypical tumor phenotypes. *Am J Hum Genet*. 2018;103(1):3-18. doi:10.1016/j.ajhg.2018.04.013
9. Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. *Genome Biol*. 2019;20(1):246. doi:10.1186/s13059-019-1828-7
10. Dremsek P, Schwarz T, Weil B, Malashka A, Laccone F, Neesen J. Optical genome mapping in routine human genetic diagnostics—its advantages and limitations. *Genes*. 2021;12(12):1958. doi:10.3390/genes12121958
11. Mantere T, Neveling K, Pebrel-Richard C, et al. Optical genome mapping enables constitutional chromosomal aberration detection. *Am J Hum Genet*. 2021;108(8):1409-1422. doi:10.1016/j.ajhg.2021.05.012
12. Brozou T, Taeubner J, Velleuer E, et al. Genetic predisposition in children with cancer—affected families' acceptance of Trio-WES. *Eur J Pediatr*. 2018;177(1):53-60.
13. Jongmans MC, Loeffen JL, Waanders E, et al. Recognition of genetic predisposition in pediatric cancer patients: an easy-to-use selection tool. *Eur J Med Genet*. 2016;59(3):116-125. doi:10.1016/j.ejmg.2016.01.008
14. Ripperger T, Bielack SS, Borkhardt A, et al. Childhood cancer predisposition syndromes—a concise review and recommendations by the

- cancer predisposition working Group of the Society for pediatric oncology and hematology. *Am J Med Genet A*. 2017;173(4):1017-1037.
15. Lappalainen I, Lopez J, Skipper L, et al. DbVar and DGVA: public archives for genomic structural variation. *Nucleic Acids Res*. 2013;41-(Database issue):D936-D941. doi:10.1093/nar/gks1213
 16. Fishilevich S, Nudel R, Rappaport N, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*. 2017;2017:bax028. doi:10.1093/database/bax028
 17. Kloosterman WP, Francioli LC, Hormozdiari F, et al. Characteristics of de novo structural changes in the human genome. *Genome Res*. 2015;25(6):792-801. doi:10.1101/gr.185041.114
 18. Muller D, Rouleau E, Schultz I, et al. An entire exon 3 germ-line rearrangement in the BRCA2 gene: pathogenic relevance of exon 3 deletion in breast cancer predisposition. *BMC Med Genet*. 2011;12:121. doi:10.1186/1471-2350-12-121
 19. Caputo SM, Léone M, Damiola F, et al. Full in-frame exon 3 skipping of BRCA2 confers high risk of breast and/or ovarian cancer. *Oncotarget*. 2018;9(25):17334-17348. doi:10.18632/oncotarget.24671
 20. Xia B, Sheng Q, Nakanishi K, et al. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol Cell*. 2006;22(6):719-729. doi:10.1016/j.molcel.2006.05.022
 21. Ji T, Zhang X, Li W. MicroRNA-543 inhibits proliferation, invasion and induces apoptosis of glioblastoma cells by directly targeting ADAM9. *Mol Med Rep*. 2017;16(5):6419-6427. doi:10.3892/mmr.2017.7332
 22. Fan X, Wang Y, Zhang C, et al. ADAM9 expression is associate with glioma tumor grade and histological type, and acts as a prognostic factor in lower-grade gliomas. *Int J Mol Sci*. 2016;17(9):1276. doi:10.3390/ijms17091276
 23. Altschuler DL, Ribeiro-Neto F. Mitogenic and oncogenic properties of the small G protein Rap1b. *Proc Natl Acad Sci U S A*. 1998;95(13):7475-7479. doi:10.1073/pnas.95.13.7475
 24. Lou L, Urbani J, Ribeiro-Neto F, Altschuler DL. cAMP inhibition of Akt is mediated by activated and phosphorylated Rap1b. *J Biol Chem*. 2002;277(36):32799-32806. doi:10.1074/jbc.M201491200
 25. Niemann JH, Du C, Morlot S, et al. De novo missense variants in the RAP1B gene identified in two patients with syndromic thrombocytopenia. *Clin Genet*. 2020;98(4):374-378. doi:10.1111/cge.13807
 26. Miller D, Saeed A, Nelson AC, Bower M, Aggarwal A. Third reported patient with RAP1B-related syndromic thrombocytopenia and novel clinical findings. *Am J Med Genet A*. 2022;188(9):2808-2814. doi:10.1002/ajmg.a.62760
 27. Wold MS. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem*. 1997;66:61-92. doi:10.1146/annurev.biochem.66.1.61
 28. Sharma R, Sahoo SS, Honda M, et al. Gain-of-function mutations in RPA1 cause a syndrome with short telomeres and somatic genetic rescue. *Blood*. 2022;139(7):1039-1051. doi:10.1182/blood.2021011980
 29. Outwin E, Carpenter G, Bi W, Withers MA, Lupski JR, O'Driscoll M. Increased RPA1 gene dosage affects genomic stability potentially contributing to 17p13.3 duplication syndrome. *PLoS Genet*. 2011;7(8):e1002247. doi:10.1371/journal.pgen.1002247

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wagener R, Brandes D, Jung M, et al. Optical genome mapping identifies structural variants in potentially new cancer predisposition candidate genes in pediatric cancer patients. *Int J Cancer*. 2024;154(4):607-614. doi:10.1002/ijc.34721

3.2.1 Original Publication I: Supplementary Methods and Figures

Relevant results for the discussion of this thesis are included in the following section. Complete supplementary methods and figures can be found online at the publisher site (refer to supporting information):

<https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fijc.34721&file=ijc34721-sup-0002-Supinfo.pdf>

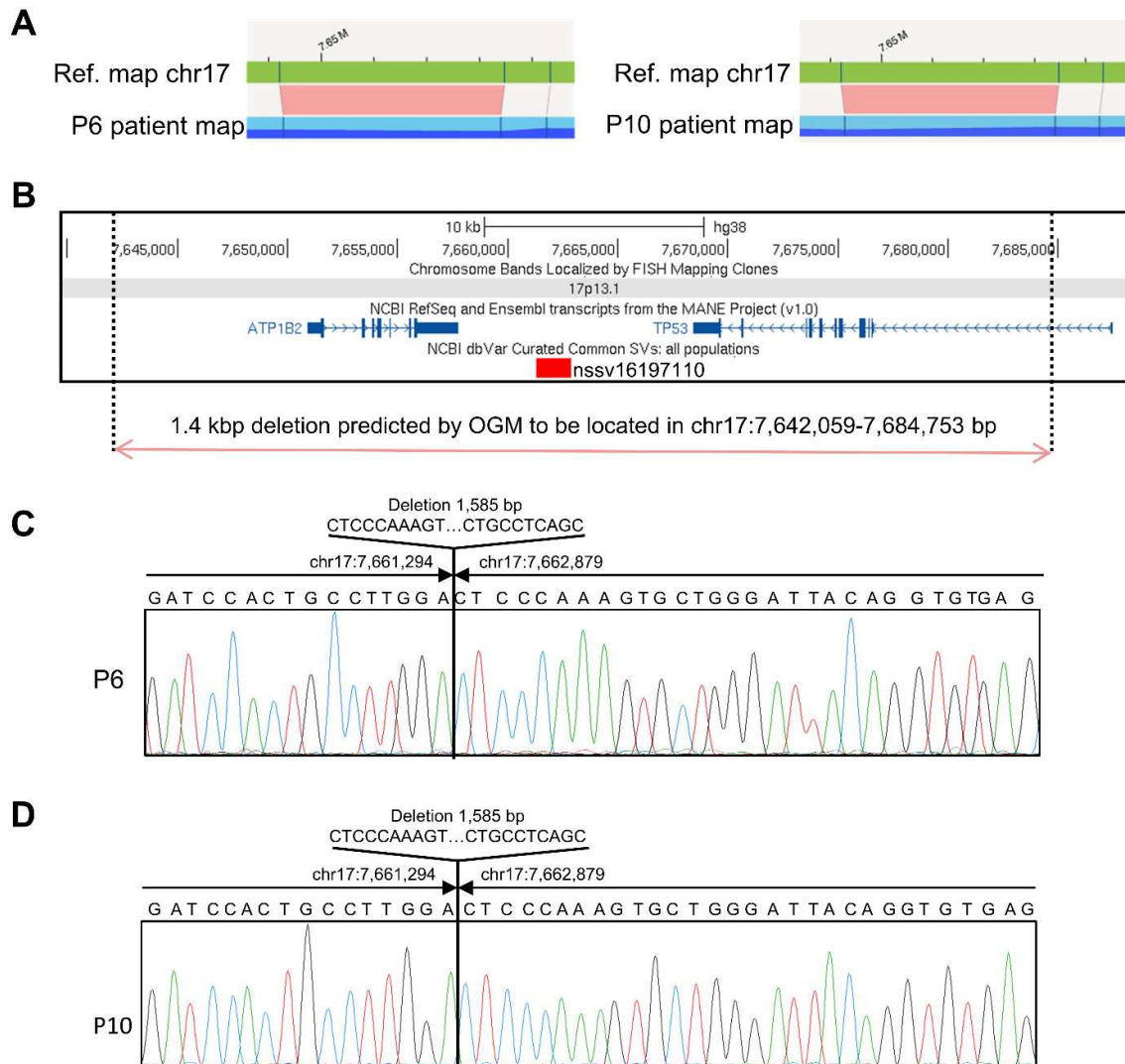
Supplementary Material and Methods

Rare Variant Analysis Pipeline

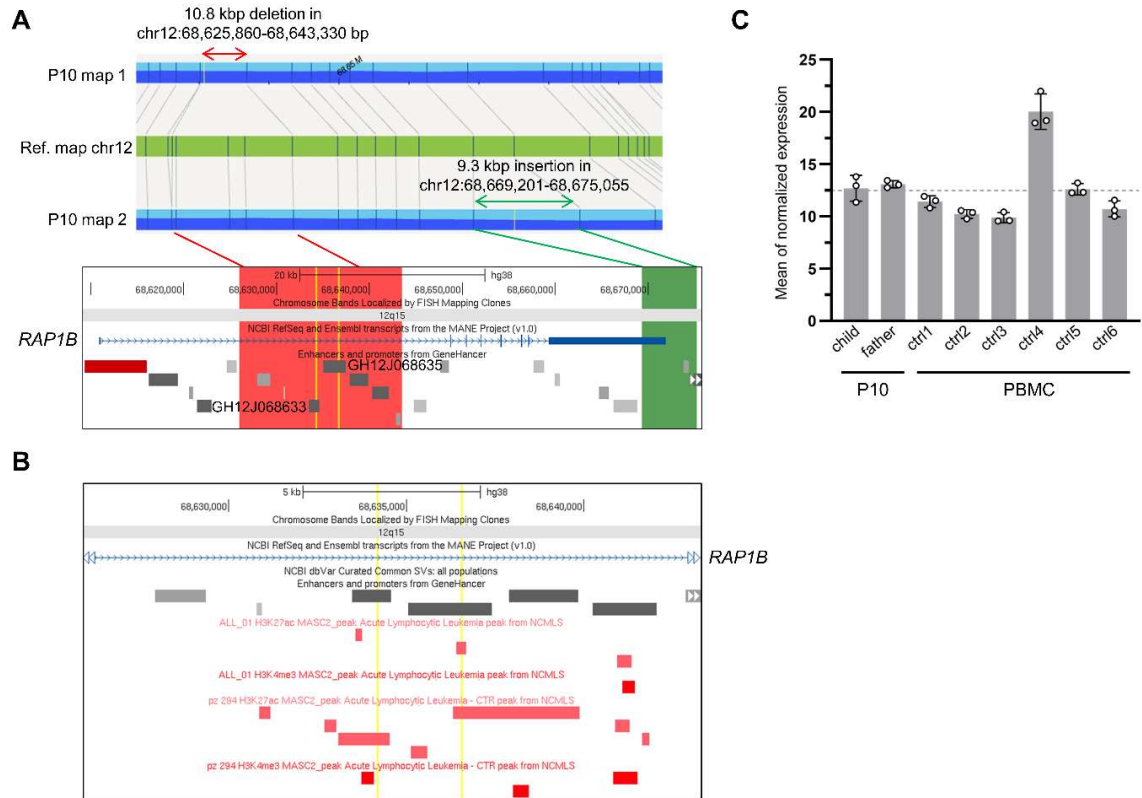
Moreover, we performed a rare variant analysis of the tumor and the germline of patient P31 in order to check for the presence of the ring chromosome r(21). We applied the algorithm using the default settings against the human reference genome GRCh38. The variants were annotated using the “Dual” annotation pipeline, allowing the identification of somatic SVs. Somatic SVs were filtered using the following settings: 1) confidence score 0 and minimum length of 500 bp for insertions and deletions; 2) confidence score >0.01 for inversions; 3) confidence score 0 for translocations; and 4) confidence score -1 for duplications.

Taken together, we (1) repeated the de novo assembly algorithm on this sample using the newest software version 1.7 and (2) we analyzed the tumor sample with the rare variant analysis pipeline of Bionano Genomics. Irrespective of the analysis algorithm applied, we were not able to detect the ring chromosome 21 in patient P31 using OGM.

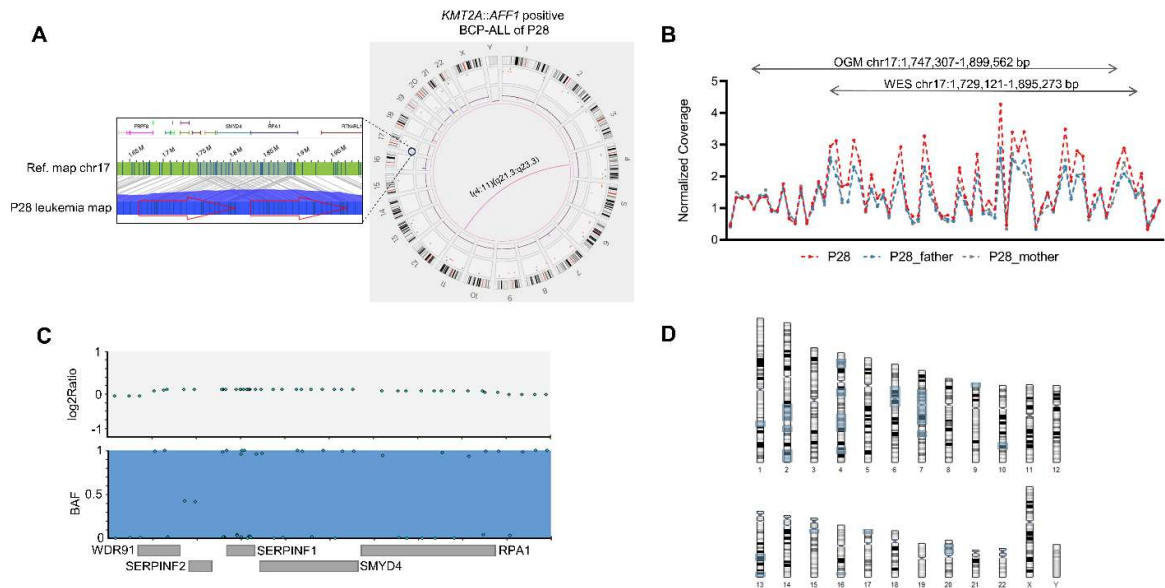
Supplementary Figures



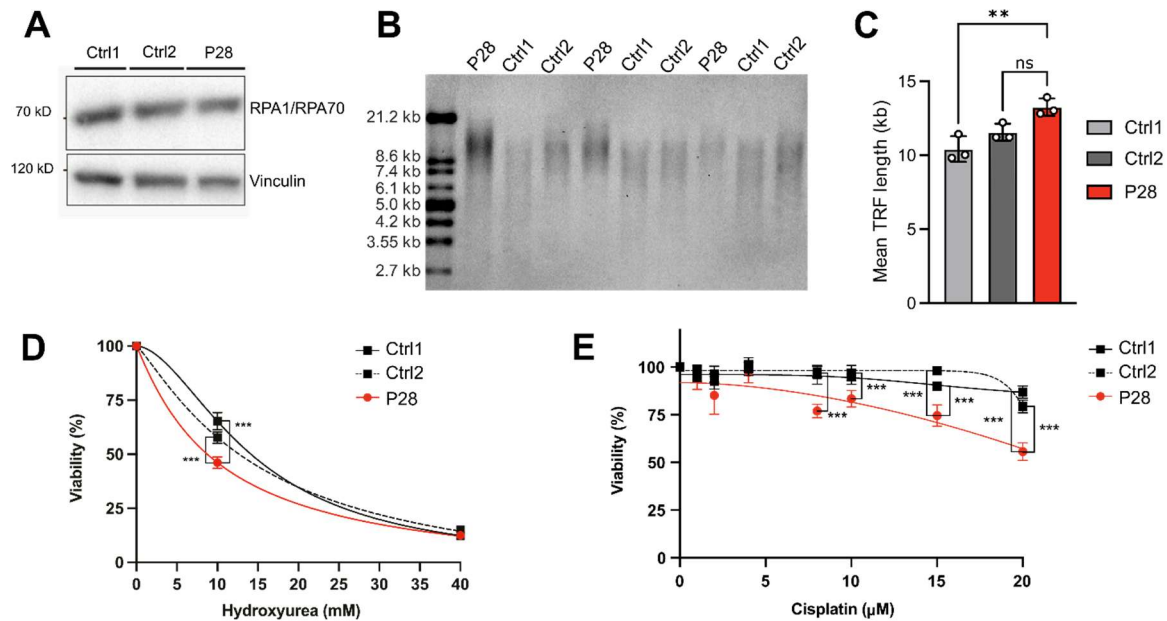
Supplementary Figure 1: Deletions detected in 17p13.1 by whole-genome optical mapping in patients P6 and P10. A) Genome maps (green bar above reference and blue bar below patient) indicating the presence of a 1.4 kbp deletion in 17p13.1 in patient P6 and P10. B) The UCSC track displays the region in 17p13.1 (chr17:7,642,059-7,684,753 bp) in which the 1.4 kbp deletion was detected by whole-genome optical mapping. The *ATP1B2* gene and the cancer-predisposition gene *TP53* (NCBI RefSeq and Ensembl transcripts track) are located in that region and, hence, could be targeted by the deletion. Moreover, a common deletion (nssv16197110) is located in this region. Performing PCR-based Sanger sequencing in patients P6 (C) and P10 (D), we confirmed the breakpoints of the deletion to be chr17:7,661,294-7,662,879 bp, yielding a 1.5 kbp deletion that overlaps with the common deletion nssv16197110. Accordingly, this deletion is a common polymorphism located upstream of the *TP53* gene.



Supplementary Figure 3: Compound heterozygous SV in *RAP1B*. A) Upper panel displays the genome map of a region in 12q15 targeted by compound heterozygous SVs affecting *RAP1B* gene in patient P10 (green bar represents reference map and blue bars the patient maps). Patient map 1 displays a 10.8 kbp deletion in chr12:68,625,860-68,643,330 bp, whereas patient map 2 shows a 9.3 kbp insertion downstream of the deletion. The lower panel shows the UCSC genome browser track of the *RAP1B* gene region, with the green highlight displaying the region with the insertion and the red highlight the region of the deletion. The yellow lines indicate the location of the labels that are deleted based on OGM. Moreover, the enhancers (grey bars) and promoters (red bars) from the GeneHancer database are displayed. B) UCSC genome browser view of region in intron 1 of *RAP1B* gene containing a 10.8 kbp deletion in patient P10 based on OGM. Highlighted in yellow are the labels that are deleted in this region, which allow us to pinpoint the deleted regions more precisely. The deleted region overlaps with several enhancers (GeneHancer track, grey bars). Interestingly, using the ChIP-seq data for H3K27ac and H3K4me3 histone markers of two B-ALL samples (ALL_01 from NCMLS and pz 294 from NCMLS), we observed an open chromatin confirmation in this deleted enhancer region, which might point to this region having a regulatory function for *RAP1B* gene expression in B-ALL. C) Analysis of *RAP1B* mRNA expression levels reveals comparable expression levels in the blood (PB) of the patient (P1_child) in comparison to the father and controls. The dashed lines represent the mean expression levels of the blood (PB) control samples.



Supplementary Figure 4: Verification of duplication in 17p13.3 identified in patient P28 using OGM. A) Circos plot of the corresponding leukemia is shown with $t(4;11)(q21.3;q23.3)$ leading to *KMT2A::AFF1* fusion. Optical map of the persistent duplication in the leukemia map (blue) compared to the reference map (green) in 17p13.3 is shown. B) WES-based coverage information of the patient and the parents in the chromosomal region chr17:1,725,175-1,937,747 bp overlapping with the predicted duplicated region. Each dot represents the normalized coverage value of a sequenced target region. The arrows above represent the duplication predicted by OGM and the copy number gain observed by WES in the patient as an increase in coverage in comparison to the parents. C) Display of the SNP-array analysis of the patient, with the upper panel showing the log2Ratio and the lower panel the B-allele frequency (BAF), confirming the presence of the duplication. Moreover, a copy number neutral loss of heterozygosity was identified in 17p13.3-p13.1 encompassing the duplicated region. D) Karyotype of patient P28 showing 22 regions of homozygosity (>3Mbp), highlighted in blue as detected by SNP-array. The autosomal genomic measure of individual autozygosity is 11.2%, in line with a third-degree parental relationship.



Supplementary Figure 5: Functional validation of cells of patient P28 with focal duplication of 17p13.3, including *RPA1*-gene locus. Cells show comparable RPA1 protein expression, a mild increase of telomere length and higher sensitivity to DNA damage and replication stress, compared to age/gender-matched controls. A) Immunoblot detecting RPA1 protein expression in dermal fibroblasts of patient P28 and age/gender-matched fibroblast controls (Ctrl1 and Ctrl2) from children without *RPA1* alterations. B) TeloTAGGG assay: southern blot showing telomere fragment lengths of dermal fibroblasts of patient P28 and age/gender-matched fibroblast controls (Ctrl1 and Ctrl2) from children without *RPA1* alterations. Telomere length is calibrated against a DIG-labelled molecular weight marker. Biological triplicates are shown. C) Quantification of the representative TeloTAGGG southern blot. Chemiluminescence signal peaks of the telomere restriction fragments (TRF) were defined as mean TRF length. Biological triplicates were analyzed. p-values were calculated using One-Way-ANOVA. Ns $p > 0.05$ and $**p < 0.01$. D) Cell-Titer-Glo viability assay: Replication stress was induced through nucleotide depletion by hydroxyurea treatment. One of two biological replicates is shown. Every biological replicate contained four technical replicates. e) Cell-Titer-Glo viability assay: DNA damage was induced by Cisplatin treatment. One of two biological replicates is shown. Every biological replicate contained four technical replicates. p-values were calculated using Two-Way-ANOVA with multiple comparisons. $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$.

3.2.2 Original Publication I: Supplementary data

Relevant results for the discussion of this thesis were included in the following section. Complete supplementary data can be found online at the publisher site (refer to supporting information):

<https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fijc.34721&file=ijc34721-sup-0001-Tables.xlsx>

Table S3: Number of SVs identified pre- and post-filtering in the germline of the 34 pediatric cancer patients.

Case ID	Pre-filtering							Post-filtering							Location of SVs		
	Total number of SVs	Number of Deletions	Number of Insertions	Number of Inversions	Number of Duplications	Number of Translocations	Total number of SVs	Number of Deletions	Number of Insertions	Number of Inversions	Number of Duplications	Number of Translocations	De novo SVs	Homozygous SVs	Intergenic	Intronic	potentially exonic
P1	6465	1772	4450	144	77	52	89	58	27	1	3	0	n.a.	2	44	17	28
P2	6447	1767	4440	137	54	49	64	42	19	1	2	0	n.a.	1	38	9	17
P3	6521	1771	4461	164	72	53	37	22	15	0	0	0	n.a.	1	20	7	10
P4	6491	1738	4475	159	65	54	27	15	11	0	1	0	0	0	14	5	8
P5	6454	1759	4440	139	60	56	44	27	16	0	1	0	n.a.	1	24	9	11
P6	6467	1721	4476	147	66	57	40	25	15	0	0	0	0	0	19	10	11
P7	6574	1758	4545	155	59	57	48	25	20	0	3	0	n.a.	1	28	6	14
P8	6370	1731	4335	153	88	63	47	30	13	0	4	0	n.a.	1	27	8	12
P9	6340	1712	4411	117	53	47	44	31	10	0	3	0	n.a.	2	27	6	11
P10	6655	1859	4489	166	65	76	59	36	21	0	2	0	0	0	38	10	11
P11	6314	1673	4416	119	56	50	48	28	20	0	0	0	n.a.	2	22	12	14
P12	6330	1687	4401	140	54	48	34	20	12	2	0	0	n.a.	1	21	5	8
P13	6367	1761	4377	123	61	45	39	22	14	0	3	0	n.a.	1	18	7	14
P14	6865	1814	4552	169	94	56	149	93	53	0	3	0	0	0	86	28	35
P15	6354	1737	4380	129	58	50	52	29	19	1	3	0	0	3	26	7	19
P16	6448	1711	4473	137	60	67	41	29	10	0	2	0	0	0	20	8	13
P17	6449	1782	4393	137	73	64	40	20	16	0	4	0	n.a.	0	22	6	12
P18	6616	1819	4521	181	43	52	62	37	22	1	2	0	0	7	33	14	15
P19	5981	1628	4133	116	43	61	52	32	18	0	1	0	0	3	27	11	13
P20	6455	1736	4465	132	54	68	43	24	18	0	1	0	0	1	26	11	6
P21	6459	1767	4422	137	68	65	57	37	16	0	4	0	0	1	28	17	12
P22	5696	1628	3903	80	45	40	41	20	18	0	3	0	0	0	20	8	13
P23	6284	1661	4379	132	58	54	36	19	13	0	4	0	n.a.	1	21	6	9
P24	6467	1770	4441	139	71	46	68	40	22	1	5	0	0	2	44	9	15
P25	6635	1781	4513	193	76	72	65	46	18	0	1	0	0	1	39	10	16
P26	6435	1732	4414	149	81	59	58	33	23	0	2	0	0	0	30	8	20
P27	6544	1745	4514	156	60	69	56	38	16	0	2	0	0	1	26	15	15
P28	6213	1664	4311	114	79	45	58	36	19	1	2	0	n.a.	5	34	15	9
P29	6533	1807	4452	153	60	61	64	34	25	0	5	0	0	1	38	11	15
P30	6482	1740	4441	169	73	59	46	29	15	0	2	0	0	2	25	7	14
P31	6531	1764	4479	140	85	63	73	44	28	0	1	0	n.a.	3	43	15	15
P32	6413	1680	4469	134	68	62	46	25	16	0	5	0	0	0	28	7	11
P33	6250	1732	4291	113	52	79	50	30	17	1	2	0	0	1	22	14	11
P34	6457	1703	4476	136	63	79	55	36	17	0	2	0	n.a.	2	34	10	11

3.3 Summary of Original Publication II

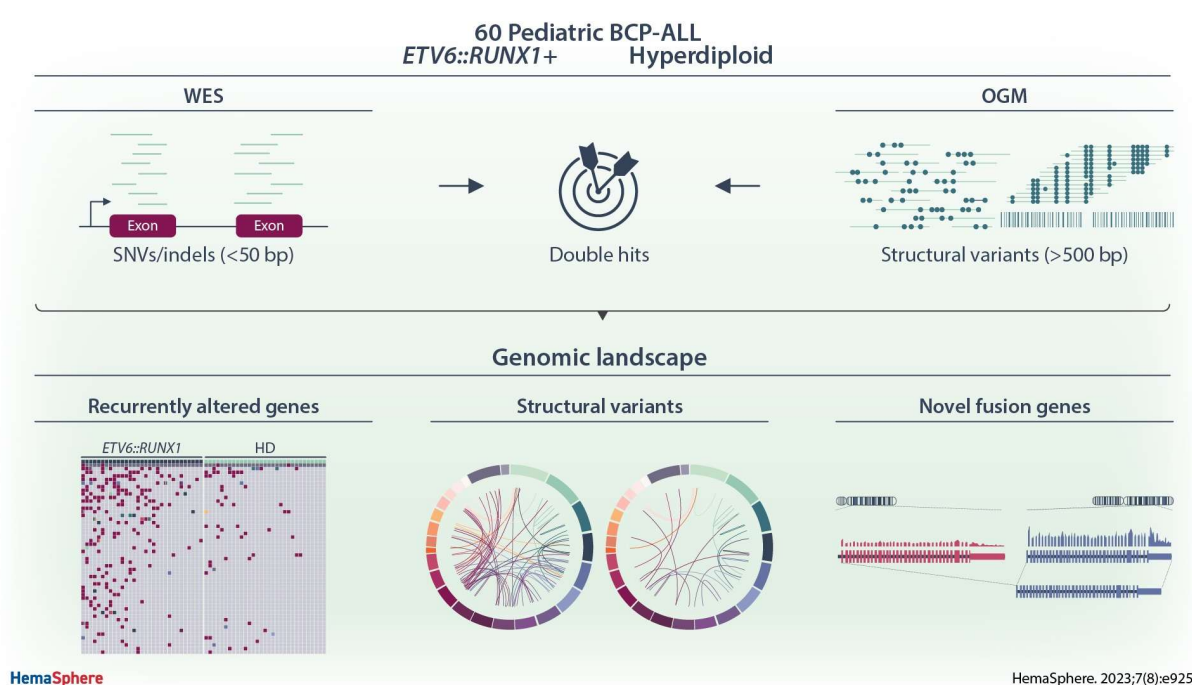
Brandes D., Yasin L., Nebral K., Ebler J., Schinnerl D., Picard D., Bergmann A.K. Alam J., Köhrer S., Haas O.A., Attarbaschi A., Marschall T., Stanulla M., Borkhardt A., Brozou T., Fischer U., Wagener R.

Optical Genome Mapping Identifies Novel Recurrent Structural Alterations in Childhood *ETV6::RUNX1*+ and High Hyperdiploid Acute Lymphoblastic Leukemia.

Hemasphere. 2023 Jul 17;7(8):e925.

PMID: 37469802; DOI: 10.1097/HS9.0000000000000925; impact factor: 14.6

Graphical abstract:



Key findings:

- OGM reliably detected structural variations that previously have been identified with conventional cytogenetic or molecular methods performed during routine diagnostics
- Paired non-tumor/tumor approach by OGM identified additional somatic structural variants that have been missed by conventional techniques
- *ETV6::RUNX1* translocated BCP-ALL carried more structural alterations compared to the hyperdiploid subtype, mainly focal deletions
- The combination of OGM and WES identified novel recurrent and subtype specific altered regions and genes in pediatric BCP-ALL
- The combination of OGM and WES detects double hits of structural variants and single nucleotide variants in the same BCP-ALL case indicating novel tumor suppressor genes

- OGM together with long-read and RNA-sequencing identifies novel fusion genes
- Long-read sequencing refines OGM detected alterations

Contribution:

I wrote and revised the manuscript. I further conceptualized the graphical abstract. I generated the molecular genetic data including DNA extraction, RNA extraction, whole-exome sequencing, optical genome mapping and long-read sequencing experiments. I further analyzed all molecular data and created data evaluation strategies. Major contribution and my own experiments are represented in Figure 1, 2, 3, 4, 5, 6, 7 & 8; Supplementary Material and Methods II&III; Supplementary Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 & 13; Supplementary Tables 2, 4, 5, 6, 7, 8, 10, 11, 12 & 13.

3.4 Original Publication II

HemaSphere



Article
Open Access

Optical Genome Mapping Identifies Novel Recurrent Structural Alterations in Childhood *ETV6::RUNX1*+ and High Hyperdiploid Acute Lymphoblastic Leukemia

Danielle Brandes^{1,2}, Laya Yasin¹, Karin Nebral^{3,4}, Jana Ebler^{5,6}, Dagmar Schinnerl⁴, Daniel Picard¹, Anke K. Bergmann⁷, Jubayer Alam¹, Stefan Köhrer^{3,4}, Oskar A. Haas⁸, Andishe Attarbaschi^{4,8}, Tobias Marschall^{5,6}, Martin Stanulla⁹, Arndt Borkhardt^{1,10}, Triantafyllia Brozou^{1,10}, Ute Fischer^{1,10}, Rabea Wagener^{1,10}

Correspondence: Arndt Borkhardt (Arndt.Borkhardt@med.uni-duesseldorf.de).

ABSTRACT

The mutational landscape of B-cell precursor acute lymphoblastic leukemia (BCP-ALL), the most common pediatric cancer, is not fully described partially because commonly applied short-read next generation sequencing has a limited ability to identify structural variations. By combining comprehensive analysis of structural variants (SVs), single-nucleotide variants (SNVs), and small insertions-deletions, new subtype-defining and therapeutic targets may be detected. We analyzed the landscape of somatic alterations in 60 pediatric patients diagnosed with the most common BCP-ALL subtypes, *ETV6::RUNX1*+ and classical hyperdiploid (HD), using conventional cytogenetics, single nucleotide polymorphism (SNP) array, whole exome sequencing (WES), and the novel optical genome mapping (OGM) technique. Ninety-five percent of SVs detected by cytogenetics and SNP-array were verified by OGM. OGM detected an additional 677 SVs not identified using the conventional methods, including (subclonal) *IKZF1* deletions. Based on OGM, *ETV6::RUNX1*+ BCP-ALL harbored 2.7 times more SVs than HD BCP-ALL, mainly focal deletions. Besides SVs in known leukemia development genes (*ETV6*, *PAX5*, *BTG1*, *CDKN2A*), we identified 19 novel recurrently altered regions (in $n \geq 3$) including 9p21.3 (*FOCAD/HACD4*), 8p11.21 (*IKBKB*), 1p34.3 (*ZMYM1*), 4q24 (*MANBA*), 8p23.1 (*MSRA*), and 10p14 (*SFMBT2*), as well as *ETV6::RUNX1*+ subtype-specific SVs (12p13.1 (*GPRC5A*), 12q24.21 (*MED13L*), 18q11.2 (*MIB1*), 20q11.22 (*NCOA6*)). We detected 3 novel fusion genes (*SFMBT2::DGKD*, *PDS5B::STAG2*, and *TDRD5::LPCAT2*), for which the sequence and expression were validated by long-read and whole transcriptome sequencing, respectively. OGM and WES identified double hits of SVs and SNVs (*ETV6*, *BTG1*, *STAG2*, *MANBA*, *TBL1XR1*, *NSD2*) in the same patient demonstrating the power of the combined approach to define the landscape of genomic alterations in BCP-ALL.

