

LUND UNIVERSITY

Uncovering the Subclonal Genetic Landscape of KMT2A-Rearranged Leukemia

Ahlgren, Louise

2026

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Ahlgren, L. (2026). *Uncovering the Subclonal Genetic Landscape of KMT2A-Rearranged Leukemia*. [Doctoral Thesis (compilation), Department of Laboratory Medicine]. Lund University, Faculty of Medicine.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

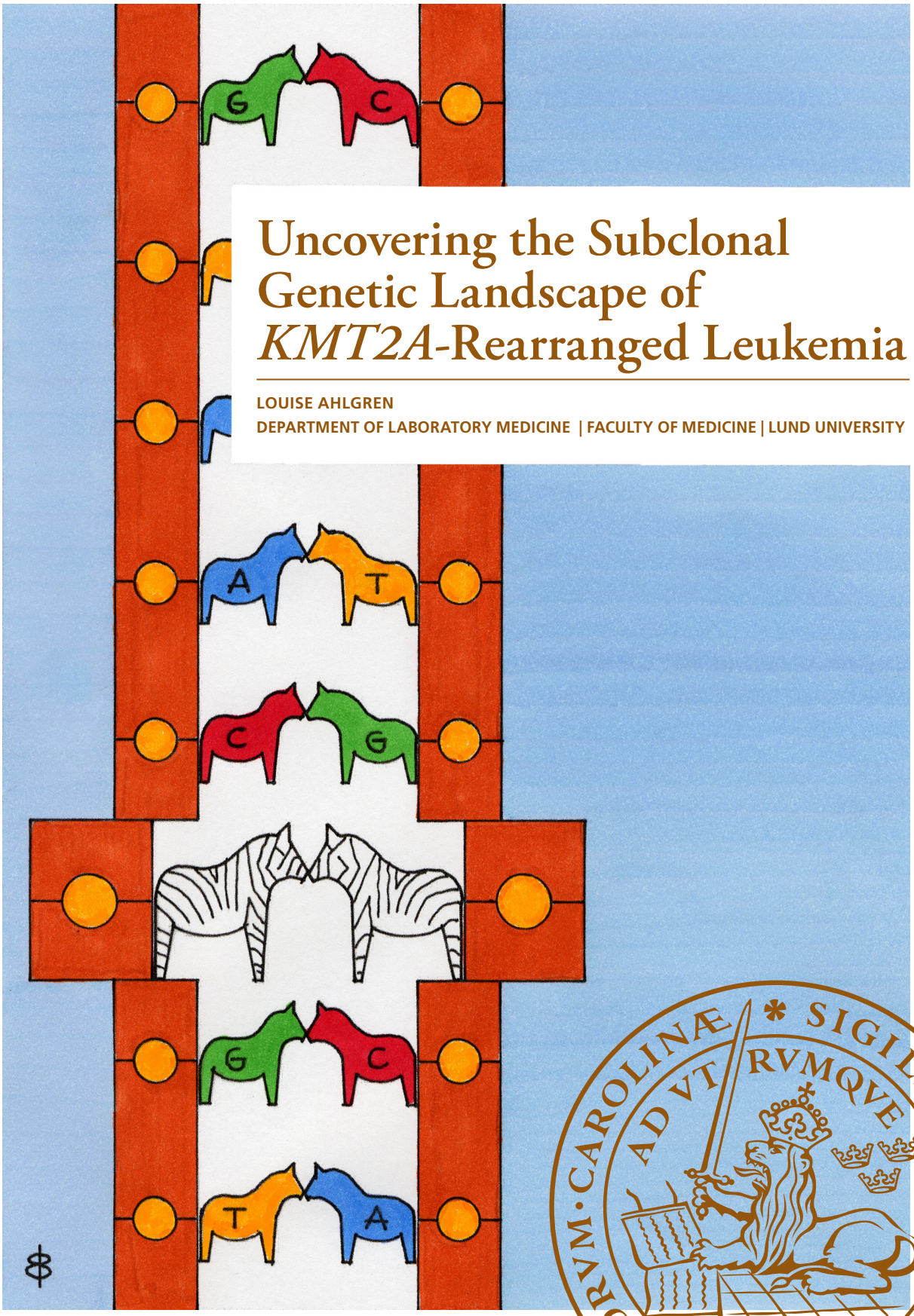
Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Download date: 28. okt.. 2026



Uncovering the Subclonal Genetic Landscape of *KMT2A*-Rearranged Leukemia

LOUISE AHLGREN
DEPARTMENT OF LABORATORY MEDICINE | FACULTY OF MEDICINE | LUND UNIVERSITY

Uncovering the Subclonal Genetic Landscape of *KMT2A*-Rearranged Leukemia

Louise Ahlgren



LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on Friday the 13th of November 2026 at 09.00 in Lundmarkssalen, Astronomihuset, Sölvegatan 27, Lund.