¹Pediatric Oncology, Hematology and Clinical Immunology, Medical Faculty, Heinrich-Heine University and University Hospital Düsseldorf, Germany

²Düsseldorf School of Oncology (DSO), Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany

³Labdia Labordiagnostik, Clinical Genetics, Vienna, Austria

⁴St. Anna Children's Cancer Research Institute (CCRI), Vienna, Austria

⁵Institute for Medical Biometry and Bioinformatics, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany

⁶Center for Digital Medicine, Heinrich-Heine University, Düsseldorf, Germany

⁷Institute of Human Genetics, Hannover Medical School (MHH), Hannover, Germany

⁸St. Anna Children's Hospital, Department of Pediatric Hematology/Oncology, Pediatric Clinic, Medical University, Vienna, Austria

⁹Pediatric Hematology and Oncology, Hannover Medical School (MHH), Hannover, Germany

¹⁰German Cancer Consortium (DKTK), partner site Essen/Düsseldorf, Germany TB, UF, and RW have contributed equally to this work.

Supplemental digital content is available for this article. Copyright © 2023 the Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the European Hematology Association. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

HemaSphere (2023) 7:8(e925).

<http://dx.doi.org/10.1097/HS9.0000000000000925>.

Received: January 31, 2023 / Accepted: June 1, 2023

INTRODUCTION

In the early 1960s and 1970s, a plethora of specific chromosomal abnormalities were identified in B-cell precursor acute lymphoblastic leukemia (BCP-ALL), the most common childhood cancer. Their association with clinical variables, response to chemotherapy, and ultimately outcome has deeply impacted trial design, patient stratification, and therapy allocation. Today, in most routine diagnostic laboratories, a combination of standard, low-cost, and easy-to-perform technologies such as cytogenetics as well as fluorescence in situ hybridization (FISH) and more sophisticated genome-wide analytical tools such as single nucleotide polymorphism (SNP) array analysis are used to determine the molecular genetic make-up of BCP-ALL. The 2 most frequent genetic subtypes, encompassing 50%–55% of all childhood BCP-ALL cases are *ETV6::RUNX1*-translocated (t[12;21][p13;q22]) BCP-ALL and classical hyperdiploid (HD) BCP-ALL, harboring about 51–67 chromosomes per leukemic cell and comprise heterozygous di-, tri-, and tetrasomies.¹ Both genetic entities usually have excellent therapy responses. Their good outcomes may identify them as candidates for therapy reduction²; however, ≈20% of patients experience a relapse, partly with unknown reason, in a subset of initially diagnosed standard-risk

patients.³ It is not yet known whether additional secondary genomic alterations in *ETV6::RUNX1*+ and HD BCP-ALL, particularly the frequently overlooked large-scale structural variations, may help to characterize and subdefine these entities and their outcome even further. For instance, studies by Enshaie et al⁴ and others highlighted that specific copy number (CN) patterns of HD leukemia may influence prognosis and enable further sub-classification in low and poor-risk groups.¹

According to the 2-step model of childhood leukemogenesis, both lesions are acquired in utero during fetal hematopoiesis, and postnatally acquired secondary genetic hits are required for the development of overt disease.⁵ Thus, a distinct constellation of secondary structural variants (SVs) and single-nucleotide variants (SNVs) together with the entity-defining key lesion likely drive the leukemic process.

Most of the mutational landscape of BCP-ALL has been analyzed by conventional cytogenetic methods, SNP or array comparative genomic hybridization and whole genome sequencing (WGS). However, it should be noted that large SVs are particularly difficult to detect by short-read DNA-sequencing methods.^{6,7} This can be attributed to the fact that SVs are likely enriched within or near repetitive DNA regions.⁸ Repetitive DNA is known to be difficult to analyze by short-read sequencing technologies because the resulting reads map ambiguously to multiple locations.^{9,10} The recently introduced genome-wide optical mapping technology analyzes long linear DNA molecules (>150 kb) and covers more repetitive and complicated regions, which allows for the detection of SVs, including inversions, insertions, deletions and translocations, with a size cut-off of 500 bp in length for de novo assembled genomes. The applicability of this approach has recently been shown for various hematological malignancies, and some researchers even advocate for further replacing the existing routine diagnostics methods of those neoplasms with optical genome mapping (OGM).^{11–17}

Here, we present a study focusing on 60 HD and *ETV6::RUNX1*+ BCP-ALL cases analyzed by whole exome sequencing (WES) and OGM to decipher the landscape of genomic alterations in both entities, with a special focus on their SVs. We report several novel recurrently altered genes, which are not described in previous studies that applied SNP-arrays or WGS.^{3,18–20} Moreover, we demonstrate that the overall genomic landscape of secondary alterations in these 2 types of BCP-ALL is profoundly different.

MATERIALS AND METHODS

Patient cohort

This study includes 60 pediatric patients diagnosed with either *ETV6::RUNX1*+ or classical HD BCP-ALL in 3 different clinical institutions (Suppl. Tables S1). Patients were enrolled in this study after informed consent was obtained following institutional guidelines in accordance with the declaration of Helsinki. We collected fresh-frozen bone marrow mononuclear cells (BMMNCs) or peripheral blood mononuclear cells (PBMCs) of all leukemia samples from the time point of initial diagnosis, as well as fibroblasts (n = 16 cases) or fresh-frozen BMMNCs/PBMCs from remission (n = 44 cases) as matched non-tumor controls. Four of the herein analyzed patients developed a relapse. From 1 *ETV6::RUNX1*+ patient, we were able to analyze initial tumor, matched non-tumor, and the relapsed leukemia.

Cytogenetics and molecular genetics

Karyotypes were determined by chromosomal G-banding. The results of the analysis for *ETV6::RUNX1* translocation by FISH of the leukemia samples obtained during routine diagnostics were compiled from the patient records for this study (Suppl. Tables S2 and S3). To detect CN alterations and CN neutral loss of heterozygosity (CNN-LOH), all 60 leukemia

samples were analyzed using SNP-arrays. For 18 leukemia cases, the CytoScan HD array (Thermo Fisher Scientific, Waltham, MA) was used as described.²¹ The data were initially aligned to GRCh37 and a lift-over to GRCh38 was performed for downstream analysis. The remaining 42 leukemia cases were analyzed using the Illumina CytoSNP-12 v2.1 array (Illumina, San Diego, CA) according to the manufacturer's protocol. The array was scanned using the scanner option on the NextSeq550 sequencing platform (Illumina). The raw data were processed and analyzed using the Beeline 2.0.3.3 and the BlueFuse Multi 4.5 software from Illumina. The human reference genome GRCh38 was used. We included SVs encompassing ≥20 probes, ≥50 kb for deletions and duplications, or ≥5 Mb for CNN-LOH. For all samples, B- and T-cell receptor rearrangements were excluded from downstream analysis and SVs overlapping with centromere or reference gap regions. The filtered molecular genetic results are summarized in Suppl. Table S4.

Whole exome sequencing

WES data was collected from 58 tumor samples and the corresponding non-leukemia, which allows for the detection of somatically acquired alterations. Unfortunately, no sufficient material was available for 2 leukemia samples. WES was performed using the SureSelect Human All Exon V5+UTR kit, All Exon V7 or All Exon V8 (Agilent). The library was paired-end sequenced on a NextSeq550 (2×150 bp) sequencer. An average on-target coverage of ≥70× for germline and ≥100× for tumor samples (Suppl. Figure S1) was generated. Refer to Suppl. Materials and Methods I for a detailed description of the bioinformatic processing of WES data. Data of filtered variants are summarized in Suppl. Table S5.

Optical genome mapping

Extraction of high-molecular-weight DNA, labeling, and chip loading

We extracted high-molecular-weight (HMW) DNA from PBMC or BMMNC at initial diagnosis using the SP Blood and Cell Culture DNA Isolation Kit, according to the manufacturer's guidelines (Bionano Genomics, San Diego, CA). In addition, HMW DNA was extracted from remission material or fibroblasts (if available) of the respective children as matching non-tumor control. Labeling of HMW DNA was performed using the DLS DNA Labeling Kit (Bionano Genomics), according to the manufacturer's protocol (Bionano Prep Direct Label and Stain [DLS] Protocol 30206). The labeled DNA was loaded on a Saphyr G2.3 chip and molecules were imaged using the Saphyr instrument (Gen 2, Bionano Genomics). Overall, we performed OGM on 60 matched tumor/non-tumor pairs (120 samples), and 1 sequential sample including tumor material from initial diagnosis and relapse in comparison to matched non-tumor material (ALL4). Refer to Suppl. Figure S2 for an overview of the applied OGM workflow.

Data collection and analysis

We collected an average of 124× effective coverage for all tested non-tumor samples (420 GB/sample) and 384× for tumor samples (1300 GB/sample) (Suppl. Figure S1). The Bionano Access Server ([BAS], Bionano Genomics, San Diego, CA) was used for data analysis and visualization. The human reference genome GRCh38 was used for alignment. SVs were called using the de novo (DN) assembly (BAS 1.5.2) and rare variant pipeline (RVP) (BAS 1.7) for each non-tumor and tumor sample. Before the OGM data of the leukemia samples could be processed using the DN pipeline, the molecule files had to be downsampled to 480 GB with a minimal molecule length of 150 kb. To differentiate between somatic and germline SVs, we performed dual annotation for the DN and RVP data. In the ALL4 case, we used the trio annotation module, allowing for the direct comparison of the data from initial diagnosis and

relapse as well as filtering for variants detected in the matching non-leukemia controls.

Data Filtering and Interpretation

We excluded all SVs that were present in the matching non-leukemia samples, which likely represent germline SVs. SVs were filtered as described in Suppl. Material and Methods II. The curated lists of SVs, including deletions, duplications, insertions, inversions, inter and intrachromosomal translocations, detected by both DN and RVP pipelines were merged. The applied software calls large inversions and deletions as intra-chromosomal translocations. Both DN and RVP pipelines contain an additional CN algorithm tool that detects CN alterations, mainly deletions and duplications ≥ 5 Mb including aneuploidy. Accordingly, the final dataset included the filtered SVs detected by the DN and RVP algorithms and the SVs ≥ 5 Mb detected with the CN tool.

For downstream analysis, we differentiated between SVs ≥ 5 Mb and SVs ≤ 5 Mb. The latter included inversions and translocations for each case, as well as deletions and duplications ≤ 5 Mb to delineate focal SVs and detect recurrently altered regions/genes as previously described.³ Moreover, we incorporated the detected CNN-LOH identified by SNP-array. Refer to Suppl. Table S6-S8 and Suppl. Figures S3 and S4 for the OGM results.

Validation of selected SVs by NGS

Long-read sequencing

PacBio (Pacific Biosciences, Menlo Park, CA) long-read sequencing was performed on 11 tumor samples where sufficient material was available to validate selected SVs that were newly detected by OGM. HMW DNA (5 μ g) was used for the continuous long-read (CLR) workflow (>25 kb size cut-off), according to the manufacturers' instructions. Each sample was submitted to sequencing on a PacBio Sequel II (performed by Genomics and Transcriptomics Laboratory, BMFZ, Dusseldorf, Germany) using 1 SMRT cell. We generated an average of 31 $\times$ coverage (~ 231 GB/sample). Raw bam files were sorted and aligned to the reference genome GRCh38 using pbmm2 (v1.3.0; <https://github.com/PacificBiosciences/pbmm2>). Sorted bam files were visualized using the Integrative Genomics Viewer (IGV, <https://www.broadinstitute.org/igv>, version 2.11.9). Refer to Suppl. Materials and Methods III for detailed analysis of the CLR data.

RNA sequencing

To validate newly detected gene fusions, data of RNA sequencing (RNA-seq) were collected from 6 tumor samples. Data of 2 samples were generated as described before.²² Total RNA of 4 cases was isolated from fresh-frozen cells and submitted for RNA-seq at the Genomics and Proteomics Core Facility (DKFZ, Heidelberg, Germany) or Genomics and Transcriptomics Laboratory (BMFZ, Dusseldorf, Germany). For detailed description of RNA-seq workflow refer to Suppl. Material and Methods IV. Fusions were detected and visualized using Arriba v2.4.0 algorithm.²³ Arriba was run with default settings against the GRCh38 reference genome with the GENECODE annotation.

RESULTS

OGM detects 1.6- to 3-fold more deletions than conventional SNP-arrays in a cohort of 60 primary pediatric BCP-ALL patients

We studied somatically acquired SNVs, insertions-deletions (indels), and SVs in 60 pediatric patients with *ETV6::RUNX1*+ and classical HD BCP-ALL (each $n = 30$) (Suppl. Table S9). The cohort consisted of 32 male (53%) and 28 (47%) female newly diagnosed patients with a median age of 3.8 years (range, 1.6–17 years). The majority of patients (87%; 52/60) were stratified as standard or medium risk, while 13% ($n =$

8/60) were classified as high risk. In total, 93.3% (56/60) of the patients remained in complete remission, whereas 6.7% (4/60) relapsed. By analyzing tumor and matched non-tumor samples of these leukemia patients using OGM and WES, we were able to detect somatically acquired alterations potentially associated to the leukemogenesis. A comprehensive set of genomic data was generated by adding diagnostically reported data from conventional cytogenetics (karyotyping and FISH) and SNP-array data (Figure 1).

Performing OGM, we yielded an average coverage of 384 $\times$ in the BCP-ALL samples and 124 $\times$ in the non-leukemia samples (Suppl. Figure S1). By applying 2 SV calling pipelines (DN and RVP) together with an integrated CN pipeline, we detected a total of 1203 somatic SVs by OGM. First, we analyzed the concordance rate of SVs detected by conventional genetic techniques and OGM. OGM reliably identified 95% (526/552) of SVs identified by karyotyping, FISH, and/or SNP-array, including all hallmark translocations *t*(12;21)(p13;q22) and chromosomal gains in HD BCP-ALL (Figure 2A). Notably, OGM resolved 3-way translocations involving *t*(12;21)(p13;q22) present in 9 of 30 *ETV6::RUNX1* BCP-ALL cases. All aberrations detected by FISH and 96% (164/171) of SVs identified by karyotyping were observed by OGM. Of note, the translocation *der*(14)*t*(10;14)(p11;p11) with breakpoints in centromeric regions was not detectable by OGM. In total, 96% (486/508) of SVs identified by SNP-array were detectable by OGM (Figure 2B). Alterations missed by OGM, even after manual reevaluation, included, among others, a gain of chromosome Y, 2 focal mosaic deletions ($\sim 30\%$ of cells) in chromosome 13q and 1 mosaic gain in 5p15.33p15.1 ($\sim 15\%$ of cells) (Suppl. Table S10).

Next, we compared the lengths of the deletions and duplications detected by OGM and SNP-array. As we used 2 different types of SNP-arrays that have different probe coverages (CytoSNP-12 array $\approx 300,000$ versus CytoScanHD $\approx 2,670,000$ SNPs targeting probes) and detection limits with respect to alteration length, we performed 2 separate analyses for each array (Figure 2C and 2D). Considering the CytoSNP-12 array's detection limit of 50 kb for the OGM data, the absolute numbers of deletions detected by OGM exceeded the number identified by CytoSNP-12 array by 3-fold (245 versus 81). By contrast, considering a detection limit of 1 kb using CytoScanHD, the number of deletions detected by OGM was only 1.6 times higher compared with the array (197 versus 122 deletions, respectively). Interestingly, the number of somatic duplications detected by OGM and both array types was not notably different (OGM $n = 28$ versus CytoSNP-12 $n = 20$ with 50 kb detection limit; OGM $n = 25$ versus CytoScanHD $n = 33$ with 1 kb detection limit), likely due to the overall rareness of these events in our data.

Compared with the above-described conventional genetic methods, OGM detected an additional 677 SVs, including complex 3-way translocations (Figure 2A). About 66.3% of those SVs consisted of deletions (449/677) with a median size of 40.5 kb (range, 1 kb to 3.6 Mb). In total, 55.2% (248/449) of the deletions detected by OGM were <50 kb, which is the lower detection limit of the CytoSNP-12 array. The remaining additional SVs found by OGM were insertions ($n = 89$), inter and intrachromosomal translocations ($n = 76$ and $n = 33$, respectively), inversions ($n = 17$), and duplications ($n = 13$); 37.2% (252/677) of those SVs were detected by both SV detection algorithms, whereas 53.5% (362/677) were solely identified by the RVP pipeline.

ETV6::RUNX1+ BCP-ALL harbor 2.7-fold more SVs, mainly deletions, compared with HD BCP-ALL as determined by OGM

We compared the frequencies of SVs detected by OGM between the 2 genetic subgroups of BCP-ALL (*ETV6::RUNX1*+ and HD leukemia, $n = 30$ each) that occurred in addition to their hallmark events. We identified 2.7 times more somatic SVs (focal

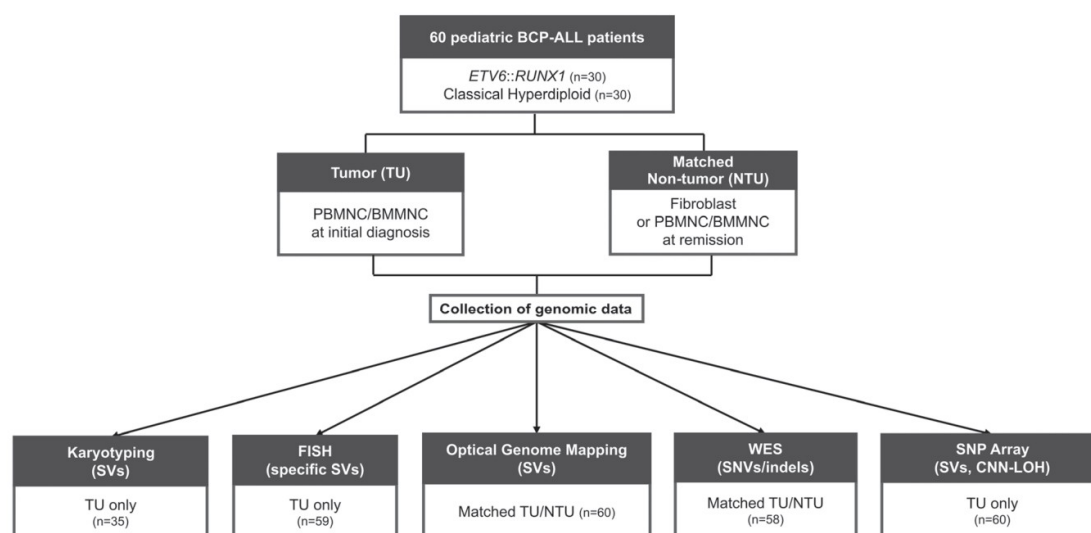


Figure 1. Overview of genomic data collection from 60 pediatric BCP-ALL patients. Tumor and matched non-tumor samples from 30 *ETV6::RUNX1* and 30 classical HD BCP-ALL patients were analyzed. A comprehensive data set was collected to detect SVs, CNN-LOHs, and SNVs/indels using karyotyping, FISH, OGM, WES, and SNP-array. BCP-ALL = B-cell precursor acute lymphoblastic leukemia; BMMNC = bone marrow mononuclear cells; CNN-LOHs = copy number neutral loss of heterozygosity; FISH = fluorescence in situ hybridization; HD = hyperdiploid; OGM = optical genome mapping; PBMNC = peripheral blood mononuclear cells; SNP = single nucleotide polymorphism; SV = structural variants; WES = whole exome sequencing.

deletions and duplications ≤ 5 Mb, insertions, inversions, inter and intrachromosomal translocations) in *ETV6::RUNX1*+ compared with HD BCP-ALL ($n = 625$ versus $n = 219$ SVs, respectively) (Figure 3A and 3B). SVs ≥ 5 Mb were twice as frequent in *ETV6::RUNX1*+ leukemia ($n = 47$ versus $n = 27$) (Figure 3A and 3B; Suppl. Table S6). In both subtypes, 68% (577/844) of all SVs ≤ 5 Mb, the most prominent type of SV in both subtypes, were focal deletions. Focal deletions were 3.4-fold more frequent in *ETV6::RUNX1*+ BCP-ALL compared with HD BCP-ALL ($n = 446$ versus $n = 131$ focal deletions). The burden of deletions in *ETV6::RUNX1* BCP-ALL was 13.5 deletions (median; range, 1–46) per tumor compared with 4 deletions (median; range, 1–14) in HD cases (Figure 3C; Suppl. Table S7).

We observed secondary interchromosomal translocations in 20 *ETV6::RUNX1* BCP-ALL cases that occurred in addition to their hallmark translocation. By contrast, secondary interchromosomal translocations were rarely detected in HD BCP-ALL cases (7 translocations in 5 cases).

We further compared patterns of aneuploidy events and SVs ≥ 5 Mb including chromosomal p/q alterations (detected by OGM), as well as CNN-LOH (observed by SNP-array) between both genetic subgroups. In terms of aneuploidy events, we detected 7 *ETV6::RUNX1*+ cases harboring a trisomy 21, 1 case with trisomy 16, and 1 female case with monosomy X (Figure 4A). Whole chromosomal CNN-LOH were detected in only 1 *ETV6::RUNX1*+ case on chrX (Figure 4A). In classical HD BCP-ALL, all cases showed typical aneuploidy patterns including tetrasomy 21 (29/30); except 1 patient with trisomy 21; trisomy 14 (24/30) or tetrasomy 14 (6/30) (Figure 4A). Whole chromosomal CNN-LOH were identified in chr2 and chr9 in 2 HD BCP-ALL and in chr1 and chrX in 2 different HD BCP-ALL (Figure 4A). Of note, to define additional potential risk factors in HD BCP-ALL as described in Enshaie et al,⁴ we analyzed OGM defined aneuploidy patterns. Specifically, HD BCP-leukemia without trisomy 17 and 18 or harboring either trisomy 17 or 18 with trisomy 5 or 20 are associated with poor risk. We detected 6 poor-risk HD leukemia (ALL31, ALL41, ALL49, ALL51, ALL52, and ALL55), based on these criteria.

Large-scale SVs ≥ 5 Mb (not affecting the whole chromosome) detected by OGM encompassed deletions and duplications (Figure 4B). Large-scale deletions were more frequent in *ETV6::RUNX1*+ than in HD BCP-ALL ($n = 31$ versus $n = 8$, respectively), including the secondary large-scale deletion (mean of 17 Mb) in chr12p that led to the loss of the second *ETV6* allele. By contrast, only 1 deletion on 1q21.1q21.2 was found to be recurrently affected in 2 of 30 HD BCP-ALL cases. Large duplications were more frequent in HD than in *ETV6::RUNX1*+ BCP-ALL ($n = 20$ versus $n = 13$, respectively). Recurrently affected regions were identified on chromosome 1q21.1q42.13 ($n = 4/30$, HD BCP-ALL) and 17p11.2q25.3 ($n = 4/30$; HD BCP-ALL). In both BCP-ALL entities, CNN-LOH on chromosome arms 9p ($n = 3$) and 9q ($n = 3$) was detected.

Moreover, we were able to analyze sequential tumor samples in 1 *ETV6::RUNX1* BCP-ALL case (ALL4) who relapsed 48 months after initial diagnosis. Comparing the diagnostic and the relapse samples in terms of common and acquired SVs detected by OGM, we observed 34 shared SVs, 34 acquired SVs, and 11 SVs that were lost in the relapse sample (Suppl. Figure S5). Acquired SVs included an *inv(X)(p11.3;q13.2)* potentially leading to a *KDM6A::XIST* gene fusion (Suppl. Figure S5B; Suppl. Table S8).

The combination of OGM and WES revealed novel commonly recurrent as well as subtype-specific altered regions and genes in *ETV6::RUNX1*+ and HD BCP-ALL

Our next aim was to identify recurrently altered regions and genes affected by either SVs ≤ 5 Mb and/or SNVs/indels, combining both OGM and WES data sets. We determined the minimal altered regions (MARs; not differentiating between the SV types) and the therein located potential target genes detected in ≥ 3 BCP-ALL cases by OGM. Next, we checked whether target genes located within these MARs were affected by SNVs/indels in the BCP-ALL cohort, that is allowing for the detection of cases with double hits (SVs and SNV/indel) or single hits harboring either an SV or SNV/indel in a target gene (Figure 5).

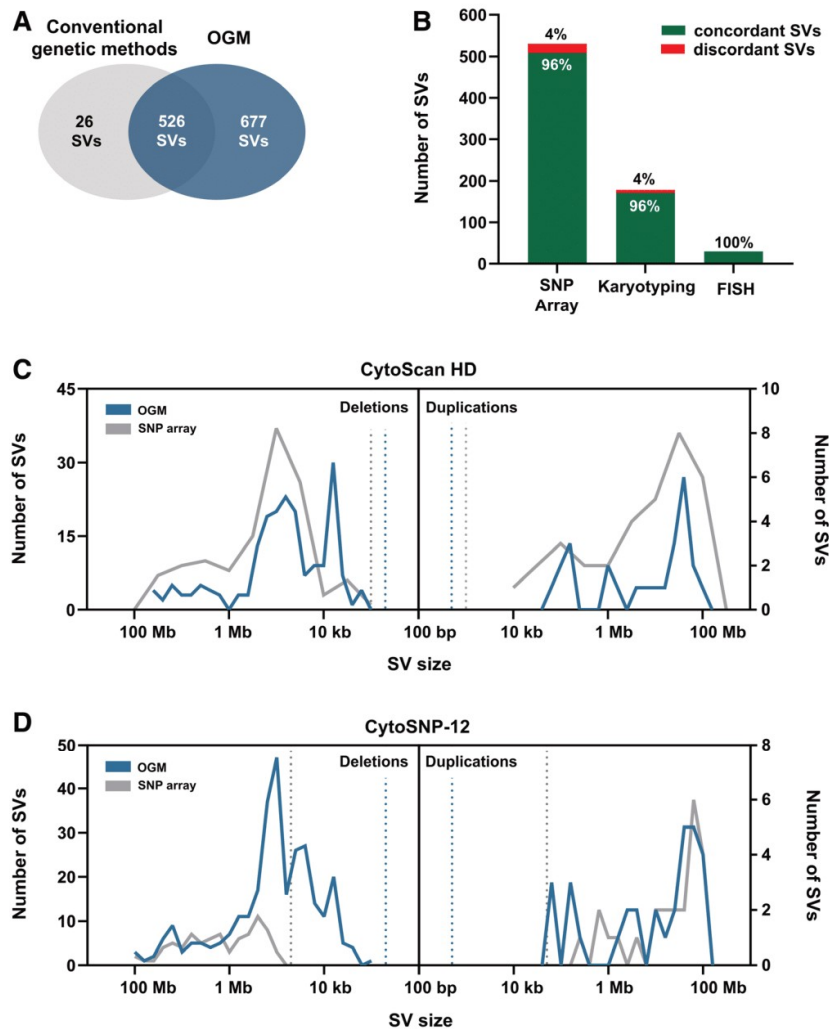


Figure 2. Comparison of SVs detected by conventional genetic techniques and OGM in 60 primary BCP-ALL cases. (A) OGM reliably detects 95% (526/552) of SVs detected by conventional genetic methods (SNP-array, karyotyping, and FISH), including all hallmark aberrations such as the *ETV6::RUNX1* translocation and hyperdiploidy. OGM detected 677 additional SVs compared with standard methods. (B) Frequencies of concordant and discordant SVs detected by karyotyping, SNP-array, and FISH compared with OGM. (C) Size distribution of deletions and duplications detected by OGM (green) and CytoScanHD (SNP-array, gray) in 18 cases. Dashed lines indicate the detection limit of 1 kbp (SNP-array, gray) and 500 bp (OGM, green). (D) Size distribution of deletions and duplications detected by OGM (green) and CytoSNP-12 (SNP-array, gray) in 42 cases. Dashed lines indicate the detection limit of 50 kbp (SNP-array, gray) and 500 bp (OGM, green). BCP-ALL = B-cell precursor acute lymphoblastic leukemia; FISH = fluorescence in situ hybridization; HD = hyperdiploid; OGM = optical genome mapping; SNP = single nucleotide polymorphism; SV = structural variants.

Overall, we detected 52 MARs, including 40 MARs harboring a potential target gene and 12 MARs, which were located in intergenic regions. We identified more *ETV6::RUNX1* cases with MARs compared with classical HD BCP-ALL cases ($n = 30$ versus $n = 24$, respectively). The most frequently detected MARs and therein located potential target genes, which were observed in $\geq 10\%$ of both BCP-ALL subtypes, were 12p13.2/*ETV6* (27%; 16/60), 9p21.3/*CDKN2A* (23%; 14/60), 9p13.2/*PAX5* (22%; 13/60), 4p16.3/*NSD2* (17%; 10/60), 11p12/*RAG2/IFTAP* (17%; 10/60), 20p12.2/*MKKS/SLX4IP* (17%; 10/60), 3q13.2/*C-D200/BTLA* (15%; 9/60), 9p13.2/*ZCCHC7* (13%; 8/60), and Xq25/*STAG2* (12%; 7/60). Less frequently detected, but recurrently altered regions/genes ($n=2$; 3%) in both *ETV6::RUNX1*+

and HD BCP-ALL included focal deletions affecting, for example 7p12.2/*IKZF1* and 3q25.1/*MBNL1* (Suppl. Figure S6).

We identified 11 novel, not yet reported to the best of our knowledge, MARs affecting 8 potential target genes that are commonly affected in both entities (Suppl. Table S11). Hence, they might be involved in general BCP-ALL leukemogenesis. These MARs and potential target genes included 9p21.3/*FOCAD/HACD4* (8%; 5/60), 8p11.21/*IKBKB* (7%; 4/60), 1p34.3/*ZMYM1/6* (5%; 3/60), 4q24/*MANBA* (5%; 3/60), 8p23.1/*MSRA* (5%; 3/60), and 10p14/*SFMBT2* (5%; 3/60). We validated selected SVs in tumors where sufficient material was available for long-read sequencing and with the available SNP-array data (Suppl. Table S11; Suppl. Figure S7).

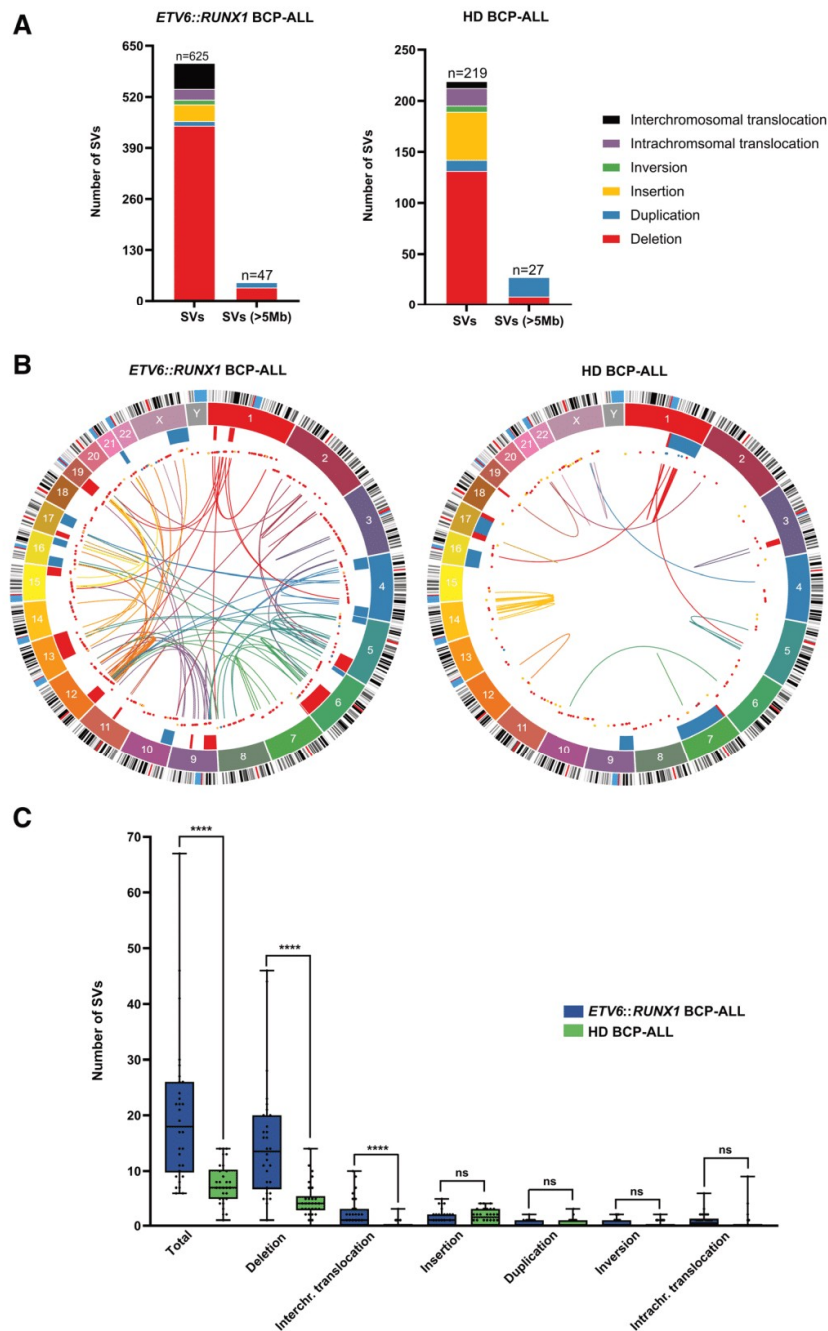


Figure 3. Frequencies of SVs detected in 60 primary BCP-ALL cases by OGM in addition to their hallmark alterations. The defining *ETV6::RUNX1* translocation and numerical whole chromosomal alterations are not included. (A) Total numbers of SVs and SVs ≥ 5 Mb in 60 primary BCP-ALLs. (B) Circos plots showing from the outer to the inner circle: the chromosomal ideogram; the chromosome number; chromosomal regions of deletions (red) and duplications (blue) ≥ 5 Mb, focal deletions (red), duplications (blue) and insertions (yellow); and intra and interchromosomal translocations and inversions of *ETV6::RUNX1* (n = 30, left circos plot) and HD (n = 30, right circos plot) BCP-ALL. (C) Box plots depicting the number of SVs per case for all SVs (total) and for each SV type in *ETV6::RUNX1* (blue) and HD (green) BCP-ALL subtypes. Box and whiskers indicate median, 25 of and 75% quartiles and $\pm$ range. Statistical significance was assessed between the 2 BCP-ALL subtypes using the Mann-Whitney U Test. **** $P < 0.0001$. BCP-ALL = B-cell precursor acute lymphoblastic leukemia; HD = hyperdiploid; ns = not significant; OGM = optical genome mapping; SV = structural variants.

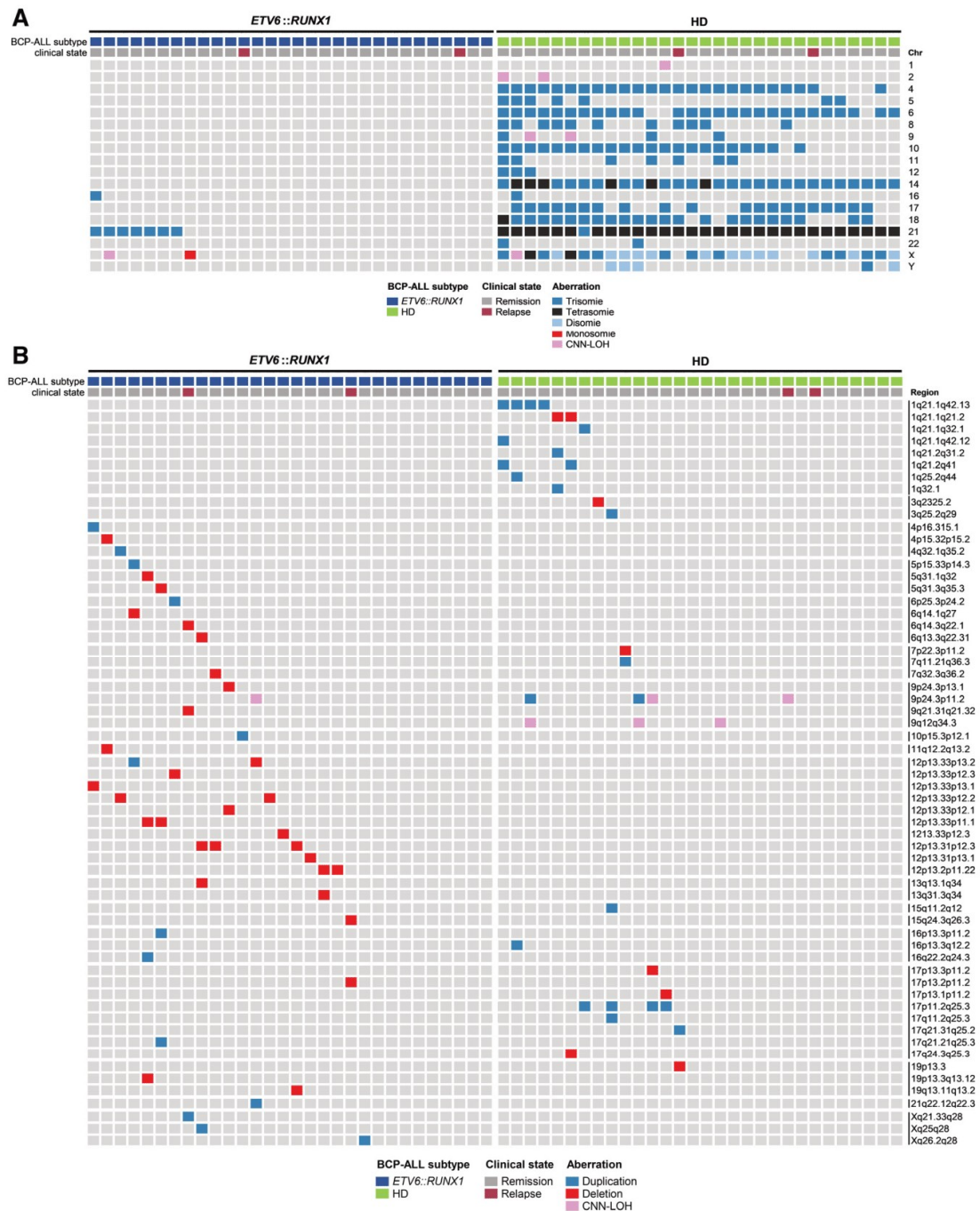


Figure 4. Landscape of aneuploidy and SVs ≥ 5 Mb in *ETV6::RUNX1*+ and HD BCP-ALL identified by OGM and SNP-array. Depicted are (A) aneuploidies and whole chromosomal CNV-LOHs and (B) SVs ≥ 5 Mb and CNV-LOHs not affecting the whole chromosome, identified *ETV6::RUNX1*+ and HD BCP-ALL cases. Each column represents a case and each line an altered region. The color code indicates the type of alteration detected. BCP-ALL = B-cell precursor acute lymphoblastic leukemia; CNV-LOHs = copy number neutral loss of heterozygosity; HD = hyperdiploid; OGM = optical genome mapping; SNP = single nucleotide polymorphism; SV = structural variants.

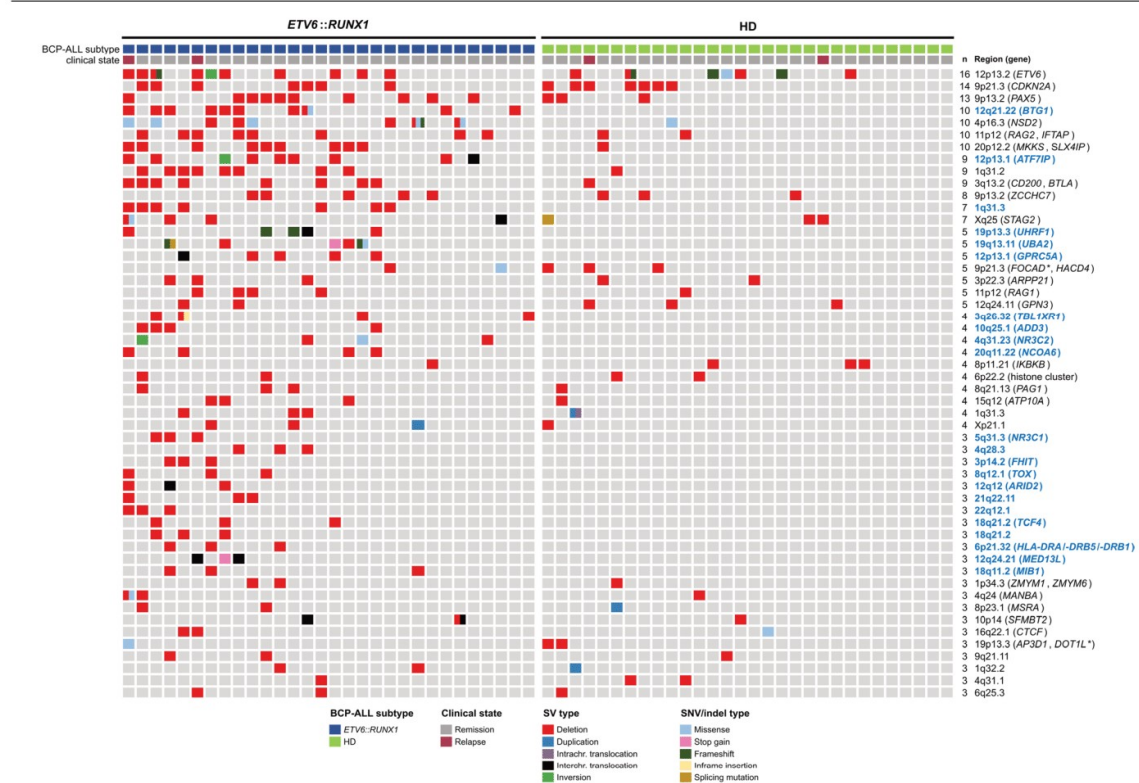


Figure 5. Combination of OGM and short-read whole exome sequencing reveals recurrently altered regions and genes in 60 primary BCP-ALL cases. The oncoprint depicting the combined dataset of MARs by SVs identified by OGM and SNVs/indels affecting the potential target gene, which are recurrently affected in ≥ 3 BCP-ALL (5%). Every column indicates a case and each line a recurrently altered region. The color code indicates the type of SV and SNV/indel alteration. The number of cases (n) with the specific region of the MARs is given on the right. Subtype-specific alterations are highlighted in blue for *ETV6::RUNX1*+ BCP-ALL. If 2 or more potential target genes are located in the MAR, an asterisk (*) indicates the gene affected by SNV/indel. BCP-ALL = B-cell precursor acute lymphoblastic leukemia; HD = hyperdiploid; MARs = minimal altered regions; OGM = optical genome mapping; SV = structural variant.

Moreover, 22 *ETV6::RUNX1*+ subtype-specific MARs (22/54) have been identified, of which 17 could be associated with potential target genes. Six of those genes were also affected by 1 or several SNVs/indels. Four of them have previously been described to be recurrently affected in this subtype: *BTG1* (33%; 10/30), *ATF7IP* (30%; 9/30), *UHRF1* (17%; 5/30), and *UBA2* (17%; 5/30).^{3,19,20} Applying long-read sequencing, we validated a deletion in ALL4 (chr19:4,804,231-4,918,731 bp) by which the promoter region of *UHRF1* was removed (Suppl. Figure S7D) and a deletion in ALL7 (chr19:34,458,211-34,460,111 bp), which removed exon 13 of *UBA2* (Suppl. Figure S7H). Strikingly, we herein identified 4 novel, that is, not yet reported to the best of our knowledge, potential target genes in the MARs that are specifically and recurrently altered in the *ETV6::RUNX1*+ BCP-ALL subgroup by SV and/or SNV (Suppl. Table S11). These included 12p13.1/*GPRC5A* (17%; 5/30), 20q11.22/*NCOA6* (13%; 4/30), 12q24.21/*MED13L* (10%; 3/30), and 18q11.2/*MIB1* (10%; 3/30). We validated selected SVs affecting a novel target gene in tumors where sufficient material was available by long-read sequencing (Suppl. Table S11; Figure 6; Suppl. Figure S7). The MAR on 18q11.2 was defined by 3 deletions detected by OGM, which affected the promoter region of *MIB1* and the first coding exon of the gene (Figure 6A and 6B). The analysis of the *MIB1* region in the existing long-read sequencing data for this case showed that the promoter region was deletion (chr18:21,725,899-21,756,226 bp) in ALL2 (Figure 6C).

Interestingly, we also detected a somatically acquired translocation t(17;18)(q12;q11.2) in the relapsed leukemia of case ALL4, with one of the breakpoints being located in the MAR of the *MIB1* gene locus (Figure 6A; Suppl. Table S8).

By combining OGM and WES datasets, we were able to identify double hits in genes targeted by SVs and SNVs/indels. The combination of deletions and truncating SNVs/indels is of particular interest, as this might indicate the presence of a tumor-suppressor gene. Mining our combined dataset, we identified 6 genes that carried such double hits. The *ETV6* gene was affected by secondary deletions and frameshift mutations in 1 *ETV6::RUNX1*+ and 1 HD BCP-ALL case. Likewise, *NSD2* was affected by a deletion and a frameshift mutation in 1 *ETV6::RUNX1* BCP-ALL case while another carried a deletion as well as frameshift and missense mutations. Further double hits of deletions and concomitant SNVs were observed in *BTG1*, *STAG2*, *MANBA*, and *TBL1XR1* in 5 *ETV6::RUNX1*+ BCP-ALL cases.

Mutational landscape of *ETV6::RUNX1*+ and classical HD BCP-ALL as assessed by WES

We analyzed the mutational landscape of the 2 BCP-ALL subgroups, considering only the SNVs/indels as detected by short-read WES. In *ETV6::RUNX1*+ BCP-ALL, we identified a median of 18 SNVs/indels (range, 2–198) per case compared with 13.5 SNVs/indels (range, 5–35) per HD BCP-ALL case (ns, $P = 0.057$, Mann-Whitney test) (Suppl. Figure S8).

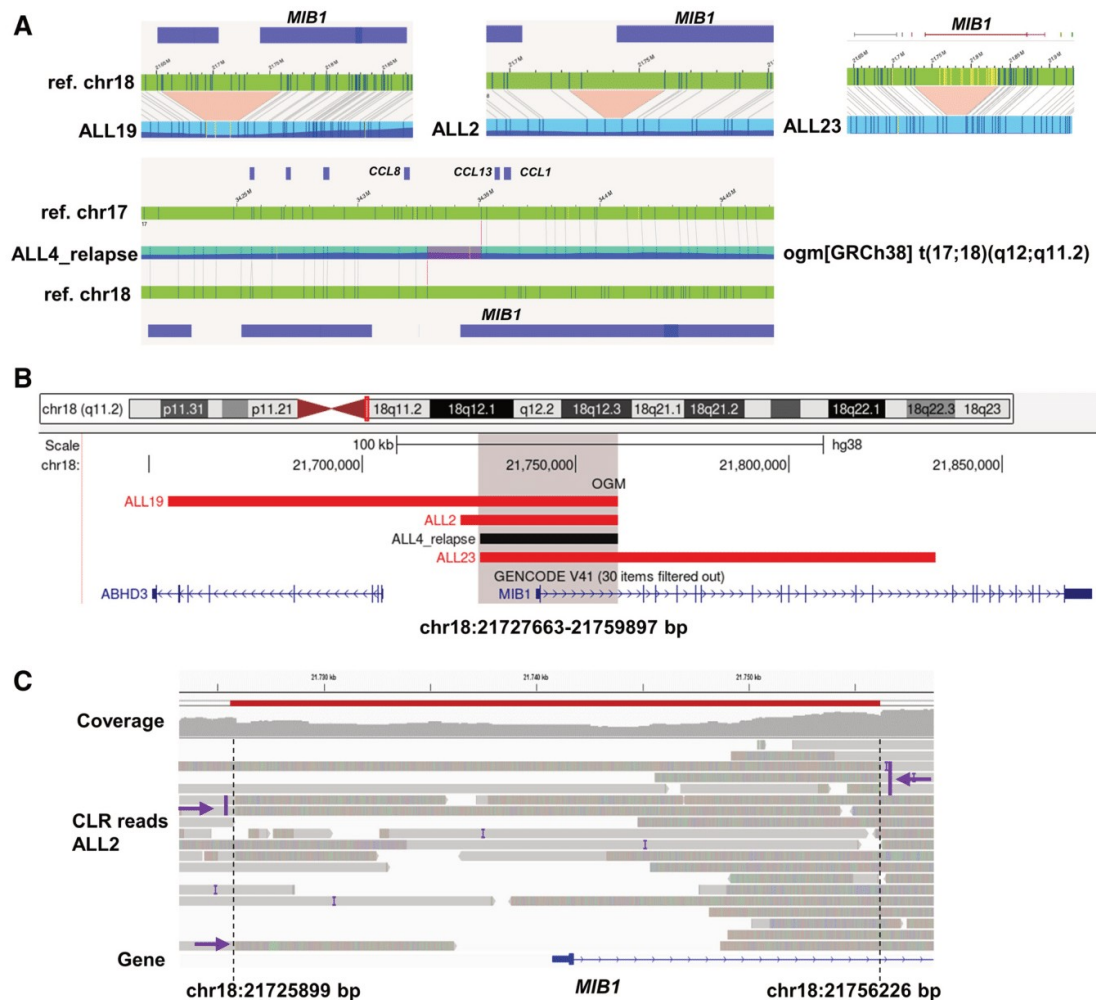


Figure 6. Examples of herein novel recurrently altered region affecting *MIB1* as potential target gene in *ETV6::RUNX1* BCP-ALL detected by OGM. (A) Optical maps (blue) of 4 *ETV6::RUNX1* BCP-ALLs (ALL19, ALL2, ALL23, ALL4_relapse) aligning to reference map of chromosome 18 (green) are shown. Top: Red highlights indicate deleted regions overlapping with the *MIB1* gene locus in 3 BCP-ALL cases. Bottom: *t*(17;18)(q12;q11.2) occurring in a relapsed leukemia with potential breakpoint in *MIB1* (indicated by pink line) is shown. (B) Detailed chromosomal locations of the SVs affecting the *MIB1* gene locus of four BCP-ALL are shown in the UCSC genome browser. Red bars indicate potentially deleted regions. Black bar shows region of the potential translocation breakpoint. Minimal altered region on 18q11.2 (chr18:21,727,663–21,759,897 bp) is highlighted in gray and overlaps with promoter region of *MIB1*. (C) CLR sequencing of ALL2 validate the deletion (chr18:21,725,899–21,756,226 bp) of the *MIB1* promoter region. CLR reads of the *MIB1* locus are shown (gray bars). Dashed lines indicate breakpoints of the deletion. Split reads supporting the deletion are marked by purple arrows. Decreased coverage of aligned reads indicates a deletion (red bar). BCP-ALL = B-cell precursor acute lymphoblastic leukemia; CLR = continuous long-read; OGM = optical genome mapping; SV = structural variants.

For the most frequently mutated genes ($n \geq 3$), we classified the recurrent mutations with regard to probable oncogenicity (Suppl. Table S12; Suppl. Materials and Methods I). The most frequently mutated genes in both subtypes were *NRAS* (24%; 14/58), *NSD2* (10%; 6/58), and *ETV6* (9%; 5/58) (Suppl. Figure S9). The herein detected *NRAS* mutations affected the well-known hotspot codons located in the GTPase domain. As previously reported,^{3,24} these mutations were more prevalent in HD than in *ETV6::RUNX1*+ BCP-ALL cases (40% versus 7%) (Suppl. Figure S9; Suppl. Table S12).

Moreover, we identified recurrent *ETV6::RUNX1*+ subtype-specific SNVs/indels in *CCDC168* (11%; 3/28) and *UBA2* (11%; 3/28), whereas in the HD BCP-ALL group, subtype-specific recurrent somatic SNVs/indels were observed in *KRAS* (27%; 8/30), *CREBBP* (20%; 6/30), *FLT3* (17%; 5/30), and *PTPN11* (10%; 3/30).

Detection of poor prognostic markers including *TP53* and *IKZF1* deletions by OGM

We examined, whether there would be an additional benefit of applying OGM alongside standard techniques in order to detect prognostic genomic alterations in BCP-ALL. Accordingly, we mined our OGM data for SVs, which have been described to be relevant to prognosis.³

In *ETV6::RUNX1*+ BCP-ALL case ALL4, which was stratified as a standard-risk case, but later developed a relapse, a somatic loss of 12 Mb length encompassing the *TP53* gene locus was detected (chr17:5,423,849–18,312,173 bp) (Suppl. Table S6). In line with this, case ALL4 had a mutator phenotype harboring 198 SNVs at initial diagnosis, which represents an excessive number of mutations compared with other *ETV6::RUNX1* cases (median 18 SNVs/indels [range, 2–60]). However, this

TP53 deletion was also present in the SNP-array data, which were retrospectively gathered in the framework of this study (Suppl. Table S4). We identified 2 double hits (combination of SV and SNV) in the genes *STAG2* and *MANBA* that were already present at initial diagnosis (Suppl. Tables S5 and S7). The 2 deletions affecting *STAG2* (chrX:123,968,231–124,001,631 bp) and *MANBA* (chr4:102,682,958–102,825,431 bp) were validated by long-read sequencing (Suppl. Figure S7A and S7E).

We could retrospectively show that *ETV6::RUNX1+* BCP-ALL case ALL13 harbored an intragenic *IKZF1* deletion. Patients characterized by *IKZF1* deletions that co-occur with deletions in *CDKN2A/B*, *PAX5*, or *PAR1* in the absence of *ERG* deletions belong to a recently described poor prognostic group, *IKZF1*^{plus}, with lower event-free survival and higher rates of relapse in an MRD-dependent manner.²⁵ ALL13 carried a 39 kb deletion (chr7:50,306,321–50,348,483 bp), which led to loss of exon 2 and 3 within the *IKZF1* gene and a concomitant 61 kb deletion (chr9:36,841,199–36,916,988 bp) affecting exon 8 and 9 of the *PAX5* gene. Deletion in *IKZF1* (chr7:50,308,978–50,348,439 bp) and *PAX5* (chr9:36,843,016–36,903,796 bp) were confirmed by long-read sequencing (Suppl. Figure S10). This case would have fulfilled the *IKZF1*^{plus} criteria, although these criteria are not clinically applied to *ETV6::RUNX1* positive cases. Neither deletion was detectable via the CytoSNP-12 v2.1 array, underscoring the benefits of applying OGM.

We detected another *IKZF1* alteration, an 88 kb deletion removing exon 2-7 (chr7:50,306,321–50,399,656 bp), which was identified by OGM and not detected by the CytoScanHD-array in an HD BCP-ALL (ALL55). OGM data suggested a low variant allele frequency (~0.06) of the *IKZF1* deletion, which is under the detection limit of the applied array. The prognostic impact of subclonal *IKZF1* deletions is still unclear and needs to be assessed.²⁶ No other deletion of the *IKZF1*^{plus} criteria occurred in this medium risk patient.

Our results show that OGM is particularly useful for the detection of SVs within the highly repetitive region of the *IKZF1* gene, where array platforms lack probe coverage.

Novel secondary translocations and potential fusion genes were frequently detected in *ETV6::RUNX1+* BCP-ALL

One of the strengths of OGM is the ability to detect inter-chromosomal translocations and complex rearrangements. We mined our data for the presence of potential secondary translocations that might lead to the genesis of novel fusion genes. Notably, due to the resolution of OGM, the exact breakpoints of the translocation could not be estimated and hence, the fusions are approximations. We identified 10 translocations where gene fusions potentially led to disruption of the juxtaposed genes, including recurrent translocations affecting the *MED13L* gene locus on 12q24.21 in 2 *ETV6::RUNX1+* BCP-ALL cases. In ALL13, a t(6;12)(p21;q24) was reported by karyotyping during routine diagnostics. By OGM, this translocation was determined to be a complex 3-way translocation t(6;8;12)(p12.1;q24.23;q24.21), which might lead to the disruption of *MED13L* and *HMGCLL1* (Suppl. Figure S11). The disruption of *MED13L* by the translocation was verified by long-read sequencing (Suppl. Figure S7M).

In this study, 9 translocations detected by OGM, which have not yet been described in the Mitelman database or by Brady et al.,³ might lead to the in-frame fusion of 2 genes. Seven of those translocations were detected in *ETV6::RUNX1+* and 2 in HD BCP-ALL cases (Suppl. Table S13) and are further described below.

OGM detected a translocation t(X;13)(q25;q13.1) in an *ETV6::RUNX1* BCP-ALL (ALL29) leading to potential fusion of exon 1 of the *PBS5B* gene to exon 3–35 of *STAG2*, which both encode for proteins of the cohesion complex (Figure 7A). By long-read sequencing and RNA-seq, we could validate the fusion of *PDS5B::STAG2* on DNA and mRNA level in the leukemic

sample (Figure 7B; Suppl. Figure S12A and S12B). In ALL10, OGM detected a translocation t(2;10)(q37.1;p14) that potentially leads to in-frame fusion of *SFMBT2* exon 1–2 and *DGKD* exon 4–30 (Figure 7C). The *SFMBT2::DGKD* fusion was confirmed by long-read sequencing and expression in the leukemic sample was verified by RNA-seq (Figure 7D; Suppl. Figure S12C and S12D). In addition, the *SFMBT2* gene locus was potentially disrupted by a translocation t(10;19)(p14;p13.3) in case ALL5, which might also alter the *UHRF1* gene located on 19p13.3. We further validated the fusion of *TPRG1::BRD8* in ALL3 by long-read sequencing (Suppl. Figure S12E and S12F) and confirmed the expression of *TDRD5::LPCAT2* in an HD BCP-ALL (ALL59) and *PEX1::FGFR1OP2* in an *ETV6::RUNX1* BCP-ALL (ALL10) by RNA-seq, respectively (Suppl. Figure S13).

An immunoglobulin heavy locus (*IGH*) rearrangement was observed in an HD BCP-ALL case, ALL51, by FISH (using an *IGH* 5' 3' break-apart probe), which indicated the presence of a t(14;14) with complex additional aberrations on the derivative chromosome 14; without determination of the translocation partner (Figure 8A). By applying OGM, we found complex intrachromosomal translocations, which potentially lead to the juxtaposition of the *IGH* region on 14q32.33 to *SNAPC1*, likely via intrachromosomal translocation of the *IGH* locus to 14q22.1 (*PTGER2*) and concomitant inversion of the *SNAPC1* locus (Figure 8B and 8C). By RNA-seq, we could detect fusion of the *IGHVII-26-2/IGHV7-27* locus to *SNAPC1* (Figure 8D). However, no gene located on 14q was highly expressed, suggesting that the *IGH* translocation is rather a bystander product of the complex rearrangement on chromosome 14 than a stand-alone driver event in that particular case.

DISCUSSION

By applying a combination of conventional cytogenetic and molecular genetic methods, WES and the novel OGM technique, we comprehensively resolved the landscapes of somatic SNVs/indels and SVs in 60 cases of the 2 most frequent subtypes of childhood BCP-ALL: *ETV6::RUNX1+* and classical HD. Our data adds to the molecular landscape of those subtypes determined by previous large-scale studies that applied SNP/micro-array and/or short-read NGS.^{3,18–20,27–29}

In our study, OGM reliably detected 95% of SVs reported by molecular cytogenetics and SNP-array. In line with this, a high concordance of karyotyping and FISH in hematological malignancies has been shown in previous studies.^{12,13,30–32} By applying this new technique, we were able to detect an additional 677 SVs not identified by conventional techniques, of which 66% (448/677) were focal deletions with a median of 40.5 kb. These included 2 cryptic deletions within the highly repetitive region of the *IKZF1* gene³³ that were missed by SNP-array likely due to poor probe coverage and low allele frequency. Other studies showed that OGM may become a particularly useful tool to identify patients belonging to the prognostically relevant *IKZF1*^{plus} group in a simple, fast, and cost-effective way.^{12,17} Of note, *IKZF1*^{plus} detection is usually assessed either by SNP-array or multiplex ligation-dependent probe amplification,²⁵ which are comparable in respect to detection limit and sensitivity of ~20%. Consequently, *IKZF1* deletions may be difficult to detect in samples with low tumor cell content or leukemic subclones by conventional techniques. However, prognostic impact of subclonal deletions still needs to be evaluated in a clinical setting.²⁶

Further, we could show that OGM is able to outperform Cyto-SNP12 array, depending on the resolution of the sensitivity of the array. OGM is a useful tool to further detect SVs in gene regions with probably low marker coverage in CytoScanHD-array, which is mainly enriched in known OMIM genes. Around 20% of leukemic blasts are required for SNP-array in a diagnostic setting for large CN alterations.^{34,35} Our study and others^{14,36,37} showed that OGM performance on samples with blast

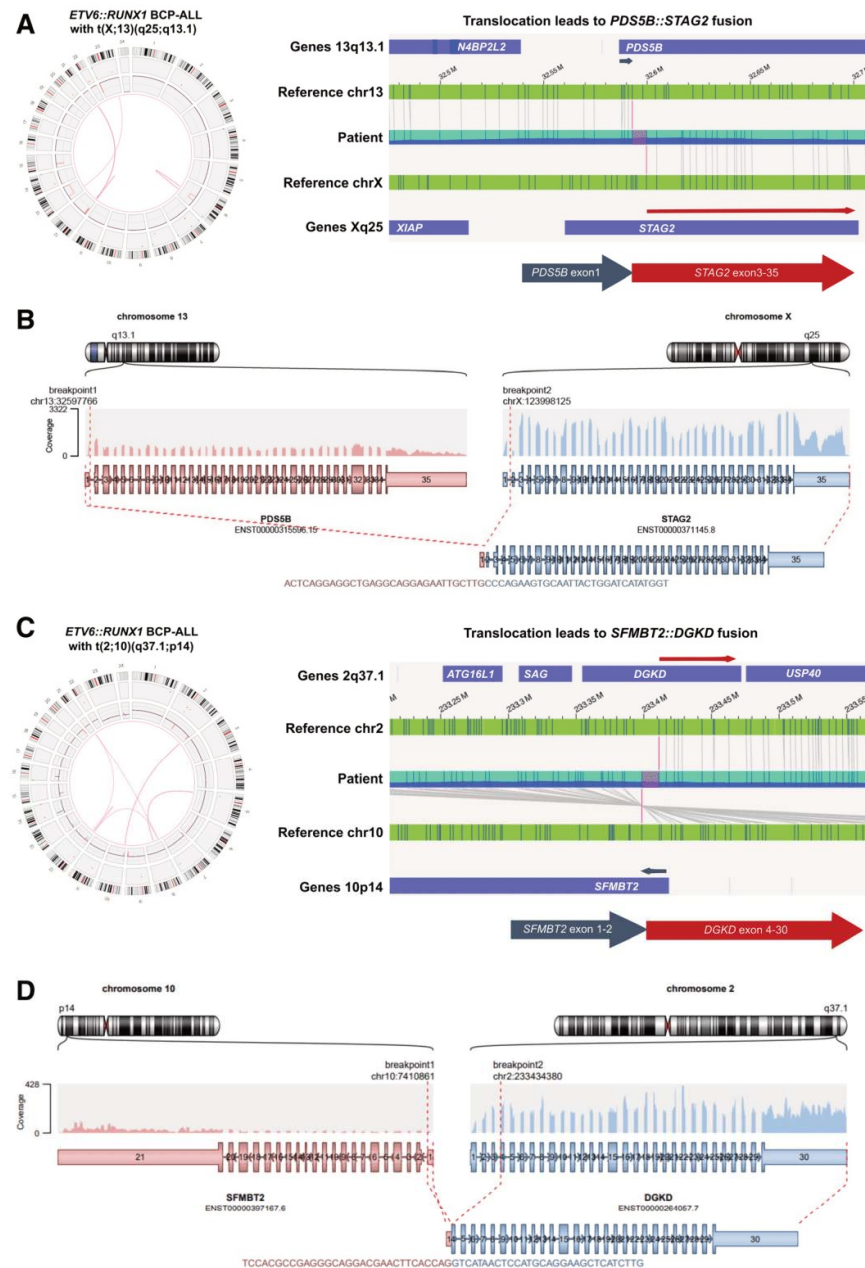


Figure 7. OGM and RNA-seq reveal expression of novel fusion genes in *ETV6::RUNX1* BCP-ALL. (A) Left: Circos plot showing somatic SVs of ALL28 with secondary translocation t(X;13)(q25;q13.1) detected by OGM (left). Right: Detailed view of patient optical map (blue) partly aligning to reference chromosomes (green) 13 and X, indicating a translocation that leads to potential fusion of *PDS5B* exon1 (gray arrow) to *STAG2* exon 3–35 (red arrow). Potential breakpoints are indicated in pink. (B) Validation of *PDS5B::STAG2* expression in ALL28 by RNA-seq. Arriba output of RNA-sequencing data is shown indicating a t(X;13)(q25;q13.1) that leads to in-frame fusion of *PDS5B* exon1 to *STAG2* exon 3–34. Coverage of the aligned reads is indicated. (C) Left: Circos plot showing somatic SVs of ALL10 with secondary translocation t(2;10)(q37.1;p14) detected by OGM (left). Right: Detailed view of patient optical map (blue) partly aligning to reference chromosomes (green) 2 and 10, indicating a translocation that leads to potential fusion of *SFMBT2* exon1–2 (gray arrow) to *DGKD* exon 4–30 (red arrow). Potential breakpoints are indicated in pink. (D) Validation of *SFMBT2::DGKD* expression in ALL10 by RNA-seq. Arriba output of RNA-sequencing data is shown indicating a t(2;10)(q37.1;p14) that leads to in-frame fusion of *SFMBT2* exon1 to *DGKD* exon 4–30. Coverage of the aligned reads is indicated. BCP-ALL = B-cell precursor acute lymphoblastic leukemia; OGM = optical genome mapping; SV = structural variants.

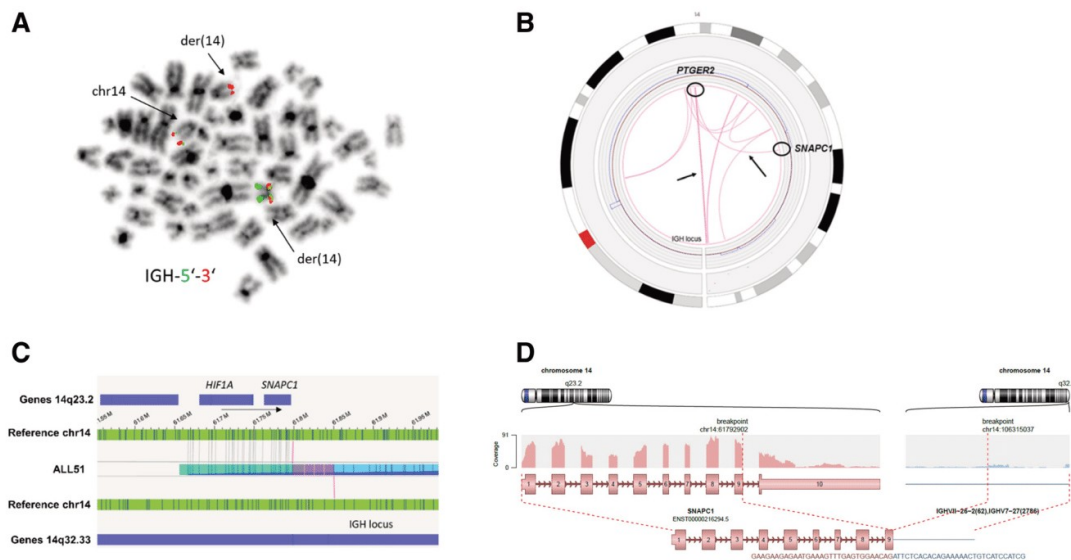


Figure 8. OGM resolved complex *IGH* rearrangements in an HD leukemia patient. (A) FISH result (with *IGH* 5' and 3' break-apart probes) is shown, indicating a t(14;14) with additional complex aberrations of the derivative chromosome 14 in ALL51. (B) Circos plot of chromosome 14 of ALL51 is depicted showing several intrachromosomal translocations including juxtaposition of the *IGH* region to the *PTGER2* loci and inversion of the *SNAPC1* region. (C) Optical map (blue) showing the juxtaposition of the *IGH* locus on 14q32.33 to the *SNAPC1* locus on 14q23.2. (D) Arriba output of RNA-sequencing data is shown indicating a t(14;14)(q23.2;p32.33) that leads to fusion of *SNAPC1* exon 1–9 to the *IGHV1-26-2/IGHV7-27* locus. Coverage of the aligned reads is indicated. HD = hyperdiploid; *IGH* = immunoglobulin heavy locus; OGM = optical genome mapping.

amounts of about 20%–30% could also detect alterations observed by conventional diagnostic methods. Hence, OGM could be an attractive option for routine diagnostic workup of hematological malignancies, as it has a fast turnaround time of ≈5 days from sample acquisition to data analysis.

We showed that the SV landscapes in *ETV6::RUNX1* and classical HD leukemia are profoundly different. Compared with HD BCP-ALL, *ETV6::RUNX1*+ BCP-ALL harbored on average 3 times more SVs (mainly focal deletions) that might be associated with hyperactive *RAG1/2* mediated recombination, which has been described as an oncogenic process in this specific subgroup.³⁸ By contrast, there are limited analyses of aberrant *RAG* activity in HD BCP-ALL, and mechanisms of secondary SVs need to be further investigated. We further identified well-known large recurrent 12p deletions (≥5 Mb) in *ETV6::RUNX1* leukemia, which are associated with biallelic inactivation of the *ETV6* gene.³⁹

By integrating OGM and WES, we identified 52 MARs affecting either 40 potential target genes or 12 intergenic regions. The latter regions are increasingly recognized in the field as possibly having functional effects on leukemia genomes through enhancer deletions, hijacking, and topically associated domain disruption, leading to gene dysregulation.^{40–42} Of note, we identified 4 intergenic MARs located in proximity (20 kb) of genes including 1q32.2 (–6 kb; *CD55*), 21q22.11 (–20 kb; *ATP5PO*), and 22q12.1 (–0.2 kb; *XBPI*). It should be noted that OGM has no exact breakpoint resolution and localization of MARs may have upstream and downstream variations dependent on label coverage of the identified region.