Faculty opponent
Professor Julie Irving, PhD, Newcastle University, England

Organization: LUND UNIVERSITY, Faculty of Medicine, Division of Clinical Genetics

Document name: Doctoral dissertation **Date of issue** 2026-11-13

Author: Louise Ahlgren **Sponsoring organization:**

Title and subtitle: Uncovering the Subclonal Genetic Landscape of *KMT2A*-Rearranged Leukemia

Abstract:

Acute leukemia occurs across all age groups and is the most common malignancy in children. It is characterized by the expansion of immature hematopoietic cells and is classified as acute lymphoblastic (ALL), acute myeloid (AML), or mixed-phenotype acute leukemia (MPAL). Despite major improvements in pediatric outcomes, prognosis remains poor for high-risk subtypes, including *KMT2A*-rearranged (*KMT2A*-r) leukemia. *KMT2A*-r occurs across all ages but is particularly common in infants, accounting for approximately 70-80% of infant ALL and 35-50% of infant AML. The overall aim of this thesis was to increase our understanding of the genetic basis of *KMT2A*-r leukemia development and relapse.

In **Article I**, whole genome, whole exome and longitudinal deep sequencing were used to study clonal dynamics and the genetic mechanisms driving relapse in *KMT2A*-r infant and childhood leukemia. This revealed different relapse mechanisms in AML and ALL, with AML relapse predominantly associated with the persistence and expansion of pre-existing resistant clones. In ALL, relapse mechanisms depended on timing. Very early relapse (<9 months) arose from multiple pre-existing clones that survived treatment, whereas early relapse (≥9 months) was characterized by greater treatment-associated evolution and the acquisition of chemoresistance-associated alterations, including *IKZF1* and *TP53*.

Article II used highly sensitive duplex sequencing to reveal a vast, previously hidden subclonal genetic landscape in pediatric *KMT2A*-r ALL, including recurrent cancer-associated mutations largely undetectable by conventional sequencing. Paired diagnosis and relapse samples showed that relapse could arise from both the expansion of pre-existing subclones and the acquisition of new alterations during treatment, including *TP53*, while many mutation-bearing subclones remained confined to minor populations, indicating that mutation acquisition alone does not determine clonal selection.

In Article III, whole genome, whole exome, whole genome bisulfite and RNA sequencing were integrated to comprehensively characterize a Nordic cohort of *KMT2A*-r leukemia. This showed age-related genetic evolution, with adults displaying a higher mutational burden and a more clonal genetic landscape. Genetic alterations were associated with leukemia lineage, *KMT2A* fusion partner and age, suggesting that cooperating lesions may reinforce lineage identity. RNA sequencing identified distinct transcriptional states linked to cellular differentiation and clinical outcome, while whole genome sequencing uncovered extensive structural complexity of *KMT2A* rearrangements, including cryptic and novel fusion genes.

In conclusion, by resolving both large-scale genomic patterns and rare subclonal populations, this thesis demonstrates how age, lineage, fusion partner, clonal architecture and treatment shape the evolution and relapse of *KMT2A*-r leukemia, with implications for improved risk stratification and treatment.

Key words: acute lymphoblastic leukemia, acute myeloid leukemia, pediatric leukemia, *KMT2A*, relapse, clonal evolution, treatment resistance, subclone, RAS, sequencing

Classification system and/or index terms (if any) Supplementary bibliographical information

Language English **Number of pages**: 152

ISSN and key title: 1652-8220 **ISBN:** 978-91-8021-944-0

Recipient's notes Price Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Date 2026-10-05

Uncovering the Subclonal Genetic Landscape of *KMT2A*-Rearranged Leukemia

Louise Ahlgren



LUND
UNIVERSITY

Copyright Louise Ahlgren
Pages 1-152 © 2026 Louise Ahlgren, ORCID: 0000-0001-7335-7535.
Paper 1 © 2025 Nature Communications
Paper 2 © 2022 HemaSphere
Paper 3 © 2026 The authors (Manuscript unpublished)

Cover art by Eva and Jan Bergfelt (licensed under CC BY 4.0)
Back cover photo by Madelaine Ahlgren