In line with previous findings, we identified SVs in both subgroups that affected several genes known to play a role in BCP-ALL development, including *ETV6*, *PAX5*, *BTG1*, *CDKN2A*, and *RAG2*.^{3,38} In addition to these well-described alterations, we identified SVs recurrently affecting *STAG2*, a member of the cohesin complex, including (1) 4 cases with focal intragenic deletions in the cohesin complex gene *STAG2*; (2) 1 case harboring a focal deletion within *STAG2* (16 kb, chrX:123,983,164–124,005,001 bp)

and a concomitant missense mutation (p.Ile118Met); and (3) 1 case with a fusion of *STAG2* to *PDS5B* gene, which also encodes for a cohesin complex protein. *STAG2* inactivation has been implicated to cause aneuploidy in human cancer⁴³; however, only 2 of the cases with intragenic *STAG2* deletion were HD tumors, and the 4 *ETV6::RUNX1* leukemia with *STAG2* alteration harbored no aneuploidy. The cohesin complex is involved in chromosome segregation in dividing cells, but is also implicated in more wide-ranging functions, such as DNA damage repair and regulation of gene expression, as indicated by the loss-of-function developmental syndrome, Cornelia de Lange.⁴⁴ Previous studies based on the short-read sequencing, conventional technologies, and OGM reported somatic alterations of *STAG2* in BCP-ALL,^{3,17,38} and showed that germline SNVs/indels in cohesin complex members, such as *NIPBL* or *RAD21*, could be involved in BCP-ALL predisposition.^{45,46} Consistently, studies in conditional knockout mice demonstrated that *Stag2* deletion in hematopoietic stem/progenitor cells increased their self-renewal capacity and impaired B-cell differentiation by decreasing chromatin accessibility and transcription of important B-cell transcription factors, such as *PAX5*.⁴⁷ Our findings of recurrent *STAG2* alterations underlines the potential relevance of the cohesin complex genes to BCP-ALL development, arguably without involvement of chromosome segregation defects.

We further showed that a combined approach of OGM and NGS, including long-read and RNA-sequencing, is able to detect novel potential driver genes in BCP-ALL. Specifically, we were able to identify 19 recurrently altered regions in BCP-ALL with novel potential leukemic drivers. Twelve of these novel regions were not detected by SNP-array, which underscores the benefits of applying the OGM technique. We identified recurrently affected genes by SVs, such as *FOCAD/HACD4*, *MANBA*, and *SFMBT2* that were validated by long-read sequencing and/or RNA-seq in at least 1 leukemia. Recurrent potentially inactivating SVs in *SFMBT2* included the following: (1) an *ETV6::RUNX1*+ case (ALL10) with translocation t(2;10)(q37.1;p14), leading to fusion of the promoter region to *DGKD* and a concomitant loss of the

whole *SFMBT2* gene; (2) an *ETV6::RUNX1*+ case (ALL5) with a translocation t(10;19)(p14;p13.3) that might lead to disruption of *SFMBT2* and *UHRF1*; and (3) a HD case with an intragenic focal deletion (45 kb; chr10:7,286,100–7,343,440 bp), which hints at a potential role in leukemogenesis. *SFMBT2* is associated with histone-binding activity and negative regulation of gene expression.⁴⁸ Its functional impact and role in BCP-ALL leukemogenesis should be addressed in the future.

We aimed to identify subtype-specific events of both genetic subgroups. In classical HD BCP-ALL, we did not find any recurrently affected region or gene by structural variation but SNVs/indels that recurrently affected genes including *KRAS*, *FLT3*, and *CREBBP*. In addition, *CREBBP* mutations are more prevalent in recurrent HD BCP-ALL.²⁴ Consistently, one of the HD BCP-ALL relapses in our study already had a *CREBBP* mutation (p.Gln1491Lys) at the time of diagnosis, which is located in the HAT domain and was previously described in relapse prone HD BCP-ALL.⁴⁹

In *ETV6::RUNX1*+ cases, we detected genes including *UHRF1* and *UBA2* recurrently altered by SV or SNV. Those genes are involved in sumoylation and ubiquitination, which is currently recognized as a new biological pathway in BCP-ALL,³ highlighting the importance of those pathways in *ETV6::RUNX1* leukemia. Interestingly, by using long-read sequencing, we detected that deletions (1) in ALL4 removed exon 1 and 2 of the *UHRF1* gene, which is coding for the ubiquitin homologues domain and (2) in ALL7 removed exon 13 of the *UBA2* gene, which is coding for the ubiquitin/SUMO-activating enzyme ubiquitin-like domain. A recent study further showed that a frameshift variant of *UBA2* was discussed to predispose to *ETV6::RUNX1*+ leukemia in monozygotic twins.⁵⁰

Novel *ETV6::RUNX1* specific regions detected in our study encompassed genes such as *GPRC5A*, *MED13L*, and *MIB1* that might play a role in leukemogenesis. The MAR of 3 cases (chr18:21,727,621–21,759,839 bp) included the *MIB1* promoter region. *MIB1* encodes for an E3 ubiquitin-protein ligase that is involved in, among other processes, NOTCH signaling and ubiquitination of centriolar satellites.^{51,52} The *MIB1* gene locus has been shown to be hypomethylated in *ETV6::RUNX1*+ BCP-ALL^{53,54} and is highly expressed in *ETV6::RUNX1* leukemia compared with other subtypes (MILE study, St. Jude Pecan Cloud).^{55,56} Interestingly, the relapsed sample of ALL4 harbored an acquired translocation t(17;18)(q12;q11.2) affecting the *MIB1* gene locus, indicating that deregulation of *MIB1* might play a role in leukemogenesis. Notably, we found 2 translocations and a stop-gain mutation disrupting *MED13L* in 3 *ETV6::RUNX1*+ cases. *MED13L* is part of a mediator complex functioning as a coactivator for RNAII transcribed genes,⁵⁷ and has been implicated in a case with constitutional haploinsufficiency in a syndromic patient who developed an ALL alongside clinical abnormalities.⁵⁸

The identification of genes such as *ETV6*, *BTG1*, *STAG2*, *MANBA*, *TBL1XR1*, and *NSD2* that were affected by double hits of SVs and SNVs/indels demonstrates the power of combining OGM and WES for comprehensive determination of somatic alterations in BCP-ALL. Moreover, double hits affecting one target gene are indicative of a potential tumor suppressor⁵⁹ that may have a strong relevance for ALL development. However, our study design with a combined approach of OGM and WES has a limited ability to detect compound heterozygosity due to lacking resolution on single-nucleotide level of OGM and short-reads surrounding the SNV. Only a scenario where for example a deletion overlaps with an SNV would suggest compound heterozygosity. That was not detected in any of the potential targets in our cohort. It is worth mentioning that secondary mutations in *TBL1XR1* have recently been implicated in improving risk stratification, especially in *ETV6::RUNX1* leukemia, where mutations led to inferior survival.³

Taken together, we have unraveled a profoundly different molecular make-up of *ETV6::RUNX1*+ and classical HD BCP-ALL in terms of their landscape of structural aberrations.

Moreover, we identified novel, recurrently altered genomic loci potentially involved in disease development and/or progression of these 2 BCP-ALL subtypes. Although OGM has obvious advantages compared with conventional technologies, comprehensive side-by-side evaluation and standardized round-robin tests must still be performed before it can be applied in routine diagnostics. And, only a fraction of genomic alterations (>500 bp) can be analyzed with OGM, so other techniques such as RNA- and long-read sequencing must be performed in parallel to cover the full spectrum of alterations. The rapidly developing third-generation high-fidelity long-read sequencing techniques might be advanced tools to decipher the whole spectrum of SVs and SNVs/indels in BCP-ALL and other tumor entities in the near future.

ACKNOWLEDGMENTS

The authors thank Silke Furlan, Daniel Scholtysik, and Katayoun Alemazkour for their excellent technical assistance. Computational support and infrastructure was provided by the Center for Information and Media Technology (ZIM) at the University of Dusseldorf (Germany). The authors greatly acknowledge the help of Bärbel Überlacker, Laboratory Zotz-Klimas, Dusseldorf. They thank Peter Ebert of the Core Unit Bioinformatics of the Medical Faculty of the Heinrich-Heine University for bioinformatic support.

AUTHORS CONTRIBUTIONS

DB generated the molecular genetic data. DB and RW analyzed the molecular genetic data. AKB and KN provided cytogenetic and molecular cytogenetic data. SK, OAH, MS, TB, and AA provided patient samples and clinical data. LY and JA performed bioinformatic analysis of WES data. TM and JE processed long-read sequencing data. DS and DP provided bioinformatic analysis of RNA-seq data. UF, RW, and AB designed and supervised the project. All authors read and approved the final article.

DISCLOSURES

The authors have no conflicts of interest to disclose.

SOURCES OF FUNDING

DB was funded by the Dusseldorf School of Oncology (DSO, Medical Faculty of the Heinrich-Heine University Dusseldorf). RW was funded by the Research Commission of the Medical Faculty of the Heinrich-Heine University Dusseldorf. AB was supported by the Tour der Hoffnung, the Katharina-Hardt-Stiftung, the German Children's Cancer Foundation, the German Federal Office for Radiation Protection (BfS, FKZ 3622S32231). UF was funded by the German Jose Carreras Foundation (DJCLS 18R/2021), the BfS (FKZ 3622S32231), the German Research Foundation (HE 8807/3-1), and the Cancer Prevention Graduate School, supported by the German Cancer Aid (grant no. 70114736) and coordinated by the German Cancer Research Center (DKFZ) (grant no. 70114766).

REFERENCES

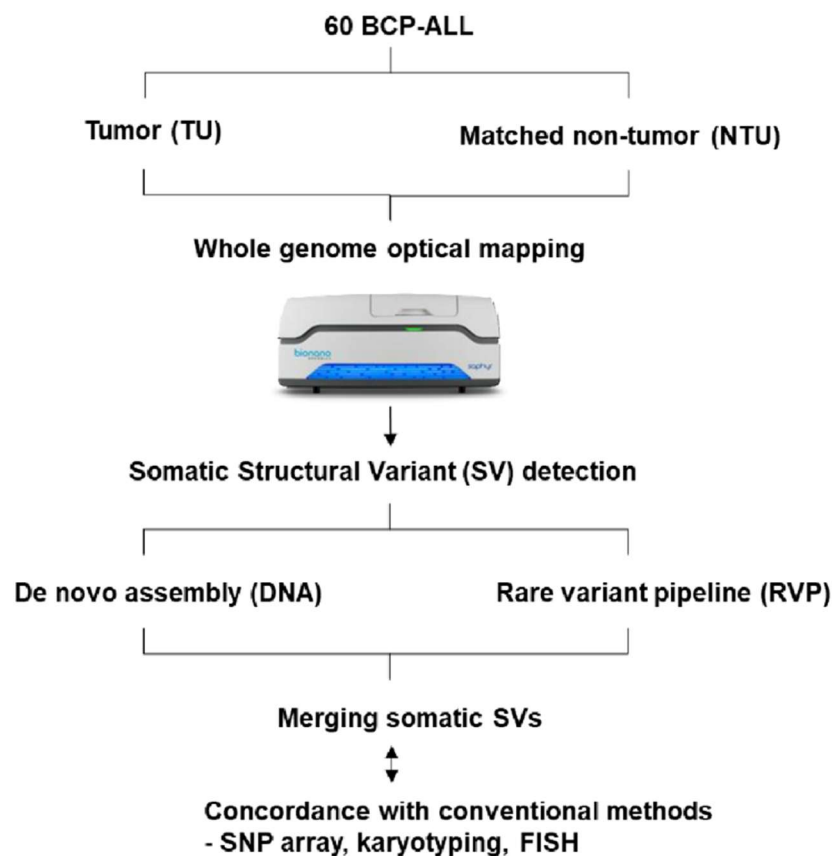
1. Haas OA, Borkhardt A. Hyperdiploidy: the longest known, most prevalent, and most enigmatic form of acute lymphoblastic leukemia in children. *Leukemia*. 2022;36:2769–2783.
2. Pui CH. Precision medicine in acute lymphoblastic leukemia. *Front Med*. 2020;14:689–700.
3. Brady SW, Roberts KG, Gu Z, et al. The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat Genet*. 2022;54:1376–1389.
4. Enshaie A, Vora A, Harrison CJ, et al. Defining low-risk high hyperdiploidy in patients with paediatric acute lymphoblastic leukaemia: a retrospective analysis of data from the UKALL97/99 and UKALL2003 clinical trials. *Lancet Haematol*. 2021;8:e828–e839.
5. Greaves M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat Rev Cancer*. 2018;18:471–484.
6. Mahmoud M, Gobet N, Cruz-Davalos DI, et al. Structural variant calling: the long and the short of it. *Genome Biol*. 2019;20:246.
7. Zhao X, Collins RL, Lee WP, et al. Expectations and blind spots for structural variation detection from long-read assemblies and short-read genome sequencing technologies. *Am J Hum Genet*. 2021;108:919–928.

8. Li Y, Roberts ND, Wala JA, et al. Patterns of somatic structural variation in human cancer genomes. *Nature*. 2020;578:112–121.
9. Salzberg SL, Yorke JA. Beware of mis-assembled genomes. *Bioinformatics*. 2005;21:4320–4321.
10. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*. 2011;13:36–46.
11. Lestrinant V, Duployez N, Penther D, et al. Optical genome mapping, a promising alternative to gold standard cytogenetic approaches in a series of acute lymphoblastic leukemias. *Genes Chromosomes Cancer*. 2021;60:657–667.
12. Luhmann JL, Stelter M, Wolter M, et al. The clinical utility of optical genome mapping for the assessment of genomic aberrations in acute lymphoblastic leukemia. *Cancers (Basel)*. 2021;13:4388.
13. Neveling K, Mantere T, Vermeulen S, et al. Next-generation cytogenetics: Comprehensive assessment of 52 hematological malignancy genomes by optical genome mapping. *Am J Hum Genet*. 2021;108:1423–1435.
14. Suttrop J, Luhmann JL, Behrens YL, et al. Optical genome mapping as a diagnostic tool in pediatric acute myeloid leukemia. *Cancers (Basel)*. 2022;14:2058.
15. Puiggros A, Ramos-Campoy S, Kamaso J, et al. Optical Genome mapping: a promising new tool to assess genomic complexity in chronic lymphocytic leukemia (CLL). *Cancers (Basel)*. 2022;14:3376.
16. Rack K, De Bie J, Amey G, et al. Optimizing the diagnostic workflow for acute lymphoblastic leukemia by optical genome mapping. *Am J Hematol*. 2022;97:548–561.
17. Gao H, Xu H, Wang C, et al. Optical Genome Mapping for Comprehensive Assessment of Chromosomal Aberrations and Discovery of New Fusion Genes in Pediatric B-Acute Lymphoblastic Leukemia. *Cancers (Basel)*. 2022;15:35.
18. Paulsson K, Forestier E, Lilljebjorn H, et al. Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A*. 2010;107:21719–21724.
19. Lilljebjorn H, Soneson C, Andersson A, et al. The correlation pattern of acquired copy number changes in 164 ETV6/RUNX1-positive childhood acute lymphoblastic leukemias. *Hum Mol Genet*. 2010;19:3150–3158.
20. Studd JB, Cornish AJ, Hoang PH, et al. Cancer drivers and clonal dynamics in acute lymphoblastic leukaemia subtypes. *Blood Cancer J*. 2021;11:177.
21. Abbasi MR, Nebral K, Haslinger S, et al. Copy number changes and allele distribution patterns of chromosome 21 in B cell precursor acute lymphoblastic leukemia. *Cancers (Basel)*. 2021;13:4597.
22. Schinnerl D, Mejsnikova E, Schumich A, et al. CD371 cell surface expression: a unique feature of DUX4-rearranged acute lymphoblastic leukemia. *Haematologica*. 2019;104:e352–e355.
23. Uhrig S, Ellermann J, Walther T, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res*. 2021;31:448–460.
24. Chen C, Bartenhagen C, Gombert M, et al. Next-generation-sequencing of recurrent childhood high hyperdiploid acute lymphoblastic leukemia reveals mutations typically associated with high risk patients. *Leuk Res*. 2015;39:990–1001.
25. Stanulla M, Dagdan E, Zaliava M, et al. IKZF1(plus) defines a new minimal residual disease-dependent very-poor prognostic profile in pediatric B-cell precursor acute lymphoblastic leukemia. *J Clin Oncol*. 2018;36:1240–1249.
26. Stanulla M, Cave H, Moorman AV. IKZF1 deletions in pediatric acute lymphoblastic leukemia: still a poor prognostic marker? *Blood*. 2020;135:252–260.
27. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*. 2007;446:758–764.
28. Paulsson K, Lilljebjorn H, Biloglav A, et al. The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Nat Genet*. 2015;47:672–676.
29. Ueno H, Yoshida K, Shiozawa Y, et al. Landscape of driver mutations and their clinical impacts in pediatric B-cell precursor acute lymphoblastic leukemia. *Blood Adv*. 2020;4:5165–5173.
30. Gerding WM, Tembrink M, Nilius-Eliliwi V, et al. Optical genome mapping reveals additional prognostic information compared to conventional cytogenetics in AML/MDS patients. *Int J Cancer*. 2022;150:1998–2011.
31. Yang H, Garcia-Manero G, Sasaki K, et al. High-resolution structural variant profiling of myelodysplastic syndromes by optical genome mapping uncovers cryptic aberrations of prognostic and therapeutic significance. *Leukemia*. 2022;36:2306–2316.
32. Sahajpal NS, Mondal AK, Tyrdik T, et al. Clinical validation and diagnostic utility of optical genome mapping for enhanced cytogenomic analysis of hematological neoplasms. *J Mol Diagn*. 2022;24:1279–1291.
33. Bashton M, Hollis R, Ryan S, et al. Concordance of copy number abnormality detection using SNP arrays and Multiplex Ligation-dependent Probe Amplification (MLPA) in acute lymphoblastic leukaemia. *Sci Rep*. 2020;10:45.
34. Peterson JF, Van Dyke DL, Hoppman NL, et al. The utilization of chromosomal microarray technologies for hematologic neoplasms: an ACLPS critical review. *Am J Clin Pathol*. 2018;150:375–384.
35. Schoumans J, Suela J, Hastings R, et al. Guidelines for genomic array analysis in acquired haematological neoplastic disorders. *Genes Chromosomes Cancer*. 2016;55:480–491.
36. Nilius-Eliliwi V, Gerding WM, Schroers R, et al. Optical genome mapping for cytogenetic diagnostics in AML. *Cancers (Basel)*. 2023;15:1684.
37. Soler G, Ouedraogo ZG, Goumy C, et al. Optical genome mapping in routine cytogenetic diagnosis of acute leukemia. *Cancers (Basel)*. 2023;15:2131.
38. Papaemmanuil E, Rapado I, Li Y, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet*. 2014;46:116–125.
39. Zhang J, Mullighan CG, Harvey RC, et al. Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood*. 2011;118:3080–3087.
40. Yang M, Safavi S, Woodward EL, et al. 13q12.2 deletions in acute lymphoblastic leukemia lead to upregulation of FLT3 through enhancer hijacking. *Blood*. 2020;136:946–956.
41. Wang H, Cui B, Sun H, et al. Aberrant GATA2 activation in pediatric B-cell acute lymphoblastic leukemia. *Front Pediatr*. 2021;9:795529.
42. Yang M, Vesterlund M, Siavelis I, et al. Proteogenomics and Hi-C reveal transcriptional dysregulation in high hyperdiploid childhood acute lymphoblastic leukemia. *Nat Commun*. 2019;10:1519.
43. Solomon DA, Kim T, Diaz-Martinez LA, et al. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science*. 2011;333:1039–1043.
44. Parenti I, Kaiser FJ. Cornelia de Lange syndrome as paradigm of chromatinopathies. *Front Neurosci*. 2021;15:774950.
45. Fazio G, Massa V, Grioni A, et al. First evidence of a paediatric patient with Cornelia de Lange syndrome with acute lymphoblastic leukaemia. *J Clin Pathol*. 2019;72:558–561.
46. Schedel A, Friedrich UA, Morcos MNE, et al. Recurrent germline variant in RAD21 predisposes children to lymphoblastic leukemia or lymphoma. *Int J Mol Sci*. 2022;23:5174.
47. Viny AD, Bowman RL, Liu Y, et al. Cohesin members stag1 and stag2 display distinct roles in chromatin accessibility and topological control of HSC self-renewal and differentiation. *Cell Stem Cell*. 2019;25:682–696.e8.
48. Lee K, Na W, Maeng JH, et al. Regulation of DU145 prostate cancer cell growth by Scm-like with four mbt domains 2. *J Biosci*. 2013;38:105–112.
49. Inthral A, Zeitlhofer P, Zeginigg M, et al. CREBBP HAT domain mutations prevail in relapse cases of high hyperdiploid childhood acute lymphoblastic leukemia. *Leukemia*. 2012;26:1797–1803.
50. Bang B, Eisfeldt J, Barbany G, et al. A somatic UBA2 variant preceded ETV6-RUNX1 in the concordant BCP-ALL of monozygotic twins. *Blood Adv*. 2022;6:2275–2289.
51. Itoh M, Kim CH, Palardy G, et al. Mind bomb is a ubiquitin ligase that is essential for efficient activation of Notch signaling by Delta. *Dev Cell*. 2003;4:67–82.
52. Douanne T, Andre-Gregoire G, Thys A, et al. CYLD regulates centriolar satellites proteostasis by counteracting the E3 ligase MIB1. *Cell Rep*. 2019;27:1657–1665.e4.
53. Nordlund J, Backlin CL, Wahlberg P, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol*. 2013;14:r105.
54. Figueroa ME, Chen SC, Andersson AK, et al. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *J Clin Invest*. 2013;123:3099–3111.
55. McLeod C, Gout AM, Zhou X, et al. St. Jude cloud: a pediatric cancer genomic data-sharing ecosystem. *Cancer Discov*. 2021;11:1082–1099.
56. Haeflrich T, Kohlmann A, Wiczorek L, et al. Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. *J Clin Oncol*. 2010;28:2529–2537.
57. Adegbola A, Musante L, Calleeaert B, et al. Redefining the MED13L syndrome. *Eur J Hum Genet*. 2015;23:1308–1317.
58. Cafiero C, Marangi G, Orteschi D, et al. Novel de novo heterozygous loss-of-function variants in MED13L and further delineation of the MED13L haploinsufficiency syndrome. *Eur J Hum Genet*. 2015;23:1499–1504.
59. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971;68:820–823.

3.4.1 Original Publication II: Supplementary Figures

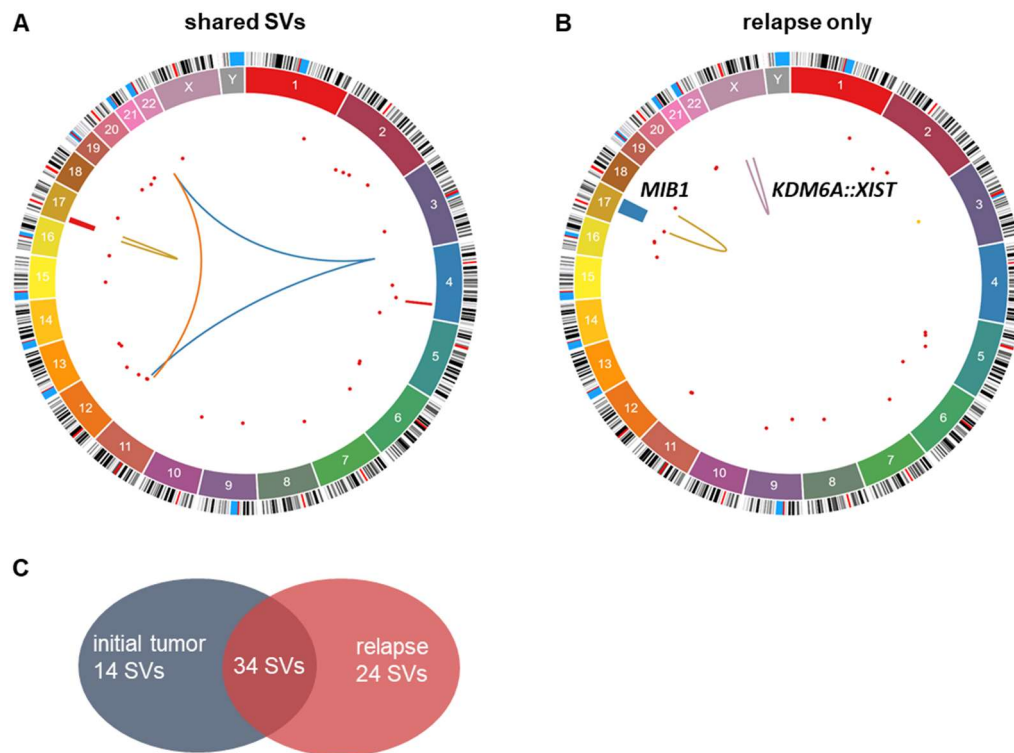
Relevant results for the discussion of this thesis were included in the following section. Complete supplementary methods and figures can be found online at the publisher site (refer to supplemental digital content): <https://links.lww.com/HS/A452>

SDC, Figure 2



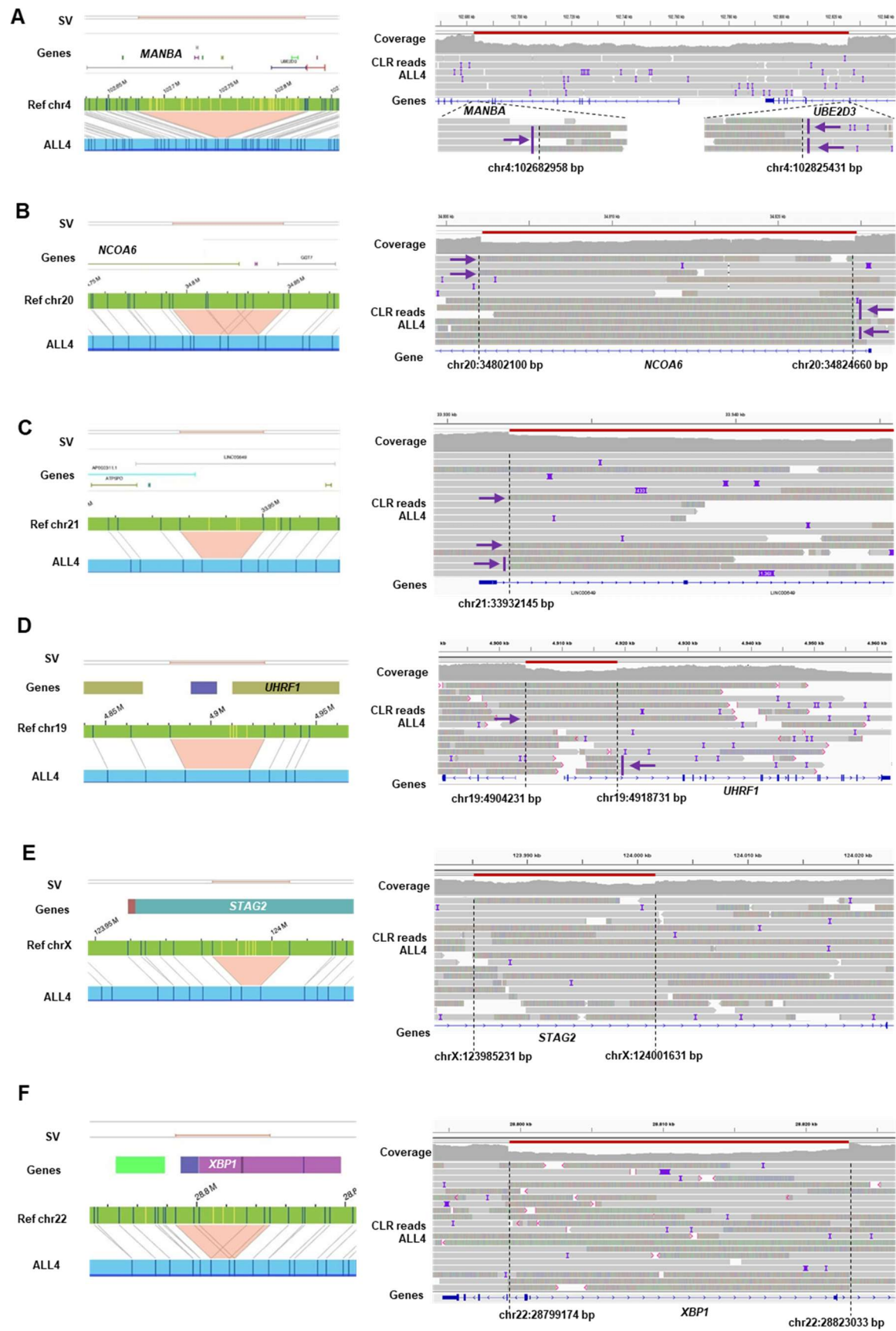
SDC, Figure 2: Optical genome mapping (OGM) workflow to detect somatically acquired SVs in 60 BCP-ALL. Tumor and matched non-tumor samples were collected from 60 BCP-ALL cases. De novo assembly and rare variant pipeline were applied for SV calling using the duo-annotation mode allowing for the detection of somatic SVs. Filtered somatic SVs of each pipeline were merged to a final set of somatic SVs. Concordance of diagnostically reported SVs (karyotyping and FISH), SVs detected by SNP array and OGM was determined.

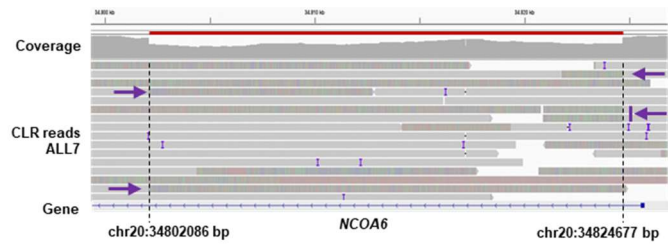
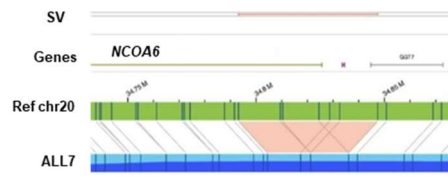
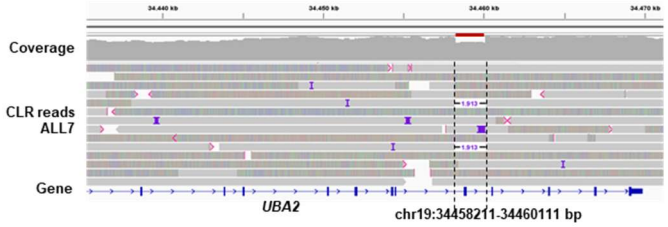
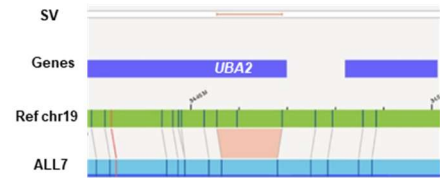
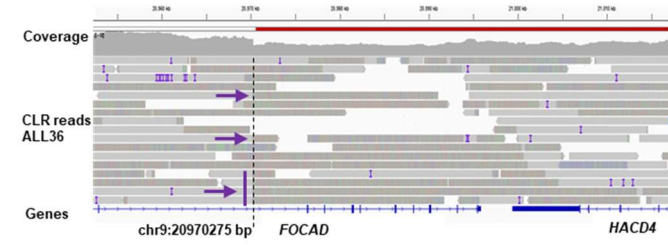
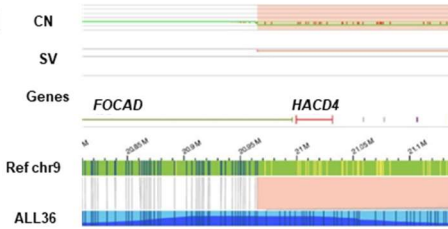
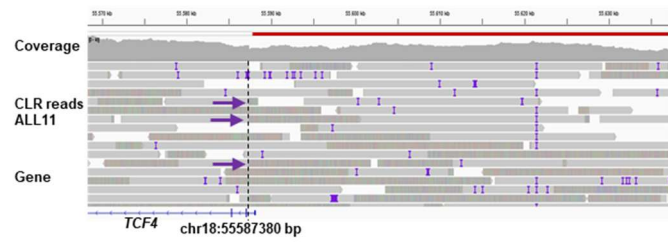
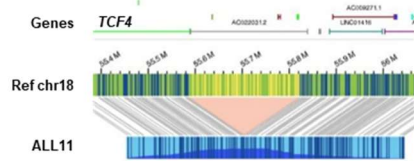
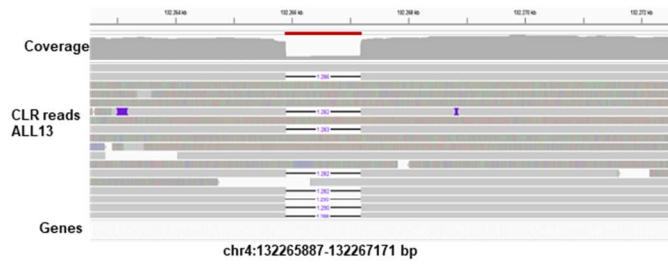
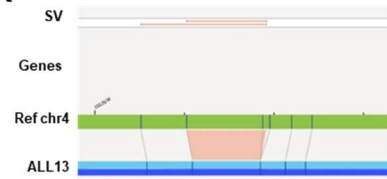
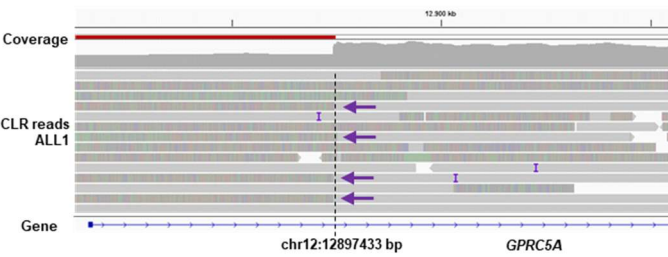
SDC, Figure 5

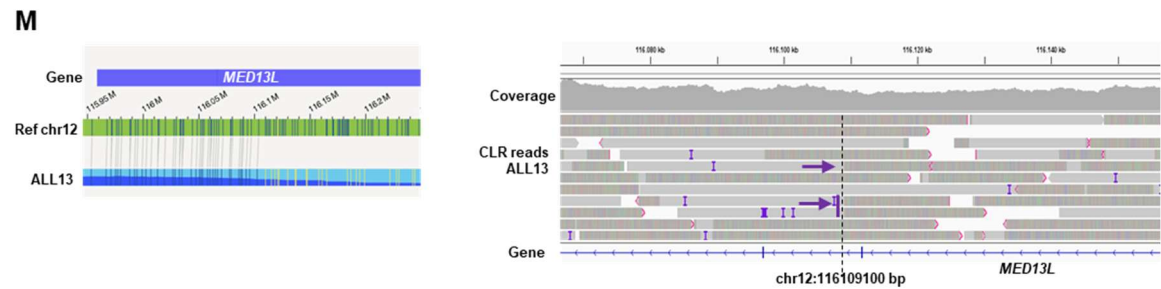


SDC, Figure 5: Tumor evolution in ETV6::RUNX1 BCP-ALL case ALL4, who developed a relapse 48 months after initial diagnosis. Circos plots showing (A) SVs shared by initial and relapsed BCP-ALL samples and (B) SVs unique to the relapse sample identified by OGM including t(17;18)(q12;q11.2) affecting the *MIB1* gene locus and inv(X)(p11.3;q13.2) leading to potential fusion of *KDM6A::XIST*. (C) Quantity chart showing that 34 SVs were shared between initial and relapse tumor samples, whereas 14 SVs were unique to the initial tumor and 26 SVs were acquired during disease progression and solely identified in the relapsed sample.

SDC, Figure 7

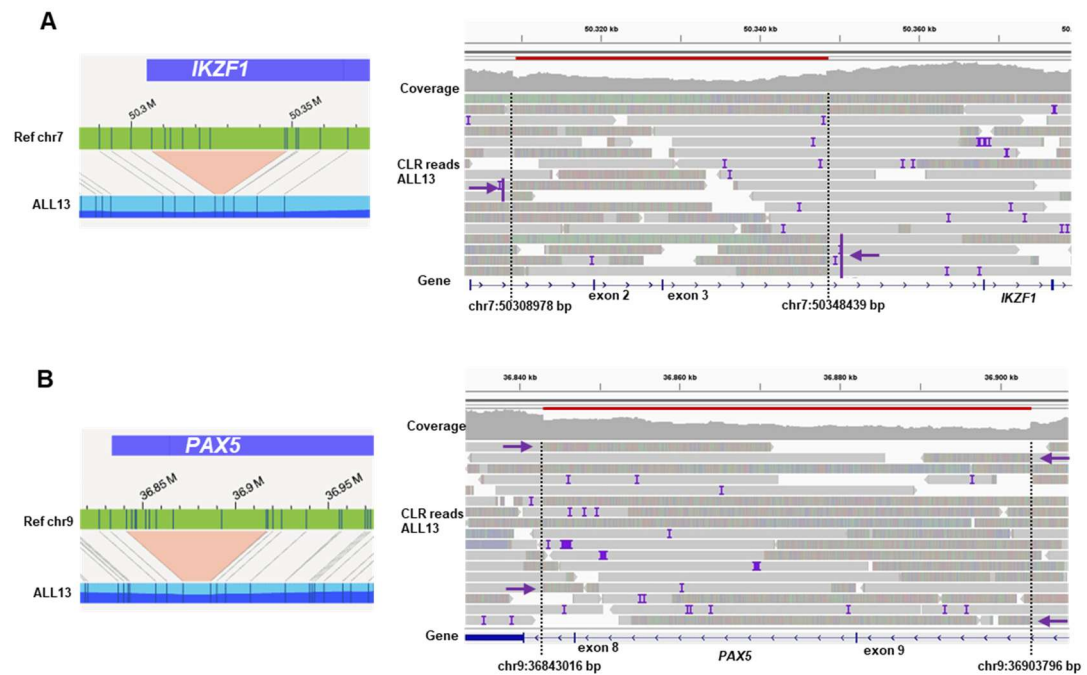


G**H****I****J****K****L**



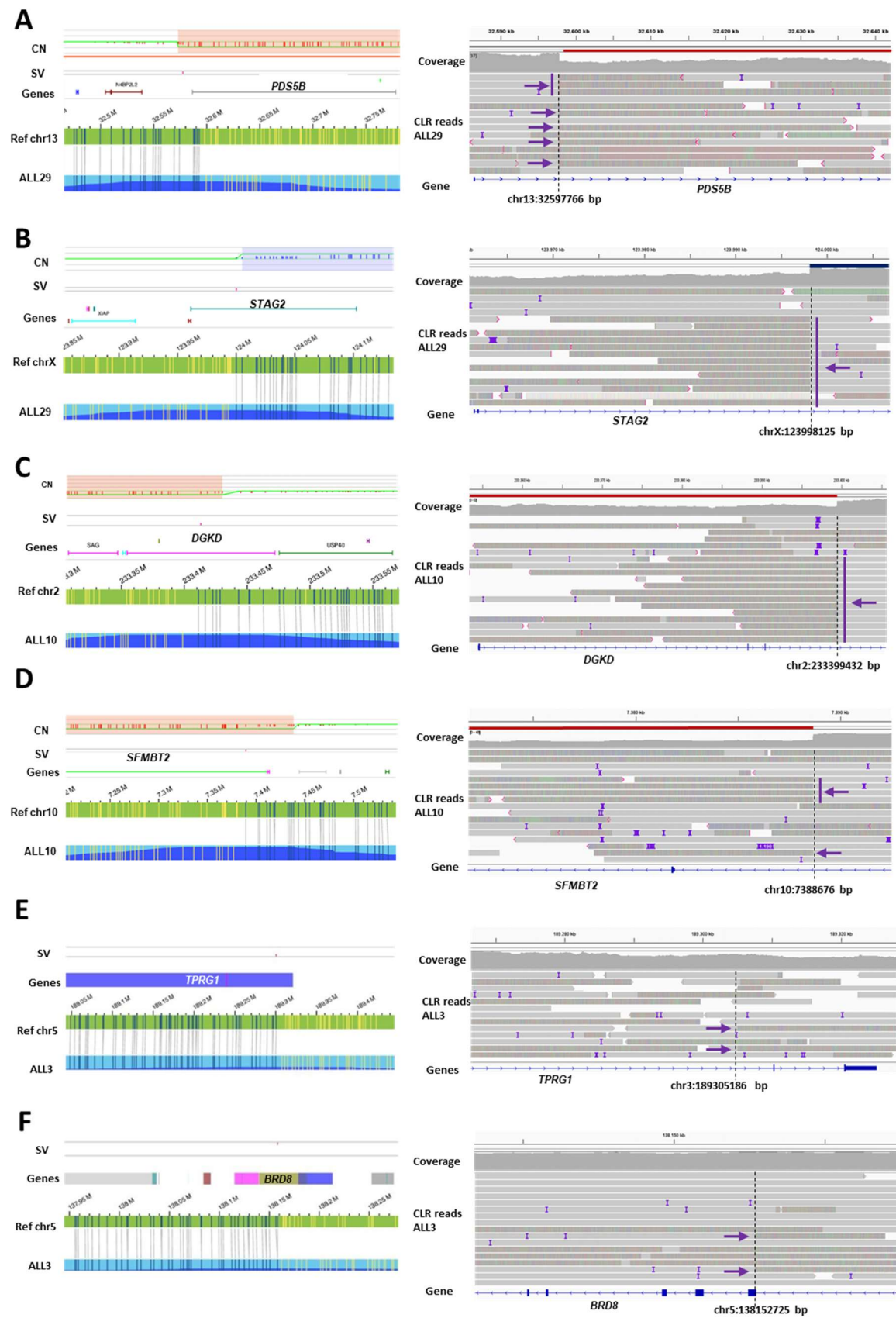
SDC, Figure 7: Selected SVs detected by OGM and cross-validation with long-read sequencing in seven BCP-ALL. Left: Optical map of the leukemia (blue) and respective reference map (green) indicating SV and overlapping gene. Red highlight indicates deletions. Split maps with aligned blue labels and unaligned yellow labels indicate translocations. Right: Continuous long-reads (CLR) aligning to the respective gene locus are shown (grey bars). Dashed lines indicate location of breakpoints. Split reads supporting the respective SV are marked by purple arrows. Decreased coverage of aligned reads support a deletion (red bar). (A) ALL4; 4q24 (MANBA) (B) ALL4; 20q11.22 (NCOA6) (C) ALL4; 21q22.11 (D) ALL4; 19p13.3 (UHRF1) (E) ALL4; Xq25 (STAG2) (F) ALL4, 22q12.1 (XBP1) (G) ALL7; 20q11.22 (NCOA6) (H) ALL7; 19q13.11 (UBA2) (I) ALL36; 9p21.3 (FOCAD; HACD4) (J) ALL11; 18q21.2 (TCF4) (K) ALL13; 4q28.3 (L) ALL1; 12p13.1 (GPRC5A) and (M) ALL13; 12q24.21 (MED13L).

SDC, Figure 10



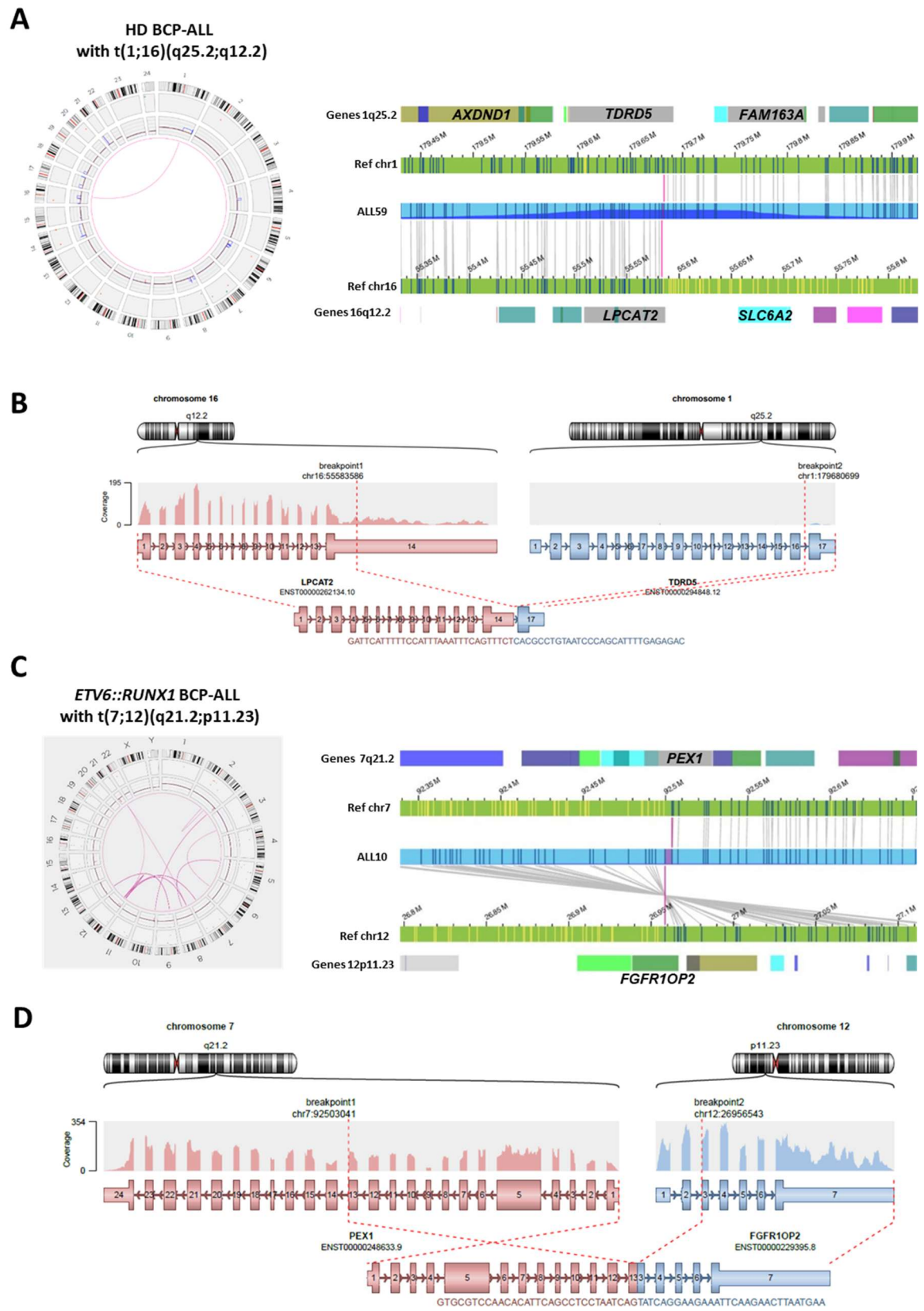
SDC, Figure 10: High risk ETV6::RUNX1 BCP-ALL case harboring concomitant IKZF1 and PAX5 deletion. (A) Left: Optical map of ALL13 (blue) compared to reference chr7 (green) indicating deletion in IKZF1. Right: ALL13 CLR-seq reads mapping to the IKZF1 locus are shown. Split reads (purple arrows) and decreased coverage (red bar) reveal a deletion on chr7:50,308,978-50,348,439 bp, removing exon 2 and 3 of IKZF1. (B) Left: Optical map of ALL13 (blue) compared to reference chr9 (green) indicating deletion in PAX5. Right: ALL13 CLR-seq reads mapping to the PAX5 locus are shown. Split reads (marked by purple arrows) and decreased coverage (red bar) reveal a deletion on chr9:36,843,016-36,903,796 bp, removing exon 8 and 9 of PAX5.

SDC, Figure 12



SDC, Figure 12: Cross-validation of potential in-frame fusion genes with long-read sequencing in three BCP-ALL. Left: Optical map of the leukemia (blue) and respective reference map (green) indicating translocation and potential breakpoint in the overlapping gene. Split maps with aligned blue labels and unaligned yellow labels indicate translocations. Right: Continuous long-reads (CLR) aligning to the respective gene locus are shown (grey bars). Dashed lines indicate location of breakpoints. Split reads supporting the respective translocation are marked by purple arrows. Decreased coverage of aligned reads support a deletion (red bar), increased coverage of aligned reads support a duplication (blue bar). (A), (B) t(X;13)(q25;q13.1) PDS5B::STAG2 (C), (D) t(2;10)(q37.1;p14) SFMBT2::DGKD and (E), (F) t(3;5)(q28;q31.2) TPGR1::BRD8.

SDC, Figure 13



SDC, Figure13: OGM and RNAseq reveal novel fusion genes in *ETV6::RUNX1* and HD BCP-ALL. (A) Left: Circos plot showing somatic SVs of an HD BCP-ALL (ALL59) with a secondary translocation t(1;6)(q25.2;q12.2) detected by OGM. Right: Patient optical map (blue) partly aligning to reference chromosomes (green) 1 and 16. Potential breakpoints (pink line) overlap with *TDRD5* and *LPCAT2*. (C) Validation of *TDRD5::LPCAT2* expression in ALL59 by RNA-seq. Arriba output of RNA-sequencing data is shown indicating t(1;16)(q25.1;q12.2) translocation leading to in-frame fusion of *LPCAT2* exon1-14 to *TDRD5* exon 17. Coverage of the aligned reads is indicated. (D) Left: Circos plot showing somatic SVs of an *ETV6::RUNX1* BCP-ALL (ALL10) with secondary translocations, including t(7;12)(q21.2;p11.23) detected by OGM. Right: Patient optical map (blue) partly aligning to reference chromosomes (green) 7 and 12. Potential breakpoints (pink line) overlap with *PEX1* and *FGFR1OP2*. (E) Validation of *PEX1::FGFR1OP2* expression in ALL10 by RNA-seq. Arriba output of RNA-sequencing data is shown indicating t(1;16)(q25.1;q12.2) translocation leading to in-frame fusion of *PEX1* exon1-13 to *FGFR1OP2* exon 3-7. Coverage of the aligned reads is indicated.

3.4.2 Original Publication II: Supplementary data

Relevant results for the discussion of this thesis were included in the following section. Complete supplementary data can be found online at the publisher site (refer to supplemental digital content): <https://links.lww.com/HS/A453>

Supplementary Table S2 was adapted from [181] by filtering for (HD).

Supplementary Table S2.		
Karyotypes determined by chromosomal G-banding from initial diagnosis obtained during clinical diagnostics of 37 BCP-ALL samples.		
Chromosomal deletions and gains not detected in respective SNP-array analysis were not included for concordance analysis (i.e. comparison conventional (cyto)genetics vs OGM).		
Abbreviations: ID: Patient's identification; E::R: ETV6::RUNX1-translocated BCP-ALL; HD: classical hyperdiploid BCP-ALL; OGM: optical genome mapping		
ID	Subtype	Karyotypes
ALL31	HD	55,XX,+X,+4,+6,i(7)(q10),+9,+10+,11+14,+21,+21
ALL32	HD	58,XX,+X,+4,+6,+8,+10,+14,+17,+18,+21,+21,+21
ALL33	HD	46 XY [3]
ALL34	HD	56,XY,+X,+der(1;9)(q10;p10),+4,+6,+10,+14,+17,+18,+21,+21 [3]
ALL35	HD	56, XY, +X, +Y, +4, +6, +10, +14, +17, +18, +21, +21 [7];46 XY [10]
ALL36	HD	no metaphases
ALL37	HD	52-57,XX,+x,+4,+5,+6,+7,10,+14,+15,+17,+18,+21,+21,inc[cp5]/46,XX[15]
ALL38	HD	56 ,XY, +X,+Y, +4, +6,+10,+14,+18,+21,+21,+22 [2] 56 ,XY, +X,+Y, +4, +6,+10,+14,i(17)(q10),+18,+21,+21,+22 [2] 46, XY, [20]
ALL39	HD	55, XY, +X,+4,+6,+10+14,+17,+18,+21,+21 [5]/46,XY[17]
ALL40	HD	51~53, XX, +X, +4, +7,10, +14, +18, +21, +21[cp3]/46,XX[9]
ALL41	HD	51~53,XY,+X+Y+6,+14,add(19)(p13),+21,+21,+mar[cp7]/46,XY[8]
ALL42	HD	55,XY,+X,+4,+6,+10,+11,+14,+18,+21,+21[4]/46,XY[11]
ALL43	HD	no metaphases
ALL44	HD	55,XX,+X,+4,+6,+8,+10,+14,+17,+21,+21[5]/46,XX[10]
ALL45	HD	46,XY[10]
ALL47	HD	57,XX,+X,+X,dup(1)(q21q43),+4,+6,+8,+10,+14,+17,+18,+21,+21,inc[2]/46,XX[1]
ALL48	HD	46,XX[13]
ALL58	HD	61,XY,+X,+4,+5,+6,+8,+10,+11,+12,+14,+14,+16,+17,+18,+21,+21[12]/46,XY[3]
ALL59	HD	56~58,XY,+X,+4,+5,+6,+8,+10,+14,+7,17,+7,18,+21,+21,+mar[cp11]
ALL60	HD	57~58,XY,+X,dup(1)(q24q44),+3,+4,+6,+8,+10,+14,+14,+21,+21,+mar1,+mar2,inc[cp15]/57~58,idem, trp(1)(q24q44)[cp2]

Supplementary Table S4 was adapted from [181] by filtering for (n).

Supplementary Table S4.											
Overview on copy number alterations detected by SNP-array and concordance of OGM data in 60 BCP-ALL samples.											
Hg38 was used as reference genome. For aneuploidy events (X) coordinates and size were not assessed (NA).											
CNN-LOH analysis was not assessed in our study using OGM thus "NA" was given at "Detected by OGM".											
A yes (y) or no (n) indicates if the copy number alterations was detected by OGM.											
If detectable by manual control but not called an "n visible" is given.											
Abbreviations: ID: Patient's identification; E::R: ETV6::RUNX1-translocated BCP-ALL; HD: classical hyperdiploid BCP-ALL;											
CNN-LOH: copy number neutral loss of heterozygosity; Cyto: cytoband; CN: copy number; SNP: single nucleotide polymorphism											
ID	Subtype	Start Cyto	End Cyto	CN Type	Chr	Genomic Start	Genomic End	Size (bp)	Aneuploidy	Detected by OGM	
ALL10	E::R	2q31.3	2q37.1	GAIN	2	180,471,069	231,336,638	50865570		n	
ALL12	E::R	Yp11.2	Yq11.223	GAIN	Y	NA	NA	NA	X		
ALL15	E::R	13q21.33	13q22.1	LOSS	13	70,293,313	74,577,622	4284309		n visible (30%)	
ALL15	E::R	13q31.1	13q31.1	LOSS	13	82,675,257	84,517,201	1841944		n visible (30%)	
ALL15	E::R	Xq28	Xq28	GAIN	X	153,456,261	153,508,273	52012		n	
ALL16	E::R	16p12.1	16p12.1	LOSS	16	27,426,261	27,478,516	52255		n	
ALL16	E::R	Xq26.3	Xq26.3	GAIN	X	134,870,853	134,891,198	20345		n	
ALL17	E::R	Xq27.1	Xq27.1	LOSS	X	139,215,290	139,219,296	4006		n	
ALL19	E::R	4q13.2	4q13.2	GAIN	4	68,500,975	68,676,245	175270		n	
ALL20	E::R	12p12.1	12p11.22	LOSS	12	23,336,874	28,430,051	5093177		n visible (20%)	
ALL21	E::R	5p15.33	5p15.1	GAIN	5	113,462	15,469,536	15356074		n visible (15%)	
ALL21	E::R	13q31.3	13q31.3	LOSS	13	92,015,933	92,037,389	21456		n	
ALL22	E::R	8p22	8p22	LOSS	8	18,994,017	19,001,298	7281		n	
ALL22	E::R	22q11.21	22q11.21	LOSS	22	18,046,575	18,079,435	32860		n	
ALL22	E::R	Xq24	Xq24	LOSS	X	119,744,652	119,754,545	9893		n visible	
ALL22	E::R	Xq26.1	Xq26.1	LOSS	X	130,241,415	130,242,491	1076		n	
ALL22	E::R	Xq27.1	Xq27.1	LOSS	X	140,722,033	140,724,090	2057		n	
ALL51	HD	1q24.2	1q24.2	LOSS	1	167,706,471	167,735,965	29494		n	
ALL51	HD	4q12	4q12	GAIN	4	54,237,750	54,255,091	17341		n	
ALL51	HD	8q24.23	8q24.23	LOSS	8	136,665,653	136,850,192	184539		n	
ALL52	HD	3q13.33	3q13.33	LOSS	3	119,657,202	119,696,984	39782		n	
ALL6	E::R	Xq22.3	Xq22.3	GAIN	X	108,733,214	109,235,639	502426		n	

Supplementary Table S11.**Novel common region/genes in 60 *ETV6::RUNX1* and classical HD BCP-ALL.**

Abbreviations: DEL: deletion; TRANS: translocation; STOPGAIN: Stopgain mutation;

n samples: number of samples with genetic alterations in this region , *by at least one case;

CLR: continuous long-read sequencing (PacBio); n.a.: not available;

CBA: chromosomal banding analysis; SNP: single nucleotide polymorphism;

MAR: minimal altered region; HD: classical hyperdiploid BCP-ALL

Common region (genes)	<i>ETV6::RUNX1</i>		HHD		Minimal altered region (MAR, hg38)	Validation by orthogonal method*
	Type	n	Type	n		
9p21.3 (<i>FOCAD*</i> , <i>HACD4</i>)	Missense DEL	1 1	DEL	3	chr9:20965099-21001153	CLR, SNP-array
8p11.21 (<i>IKBKB</i>)	DEL	1	DEL	3	chr8:42308052-42337440	n.a.
Xp21.1	DEL DUP	2 1	DEL	1	chrX:33548160-33612296	n.a.
1p34.3 (<i>ZMYM1</i> , <i>ZMYM6</i>)	DEL	2	DEL	1	chr1:35029474-35100899	n.a.
4q24 (<i>MANBA</i>)	Missense DEL	1 1	DEL	1	chr4:102716969-102793507	CLR, SNP-array
8p23.1 (<i>MSRA</i>)	DEL	2	DUP	1	chr8:10056881-10097222	SNP-array
10p14 (<i>SFMBT2</i>)	TRANS DEL	2 1	DEL	1	-	CLR, SNP-array
9q21.11	DEL	2	DEL	1	chr9:67478734-67489803	n.a.
1q32.2	DEL	2	DUP	1	chr1:207366708-207396170	SNP-array
4q31.1	DEL	1	DEL	2	chr4:138546463-138553645	n.a.
6q25.3	DEL	2	DEL	1	chr6:156271736-156392444	SNP-array
12p13.1 (<i>GPRC5A</i>)	DEL TRANS	4 1			-	CLR, SNP-array
20q11.22 (<i>NCOA6</i>)	DEL	4			chr20:34793163-34828648	CLR
4q28.3	DEL	3			chr4:132260249-132268752	CLR
21q22.11	DEL	3			chr21:33935406-33950171	CLR
22q12.1	DEL	3			chr22:28800807-28812628	CLR
18q21.2	DEL	3			chr18:55584361-55831523	CLR
12q24.21 (<i>MED13L</i>)	TRANS STOPGAIN	2 1			-	CLR, CBA
18q11.2 (<i>MIB1</i>)	DEL	3			chr18:21727664-21759897	CLR

4. DISCUSSION

In this thesis, the application of OGM was leveraged to analyze two pediatric cancer patient cohorts in distinct contexts: (i) inherited cancer predisposition (Original Publication I), and (ii) somatic mutational landscapes of BCP-ALL cases (Original Publication II). Both publications demonstrate and discuss the potential of OGM as a new tool for the genomic analysis of structural variation and copy number alteration in patient samples, in addition to or interlinked with NGS-based approaches. Both Original Publications conclude that a multi-platform strategy utilizing OGM and NGS effectively captures a broader spectrum of germline or somatic genetic alterations [180, 181]. This strategy can uncover previously uncharacterized or inaccessible sites of the genome. Integration of novel technologies enabled the generation of genomic portraits underlying inherited cancer predisposition and pediatric tumorigenesis, while revealing candidate regions/genes and establishing a research platform for further functional and clinical investigations.

4.1 Utility of OGM in clinical routine testing

Original Publication I and II, along with other studies, highlight the advantage of applying OGM in both constitutional disease [39, 46] and hematological disorders [40, 45, 48, 52, 54, 182]. These include the high concordance of OGM and conventional methods and its added value as genetic follow-up tool (refer to section 4.3.1 and 4.5.1).

OGM is a DNA-based approach that requires no cell cultivation and directly utilizes the input of 1.5 million cells for DNA extraction. Unlike short-read NGS or classical cytogenetic approaches, OGM does not rely on DNA amplification or cell cultivation, thereby preventing false-positive results due to PCR-related bias or culturing artifact [41]. In addition, OGM detects SVs with a higher resolution compared to conventional karyotyping methods (>500bp vs ~5-10Mb) [41]. Compared with FISH and array-based methods, OGM is a non-targeted, single-assay approach that enables genome-wide detection of both balanced and unbalanced SVs. From a practical perspective, OGM is a cost- and time-effective assay that can be streamlined in routine clinical workflows, achieving results after approximately 4-6 days [40, 44]. The turnaround time is composed of hands-on laboratory experiments (2 days), instrument runtime (1-2 days) and automated data analysis (1-2 days). In contrast to NGS workflows, where costs per sample vary depending on genome coverage and batch size, OGM maintains a consistent per-sample cost [41]. For instance, the high sequencing depth required for the detection of subclonal mutations

in cancer diagnostics substantially increase the costs of NGS [183]. In contrast, OGM generates more than 300x genome coverage per run and potentially detect variants with a VAF as low as 5% [41], offering a feasible alternative in a diagnostic context. The analysis of OGM data is supported by automated pipelines, reducing the need for extensive bioinformatic expertise and can be implemented relatively easy compared to NGS-based methods [41].

Overall, the high reliability and detection sensitivity of OGM, combined with its streamlined lab and data analysis workflow, underline the novel technology as powerful tool in routine clinical diagnostics.

4.2 Limitations of OGM in clinical routine testing

OGM has several technical and biological limitations that need to be considered while evaluating its clinical utility. First, OGM requires HMW which needs to be extracted from fresh or frozen material. Standard DNA preparations commonly used for array-based methods or short-read sequencing are not suitable. However, obtaining good quality HMW DNA from clinical samples is often more challenging compared to requirements for standard methods.

Second, OGM has some limitations due to its non-sequencing approach, depending on the genomic distribution of the fluorescence marker recognition sites. Fluorescence labeling occurs at an average label density of 15 labels per 100kbp [40], resulting in a detection threshold of approximately 500bp and limited breakpoint resolution. Consequently, breakpoints detected by OGM are approximations and require validation by orthogonal methods. In the Original Publications presented in the thesis, PCR-based Sanger-Sequencing or NGS-based approaches were applied to validate and refine the OGM detected SVs [180, 181]. In Original Publication II, CLR sequencing was used to refine breakpoints of selected SVs, and RNA-sequencing confirmed potential fusion genes predicted by OGM (SDC Figure 7, 12, 13 [181]). Other studies have applied whole-genome short-read sequencing to close OGM gaps [184] or Cas9-directed long-read sequencing to resolve OGM findings [185]. Notably, the Cas9-directed approach confirmed all alterations and pinpointed the actual location of the breakpoint between 0.2 and 5.5kbp of the OGM labels [185]. In Original Publication I, clinically relevant deletions overlapping with the *TP53* locus were initially identified in two cases with suspected cancer predisposition [180]. Subsequent PCR-based Sanger sequencing demonstrated that the actual deletion affected intergenic region overlapping with benign polymorphism (Suppl. Figure 1 [180]). This example highlights the importance of precise breakpoint characterization, especially in clinically relevant genes such as *TP53*.

Third, the interpretation of OGM results is often limited by the lack comprehensive OGM reference databases. Extended internal and external databases of OGM detected SVs in healthy and disease samples would facilitate accurate SV interpretation. More studies are needed to improve SV classification, to filter common benign variants, and to assess their phenotypic impact. However, such databases remain scarce for OGM and other advanced technologies such as long-read sequencing [48, 59].

Other technical challenges of OGM involve the resolving of specific genomic regions, including highly repetitive telomere and centromere regions, and areas with insufficient label coverage. OGM detected SVs, especially in these regions, require manual review which demands training and specific expertise [42]. Even after manual re-evaluation, OGM missed several SVs in the cohort of Original Publication I and II that were previously detected with conventional methods [180, 181]. These included a constitutional ring chromosome 21 (Original Publication I, Suppl. Material and Methods [180]), centromeric translocations, gain of chromosome Y and focal variants with low cell fractions of 15-30% (Original Publication II, Suppl. Table S4 [181]). Other OGM studies have reported similar limitations, such as missed constitutional Robertsonian translocation [39], (peri-) centromeric rearrangements, and low VAF (~10%) SVs in hematological malignancies [40].

Finally, both Original Publications and other studies discuss that improvements of the Bionano reference database, as well as application of the T2T reference genome, may help to overcome these limitations [39, 49, 51, 53, 180, 181, 186, 187].

4.3 OGM and NGS-based approaches in constitutional disease

4.3.1 Clinical utility of OGM in constitutional disorders

In the field of constitutional disease diagnostics, Original Publication I together with other studies advocate the use of advanced methods such as OGM, particularly for cases that remain unresolved after conventional cytogenetic testing or short-read sequencing [39, 46, 180, 187]. The first significant large-scale application of OGM in a cohort of developmental disorders (n=85) by Mantere *et al.* demonstrated high concordance of OGM and standard methods while identifying or refining additional pathogenic SVs relevant for the genetic disease [39]. The study further highlighted that OGM had a higher resolution than karyotyping and could resolve regions refractory to NGS analysis [39]. A multi-center study of constitutional postnatal cases (n=359) confirmed the reliable detection of chromosomal aberrations with a high technical

concordance of 99.5% [187]. Schrauwen *et al.* analyzed families with developmental disorders and a negative WES result and found (likely) pathogenic SVs in 10.6% (5/47) of cases, using OGM [46]. Following a similar approach, Original Publication I identified (likely) pathogenic SVs in 6% (2/34) of pediatric cancer samples with a suspected CPS and a negative WES result [180]. Notably, integration of OGM and long-read sequencing resolved 100% (6/6) of cases with a suspected genetic disease where previous genetic testing remained negative [186]. Hiatt *et al.* reported an additional diagnostic yield of 7.3% (7/96 cases) when combining long-read sequencing with short-read data in a neurodevelopmental disorder cohort [188].

These studies propose OGM as a complementary or alternative approach in constitutional disorder diagnostics and further suggest that integration of OGM and long-read sequencing methods may be the most effective strategy to resolve unresolved genetic cases in the future.

4.3.2 Advantages of long-read sequencing

In comparison to OGM, long-read sequencing methods offer single-base resolution for both SVs and SNVs, eliminating the need for complementary methods to refine breakpoints. The long-reads enable phasing of SNVs and SVs even when parental samples are missing, distinguishing genetic variation on the same chromosome (*cis*) or on different chromosomes (*trans*) [189]. Phasing is particularly relevant for autosomal recessive disorders, as detection of variants in *trans* with a known second pathogenic alteration can provide moderate evidence for pathogenicity and support reclassification of a VUS in some cases [61]. For example, OGM *de novo* assembly can also provide haplotype-resolved SVs as shown in Original Publication I where compound heterozygous SVs in a leukemia patient with developmental delay was described (Suppl. Figure 3 [180]). The SVs, a deletion and insertion, potentially overlapped with regulatory elements of the *RAP1B* gene locus [180]; long-read sequencing could clarify the precise breakpoints and the sequence of the insertion and subsequent pathogenicity classification and functional impact. Haplotype phasing using long-reads has been successfully applied in clinical settings [190-192]. However, the phasing accuracy is dependent on the scope of the reference genome used which is currently underrepresented in many subpopulations and genetic ancestries [60, 193].

Further, a study by Noyes *et al.* demonstrated that familial long-read sequencing increased *de novo* mutation detection by over 25% compared to short-read sequencing [194]. *De novo* mutations play a significant role in the etiology of genetic disorders [195, 196], including pediatric cancer risk [27, 88, 89, 197, 198]. In Original Publication I, analyses of child-parent

trios revealed *de novo* SVs in 17% of the children, demonstrating the applicability of OGM in this context [180]. However, the additional yield of *de novo* SVs detected by OGM compared to short-read methods was not analyzed in this cohort.