Department of Laboratory Medicine, Lund
Division of Clinical Genetics
Faculty of Medicine
Lund University

ISBN 978-91-8021-944-0
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University,
Lund, 2026



Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se
MADE IN SWEDEN

I am so clever that sometimes I do not understand a single word of what I am saying

Oscar Wilde

Table of Contents

Original articles and manuscripts **9**

Abbreviations **10**

Introduction **12**

Cancer 12
Cancer cell evolution 12
DNA structure, genetic alterations and gene regulation 14
Methods to study cancer 22
Bioinformatics 24
Normal hematopoiesis 25
Hematopoietic hierarchy 26
Regulation of hematopoiesis 27
Hematological malignancies 29
Leukemia 29
Leukemia development 30
Prenatal origin of acute leukemia 32
Pediatric acute lymphoblastic leukemia 33
Pediatric acute myeloid leukemia 38
Mixed phenotype acute leukemia 41
Secondary genetic alterations in leukemia 42
Genetic predisposition 48
Pediatric versus adult malignancies 48
KMT2A-rearranged leukemia 50
Prevalence 50
The normal function of the *KMT2A* gene 51
Fusion partners in *KMT2A*-r leukemia 52
The cellular origin of *KMT2A*-r leukemia 56
Gene expression and cellular states in *KMT2A*-r leukemia 57
Mouse models of *KMT2A*-r leukemia 59
The clinic 63
Symptoms and diagnosis 63
Measurable residual disease 64
Risk classification in pediatric acute leukemia 66

Prognosis 67
Treatment 69
Treatment sequelae 75
Relapse 75

The present investigation 77

Aims of the study 77

Materials and methods 79
Sample collection and processing 79
Polymerase chain reaction 79
Whole genome sequencing 80
Whole exome sequencing 81
Whole genome bisulfite sequencing 81
RNA sequencing 81
Single cell sequencing 82
Targeted sequencing 82
Duplex sequencing 83
Digital droplet PCR 84
CRISPR/Cas9 genome editing 84
Plasmid construction and lentivirus production 85
Western blotting 85
Drug sensitivity assessment 85
Mutational signatures 85
Mouse experiments 86
Statistical considerations 86
Alternative methodological approaches 87

Ethical considerations 87

Summary of articles 88
Article I 88
Article II 92
Article III 96

Conclusions 101

Discussion 103
Future perspectives 114
Clinical impact of this thesis 115
Limitations of the study 116

AI tools and software for illustrations 117

Populärvetenskaplig sammanfattning 118

Acknowledgements 122

References 128

Original articles and manuscripts

The thesis is based on the following articles:

Article I

Ahlgren, L., Pilheden, M., Sturesson, H., Song, G., Walsh, M.P., Yang, M., Maillard, M., Zhao, H., Cheng, Z., Singh, V., Castor, A., Pronk, C.J., Marquart, H.V., Lausen, B., Schneider, P., Barbany, G., Pokrovskaja Tamm, K., Abrahamsson, J., Lohi, O., Fogelstrand, L., Menendez, P., Pieters, R., Zhang, J., Lindkvist-Petersson, K., Yang, J.J., Gruber, T.A., Stam, R.W., Ma, J. & Hagström-Andersson, A.K. **The genomic landscape of relapsed infant and childhood *KMT2A*-rearranged acute leukemia.** *Nature Communications. 16, 8964 (2025).*

Article II

Pilheden, M., Ahlgren, L., Hyrenius-Wittsten, A., Gonzalez-Pena, V., Sturesson, H., Marquart, H.V.H., Lausen, B., Castor, A., Pronk, C.J., Barbany, G., Pokrovskaja Tamm, K., Fogelstrand, L., Lohi, O., Norén-Nyström, U., Asklin, J., Chen, Y., Song, G., Walsh, M., Ma, J., Zhang, J., Saal, L.H., Gawad, C. & Hagström-Andersson, A.K. **Duplex sequencing uncovers recurrent low-frequency cancer-associated mutations in infant and childhood *KMT2A*-rearranged acute leukemia.** *HemaSphere. 2022;6(10):e785.*

Article III

Pilheden, M.* & Ahlgren, L.*, Falques-Costa, T., Sturesson, H., Walsh, M.P., Song, G., Lahnalampi, M., Hafiichuk, K., Castor, A., Pronk, C.J., Marquart, H.V.H., Lausen, B., Barbany, G., Pokrovskaja Tamm, K., Hansson, M., Orsmark-Pietras, C., Liljebjörn, H., Norén-Nyström, U., Oskolkov, N., Cakirca, G., Uddin, K., Scheding, S., Heyman, M., Ehinger, M., Abrahamsson, J., Fogelstrand, L., Lohi, O., Hyrenius-Wittsten, A., Fioretos, T., Zhang, J., Downing, J.R., Yang, M., Paulsson, K., Ma, J. & Hagström-Andersson, A.K. **A genomic atlas of *KMT2A*-rearranged leukemia from infancy to adulthood.** Manuscript in preparation.