Another advantage of long-read based approaches (Nanopore and PacBio technologies) is their ability to detect base modifications, such as CpG methylation, which is crucial for diagnosing imprinting disorders [199]. For example, haplotype-resolved HiFi sequencing combined with methylation profiling identified rare hypermethylation events resulting in allele-specific loss of regulatory element activity in a rare disease cohort [200].

The clinical utility of whole-genome long-read sequencing and potential increased diagnostic yield over short-read WGS in constitutional disease, including pediatric cancer predisposition, still remains uncertain and warrants further cohort screenings and clinical investigation.

4.3.3 Current limitations of advanced technologies in constitutional diagnostics

Despite major advances, several barriers remain for clinical implementation of advanced technologies, such as OGM and long-read sequencing. Clinical laboratories require extensive quality control and validation before adopting such technologies to ensure high specificity and minimize false-positive results [189].

Comprehensive genome-wide testing using these novel technologies will generate a larger number of genetic findings, many of which are difficult to interpret [201]. Pagnamenta *et al.* showed that WGS increased the diagnostic yield in a rare disease cohort, mainly due to SVs and non-coding alterations and interpretation of such is challenged by incomplete reference databases, sparse annotations, and insufficient understanding of the functional impact [202]. This missing information often leads to identification of VUS without knowing their actual impact on the diagnosed disease. In Original Publication I, the majority of identified rare SVs (74%) were located in non-coding regions, potentially affecting regulatory elements, yet their functional relevance remained unclear (Suppl. Table S3 [180]). A study by Chiang *et al.* showed a larger impact of SVs on gene expression compared to SNVs/indels and that 88.3% of those resided in non-coding regions enriched at CREs [15]. However, Collins *et al.* could show that SVs in 14,891 genomes of the general population are causing around one quarter of protein truncating events, and that non-coding SVs, mostly deletions of the full regulatory element, show only a modest selection [203]. Thus, there could be risk of overestimating the pathogenicity of SVs identified by novel technologies. To enable clinical adoption, continued

expansion of population databases, reference genomes and family segregation studies together with analytical strategies is crucial for accurately distinguishing pathogenic SVs and non-coding variants from benign polymorphisms and technical artifacts [201, 204]. Computational approaches could further help overcoming those limitations. For example, Cohen *et al.* integrated HiFi long-read sequencing with machine learning to analyze a pediatric rare disease cohort, showing that computational approaches are efficient for diagnostic SNV prioritization [205].

With decreasing costs, improved data analysis and interpretation, advanced technologies are expected to transition from follow-up genetic-test to standalone clinical methods in the future, proving a one-step workflow for comprehensive genetic testing thereby replacing current multi-step diagnostic pipelines [39, 189, 206].

4.4. Genetic germline variation in pediatric cancer predisposition

4.4.1 Genotyping approaches in pediatric cancer patients

Germline genetic predisposition is a significant risk factor in pediatric cancer, with approximately 8.5–12% of cases attributed to pathogenic variants in known cancer-predisposition genes [25, 26, 29-31]. Original Publication I highlights the increasing importance and complexity of germline testing in children with cancer, including individuals with a high suspicion for a cancer predisposition syndrome [26]. The cohort was selected based on clinical phenotypic criteria and a negative WES result, so that children with (likely) pathogenic SVs without overt clinical signs would have been missed. Several working groups recommend phenotype-agnostic genetic testing of all children with cancer to evaluate their risk for an underlying CPS and to establish guidelines to accurately inform diagnosis, therapy guidance, and support cancer prevention [26, 29, 30, 180, 207-209]. Multidimensional germline testing should include aspects such as (i) genetic testing of all affected individuals and if possible their parents to assess inherited vs *de novo* events [128][197] as well as (ii) comprehensive tumor-germline testing of SNVs/indels, SVs and epigenetic considerations.

A multi-center trial identified a pathogenic variant in a CPG in 75% of pediatric patients with a rare cancer entity and could be newly diagnosed due to study participation [207]. This study concluded that germline testing can identify the first patient of a hereditary cancer family through a genotyping approach. The advantage of parent-child sequencing is the systematic testing of *de novo* variation in pediatric cancer cohorts [89], that could bring novel insights into

novel genes and mechanisms in cancer predisposition as well as relevant decision making for families with potential hereditary cancer. Newman *et al.* emphasized the relevance of paired tumor-normal testing, showing that 55% of pathogenic germline variants contribute to tumorigenesis in a pediatric cancer cohort, particularly when combined with RNA sequencing and WGS analyses to investigate impacts on splicing and mutational signatures [210]. Further, emerging evidence suggests an underestimation of the contribution of genomic cancer predisposition due to limitations of conventional testing that primarily focus on SNVs/indels or predefined gene-panels (refer to section 4.4.2). Increasing genetic testing of children and their families requires extended research to evaluate the consequences of CPS identification and impact on treatment, patient care and tumor surveillance [208, 209].

4.4.2 Contribution of structural variation and multi-omics profiling

Recent advances in SV detection and systematic WGS studies, including Original Publication I, support that large genomic alteration contribute to additional risk mechanisms in pediatric cancer patients [59, 60, 180, 210-212].

The importance of the contribution of germline SVs to various genetic diseases have been previously described. For example, large-scale studies reported additional impacts on diagnostic yield from 8 to 30% in patients with genetic disease, mainly due to SVs and deep intronic variants by applying WGS [92, 204]. Notably, two studies screened germline SVs in distinct cancer cohort by applying long-read sequencing, showing that pathogenic SVs can be found in ~1.5% of cases in an adult cancer cohort [59] or in suspected inherited breast cancer predisposition [60]. Studies are increasingly investigating the contribution of SVs in pediatric cancer cohorts, including Original Publication I that identified a potential cancer predisposing SV in 6% (2/34) of pediatric cases, however in a relatively small and preselected cohort of clinically suspicious pediatric cases [180]. By combining OGM and long-read sequencing, Sabatella *et al.* identified a constitutional retrotransposon insertion in *SMARCB1*, functionally disrupting its splicing, that could explain the atypical teratoid rhabdoid tumors diagnosed in two siblings [213]. To understand the mechanistic role of germline SVs on tumorigenesis, studies are increasingly applying multi-omics profiling [211, 212, 214, 215]. Gillani *et al.* showed that rare germline SVs, affecting not only mutationally intolerant genes but also tissue-relevant 3D chromatin structures, increase the risk for pediatric solid tumors [214]. The Pan-Cancer Analysis of Whole-genomes consortium integrated whole-genome and transcriptome data to identify germline SVs impacting genes that might contribute to cancer [215]. Wong *et al.*

analyzed 245 high-risk pediatric patients using integration of genome, transcriptome and mutational signature analysis, reporting a higher incidence of pathogenic germline alterations in 16.2% of cases [212] compared to other pediatric cohorts [26, 29]. Chen *et al.* recently combined genotyping of germline SVs with tumor transcriptomes and global DNA methylation of corresponding pediatric brain tumors, showing that germline SVs modulate cancer predisposition and progression through epigenetic dysregulation [211]. Long-read sequencing is able to detect whole-genome variation and methylome changes in one single assay, stressing that this application could be a valuable research tool to study SV-driven risk mechanisms in pediatric oncology.

Differences in SV occurrence in pediatric cancer cohorts may result from combined analysis of germline and somatic WGS, cancer subtype constellations, risk profiles and patient ethnic backgrounds. Extended studies and multi-platform approaches, including whole-genome sequencing, will be essential to uncover the full spectrum of SVs and genetic risk mechanisms in hereditary cancer susceptibility. For reliable comparison across studies and clinical implementation, the application of harmonized bioinformatic pipelines, standardized variant detection and classification, including SVs and non-coding alterations will be crucial.

4.4.3 Adult-onset CPGs and pediatric cancer

Recent studies increasingly suggest a potential association between heterozygous pathogenic variants in adult-onset CPGs and pediatric cancer risk [28, 216], including genes classically linked to hereditary breast and ovarian cancer (HBOC) [217, 218]. Most of these studies considered only small variations, while the contribution of SVs remains underexplored. Comprehensive testing of SVs and SNVs could potentially expand the understanding of adult-onset CPGs and risk association in childhood malignancies.

Original Publication I identified a heterozygous likely pathogenic deletion in the DNA-damage repair regulator *BRCA2* that segregated in a family with breast cancer and a child with leukemia [180]. *BRCA2* is grouped among HBOC genes, but is additionally frequently altered in pediatric cancer patients [29, 71]. Further, pathogenic variants in adult-onset CPGs and HBOC genes are associated with an increased risk for secondary malignancies [217-220], warranting further research that may impact clinical surveillance for affected individuals, including the child with *BRCA2* deletion [180].

Bi-allelic mutations in *BRCA2* are well established as causative of the CPS called Fanconi anemia with an increased risk of leukemia [221]. In contrast, the association of pediatric cancer,

including hematological disease, and heterozygous carriers of pathogenic variants in *BRCA2* remains debated [222-224]. In a pediatric cancer cohort, Stoltze *et al.* systematically analyzed mutational constraint that reflects the depletion of genetic variation due to negative selection and acting against deleterious mutation [225]. Childhood mortality and genetic cancer risk should be linked to this evolutionary theory [27]. However, analyzing monoallelic loss-of-function variants (only SNVs) of classically recessive childhood cancer risk genes, the authors found no evidence of constraint for monoallelic *BRCA2* variants [27], suggesting no or a potential more complex relationship in pediatric cancers. Recent recommendations discourage the reporting of heterozygous pathogenic variants in recessive disease [208]. However, heterozygous variation combined with additional genetic layers, including SVs and haplotypic changes, remains insufficiently studied. Comprehensive whole-genome sequencing coupled with functional testing may help clarify the contribution of adult on-set CPGs, such as *BRCA2*, to pediatric cancer predisposition.

4.4.4 Candidate cancer predisposition genes and functional characterization

Comprehensive germline testing of pediatric cancer patients, using OGM and long-read technologies, will lead to more genetic findings including novel candidate CPGs.

Consistent with that, Original Publication I identified a *de novo* duplication in *RPA1*, a gene involved in DNA replication and DNA repair gene [226], as a potentially new CPG in a leukemia patient (Suppl. Figure 4 [180]). Interestingly, gain-of-function germline variants in *RPA1* has recently been implicated in germline predisposition of a short telomere syndrome, including hematopoietic malfunctions [227]. Functional studies of the *RPA1* mutation in induced pluripotent stem cells (iPSCs) exhibited telomere shortening and defects in hematopoietic differentiation [227]. In addition, a systematic study reported rare heterozygous *RPA1* mutations in 1% (75/5993) of children with cancer [228], supporting its role as candidate CPG and emphasizing the need for further functional investigation. In Original Publication I, functional assays on patient fibroblast cells carrying the germline *RPA1* duplication revealed increased sensitivity to DNA-damaging reagents, but no telomere shortening (Suppl. Figure 5 [180]). Functional modeling by introducing relevant novel genetic aberrations, such as *BRCA2* and *RPA1* SVs detected in Original Publication I, into iPSCs provides a valuable tool to delineate the potential predisposing role of germline variation of childhood malignancies in a cell-type specific context [229]. An iPSC model of patient-derived cells of a Li-Fraumeni patient could recapitulate features of the diagnosed osteosarcoma and identify dysregulated genes with

clinical implications [230]. Babbe *et al.* generated patient-derived iPSCs, carrying a germline *PTPN11* mutation, which could recapitulate the hematopoietic abnormalities of the diagnosed juvenile myelomonocytic leukemia and increased cell proliferation [231]. Importantly, this method was also used for drug-screening of differentiated myeloid cells in for deciphering new treatment options.

Functional characterization could be also relevant for SVs affecting the non-coding candidate regions overlapping with regulatory elements identified in Original Publication I [180], as CREs especially function in a tissue and cell-type-specific manner [232-234]. Kellman *et al.* identified cancer-associated germline SNVs by integrating reporter, expression and chromatin accessibility assays with CRISPR knockout screenings, thereby linking germline variants to regulatory mechanisms in cancer [235]. Further, the role of non-coding mutations in leukemogenesis has been shown by Yang *et al.* showing that germline non-coding *GATA3* mutation leads to enhancer activation and reshaped chromatin accessibility and 3D genome organization in ALL [236].

Both Original Publications showed that applying integrative approaches have the potential to uncover novel CPGs and predisposing mechanisms in the future. This will include the detection of novel so far underrepresented genomic alterations like SVs, *de novo* variation as well as regulatory perturbations. Comprehensive genetic studies and multi-omics testing of pediatric cancer cohorts are expected to expand the currently acknowledged contribution of inherited cancer predisposition.

4.5. OGM and NGS-based approaches in hematological malignancies

4.5.1 Clinical utility of OGM in hematological malignancies

Due to emerging novel subtype-defining genetic characteristics established by WHO [237] and ICC recommendations [238] for accurate diagnosis, risk assessment and treatment selection of hematological neoplasms, multiple complementary diagnostic assays are often required to capture the full spectrum of alterations. Not yet standardized, the application of advanced diagnostic tools includes transcriptome sequencing, short-read whole-genome sequencing and OGM, which offer insights into expression profiles as well as structural and sequence-level genomic alterations [126]. Up to date, OGM has been extensively adopted in diagnostic laboratories to identify genomic alteration in hematological malignancies thereby replacing standard of care methods [40, 45, 48, 52, 54, 182].

The first clinical adoption of OGM as a DNA-based single assay was tested in large-scale studies of distinct hematological neoplasms [40] including AML and myelodysplastic syndromes (MDS) patients [45, 50, 52, 182]. Interestingly, Original Publication II and these studies demonstrated a high concordance of OGM and conventional cytogenetic methods, such as FISH and CBA, ranging from 93-100% [40, 45, 50, 52, 181]. As stand-alone comparison, Original Publication II analyzed the concordance of OGM and SNP-array, revealing a higher sensitivity of OGM for the detection of submicroscopic deletions [181]. Further, all studies revealed additional SVs, partly relevant for the refinement of risk stratification, underlining the advantage of OGM compared to conventional cytogenetics in the diagnostic work-up of hematological cancer genomes [40, 45, 50, 52, 181]. Original Publication II detected novel cytogenetic markers and clinically actionable results including co-occurring deletions affecting *IKZF1* and *PAX5* in one BCP-ALL patient (SDC Figure 10, [181], potentially defining a HR prognostic subgroup in BCP-ALL with adverse outcomes [126, 239, 240]. A systematically conducted OGM study by Lühmann *et al.* categorized 142 BCP-ALL cases into *IKZF1*^{del} and *IKZF1*^{plus} subgroups, where OGM led to reclassification due to previously missed cryptic deletions [241], highlighting the advantage of OGM to provide broader insights into prognostic subgroup classification in pediatric BCP-ALL. Of note, E::R and hyperdiploid BCP-ALL were more common in the *IKZF1*^{del} group [241].

The broad clinical application of OGM has led to the formation of international frameworks as the International Consortium for OGM which established guidelines to standardize the use across laboratories [42, 51], thereby discussing clinical indications, validation and interpretation guidelines. OGM was supported as first-line test for pediatric leukemia such as ALL, AML and BCP-ALL for diagnosis and relapsed samples [51]. Original Publication II is a practical demonstration of the application of these guidelines in BCP-ALL samples. OGM was able to detect all subtype-defining lesions, E::R translocation and hyperdiploidy, of the two most common BCP-ALL subtypes [181]. For characterization of the most common HD BCP-ALL subtype, conventional karyotyping often fails as the hyperdiploid leukemia cells grow slowly due to mitotic arrest, which leads to missing chromosomal metaphases that are required for routine karyotyping [128]. OGM as DNA-based approach can overcome these limitations. Exemplarily, the aneuploid profiles of 30 analyzed HD BCP-ALL cases from Original Publication II could be resolved by OGM, even in 10% (2/20) of cases where CBA failed due to missing or low number of metaphases (Suppl. Table S2 [181]). The application of OGM further identified poor-risk profiles in 20% (6/30) of HD BCP-ALL cases, based on specific copy number profiles that were established by Enshaei *et al.* [133]. However, there is a proportion of ~9% of BCP-ALL

subtypes where WES or transcriptome analysis is needed for subtype identification [99], which would be missed by OGM. In addition, a significant amount of BCP-ALL, called B-other, remains without a defining genetic lesion after conventional diagnostic testing [67, 130]. Exemplarily, a recent study by Hidalgo-Gomez *et al.* showed that seven distinct subgroups can be identified within the B-other subgroup when characterizing them using OGM [242]. Further, current detection threshold of ~5% VAF and non-targeted use restricts the application of OGM for MRD testing in hematological malignancies that is currently tested with short-read sequencing approaches [175, 243].

All contributing OGM studies highlight the applicability of OGM for reliable detection of relevant genomic alteration for diagnosis of hematological malignancies, including pediatric BCP-ALL. Comprehensive diagnostic screening will reveal detailed genomic profiles, refined risk stratification and treatment-relevant alterations in the future.

4.5.2 Strategies of OGM data integration and analysis

Several studies, including Original Publication II, highlight the clinical value of integrating OGM and NGS-based methods [50, 52, 181]. Sahajpal *et al.* used the integration of OGM and NGS-panels as clinical work-up for myeloid leukemia, leading to improved clinical management in 20% (6/30) of myeloid leukemia cases [50]. Similarly, a large-scale study of 101 MDS cases achieved an improved prognostication in 17% (17/101) of cases [52]. As a novel integrative approach, Original Publication II combined short-read WES and OGM data to identify recurrently altered regions and gene targets in diagnostic BCP-ALL cases [181]. This approach identified minimal altered regions based on the smallest common overlap region of OGM detected SVs (Figure 9). Minimal overlapping region can be used to indicate key genes involved in the biology and outcome of pediatric leukemia. For example, Bokemeyer *et al.* showed that recurrent *NR3C1* deletions are associated with poor response to induction treatment in relapsed BCP-ALL [173]. With this strategy, Original Publication II further identified novel potential tumor suppressor genes by integrating somatic mutations based on WES in order to identify double hits (SV and SNV/indel) affecting the same target gene [181].

The OGM frameworks for standardized laboratory usage recommend the application of the bioinformatic analysis using the RVP for cancer genomes [42, 51]. In contrast, Original Publication II applied both downstream pipelines to check for performance and analyze the most comprehensive SV landscape, including smaller (>500bp, DN) and larger subclonal (>5kbp, RVP) alterations, in the pediatric BCP-ALL samples (SDC Figure 2 [181]).

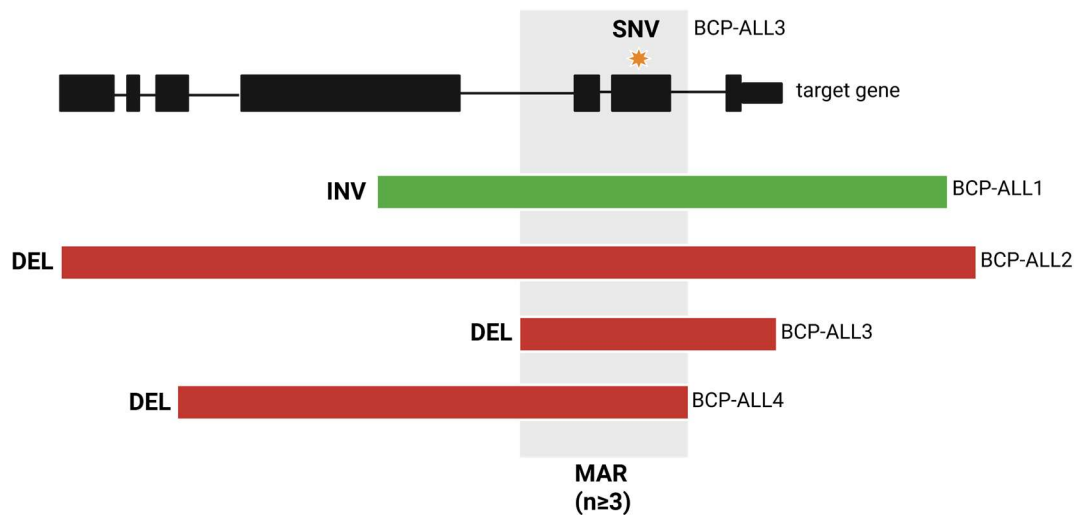


Figure 9: Integration of OGM and NGS-based methods to identify recurrently altered genes and double hits in potential tumor suppressor genes.

The identification of smallest overlap regions in multiple samples with the same disease narrow down candidate regions to understand the functional consequences of the SVs. Minimal altered regions (grey) are determined based on boundaries of OGM detected structural variants as shown in the exemplarily figure with deletions (red) and an inversion (green) occurring in at least three of the leukemia cases (BCP-ALL1-4). SNV variants (orange) in the target gene affected by the MAR are integrated to determine recurrently altered genes as well as double hits when SV and SNV occur in the same sample (here BCP-ALL3). Abbreviations: MAR=minimal altered region; BCP-ALL= B-cell precursor acute lymphoblastic leukemia; INV= inversion; DEL= deletion. Figure was created in <https://Biorender.de>.

Using the combinational approach in the BCP-ALL cohort of Original Publication II (Suppl. Figure 1 [181]) revealed overlapping SV detection in around half of the somatic SVs (47.6%, 416/874). Notably, 7.9% of somatic SVs (69/874) would have been missed by the RVP alone, including one subtype defining t(12;21) translocation and the deletion of the second ETV6 allele in two cases [181]. The DN pipeline also enables the detection of loss-of-heterozygosity (LOH) due to assembly of haplotype-specific maps [42]. This is particularly relevant in aneuploid BCP-ALL diagnosis and risk stratification, where LOH detection identifies high-risk cases. Masked hypodiploidy, resulting from endoreduplication of the haploid genome, could potentially lead to misclassification in a lower hyperdiploid risk group and treatment failure, when LOH detection is not considered [244]. The clinical use and sensitivity of OGM classifying aneuploid BCP-ALL should be tested in the future. Further, systematic testing of both pipelines increases the computational analysis time from one to three days, but would potentially improve the diagnostic yield and risk stratification of a broader BCP-ALL spectrum.

4.5.3 Long-read sequencing in hematological malignancies and other cancers

In addition to OGM, long-read sequencing is emerging as novel tool with diagnostic potential in hematological malignancies [245, 246]. Geyer *et al.* showed that nanopore long-read sequencing with adaptive sampling reliably defined 57 pediatric ALL cases, while reducing both cost and turnaround time [246]. The largest somatic long-read dataset to date was established by the Personalized OncoGenomics consortium, which analyzed 189 adult solid tumors with nanopore long-read sequencing [247]. This effort illustrated the ability of long-reads to resolve complex SVs, viral integrations and extrachromosomal DNA [247]. Phased methylation analysis provided by the long-read technologies further allowed detection of differentially methylated regions and allele-specific inactivation of cancer hallmark genes, which was driving homologous recombination deficiency without coding mutations [247]. Aberrant DNA methylation is a hallmark of pediatric ALL, where methylation profiles often correlate with genetic subtypes [248] and clinical outcomes [249, 250]. Novel integrational approaches that combine (epi)-genomes, transcriptome and methylome analysis may therefore accelerate the identification of driver genes, tailor therapies and resistance mechanisms [251]. For example, a multi-platform approach of whole-genome, transcriptome and methylome applied by Wong *et al.* identified actionable targets in 71% of cases of a high-risk pediatric cancer cohort [212].

Despite these initial advances, several barriers limit the widespread adoption of whole-genome long-read sequencing in clinical tumor diagnostics, including high costs compared to short-read sequencing, clinical validation, the requirement for deep coverage to confidently call somatic variants and the need for specialized data expertise. As costs continue to decrease and bioinformatic pipelines mature, clinical use of long-read sequencing is expected to expand in the future.

4.6. Contribution of structural variation in somatic landscapes of pediatric ALL

Somatic tumor profiling has led to tailored targeted therapy approaches in various pediatric cancers [71, 252, 253]. The somatic landscape of pediatric BCP-ALL is characterized by a low overall mutational burden but a remarkable genomic diversity, including SVs as central drivers of leukemogenesis, therapy response and relapse [119].

Comprehensive studies, including Original Publication II and others, consistently showed that focal SVs represent the key genetic alterations in pediatric ALL [99, 181]. Advances in WGS technologies have enhanced the detection of these focal events and other SVs, improving the understanding of subtype specificity and clinical significance [99, 254-256]. However, most of

these studies focused primarily on SNVs and focal deletions to identify driver and potential target genes, whereas the approach of Original Publication II captured and analyzed other alterations, such as inversions and translocations, alongside the integration of SNV profiles for target discovery in BCP-ALL. This resulted in the detection of 19 novel recurrent MAR that affected potential novel target genes, including *IKBKB* and *MIB1*, and potential tumor suppressor genes, including *STAG2* and *MANBA* [181]. Together, these findings highlight the added value of OGM profiling for uncovering novel candidate genes that can be functionally investigated for their role in leukemogenesis and as potential therapeutic targets in pediatric BCP-ALL.

4.6.1 Recurrent alterations and mechanistic insights

Recurrent alterations in BCP-ALL are frequently targeting genes involved in RNA processing, lymphoid development, cell-cycle control, and epigenetic regulation [99]. Focal deletions arise through mechanisms such as aberrant recombination by off-target RAG activity, predominantly in the *E::R*-positive subtype [169]. Original Publication II confirmed focal deletions are the most common type of alteration in *E::R* BCP-ALL, though without analyzing the mechanistic basis of SV formation [181]. Because OGM lacks nucleotide-level resolution, potential RAG recognition motifs flanking the identified SVs could not be analyzed, but could be integrated with long-read data as follow-up experiment. Mapping RAG off-target-mediated events could give insights on biological mechanism of novel candidate SVs in leukemogenesis, as RAG activity is an early event in B-cell development [257]. A recent systematic analysis of RAG-mediated recurrent alterations in 1616 pediatric cancer cases established a comprehensive resource of such sites for further functional studies, further showing that RAG-driven hotspots are specific to pediatric ALL [258].

Original Publication II identified recurrent MARs in *E::R* translocated BCP-ALL cases that affected genes involved in protein modification pathways such as SUMOylation (*UBA2*) and ubiquitination (*MIB1*, *UHRF1*) [181]. Increasing studies support that specific E3 ubiquitin ligases play a role in B-cell development and malignancy [259-261]. SUMOylation has been implicated as a predisposing factor in *E::R* leukemia [262], as a potential target in AML [263] and as a modulator of dexamethasone sensitivity in ALL [264]. HypoSUMOylation thereby increased the sensitivity in leukemic cells through altered chromatin openness [264]. Together with other genomic studies [99, 181], the recurrence of mutations in protein modification genes implicate

post-translational modification pathways as potential mechanistic drivers for leukemia development.

Further, the novel candidate target *IKBKB*, identified in Original Publication II, is a I κ B kinase that plays a central role in the canonical NF- κ B signaling pathway, regulating inflammation, cell survival and proliferation [265]. Interestingly, it has been shown that NF- κ B activation contributes to myeloid and lymphoid leukemogenesis in *BCR::ABL1* expressing cells, indicating the pathway as potential pharmacological targets for the fusion-driven ALL that is characterized by a poor prognosis and the need for intensive therapy [266, 267]. However, how the recurrent alteration contributes to other types of pediatric BCP-ALL still needs to be elucidated.

Overall, the contribution of these identified pathways to pediatric ALL biology and their potential target exploitation remains unclear and supports the need for further investigations.

4.6.2 Relapse-associated alterations

Although *E::R* and HD BCP-ALL have a favorable prognosis, they still make up around ~20% of relapsed cases, warranting further insights into somatic landscapes and risk mechanisms of relapsed cases, which often initially are diagnosed as SR patients [99]. In Original Publication II, relapse-associated focal deletions were detected in genes involved in B-cell development (*CDKN2A*[268]) or glucocorticoid signaling pathway (*NR3C1*[269]) [181]. Additionally, two relapsed cases carried deletions in *STAG2* [181] which is involved in the cohesin complex, regulating chromosome segregation, DNA damage repair and gene expression [270]. Recurrent alterations in epigenetic modifiers have been observed in large-scale cohorts of initial and relapsed BCP-ALL [99, 256]. The recurrency of *STAG2* alterations, the functional fusion of two cohesin complex members and double-hit analysis [181] further points towards an important role of the cohesion complex in B-ALL biology. Generally, it has been shown that genome architecture and histone modification play a critical role in cell fate decision and dynamic gene expression during lineage differentiation and commitment in hematopoietic development [271, 272]. Functional studies by Viny *et al.* further revealed that loss of Stag2 enhances stem cell self-renewal, impairs differentiation and is involved in regulatory loop interaction of lineage genes [273].

OGM identified relapse-specific alterations in one *E::R* BCP-ALL, including a translocation involving the histone demethylase *KDM6A* (SDC Figure 5 [181]), which has previously shown to be enriched in relapsed ALL among other epigenetic regulators [274]. *KDM6A* and has been

implicated in several treatment resistance mechanisms in AML [275, 276], but the role in pediatric BCP-ALL still needs to be elucidated.

Recurrent deletions and the novel expressed fusion genes identified in Original Publication II provide an opportunity to functionally investigate genomic architecture perturbation in relapsed ALL, modeling their role in leukemogenesis, therapy resistance and relapse.

4.6.3 Non-coding variation and functional characterization

Intergenic regions are recurrently altered in BCP-ALL with Original Publication II detecting 19 novel recurrently altered regions with around 50% (9/19) affecting non-coding regions (Suppl. Table S11 [181]). In ALL, non-coding and intergenic SVs are emerging as functionally significant alterations, disrupting CREs or regulatory loops that lead to oncogenic gene dysregulation [277-280]. This suggests, that SVs outside of coding regions can contribute to leukemia development, highlighting the importance of functionally investigating such events for potential driver event and target detection.

Particularly in T-ALL, a recent large-scale genomic study by Polonen *et al.* redefined the genomic landscape by integration of genomic and transcriptome data, showing that more than half of the driving genetic alterations are non-coding changes [280]. These changes involve common enhancer perturbation events such as enhancer hijacking or creation of neo-enhancers that grouped T-ALL cases into distinct subtypes with distinct treatment outcomes [280]. In BCP-ALL, Yang *et al.* could show that focal deletions in 13q12.2 led to loss of TAD boundaries and *FLT3* upregulation through enhancer hijacking in [277]. Yang *et al.* showed that chromatin architectural changes, including TAD disruption, also play a role in the biology of pediatric hyperdiploid BCP-ALL [138]. Genome architecture analysis of pediatric ALL cases further showed that fusion genes, aneuploidy and SVs can remodel the 3D genome, affect chromatin compartments and enhancer-promotor loops, thereby dysregulating driver genes such as *FLT3*, *TP53* and *IKZF1* [279]. Since chromosomal alterations can weaken TAD boundaries and lead to widespread genome dysregulation, analyzing the genomic architecture in other poor prognostic aneuploid BCP-ALL, such as hyperhaploid and hypodiploid cases, could give valuable insights on target discovery.

Evaluating the biological consequences of recurrent non-coding alterations identified in Original Publication II, requires the integration of transcriptome data and chromosome confirmation capture assays. Functional perturbation approaches provide a strategy to assess the perturbations of transcription and genomic 3D structure of regulatory DNA variants and

their role leukemogenesis. Li *et al.* used targeted sequencing of CREs, relevant for hematopoietic lineage development, to detect recurrently mutated regions that were further characterized by CRISPR-Cas9-mediated perturbation screening in leukemia samples [281]. Discovery of novel therapeutic targets could be especially relevant for BCP-ALL subtypes with high-risk features, such as *KMT2A*-translocated and hypodiploid leukemia, where novel blinatumomab treatment approaches generally improved disease-free survival compared to chemotherapy alone, but still is poor compared to other BCP-ALL subgroups [179].

Multi-platform studies increasingly underscore the importance of integrative analysis. Original publication II used transcriptomic and CLR data for validation and refinement of the OGM-detected results [181], but follow-up integration of transcriptomic and long-read based methylation analysis could give more insights on detected candidate targets. CLR is based on larger library insert sizes and may be therefore more effective compared to HiFi sequencing to detect larger SVs identified by OGM. However, CLR-sequencing generally has a higher error-rate. For comprehensive and reliably detection of haplotype-resolved alterations and genome assembly, HiFi sequencing may represent the preferential method for follow-up experiments [282]. The Genomes for Kids study used WGS and RNA sequencing to identify activating gene fusions and enhancer hijacking in pediatric cancer patients [210]. The combination of non-coding mutations together with RNA sequencing and assessment of accessible chromatin sites identified 128 mutational hotspots with differential gene expression of nearby genes in 345 pediatric ALL cases [278].

The work presented in this thesis together with other studies, highlights the pivotal role of SVs and non-coding variation in shaping the somatic landscapes of pediatric ALL. Detailed future research, using long-read sequencing, epigenomic profiling and functional modeling, will be essential to fully delineate their contributions to ALL pathogenesis and clinical outcome.

5. CONCLUSION AND OUTLOOK

The Original Publications presented in this thesis contribute to the development of a genomic platform for the identification and annotation of structural variation in both germline and somatic genomes of children with cancer. Original Publication I and II demonstrate the advantage of integrating advanced genomic technologies such as OGM with NGS-based approaches to offer novel insights into pediatric cancer predisposition and mechanisms underlying pediatric BCP-ALL, thereby bridging gaps in clinical diagnostics [180, 181].

This work provides insights on genetic backgrounds and novel gene targets in children with cancer establishing a foundation for future functional studies and clinical investigations. As whole-genome approaches become faster, affordable and clinically accessible, it is crucial to develop standardized guidelines for informed genetic results that ensure appropriate clinical implementation together with ethical considerations for the pediatric patients and their families [283, 284].

Future perspectives on decoding germline and somatic genomics should include the improvement of genomic reconstruction and SV detection by using advanced references genomes such as the T2T reference [10] and pangenome graphs [285] as well as long-read sequencing [286]. While long-reads, *de novo* assemblies, T2T reference and pangenomes improve variant detection in newly accessible repetitive regions and complex variants as well as considering haplotypes of diverse ancestry, benchmarking of such alterations is a future task that is especially important in medically relevant regions [286, 287]. For clinical adoption, novel reference genomes, such as T2T or pangenomes, have to overcome some hurdles to be adopted in research and clinical settings. These limitations include the need for extensive re-annotation of new genomic areas, which are conservatively established for older references (GRCH37/38) and the need for high computational power and storage capacity [286]. Further challenges represent the interpretation of the large number of VUS that would be discovered [286]. Emerging studies applying long-read sequencing in population cohorts or haplotype-resolved assemblies provide valuable foundations for novel reference and variant annotation tools [288, 289], supporting clinical interpretation. For the clinical adoption of advanced genomic methods, including long-read sequencing, standardized ACMG criteria need to be established for comprehensive SV detection, that are mainly established for SNVs/indels [61] and copy-number alterations [64]. Extended studies on frequency and functionality of SVs in

coding and non-coding regions in the context of health and disease will lead to novel curated databases and an improved annotation for clinical interpretation.

Despite significant progress with advanced sequencing methods and improved bioinformatic processing in recent years, the complete (non)-coding genetic landscape together with regulatory networks remains to be fully explored. Increasing multi-omics data integration from both germline and tumor samples will provide a broader understanding of how germline variation impacts pediatric cancer risk and development. When combining clinical information, such multi-factorial approaches will generate vast and complex datasets that can be leveraged for the application of artificial intelligence (AI) in order to construct comprehensive risk profiles and reveal novel disease associations [290]. However, the robust integration of machine learning tools into health care practices requires access to consistent and harmonized data, together with multi-disciplinary collaborations and technical strategies, to ensure both effective and ethical implementation [291].

By providing comprehensive insights into the genetic backgrounds of children with cancer, it may shape a future with improved cancer diagnosis, refined risk assessment and targeted therapeutic interventions in pediatric oncology (Figure 10).

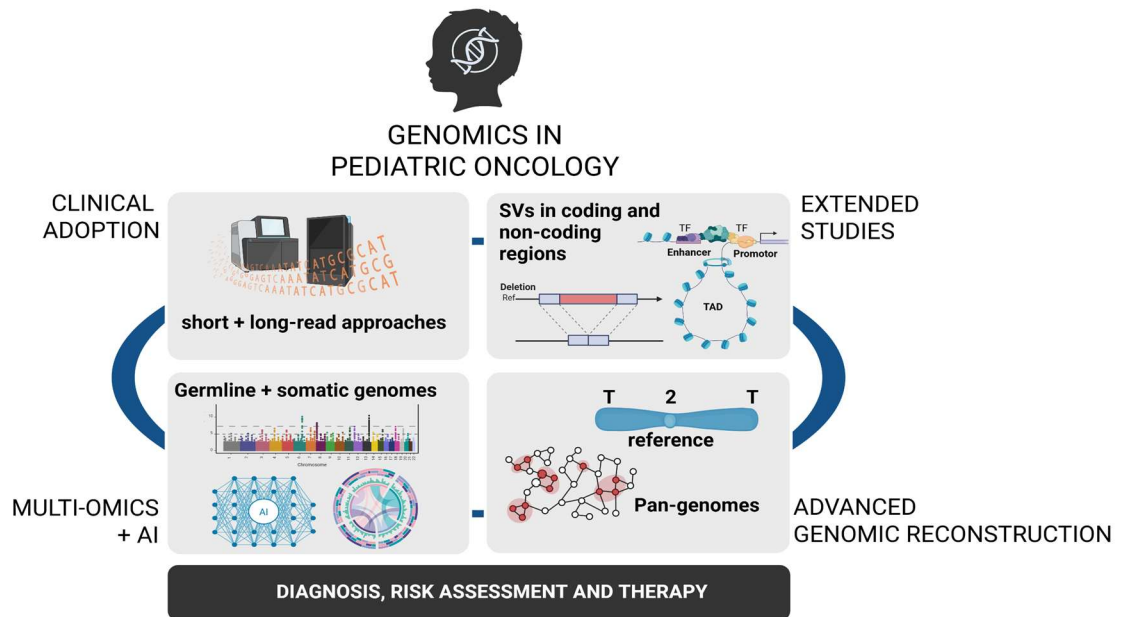


Figure 10: Future era of genomics in pediatric oncology.

The illustration outlines the role of multi-faceted genomics to enhance diagnostic precision, risk assessment and tailored therapeutic strategies for children with cancer. The integration of short- and long-read technologies enables improved resolution of small variants and complex SVs across coding and non-coding regions, including regulatory elements such as enhancers, promoters or TADs. Broader clinical implementation and extended research studies will help to understand their contribution to cancer predisposition and their oncogenic potential. Comprehensive profiling of germline and somatic genomes combined with multi-omics data and AI-driven analytics will facilitate the detection of disease-driving alteration and predictive biomarkers. The transition to gap-less T2T reference genomes and population-representative pan-genomes will enable a more accurate genomic reconstruction and the discovery of novel variation relevant to pediatric cancer susceptibility. Abbreviations: telomere-2-telomere (T2T); structural variants (SVs); transcription factor (TF); artificial intelligence (AI); topologically associated domain (TAD). Figure was created in <https://Biorender.de>.

REFERENCES

1. Schueler, M.G. and B.A. Sullivan, *Structural and functional dynamics of human centromeric chromatin*. Annu Rev Genomics Hum Genet, 2006. **7**: p. 301-13.
2. O'Sullivan, R.J. and J. Karlseder, *Telomeres: protecting chromosomes against genome instability*. Nat Rev Mol Cell Biol, 2010. **11**(3): p. 171-81.
3. Chen, Y., et al., *Chromatin accessibility: biological functions, molecular mechanisms and therapeutic application*. Signal Transduct Target Ther, 2024. **9**(1): p. 340.
4. Lu, Q., et al., *A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data*. Sci Rep, 2015. **5**: p. 10576.
5. Maston, G.A., S.K. Evans, and M.R. Green, *Transcriptional regulatory elements in the human genome*. Annu Rev Genomics Hum Genet, 2006. **7**: p. 29-59.
6. Rowley, M.J. and V.G. Corces, *Organizational principles of 3D genome architecture*. Nat Rev Genet, 2018. **19**(12): p. 789-800.
7. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
8. Venter, J.C., et al., *The sequence of the human genome*. Science, 2001. **291**(5507): p. 1304-51.
9. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome*. Nature, 2004. **431**(7011): p. 931-45.
10. Nurk, S., et al., *The complete sequence of a human genome*. Science, 2022. **376**(6588): p. 44-53.
11. Secomandi, S., et al., *Pangenome graphs and their applications in biodiversity genomics*. Nat Genet, 2025. **57**(1): p. 13-26.
12. Yu, Z., et al., *Genetic variation across and within individuals*. Nat Rev Genet, 2024. **25**(8): p. 548-562.
13. Weischenfeldt, J., et al., *Phenotypic impact of genomic structural variation: insights from and for human disease*. Nat Rev Genet, 2013. **14**(2): p. 125-38.
14. Eichler, E.E., *Genetic Variation, Comparative Genomics, and the Diagnosis of Disease*. N Engl J Med, 2019. **381**(1): p. 64-74.
15. Chiang, C., et al., *The impact of structural variation on human gene expression*. Nat Genet, 2017. **49**(5): p. 692-699.
16. Stankiewicz, P. and J.R. Lupski, *Structural variation in the human genome and its role in disease*. Annu Rev Med, 2010. **61**: p. 437-55.
17. Preissl, S., K.J. Gaulton, and B. Ren, *Characterizing cis-regulatory elements using single-cell epigenomics*. Nat Rev Genet, 2023. **24**(1): p. 21-43.
18. Szabo, Q., F. Bantignies, and G. Cavalli, *Principles of genome folding into topologically associating domains*. Sci Adv, 2019. **5**(4): p. eaaw1668.
19. Kumar, S. and M. Gerstein, *Unified views on variant impact across many diseases*. Trends Genet, 2023. **39**(6): p. 442-450.
20. Rheinbay, E., et al., *Analyses of non-coding somatic drivers in 2,658 cancer whole genomes*. Nature, 2020. **578**(7793): p. 102-111.
21. Zhang, X. and M. Meyerson, *Illuminating the noncoding genome in cancer*. Nat Cancer, 2020. **1**(9): p. 864-872.
22. International HapMap, C., *A haplotype map of the human genome*. Nature, 2005. **437**(7063): p. 1299-320.
23. Rahal, Z., P. Scheet, and H. Kadara, *Somatic Mutations in Normal Tissues: Calm before the Storm*. Cancer Discov, 2024. **14**(4): p. 605-609.
24. Liu, Y., et al., *Somatic mutational profiles and germline polygenic risk scores in human cancer*. Genome Med, 2022. **14**(1): p. 14.
25. Akhavanfard, S., et al., *Comprehensive germline genomic profiles of children, adolescents and young adults with solid tumors*. Nat Commun, 2020. **11**(1): p. 2206.
26. Wagener, R., et al., *Comprehensive germline-genomic and clinical profiling in 160 unselected children and adolescents with cancer*. Eur J Hum Genet, 2021. **29**(8): p. 1301-1311.
27. Stoltze, U.K., et al., *The evolutionary impact of childhood cancer on the human gene pool*. Nat Commun, 2024. **15**(1): p. 1881.

28. Kratz, C.P., et al., *Adult-Onset Cancer Predisposition Syndromes in Children and Adolescents-To Test or not to Test?* Clin Cancer Res, 2024. **30**(9): p. 1733-1738.
29. Zhang, J., et al., *Germline Mutations in Predisposition Genes in Pediatric Cancer.* N Engl J Med, 2015. **373**(24): p. 2336-2346.
30. Bakhuizen, J.J., et al., *Assessment of Cancer Predisposition Syndromes in a National Cohort of Children With a Neoplasm.* JAMA Netw Open, 2023. **6**(2): p. e2254157.
31. Byrjalsen, A., et al., *Nationwide germline whole genome sequencing of 198 consecutive pediatric cancer patients reveals a high incidence of cancer prone syndromes.* PLoS Genet, 2020. **16**(12): p. e1009231.
32. Scherer, S.W., et al., *Challenges and standards in integrating surveys of structural variation.* Nat Genet, 2007. **39**(7 Suppl): p. S7-15.
33. Goodwin, S., J.D. McPherson, and W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies.* Nat Rev Genet, 2016. **17**(6): p. 333-51.
34. Sedlazeck, F.J., et al., *Piercing the dark matter: bioinformatics of long-range sequencing and mapping.* Nat Rev Genet, 2018. **19**(6): p. 329-346.
35. Sharp, A.J., et al., *Segmental duplications and copy-number variation in the human genome.* Am J Hum Genet, 2005. **77**(1): p. 78-88.
36. Logsdon, G.A., M.R. Vollger, and E.E. Eichler, *Long-read human genome sequencing and its applications.* Nat Rev Genet, 2020. **21**(10): p. 597-614.
37. Mahmoud, M., et al., *Structural variant calling: the long and the short of it.* Genome Biol, 2019. **20**(1): p. 246.
38. Chaisson, M.J.P., et al., *Multi-platform discovery of haplotype-resolved structural variation in human genomes.* Nat Commun, 2019. **10**(1): p. 1784.
39. Mantere, T., et al., *Optical genome mapping enables constitutional chromosomal aberration detection.* Am J Hum Genet, 2021. **108**(8): p. 1409-1422.
40. Neveling, K., et al., *Next-generation cytogenetics: Comprehensive assessment of 52 hematological malignancy genomes by optical genome mapping.* Am J Hum Genet, 2021. **108**(8): p. 1423-1435.
41. Smith, A.C., K. Neveling, and R. Kanagal-Shamanna, *Optical genome mapping for structural variation analysis in hematologic malignancies.* Am J Hematol, 2022. **97**(7): p. 975-982.
42. Levy, B., et al., *A framework for the clinical implementation of optical genome mapping in hematologic malignancies.* Am J Hematol, 2024. **99**(4): p. 642-661.
43. Goldrich, D.Y., et al., *Identification of Somatic Structural Variants in Solid Tumors by Optical Genome Mapping.* J Pers Med, 2021. **11**(2).
44. Sahajpal, N.S., et al., *Optical Genome Mapping as a Next-Generation Cytogenomic Tool for Detection of Structural and Copy Number Variations for Prenatal Genomic Analyses.* Genes (Basel), 2021. **12**(3).
45. Gerding, W.M., et al., *Optical genome mapping reveals additional prognostic information compared to conventional cytogenetics in AML/MDS patients.* Int J Cancer, 2022. **150**(12): p. 1998-2011.
46. Schrauwen, I., et al., *Optical genome mapping unveils hidden structural variants in neurodevelopmental disorders.* Sci Rep, 2024. **14**(1): p. 11239.
47. Dharmadhikari, A.V., et al., *Optical genome mapping improves clinical interpretation of constitutional copy-number gains and reduces their VUS burden.* Genet Med, 2025. **27**(8): p. 101452.
48. Luhmann, J.L., et al., *The Clinical Utility of Optical Genome Mapping for the Assessment of Genomic Aberrations in Acute Lymphoblastic Leukemia.* Cancers (Basel), 2021. **13**(17).
49. Rack, K., et al., *Optimizing the diagnostic workflow for acute lymphoblastic leukemia by optical genome mapping.* Am J Hematol, 2022. **97**(5): p. 548-561.
50. Sahajpal, N.S., et al., *Clinical Utility of Optical Genome Mapping and 523-Gene Next Generation Sequencing Panel for Comprehensive Evaluation of Myeloid Cancers.* Cancers (Basel), 2023. **15**(12).
51. Kanagal-Shamanna, R., et al., *Integration of Optical Genome Mapping in the Cytogenomic and Molecular Work-Up of Hematological Malignancies: Expert Recommendations From the*

- International Consortium for Optical Genome Mapping*. Am J Hematol, 2025. **100**(6): p. 1029-1048.
52. Yang, H., et al., *High-resolution structural variant profiling of myelodysplastic syndromes by optical genome mapping uncovers cryptic aberrations of prognostic and therapeutic significance*. Leukemia, 2022. **36**(9): p. 2306-2316.
 53. Cocco, N., et al., *Feasibility of Optical Genome Mapping in Cytogenetic Diagnostics of Hematological Neoplasms: A New Way to Look at DNA*. Diagnostics (Basel), 2023. **13**(11).
 54. Levy, B., et al., *Optical genome mapping in acute myeloid leukemia: a multicenter evaluation*. Blood Adv, 2023. **7**(7): p. 1297-1307.
 55. Broeckel, U., et al., *Detection of Constitutional Structural Variants by Optical Genome Mapping: A Multisite Study of Postnatal Samples*. J Mol Diagn, 2024. **26**(3): p. 213-226.
 56. Warburton, P.E. and R.P. Sebra, *Long-Read DNA Sequencing: Recent Advances and Remaining Challenges*. Annu Rev Genomics Hum Genet, 2023. **24**: p. 109-132.
 57. Olivucci, G., et al., *Long read sequencing on its way to the routine diagnostics of genetic diseases*. Front Genet, 2024. **15**: p. 1374860.
 58. Ebert, P., et al., *Haplotype-resolved diverse human genomes and integrated analysis of structural variation*. Science, 2021. **372**(6537).
 59. Thibodeau, M.L., et al., *Improved structural variant interpretation for hereditary cancer susceptibility using long-read sequencing*. Genet Med, 2020. **22**(11): p. 1892-1897.
 60. Dixon, K., et al., *Defining the heterogeneity of unbalanced structural variation underlying breast cancer susceptibility by nanopore genome sequencing*. Eur J Hum Genet, 2023. **31**(5): p. 602-606.
 61. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology*. Genet Med, 2015. **17**(5): p. 405-24.
 62. Horak, P., et al., *Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): Joint recommendations of Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC), and Variant Interpretation for Cancer Consortium (VICC)*. Genet Med, 2022. **24**(9): p. 1991.
 63. Green, R.C., et al., *ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing*. Genet Med, 2013. **15**(7): p. 565-74.
 64. Brandt, T., et al., *Adapting ACMG/AMP sequence variant classification guidelines for single-gene copy number variants*. Genet Med, 2020. **22**(2): p. 336-344.
 65. Riggs, E.R., et al., *Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen)*. Genet Med, 2020. **22**(2): p. 245-257.
 66. Organization, W.H. *Childhood cancer*. 2024; Available from: <https://www.who.int/news-room/fact-sheets/detail/cancer-in-children>.
 67. Erdmann F, K.P., Grabow D, Spix C. *German Childhood Cancer Registry - Annual Report 2019(1980-2018)*. Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI) at the University Medical Center of the Johannes Gutenberg University Mainz. 2020.
 68. Navarrete-Meneses, M.D.P., et al., *Environmental Pollution and Risk of Childhood Cancer: A Scoping Review of Evidence from the Last Decade*. Int J Mol Sci, 2024. **25**(6).
 69. Plon, S.E. and P.J. Lupo, *Genetic Predisposition to Childhood Cancer in the Genomic Era*. Annu Rev Genomics Hum Genet, 2019. **20**: p. 241-263.
 70. McGee, R.B. and K.E. Nichols, *Introduction to cancer genetic susceptibility syndromes*. Hematology Am Soc Hematol Educ Program, 2016. **2016**(1): p. 293-301.
 71. Grobner, S.N., et al., *The landscape of genomic alterations across childhood cancers*. Nature, 2018. **555**(7696): p. 321-327.
 72. Ueno, S., T. Sudo, and A. Hirasawa, *ATM: Functions of ATM Kinase and Its Relevance to Hereditary Tumors*. Int J Mol Sci, 2022. **23**(1).
 73. Subramanian, V., et al., *Bloom syndrome DNA helicase deficiency is associated with oxidative stress and mitochondrial network changes*. Sci Rep, 2021. **11**(1): p. 2157.

74. Sirover, M.A., T.M. Vollberg, and G. Seal, *DNA repair and the molecular mechanisms of Bloom's syndrome*. Crit Rev Oncog, 1990. **2**(1): p. 19-33.
75. Niraj, J., A. Farkkila, and A.D. D'Andrea, *The Fanconi Anemia Pathway in Cancer*. Annu Rev Cancer Biol, 2019. **3**: p. 457-478.
76. Tauchi, H., et al., *Nijmegen breakage syndrome gene, NBS1, and molecular links to factors for genome stability*. Oncogene, 2002. **21**(58): p. 8967-80.
77. Roudko, V., et al., *Lynch Syndrome and MSI-H Cancers: From Mechanisms to "Off-The-Shelf" Cancer Vaccines*. Front Immunol, 2021. **12**: p. 757804.
78. Peltomaki, P., *Role of DNA mismatch repair defects in the pathogenesis of human cancer*. J Clin Oncol, 2003. **21**(6): p. 1174-9.
79. Jongmans, M.C., et al., *Recognition of genetic predisposition in pediatric cancer patients: An easy-to-use selection tool*. Eur J Med Genet, 2016. **59**(3): p. 116-25.
80. Ripperger, T., et al., *Childhood cancer predisposition syndromes-A concise review and recommendations by the Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology*. Am J Med Genet A, 2017. **173**(4): p. 1017-1037.
81. Elitzur, S., et al., *ATM germ line pathogenic variants affect outcomes in children with ataxia-telangiectasia and hematological malignancies*. Blood, 2024. **144**(11): p. 1193-1205.
82. Bernsen, E.C., et al., *Pharmacogenomics as a Tool to Limit Acute and Long-Term Adverse Effects of Chemotherapeutics: An Update in Pediatric Oncology*. Front Pharmacol, 2020. **11**: p. 1184.
83. Kratz, C.P., et al., *Predisposition to cancer in children and adolescents*. Lancet Child Adolesc Health, 2021. **5**(2): p. 142-154.
84. Thariat, J., et al., *Avoidance or adaptation of radiotherapy in patients with cancer with Li-Fraumeni and heritable TP53-related cancer syndromes*. Lancet Oncol, 2021. **22**(12): p. e562-e574.
85. Blake, A., et al., *Performance of Tumor Surveillance for Children With Cancer Predisposition*. JAMA Oncol, 2024. **10**(8): p. 1060-1067.
86. Waespe, N., et al., *Cancer predisposition syndromes as a risk factor for early second primary neoplasms after childhood cancer - A national cohort study*. Eur J Cancer, 2021. **145**: p. 71-80.
87. Kuhlen, M. and A. Borkhardt, *Trio sequencing in pediatric cancer and clinical implications*. EMBO Mol Med, 2018. **10**(4).
88. Kuhlen, M., et al., *Family-based germline sequencing in children with cancer*. Oncogene, 2019. **38**(9): p. 1367-1380.
89. Brozou, T., et al., *Resolving inherited and de novo germline predisposing sequence variants by means of whole exome trio analyses in childhood hematological malignancies*. Front Pediatr, 2022. **10**: p. 1080347.
90. Neglia, J.P., et al., *Second malignant neoplasms in five-year survivors of childhood cancer: childhood cancer survivor study*. J Natl Cancer Inst, 2001. **93**(8): p. 618-29.
91. Shahani, S.A. and E.L. Marcotte, *Landscape of germline cancer predisposition mutations testing and management in pediatrics: Implications for research and clinical care*. Front Pediatr, 2022. **10**: p. 1011873.
92. Wojcik, M.H., et al., *Genome Sequencing for Diagnosing Rare Diseases*. N Engl J Med, 2024. **390**(21): p. 1985-1997.
93. Pocza, T., et al., *Germline Structural Variations in Cancer Predisposition Genes*. Front Genet, 2021. **12**: p. 634217.
94. Cosenza, M.R., B. Rodriguez-Martin, and J.O. Korbel, *Structural Variation in Cancer: Role, Prevalence, and Mechanisms*. Annu Rev Genomics Hum Genet, 2022. **23**: p. 123-152.
95. Sawyers, C.L., C.T. Denny, and O.N. Witte, *Leukemia and the disruption of normal hematopoiesis*. Cell, 1991. **64**(2): p. 337-50.
96. Chennamadhavuni, A., et al., *Leukemia*, in StatPearls. 2025: Treasure Island (FL).
97. Khoury, J.D., et al., *The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms*. Leukemia, 2022. **36**(7): p. 1703-1719.
98. Pagliaro, L., et al., *Acute lymphoblastic leukaemia*. Nat Rev Dis Primers, 2024. **10**(1): p. 41.

99. Brady, S.W., et al., *The genomic landscape of pediatric acute lymphoblastic leukemia*. Nat Genet, 2022. **54**(9): p. 1376-1389.
100. Hunger, S.P. and C.G. Mullighan, *Acute Lymphoblastic Leukemia in Children*. N Engl J Med, 2015. **373**(16): p. 1541-52.
101. Nesbit, M.E., Jr., et al., *Testicular relapse in childhood acute lymphoblastic leukemia: association with pretreatment patient characteristics and treatment. A report for Childrens Cancer Study Group*. Cancer, 1980. **45**(8): p. 2009-16.
102. Onciu, M., *Acute lymphoblastic leukemia*. Hematol Oncol Clin North Am, 2009. **23**(4): p. 655-74.
103. Redaelli, A., et al., *A systematic literature review of the clinical and epidemiological burden of acute lymphoblastic leukaemia (ALL)*. Eur J Cancer Care (Engl), 2005. **14**(1): p. 53-62.
104. Eden, T., *Aetiology of childhood leukaemia*. Cancer Treat Rev, 2010. **36**(4): p. 286-97.
105. Belson, M., B. Kingsley, and A. Holmes, *Risk factors for acute leukemia in children: a review*. Environ Health Perspect, 2007. **115**(1): p. 138-45.
106. Greaves, M.F., *Aetiology of acute leukaemia*. Lancet, 1997. **349**(9048): p. 344-9.
107. Greaves, M., *A causal mechanism for childhood acute lymphoblastic leukaemia*. Nat Rev Cancer, 2018. **18**(8): p. 471-484.
108. Mori, H., et al., *Chromosome translocations and covert leukemic clones are generated during normal fetal development*. Proc Natl Acad Sci U S A, 2002. **99**(12): p. 8242-7.
109. Greaves, M., *In utero origins of childhood leukaemia*. Early Hum Dev, 2005. **81**(1): p. 123-9.
110. Greaves, M.F. and J. Wiemels, *Origins of chromosome translocations in childhood leukaemia*. Nat Rev Cancer, 2003. **3**(9): p. 639-49.
111. Kinlen, L., *Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain*. Lancet, 1988. **2**(8624): p. 1323-7.
112. Greaves, M., *Infection, immune responses and the aetiology of childhood leukaemia*. Nat Rev Cancer, 2006. **6**(3): p. 193-203.
113. Gocho, Y. and J.J. Yang, *Genetic defects in hematopoietic transcription factors and predisposition to acute lymphoblastic leukemia*. Blood, 2019. **134**(10): p. 793-797.
114. Trevino, L.R., et al., *Germline genomic variants associated with childhood acute lymphoblastic leukemia*. Nat Genet, 2009. **41**(9): p. 1001-5.
115. Prasad, R.B., et al., *Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood*. Blood, 2010. **115**(9): p. 1765-7.
116. Klcio, J.M. and C.G. Mullighan, *Advances in germline predisposition to acute leukaemias and myeloid neoplasms*. Nat Rev Cancer, 2021. **21**(2): p. 122-137.
117. Moriyama, T., M.V. Relling, and J.J. Yang, *Inherited genetic variation in childhood acute lymphoblastic leukemia*. Blood, 2015. **125**(26): p. 3988-95.
118. Pui, C.H., K.E. Nichols, and J.J. Yang, *Somatic and germline genomics in paediatric acute lymphoblastic leukaemia*. Nat Rev Clin Oncol, 2019. **16**(4): p. 227-240.
119. Inaba, H. and C.G. Mullighan, *Pediatric acute lymphoblastic leukemia*. Haematologica, 2020. **105**(11): p. 2524-2539.
120. Rack, K.A., et al., *European recommendations and quality assurance for cytogenomic analysis of haematological neoplasms*. Leukemia, 2019. **33**(8): p. 1851-1867.
121. Mullighan, C.G., *Molecular genetics of B-precursor acute lymphoblastic leukemia*. J Clin Invest, 2012. **122**(10): p. 3407-15.
122. Roberts, K.G. and C.G. Mullighan, *The Biology of B-Progenitor Acute Lymphoblastic Leukemia*. Cold Spring Harb Perspect Med, 2020. **10**(7).
123. Li, J.F., et al., *Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases*. Proc Natl Acad Sci U S A, 2018. **115**(50): p. E11711-E11720.
124. Iacobucci, I., S. Kimura, and C.G. Mullighan, *Biologic and Therapeutic Implications of Genomic Alterations in Acute Lymphoblastic Leukemia*. J Clin Med, 2021. **10**(17).
125. Li, J., et al., *Emerging molecular subtypes and therapeutic targets in B-cell precursor acute lymphoblastic leukemia*. Front Med, 2021. **15**(3): p. 347-371.

126. Stanulla, M., et al., *IKZF1(plus) Defines a New Minimal Residual Disease-Dependent Very-Poor Prognostic Profile in Pediatric B-Cell Precursor Acute Lymphoblastic Leukemia*. J Clin Oncol, 2018. **36**(12): p. 1240-1249.
127. Purvis, K., et al., *Outcomes in patients with ETV6::RUNX1 or high-hyperdiploid B-ALL treated in the St. Jude Total Therapy XV/XVI studies*. Blood, 2025. **145**(2): p. 190-201.
128. Haas, O.A. and A. Borkhardt, *Hyperdiploidy: the longest known, most prevalent, and most enigmatic form of acute lymphoblastic leukemia in children*. Leukemia, 2022. **36**(12): p. 2769-2783.
129. Zeckanovic, A., et al., *Update on long-term outcomes of a cohort of patients with TCF3::HLF-positive acute lymphoblastic leukemia treated with blinatumomab and stem cell transplantation*. Haematologica, 2025. **110**(6): p. 1373-1378.
130. Zalioua, M., et al., *Genomic landscape of pediatric B-other acute lymphoblastic leukemia in a consecutive European cohort*. Haematologica, 2019. **104**(7): p. 1396-1406.
131. Paulsson, K., et al., *Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia*. Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21719-24.
132. Djokic, M., et al., *Overexpression of CD123 correlates with the hyperdiploid genotype in acute lymphoblastic leukemia*. Haematologica, 2009. **94**(7): p. 1016-9.
133. Enshaei, A., et al., *Defining low-risk high hyperdiploidy in patients with paediatric acute lymphoblastic leukaemia: a retrospective analysis of data from the UKALL97/99 and UKALL2003 clinical trials*. Lancet Haematol, 2021. **8**(11): p. e828-e839.
134. Paulsson, K. and B. Johansson, *High hyperdiploid childhood acute lymphoblastic leukemia*. Genes Chromosomes Cancer, 2009. **48**(8): p. 637-60.
135. Paulsson, K., et al., *Evidence for a single-step mechanism in the origin of hyperdiploid childhood acute lymphoblastic leukemia*. Genes Chromosomes Cancer, 2005. **44**(2): p. 113-22.
136. Maia, A.T., et al., *Identification of preleukemic precursors of hyperdiploid acute lymphoblastic leukemia in cord blood*. Genes Chromosomes Cancer, 2004. **40**(1): p. 38-43.
137. Bateman, C.M., et al., *Evolutionary trajectories of hyperdiploid ALL in monozygotic twins*. Leukemia, 2015. **29**(1): p. 58-65.
138. Yang, M., et al., *Proteogenomics and Hi-C reveal transcriptional dysregulation in high hyperdiploid childhood acute lymphoblastic leukemia*. Nat Commun, 2019. **10**(1): p. 1519.
139. Andersson, A., et al., *Molecular signatures in childhood acute leukemia and their correlations to expression patterns in normal hematopoietic subpopulations*. Proc Natl Acad Sci U S A, 2005. **102**(52): p. 19069-74.
140. Sakamoto, K., et al., *Acute lymphoblastic leukemia developing in a patient with Noonan syndrome harboring a PTPN11 germline mutation*. J Pediatr Hematol Oncol, 2014. **36**(2): p. e136-9.
141. Cave, H., et al., *Acute lymphoblastic leukemia in the context of RASopathies*. Eur J Med Genet, 2016. **59**(3): p. 173-8.
142. Moriyama, T., et al., *Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study*. Lancet Oncol, 2015. **16**(16): p. 1659-66.
143. de Smith, A.J., et al., *Predisposing germline mutations in high hyperdiploid acute lymphoblastic leukemia in children*. Genes Chromosomes Cancer, 2019. **58**(10): p. 723-730.
144. Zipper, L., et al., *Hyperdiploid acute lymphoblastic leukemia in children with LZTR1 germline variants*. Hemasphere, 2024. **8**(1): p. e26.
145. Wiemels, J.L., et al., *A functional polymorphism in the CEBPE gene promoter influences acute lymphoblastic leukemia risk through interaction with the hematopoietic transcription factor Ikaros*. Leukemia, 2016. **30**(5): p. 1194-7.
146. Walsh, K.M., et al., *Novel childhood ALL susceptibility locus BMI1-PIP4K2A is specifically associated with the hyperdiploid subtype*. Blood, 2013. **121**(23): p. 4808-9.
147. Malinowska-Ozdowy, K., et al., *KRAS and CREBBP mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia*. Leukemia, 2015. **29**(8): p. 1656-67.

148. Chen, C., et al., *Next-generation-sequencing of recurrent childhood high hyperdiploid acute lymphoblastic leukemia reveals mutations typically associated with high risk patients*. *Leuk Res*, 2015. **39**(9): p. 990-1001.
149. Inthal, A., et al., *CREBBP HAT domain mutations prevail in relapse cases of high hyperdiploid childhood acute lymphoblastic leukemia*. *Leukemia*, 2012. **26**(8): p. 1797-803.
150. Mullighan, C.G., et al., *CREBBP mutations in relapsed acute lymphoblastic leukaemia*. *Nature*, 2011. **471**(7337): p. 235-9.
151. Woodward, E.L., et al., *Clonal origin and development of high hyperdiploidy in childhood acute lymphoblastic leukaemia*. *Nat Commun*, 2023. **14**(1): p. 1658.
152. Sood, R., Y. Kamikubo, and P. Liu, *Role of RUNX1 in hematological malignancies*. *Blood*, 2017. **129**(15): p. 2070-2082.
153. Hock, H. and A. Shimamura, *ETV6 in hematopoiesis and leukemia predisposition*. *Semin Hematol*, 2017. **54**(2): p. 98-104.
154. Fuka, G., et al., *The leukemia-specific fusion gene ETV6/RUNX1 perturbs distinct key biological functions primarily by gene repression*. *PLoS One*, 2011. **6**(10): p. e26348.
155. Hiebert, S.W., et al., *The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription*. *Mol Cell Biol*, 1996. **16**(4): p. 1349-55.
156. Boiers, C., et al., *A Human IPS Model Implicates Embryonic B-Myeloid Fate Restriction as Developmental Susceptibility to B Acute Lymphoblastic Leukemia-Associated ETV6-RUNX1*. *Dev Cell*, 2018. **44**(3): p. 362-377 e7.
157. Inaba, H., M. Greaves, and C.G. Mullighan, *Acute lymphoblastic leukaemia*. *Lancet*, 2013. **381**(9881): p. 1943-55.
158. Uckun, F.M., et al., *Expression of TEL-AML1 fusion transcripts and response to induction therapy in standard risk acute lymphoblastic leukemia*. *Leuk Lymphoma*, 2001. **42**(1-2): p. 41-56.
159. Sun, C., L. Chang, and X. Zhu, *Pathogenesis of ETV6/RUNX1-positive childhood acute lymphoblastic leukemia and mechanisms underlying its relapse*. *Oncotarget*, 2017. **8**(21): p. 35445-35459.
160. Pine, S.R., et al., *TEL-AML1 fusion precedes differentiation to pre-B cells in childhood acute lymphoblastic leukemia*. *Leuk Res*, 2003. **27**(2): p. 155-64.
161. Hubner, S., et al., *High incidence and unique features of antigen receptor gene rearrangements in TEL-AML1-positive leukemias*. *Leukemia*, 2004. **18**(1): p. 84-91.
162. Zuna, J., et al., *ETV6/RUNX1 (TEL/AML1) is a frequent prenatal first hit in childhood leukemia*. *Blood*, 2011. **117**(1): p. 368-9; author reply 370-1.
163. Schafer, D., et al., *Five percent of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening*. *Blood*, 2018. **131**(7): p. 821-826.
164. Hein, D., A. Borkhardt, and U. Fischer, *Insights into the prenatal origin of childhood acute lymphoblastic leukemia*. *Cancer Metastasis Rev*, 2020. **39**(1): p. 161-171.
165. Ellinghaus, E., et al., *Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia*. *Leukemia*, 2012. **26**(5): p. 902-9.
166. Jakobczyk, H., et al., *ETV6-RUNX1 and RUNX1 directly regulate RAG1 expression: one more step in the understanding of childhood B-cell acute lymphoblastic leukemia leukemogenesis*. *Leukemia*, 2022. **36**(2): p. 549-554.
167. Chen, D., et al., *RAG1 co-expression signature identifies ETV6-RUNX1-like B-cell precursor acute lymphoblastic leukemia in children*. *Cancer Med*, 2021. **10**(12): p. 3997-4003.
168. Bassing, C.H., W. Swat, and F.W. Alt, *The mechanism and regulation of chromosomal V(D)J recombination*. *Cell*, 2002. **109 Suppl**: p. S45-55.
169. Papaemmanuil, E., et al., *RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia*. *Nat Genet*, 2014. **46**(2): p. 116-25.
170. Nishii, R., et al., *Molecular basis of ETV6-mediated predisposition to childhood acute lymphoblastic leukemia*. *Blood*, 2021. **137**(3): p. 364-373.
171. Forestier, E., et al., *Outcome of ETV6/RUNX1-positive childhood acute lymphoblastic leukaemia in the NOPHO-ALL-1992 protocol: frequent late relapses but good overall survival*. *Br J Haematol*, 2008. **140**(6): p. 665-72.

172. Konrad, M., et al., *Late relapses evolve from slow-responding subclones in t(12;21)-positive acute lymphoblastic leukemia: evidence for the persistence of a preleukemic clone*. *Blood*, 2003. **101**(9): p. 3635-40.
173. Bokemeyer, A., et al., *Copy number genome alterations are associated with treatment response and outcome in relapsed childhood ETV6/RUNX1-positive acute lymphoblastic leukemia*. *Haematologica*, 2014. **99**(4): p. 706-14.
174. van Delft, F.W., et al., *Clonal origins of relapse in ETV6-RUNX1 acute lymphoblastic leukemia*. *Blood*, 2011. **117**(23): p. 6247-54.
175. Stefan, A.I., et al., *The emerging role of next-generation sequencing in minimal residual disease assessment in acute lymphoblastic leukemia: a systematic review of current literature*. *Front Med (Lausanne)*, 2025. **12**: p. 1570041.
176. Thastrup, M., et al., *Central nervous system involvement in childhood acute lymphoblastic leukemia: challenges and solutions*. *Leukemia*, 2022. **36**(12): p. 2751-2768.
177. Saygin, C., et al., *Measurable residual disease in acute lymphoblastic leukemia: methods and clinical context in adult patients*. *Haematologica*, 2022. **107**(12): p. 2783-2793.
178. Leoni, V. and A. Biondi, *Tyrosine kinase inhibitors in BCR-ABL positive acute lymphoblastic leukemia*. *Haematologica*, 2015. **100**(3): p. 295-9.
179. Hall, A.G. and R.E. Rau, *Blinatumomab use in pediatric B-ALL: where are we now?* *Blood Adv*, 2025. **9**(15): p. 3946-3954.
180. Wagener, R., et al., *Optical genome mapping identifies structural variants in potentially new cancer predisposition candidate genes in pediatric cancer patients*. *Int J Cancer*, 2024. **154**(4): p. 607-614.
181. Brandes, D., et al., *Optical Genome Mapping Identifies Novel Recurrent Structural Alterations in Childhood ETV6::RUNX1+ and High Hyperdiploid Acute Lymphoblastic Leukemia*. *Hemasphere*, 2023. **7**(8): p. e925.
182. Balducci, E., et al., *Optical genome mapping refines cytogenetic diagnostics, prognostic stratification and provides new molecular insights in adult MDS/AML patients*. *Blood Cancer J*, 2022. **12**(9): p. 126.
183. Petrackova, A., et al., *Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics*. *Front Oncol*, 2019. **9**: p. 851.
184. Tsai, M.M., et al., *Optical genome mapping with whole genome sequencing identifies complex chromosomal structural variations in acute leukemia*. *Front Genet*, 2025. **16**: p. 1496847.
185. de Boer, E.N., et al., *Cas9-directed long-read sequencing to resolve optical genome mapping findings in leukemia diagnostics*. *Sci Rep*, 2024. **14**(1): p. 8508.
186. De Clercq, G., et al., *Full characterization of unresolved structural variation through long-read sequencing and optical genome mapping*. *Sci Rep*, 2024. **14**(1): p. 29142.
187. Iqbal, M.A., et al., *Multisite Assessment of Optical Genome Mapping for Analysis of Structural Variants in Constitutional Postnatal Cases*. *J Mol Diagn*, 2023. **25**(3): p. 175-188.
188. Hiatt, S.M., et al., *Long-read genome sequencing and variant reanalysis increase diagnostic yield in neurodevelopmental disorders*. *Genome Res*, 2024. **34**(11): p. 1747-1762.
189. Conlin, L.K., et al., *Long-read sequencing for molecular diagnostics in constitutional genetic disorders*. *Hum Mutat*, 2022. **43**(11): p. 1531-1544.
190. Maestri, S., et al., *A Long-Read Sequencing Approach for Direct Haplotype Phasing in Clinical Settings*. *Int J Mol Sci*, 2020. **21**(23).
191. Miller, D.E., et al., *Targeted long-read sequencing identifies missing pathogenic variants in unsolved Werner syndrome cases*. *J Med Genet*, 2022. **59**(11): p. 1087-94.
192. Xu, L., et al., *Long-Molecule Sequencing: A New Approach for Identification of Clinically Significant DNA Variants in alpha-Thalassemia and beta-Thalassemia Carriers*. *J Mol Diagn*, 2020. **22**(8): p. 1087-1095.
193. Choi, Y., et al., *Comparison of phasing strategies for whole human genomes*. *PLoS Genet*, 2018. **14**(4): p. e1007308.
194. Noyes, M.D., et al., *Familial long-read sequencing increases yield of de novo mutations*. *Am J Hum Genet*, 2022. **109**(4): p. 631-646.

195. de Ligt, J., J.A. Veltman, and L.E. Vissers, *Point mutations as a source of de novo genetic disease*. *Curr Opin Genet Dev*, 2013. **23**(3): p. 257-63.
196. Lopes-Marques, M., et al., *Meta-analysis of 46,000 germline de novo mutations linked to human inherited disease*. *Hum Genomics*, 2024. **18**(1): p. 20.
197. Rashed, W.M., E.L. Marcotte, and L.G. Spector, *Germline De Novo Mutations as a Cause of Childhood Cancer*. *JCO Precis Oncol*, 2022. **6**: p. e2100505.
198. Sherwood, K., et al., *Germline de novo mutations in families with Mendelian cancer syndromes caused by defects in DNA repair*. *Nat Commun*, 2023. **14**(1): p. 3636.
199. Soejima, H. and K. Higashimoto, *Epigenetic and genetic alterations of the imprinting disorder Beckwith-Wiedemann syndrome and related disorders*. *J Hum Genet*, 2013. **58**(7): p. 402-9.
200. Cheung, W.A., et al., *Direct haplotype-resolved 5-base HiFi sequencing for genome-wide profiling of hypermethylation outliers in a rare disease cohort*. *Nat Commun*, 2023. **14**(1): p. 3090.
201. Raca, G., et al., *Points to consider in the detection of germline structural variants using next-generation sequencing: A statement of the American College of Medical Genetics and Genomics (ACMG)*. *Genet Med*, 2023. **25**(2): p. 100316.
202. Pagnamenta, A.T., et al., *Structural and non-coding variants increase the diagnostic yield of clinical whole genome sequencing for rare diseases*. *Genome Med*, 2023. **15**(1): p. 94.
203. Collins, R.L., et al., *A structural variation reference for medical and population genetics*. *Nature*, 2020. **581**(7809): p. 444-451.
204. Bertoli-Avella, A.M., et al., *Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort*. *Eur J Hum Genet*, 2021. **29**(1): p. 141-153.
205. Cohen, A.S.A., et al., *Genomic answers for children: Dynamic analyses of >1000 pediatric rare disease genomes*. *Genet Med*, 2022. **24**(6): p. 1336-1348.
206. Damaraju, N., A.L. Miller, and D.E. Miller, *Long-Read DNA and RNA Sequencing to Streamline Clinical Genetic Testing and Reduce Barriers to Comprehensive Genetic Testing*. *J Appl Lab Med*, 2024. **9**(1): p. 138-150.
207. Jahn, A., et al., *Comprehensive cancer predisposition testing within the prospective MASTER trial identifies hereditary cancer patients and supports treatment decisions for rare cancers*. *Ann Oncol*, 2022. **33**(11): p. 1186-1199.
208. Jette J. Bakhuizen, F.B., Karin A.W. Wadt, Christian P. Kratz, Marjolijn C.J. Jongmans, Nicolas Waespe, *Genetic testing for childhood cancer predisposition syndromes: Controversies and recommendations from the SIOPE Host Genome Working Group meeting 2022*. *EJC Paediatric Oncology*, 2024. **4**.
209. Bakhuizen, J.J., et al., *Comparison of clinical selection-based genetic testing with phenotype-agnostic extensive germline sequencing to diagnose genetic predisposition in children with cancer: a prospective diagnostic study*. *Lancet Child Adolesc Health*, 2024. **8**(10): p. 751-761.
210. Newman, S., 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): p. 3008-3027.
211. Chen, F., et al., *Global DNA methylation differences involving germline structural variation impact gene expression in pediatric brain tumors*. *Nat Commun*, 2025. **16**(1): p. 4713.
212. Wong, M., et al., *Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer*. *Nat Med*, 2020. **26**(11): p. 1742-1753.
213. Sabatella, M., et al., *Optical genome mapping identifies a germline retrotransposon insertion in SMARCB1 in two siblings with atypical teratoid rhabdoid tumors*. *J Pathol*, 2021. **255**(2): p. 202-211.
214. Gillani, R., et al., *Rare germline structural variants increase risk for pediatric solid tumors*. *Science*, 2025. **387**(6729): p. eadq0071.
215. Consortium, I.T.P.-C.A.o.W.G., *Pan-cancer analysis of whole genomes*. *Nature*, 2020. **578**(7793): p. 82-93.

216. McGee, R.B., et al., *Pathogenic Variants in Adult-Onset Cancer Predisposition Genes in Pediatric Cancer: Prevalence and Impact on Tumor Molecular Features and Clinical Management*. Clin Cancer Res, 2023. **29**(7): p. 1243-1251.
217. Daus, K., et al., *Germline variants observed in pediatric cancer patients related to hereditary breast and ovarian cancer in adults*. Int J Cancer, 2025.
218. Qin, N., et al., *Pathogenic Germline Mutations in DNA Repair Genes in Combination With Cancer Treatment Exposures and Risk of Subsequent Neoplasms Among Long-Term Survivors of Childhood Cancer*. J Clin Oncol, 2020. **38**(24): p. 2728-2740.
219. Wang, Z., et al., *Genetic Risk for Subsequent Neoplasms Among Long-Term Survivors of Childhood Cancer*. J Clin Oncol, 2018. **36**(20): p. 2078-2087.
220. Junk, S.V., et al., *Germline variants in patients developing second malignant neoplasms after therapy for pediatric acute lymphoblastic leukemia-a case-control study*. Leukemia, 2024. **38**(4): p. 887-892.
221. Altintas, B., et al., *Genotype-phenotype and outcome associations in patients with Fanconi anemia: the National Cancer Institute cohort*. Haematologica, 2023. **108**(1): p. 69-82.
222. Stubbins, R.J., et al., *Germline loss-of-function BRCA1 and BRCA2 mutations and risk of de novo hematopoietic malignancies*. Haematologica, 2024. **109**(1): p. 351-356.
223. Kratz, C.P., et al., *Heterozygous BRCA1 and BRCA2 and Mismatch Repair Gene Pathogenic Variants in Children and Adolescents With Cancer*. J Natl Cancer Inst, 2022. **114**(11): p. 1523-1532.
224. Wang, Z., et al., *Association of Germline BRCA2 Mutations With the Risk of Pediatric or Adolescent Non-Hodgkin Lymphoma*. JAMA Oncol, 2019. **5**(9): p. 1362-1364.
225. Chen, S., et al., *A genomic mutational constraint map using variation in 76,156 human genomes*. Nature, 2024. **625**(7993): p. 92-100.
226. Haring, S.J., et al., *Cellular functions of human RPA1. Multiple roles of domains in replication, repair, and checkpoints*. J Biol Chem, 2008. **283**(27): p. 19095-111.
227. Sharma, R., et al., *Gain-of-function mutations in RPA1 cause a syndrome with short telomeres and somatic genetic rescue*. Blood, 2022. **139**(7): p. 1039-1051.
228. Sharma, R., et al., *Germline landscape of RPA1, RPA2 and RPA3 variants in pediatric malignancies: identification of RPA1 as a novel cancer predisposition candidate gene*. Front Oncol, 2023. **13**: p. 1229507.
229. Marin Navarro, A., et al., *Modeling cancer using patient-derived induced pluripotent stem cells to understand development of childhood malignancies*. Cell Death Discov, 2018. **4**: p. 7.
230. Lee, D.F., et al., *Modeling familial cancer with induced pluripotent stem cells*. Cell, 2015. **161**(2): p. 240-54.
231. Gandre-Babbe, S., et al., *Patient-derived induced pluripotent stem cells recapitulate hematopoietic abnormalities of juvenile myelomonocytic leukemia*. Blood, 2013. **121**(24): p. 4925-9.
232. Roadmap Epigenomics, C., et al., *Integrative analysis of 111 reference human epigenomes*. Nature, 2015. **518**(7539): p. 317-30.
233. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome*. Nature, 2012. **489**(7414): p. 57-74.
234. Donohue, L.K.H., et al., *A cis-regulatory lexicon of DNA motif combinations mediating cell-type-specific gene regulation*. Cell Genom, 2022. **2**(11).
235. Kellman, L.N., et al., *Functional analysis of cancer-associated germline risk variants*. Nat Genet, 2025. **57**(3): p. 718-728.
236. Yang, H., et al., *Noncoding genetic variation in GATA3 increases acute lymphoblastic leukemia risk through local and global changes in chromatin conformation*. Nat Genet, 2022. **54**(2): p. 170-179.
237. Li, W., *The 5(th) Edition of the World Health Organization Classification of Hematolymphoid Tumors*, in Leukemia, W. Li, Editor. 2022: Brisbane (AU).
238. Arber, D.A., et al., *International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data*. Blood, 2022. **140**(11): p. 1200-1228.

239. Mullighan, C.G., et al., *Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia*. N Engl J Med, 2009. **360**(5): p. 470-80.
240. Kuiper, R.P., et al., *IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL*. Leukemia, 2010. **24**(7): p. 1258-64.
241. Luhmann, J.L., et al., *Deciphering the molecular complexity of the IKZF1(plus) genomic profile using Optical Genome Mapping*. Haematologica, 2024. **109**(5): p. 1582-1587.
242. Hidalgo-Gomez, G., et al., *How to combine multiple tools for the genetic diagnosis work-up of pediatric B-cell acute lymphoblastic leukemia*. Ann Hematol, 2025. **104**(4): p. 2387-2402.
243. Reyes-Barron, C., et al., *Next-Generation Sequencing for Minimal Residual Disease Surveillance in Acute Lymphoblastic Leukemia: An Update*. Crit Rev Oncog, 2017. **22**(5-6): p. 559-567.
244. Carroll, A.J., et al., *Masked hypodiploidy: Hypodiploid acute lymphoblastic leukemia (ALL) mimicking hyperdiploid ALL in children: A report from the Children's Oncology Group*. Cancer Genet, 2019. **238**: p. 62-68.
245. Lansdon, L.A., et al., *Successful classification of clinical pediatric leukemia genetic subtypes via structural variant detection using HiFi long-read sequencing*. medRxiv, 2024.
246. Geyer, J., et al., *Real-time genomic characterization of pediatric acute leukemia using adaptive sampling*. bioRxiv, 2025.
247. O'Neill, K., et al., *Long-read sequencing of an advanced cancer cohort resolves rearrangements, unravels haplotypes, and reveals methylation landscapes*. Cell Genom, 2024. **4**(11): p. 100674.
248. Krali, O., et al., *Multimodal classification of molecular subtypes in pediatric acute lymphoblastic leukemia*. NPJ Precis Oncol, 2023. **7**(1): p. 131.
249. Schafer Hackenhaar, F., et al., *CpG island methylator phenotype classification improves risk assessment in pediatric T-cell acute lymphoblastic leukemia*. Blood, 2025. **145**(19): p. 2161-2178.
250. Mosquera Orgueira, A., et al., *Refining risk prediction in pediatric acute lymphoblastic leukemia through DNA methylation profiling*. Clin Epigenetics, 2024. **16**(1): p. 49.
251. Sakamoto, Y., S. Sereewattanawoot, and A. Suzuki, *A new era of long-read sequencing for cancer genomics*. J Hum Genet, 2020. **65**(1): p. 3-10.
252. Mody, R.J., et al., *Integrative Clinical Sequencing in the Management of Refractory or Relapsed Cancer in Youth*. JAMA, 2015. **314**(9): p. 913-25.
253. Worst, B.C., et al., *Next-generation personalised medicine for high-risk paediatric cancer patients - The INFORM pilot study*. Eur J Cancer, 2016. **65**: p. 91-101.
254. Krentz, S., 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): p. 295-304.
255. Fang, Q., et al., *Gene Deletions and Prognostic Values in B-Lineage Acute Lymphoblastic Leukemia*. Front Oncol, 2021. **11**: p. 677034.
256. Waanders, E., et al., *Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia*. Blood Cancer Discov, 2020. **1**(1): p. 96-111.
257. Oettinger, M.A., et al., *RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination*. Science, 1990. **248**(4962): p. 1517-23.
258. Greenhalgh, R., et al., *The landscape of structural variation in pediatric cancer*. bioRxiv, 2025: p. 2025.04.24.650284.
259. King, B., et al., *The ubiquitin ligase Huwe1 regulates the maintenance and lymphoid commitment of hematopoietic stem cells*. Nat Immunol, 2016. **17**(11): p. 1312-1321.
260. Yang, Y., et al., *The endoplasmic reticulum-resident E3 ubiquitin ligase Hrd1 controls a critical checkpoint in B cell development in mice*. J Biol Chem, 2018. **293**(33): p. 12934-12944.
261. Spanjaard, A., et al., *Huwe1 supports B-cell development, B-cell-dependent immunity, somatic hypermutation and class switch recombination by regulating proliferation*. Front Immunol, 2022. **13**: p. 986863.
262. Bang, B., et al., *A somatic UBA2 variant preceded ETV6-RUNX1 in the concordant BCP-ALL of monozygotic twins*. Blood Adv, 2022. **6**(7): p. 2275-2289.
263. Bossis, G., et al., *The ROS/SUMO axis contributes to the response of acute myeloid leukemia cells to chemotherapeutic drugs*. Cell Rep, 2014. **7**(6): p. 1815-23.

264. Valima, E., et al., *SUMOylation inhibition potentiates the glucocorticoid receptor to program growth arrest of acute lymphoblastic leukemia cells*. *Oncogene*, 2025. **44**(18): p. 1259-1271.
265. Yu, H., et al., *Targeting NF-kappaB pathway for the therapy of diseases: mechanism and clinical study*. *Signal Transduct Target Ther*, 2020. **5**(1): p. 209.
266. Hsieh, M.Y. and R.A. Van Etten, *IKK-dependent activation of NF-kappaB contributes to myeloid and lymphoid leukemogenesis by BCR-ABL1*. *Blood*, 2014. **123**(15): p. 2401-11.
267. Pane, F., et al., *BCR/ABL genes and leukemic phenotype: from molecular mechanisms to clinical correlations*. *Oncogene*, 2002. **21**(56): p. 8652-67.
268. Feng, J., et al., *Childhood Acute B-Lineage Lymphoblastic Leukemia With CDKN2A/B Deletion Is a Distinct Entity With Adverse Genetic Features and Poor Clinical Outcomes*. *Front Oncol*, 2022. **12**: p. 878098.
269. Liu, H., et al., *Association Between NR3C1 Mutations and Glucocorticoid Resistance in Children With Acute Lymphoblastic Leukemia*. *Front Pharmacol*, 2021. **12**: p. 634956.
270. Mondal, G., et al., *A requirement for STAG2 in replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers*. *Nat Commun*, 2019. **10**(1): p. 1686.
271. Haas, S., A. Trumpp, and M.D. Milsom, *Causes and Consequences of Hematopoietic Stem Cell Heterogeneity*. *Cell Stem Cell*, 2018. **22**(5): p. 627-638.
272. Qiu, Y. and S. Huang, *CTCF-mediated genome organization and leukemogenesis*. *Leukemia*, 2020. **34**(9): p. 2295-2304.
273. Viny, A.D., et al., *Cohesin Members Stag1 and Stag2 Display Distinct Roles in Chromatin Accessibility and Topological Control of HSC Self-Renewal and Differentiation*. *Cell Stem Cell*, 2019. **25**(5): p. 682-696 e8.
274. Mar, B.G., et al., *Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia*. *Nat Commun*, 2014. **5**: p. 3469.
275. Wang, N., T. Ma, and B. Yu, *Targeting epigenetic regulators to overcome drug resistance in cancers*. *Signal Transduct Target Ther*, 2023. **8**(1): p. 69.
276. Stief, S.M., et al., *Loss of KDM6A confers drug resistance in acute myeloid leukemia*. *Leukemia*, 2020. **34**(1): p. 50-62.
277. Yang, M., et al., *13q12.2 deletions in acute lymphoblastic leukemia lead to upregulation of FLT3 through enhancer hijacking*. *Blood*, 2020. **136**(8): p. 946-956.
278. Aydin, E., et al., *Discovery of Cis-Regulatory Mechanisms via Non-Coding Mutations in Acute Lymphoblastic Leukemia*. *Genes Chromosomes Cancer*, 2025. **64**(3): p. e70045.
279. Moura-Castro, L.H., et al., *The 3D genome of pediatric B-cell precursor acute lymphoblastic leukemia*. *bioRxiv*, 2025: p. 2025.07.22.666073.
280. Polonen, P., et al., *The genomic basis of childhood T-lineage acute lymphoblastic leukaemia*. *Nature*, 2024. **632**(8027): p. 1082-1091.
281. Li, K., et al., *Noncoding Variants Connect Enhancer Dysregulation with Nuclear Receptor Signaling in Hematopoietic Malignancies*. *Cancer Discov*, 2020. **10**(5): p. 724-745.
282. Vollger, M.R., et al., *Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads*. *Ann Hum Genet*, 2020. **84**(2): p. 125-140.
283. Pandey, R., et al., *A meta-analysis of diagnostic yield and clinical utility of genome and exome sequencing in pediatric rare and undiagnosed genetic diseases*. *Genet Med*, 2025. **27**(6): p. 101398.
284. Hodder, A., et al., *Benefits for children with suspected cancer from routine whole-genome sequencing*. *Nat Med*, 2024. **30**(7): p. 1905-1912.
285. Liao, W.W., et al., *A draft human pangenome reference*. *Nature*, 2023. **617**(7960): p. 312-324.
286. Porubsky, D. and E.E. Eichler, *A 25-year odyssey of genomic technology advances and structural variant discovery*. *Cell*, 2024. **187**(5): p. 1024-1037.
287. Olson, N.D., et al., *Variant calling and benchmarking in an era of complete human genome sequences*. *Nat Rev Genet*, 2023. **24**(7): p. 464-483.
288. Schloissnig, S., et al., *Structural variation in 1,019 diverse humans based on long-read sequencing*. *Nature*, 2025.
289. Logsdon, G.A., et al., *Complex genetic variation in nearly complete human genomes*. *Nature*, 2025. **644**(8076): p. 430-441.

-
290. Mani, S., S.R. Lalani, and M. Pammi, *Genomics and multiomics in the age of precision medicine*. *Pediatr Res*, 2025. **97**(4): p. 1399-1410.
 291. Esmailzadeh, P., *Challenges and strategies for wide-scale artificial intelligence (AI) deployment in healthcare practices: A perspective for healthcare organizations*. *Artif Intell Med*, 2024. **151**: p. 102861.

List of figures

Figure 1: Types of human genetic variation.....	14
Figure 2: Technical approaches to detect genomic variation: from SNV to aneuploidy.....	16
Figure 3: Technical workflow of optical genome mapping.....	18
Figure 4: Optical genome mapping pipelines for detection of structural variants and copy number alterations.....	19
Figure 5: Technical workflow to generate SMRT long-read sequencing data.....	20
Figure 6: Frequency, incidence and survival of pediatric tumor diagnoses in Germany.....	23
Figure 7: Leukemia subtypes resulting from distinct hematopoietic compartments.....	27
Figure 8: Prevalence and prognosis of genetic subgroups in pediatric BCP-ALL.....	29
Figure 9: Integration of OGM and NGS-based methods to identify recurrently altered genes and double hits in potential tumor suppressor genes.....	93
Figure 10: Future era of genomics in pediatric oncology.....	101

List of tables

Table 1: Cancer predisposition genes in pediatric malignancies.....	25
Table 2: Prognostic factors influencing favorable and unfavorable outcomes in ALL.....	34

APPENDIX

PUBLICATIONS

Included in the thesis:

Wagener R*, **Brandes D***, Jung M, Huetzen MA, Bergmann AK, Panier S, Picard D, Fischer U, Jachimowicz RD, Borkhardt A, Brozou T. Optical genome mapping identifies structural variants in potentially new cancer predisposition candidate genes in pediatric cancer patients. *Int J Cancer*. 2024 Feb 15;154(4):607-614. doi: 10.1002/ijc.34721. Epub 2023 Sep 30. PMID: 37776287.

Brandes D, Yasin L, Nebral K, Ebler J, Schinnerl D, Picard D, Bergmann AK, Alam J, Köhrer S, Haas OA, Attarbaschi A, Marschall T, Stanulla M, Borkhardt A, Brozou T, Fischer U, Wagener R. Optical Genome Mapping Identifies Novel Recurrent Structural Alterations in Childhood *ETV6::RUNX1+* and High Hyperdiploid Acute Lymphoblastic Leukemia. *Hemasphere*. 2023 Jul 17;7(8):e925. doi: 10.1097/HS9.0000000000000925. PMID: 37469802; PMCID: PMC10353714.

Other publications:

Daugis K, **Brandes D**, Wagener R, Yasin L; Anwar A, Alam J, Prasad Y, Bartrina y Manns J, Mescher M, Fischer U, Borkhardt A, Brozou T, Junk, SV. Germline variants observed in pediatric cancer patients related to hereditary breast and ovarian cancer in adults. *Int J Cancer*. 2025 Aug 26. doi: 10.1002/ijc.70097. Epub ahead of print. PMID: 40857049.

Schliehe-Diecks J*, Tu J*, Stachura P, Schaal K, Kemkes M, Vasileiou E, Rüchel N, **Brandes D**, Vogt M, Lenz T, Nair A, Scheu S, Dominguez PM, Pastorczak A, Nebral K, Stühler K, Fischer U, Pandya A; Borkhardt A, Bhatia S. Epigenetic remodeling via HDAC6 inhibition amplifies anti-tumoral immune responses in myeloid leukemia cells. *Cell Death&Disease*. 2025 (accepted).

Wagener R, Walter C, Surowy HM, **Brandes D**, Soura S, Alzoubi D, Yasin L, Fischer U, Dugas M, Borkhardt A, Brozou T. Noncancer-related Secondary Findings in a Cohort of 231 Children With Cancer and Their Parents. *J Pediatr Hematol Oncol*. 2023 Mar 1;45(2):e244-e248. doi: 10.1097/MPH.0000000000002475. Epub 2022 May 3. PMID: 35537032.

Brozou T, Yasin L, **Brandes D**, Picard D, Walter C, Varghese J, Dugas M, Fischer U, Borkhardt A, Haas OA. Resolving inherited and *de novo* germline predisposing sequence variants by means of whole-exome trio analyses in childhood hematological malignancies. *Front Pediatr*. 2023 Feb 7;10:1080347. doi: 10.3389/fped.2022.1080347. PMID: 36824296; PMCID: PMC9941195.

Zipper L, Wagener R, Fischer U, Hoffmann A, Yasin L, **Brandes D**, Soura S, Anwar A, Walter C, Varghese J, Hauer J, Auer F, Bhatia S, Dugas M, Junk SV, Stanulla M, Haas OA, Borkhardt A, Reiff T, Brozou T. Hyperdiploid acute lymphoblastic leukemia in children with *LZTR1* germline variants. *Hemasphere*. 2024 Jan 26;8(1):e26. doi: 10.1002/hem3.26. PMID: 38434521; PMCID: PMC10878188.

Dietz A, Subedi P, Azimzadeh O, Duchrow L, Kaestle F, Paetzold J, Katharina Payer S, Hornhardt S, von Toerne C, Hauck SM, Kempkes B, Kuklik-Roos C, **Brandes D**, Borkhardt A, Moertl S, Gomolka M. The Chromosome Passenger Complex (CPC) Components and Its Associated Pathways Are Promising Candidates to Differentiate Between Normosensitive and Radiosensitive ATM-Mutated Cells. *Biomark Insights*. 2024 Oct 30;19:11772719241274017. doi: 10.1177/11772719241274017. PMID: 39493730; PMCID: PMC11528597.

SCIENTIFIC PRESENTATIONS

Poster – Childhood Leukemia and Lymphoma Symposium Valencia (CLLS) Budapest, “Decoding pediatric hyperdiploid B-cell precursor acute lymphoblastic leukemia through long-read sequencing”, 05/2025

Talk – GPOH Meeting Frankfurt, „*Ataxia telangiectasia mutated (ATM)* Gen Veränderungen in krebskranken Kindern mit lymphatischen Malignomen“, 11/2024

Talk – Bionano Genomics Webinar *The Pathologist*, “Addressing the incomplete mutational landscape in pediatric cancers with optical genome mapping”, 03/2024

Poster - EMBO conference Düsseldorf, “Exploring *ATM* alterations in pediatric cancer patients with lymphoid malignancies by next-generation sequencing and optical genome mapping”, 05/2024

Talk – Bionano User Meeting Berlin, “Die Rolle von Optical Mapping in der Identifizierung von strukturellen Keimbahnvarianten bei pädiatrischen Patienten mit Malignomen“, 11/2023

Poster – CLLS Valencia, “Genomic landscape of pediatric B-cell precursor acute lymphoblastic leukemia analyzed by optical genome mapping and next-generation sequencing”, 05/2023

Talk – DSO-Retreat Düsseldorf, “The genomic landscape of pediatric B-cell precursor acute lymphoblastic leukemia deciphered by optical genome mapping and next-generation sequencing“, 05/2023

Poster – American Society of Hematology (ASH) Annual Meeting New Orleans, “Deciphering the somatic and germline structural variation landscape in pediatric B-cell precursor acute lymphoblastic leukemia by whole genome optical mapping“, 12/2022

Poster – European Hematology Association (EHA) Annual Meeting Vienna, “Employing whole-genome optical mapping to characterize the landscape of structural variants in B precursor acute lymphoblastic leukemia“, 06/2022

Talk – DSO-Retreat Düsseldorf, “Employing optical mapping and long-read Sequencing to identify structural variations in acute lymphoblastic leukemia predisposition and pathogenesis“, 05/2022

Talk – Society for Paediatric Oncology (SIOP) Europe Annual Meeting online, “Employing optical mapping and long-read Sequencing to identify structural variations in acute lymphoblastic leukemia predisposition and pathogenesis“, 03/2022

Talk – Leukemia Day Düsseldorf, “Employing optical mapping and long-read sequencing to identify structural variations involved in cancer pathogenesis and predisposition“, 09/2021

Talk – DSO-Retreat online, “Status quo and perspective“, 05/2021

Talk – DTK Meeting Precision Prevention in Pediatric Oncology Dresden, “Deciphering structural variants predisposing to cancer by optical mapping and long-read sequencing“, 09/2020

FUNDING & AWARDS

Research project funding, Gesellschaft für Kinderkrebsforschung (GfKKF), „Long-read Sequenzierung und Medikamentenscreening zur Identifikation neuer therapeutischer Targets bei der aneuploiden Leukämie des Kindesalters“, 01/2025

Young investigator award, Gesellschaft für pädiatrische Onkologie und Hämatologie (GPOH), „*Ataxia telangiectasia mutated (ATM)* Gen Veränderungen in krebskranken Kindern mit lymphatischen Malignomen“, 11/2024

Grand Poster Prize Winner, Bionano Genomics, “Comprehensive Analysis of B-Cell Precursor Acute Lymphoblastic Leukemia: Insights from Optical Genome Mapping and Next-Generation Sequencing”, 02/2024

Young investigator award, Elterninitiative Kinderkrebsklinik e.V. Düsseldorf, “Optical Genome Mapping Identifies Novel Recurrent Structural Alterations in Childhood *ETV6::RUNX1+* and High Hyperdiploid Acute Lymphoblastic Leukemia”, 11/2023

HeRa travel grant, Heinrich-Heine-Universität Düsseldorf, 12/2022

Clinical Research SMRT Grant, Pacific Biosciences, 07/2021

Graduate program, Düsseldorf School of Oncology (DSO), Uniklinik Düsseldorf, 2020-2023

ACKNOWLEDGMENTS

Prof. Dr. med. Arndt Borkhardt – I would like to express my gratitude for the opportunity to work on this project; and for his supervision, guidance, encouragement and support throughout my PhD journey. His expertise and feedback have been essential in shaping this work, and fostering my academic growth.

Prof. Dr. Sebastian Wesselborg – I would like to express my gratitude for being the second supervisor of this PhD thesis.

Dr. Ute Fischer – I would like to express my gratitude for her guidance and encouragement throughout my PhD time. I am thankful for her insights and constructive feedback that contributed to the quality of this thesis. Ute, thank you for all the shared memories including dog walks and plant exchanges.

Dr. Rabea Wagener – I am sincerely grateful for her excellent guidance in the beginning of my PhD journey. My scientific knowledge is based on her insights, teaching skills and honest feedback. Working together with her shaped the researcher I've become. Rabea, thank you for being such an inspiring example and pushing through (m)any challenges together with me.

I would like to thank the Germline Study working group with all present and former members - Filio, Rabea, Layal, Stefanie, Katharina, Jil, Ammarah, Jubayer, Stavrieta, Marie and Anna - for the fruitful conversations and collaboration on cancer predisposition research. Thank you for being part of my PhD experience and contributing to my PhD work.

I would like to express my gratitude to all collaboration partners that contributed to the work of this thesis – especially Dr. Karin Nebral, Dr. Ron Jachimowicz and Maxim Hützen.

I would like to thank the Fischer working group and all other KMT-lab members, I am grateful for the amazing and collegial work atmosphere we built together. It is very special to have such a welcoming group that supports everyone throughout good and bad times. Special thanks to Silke, Daniel and Kati for assisting with any technical lab work that contributed to my thesis.

I would like to thank my favorite OG3 office with all current and former members - Vera, Ersen, Nadine, Annika and Ciro – for the welcoming atmosphere, endless laughs, scientific discussions, emotional support and nurturing coffee/cake breaks. Thanks to you my PhD time felt like coming home.

Vera – I especially want to thank you for being my emotional support animal. I don't want to miss any shared moment with you throughout the past years. I am happy to have gained such a valuable friendship.

Finally, I would like to thank my sister, (grand)parents, friends and partner for being patient and understanding during my PhD time. You helped keeping my work-life balance (relatively) intact. Thank you for loving me as I am and being such wonderful supporters the past years. This thesis is dedicated to my grandmother, who taught me optimism and resilience, qualities that have guided me throughout my PhD journey.