*These authors contributed equally and share first authorship.

Abbreviations

6-MP	6-mercaptopurine
A	adenine
ADC	antibody drug conjugate
AMKL	acute megakaryoblastic leukemia
ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
AML-MR	AML myelodysplasia-related
APL	acute promyelocytic leukemia
ATAC	assay for transposase-accessible chromatin
ATRA	all-trans retinoic acid
B-ALL	B-cell acute lymphoblastic leukemia
BAM	binary alignment map
BCL	binary base call
BCR1	major breakpoint cluster region
BCR2	minor breakpoint cluster region
BM	bone marrow
bp	base pair
C	cytosine
CAR	chimeric antigen receptor
CBF	core binding factor
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
CHIP	clonal hematopoiesis of indeterminate potential
CLP	common lymphoid progenitor
CML	chronic myeloid leukemia
CMP	common myeloid progenitor
CNA	copy number alteration
CN-LOH	copy-neutral loss of heterozygosity
CNS	central nervous system
CR	complete remission
CTP	cytoplasmic translocation partner
ddPCR	digital droplet PCR
DNA	deoxyribonucleic acid
DC	dendritic cell
ELP	early lymphocyte precursor
EPO	erythropoietin
ETP	early T-cell precursor
FC	flow cytometry
FISH	fluorescence in situ hybridization
G	guanine
G-CSF	granulocyte colony-stimulating factor
GMP	granulocyte monocyte progenitor
gRNA	guide RNA
GSEA	gene set enrichment analysis

HSC	hematopoietic stem cell
HSCT	hematopoietic stem cell transplantation
iAMP21	intrachromosomal amplification of chromosome 21
INDEL	insertion or deletion
KMT2A-r	*KMT2A*-rearranged
LSC	leukemic stem cell
LOH	loss of heterozygosity
MAPK	mitogen-activated protein kinase
MEP	megakaryocyte erythroid progenitor
MIF	macrophage migration inhibitory factor
MLL	mixed lineage leukemia
MPAL	mixed phenotype acute leukemia
MPN	myeloproliferative neoplasm
MPO	myeloperoxidase
MPP	multipotent progenitor
MRD	measurable residual disease
mRNA	messenger RNA
NGS	next generation sequencing
NHEJ	non-homologous end joining
NK-cell	natural killer cell
NOPHO	Nordic Society of Paediatric Haematology and Oncology
NTP	nuclear translocation partner
PAFc	polymerase-associated factor complex
PB	peripheral blood
PCR	polymerase chain reaction
p-ERK	phosphorylated ERK
pre-mRNA	precursor messenger RNA
PTA	primary template-directed amplification
PTD	partial tandem duplication
RB	retinoblastoma
RNA	ribonucleic acid
RNA-seq	RNA sequencing
SAM	sequence alignment map
SARs	scaffold attachment regions
SBCR	Swedish Childhood Cancer Registry
SBS	single base substitution
scDNA-seq	single cell DNA sequencing
scRNA-seq	single cell RNA sequencing
SEC	super elongation complex
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SV	structural variant
T	thymine
T-ALL	T-cell lineage ALL
TCGA	The Cancer Genome Atlas
TCR	T-cell receptor
TPO	thrombopoietin
TKI	tyrosine kinase inhibitor
UMI	unique molecular identifier
VAF	variant allele frequency
VCF	variant call format
WES	whole exome sequencing
WGBS	whole genome bisulfite sequencing
WGS	whole genome sequencing
WHO	World Health Organization

Introduction

Cancer

Life, in all its complexity, begins and continues through the most fundamental biological process: the division of a single cell. From the earliest stages of embryogenesis to the daily renewal of tissues, the human body relies on an astonishing number of cell divisions, each requiring the precise replication and division of genetic material. Every time this process occurs, the entire human genome, containing more than six billion base pairs and 20 000 genes, must be copied with remarkable accuracy. Although sophisticated molecular safeguards exist to correct errors, they are not perfect. Most mistakes are harmless or corrected, while some persist and accumulate. This biological plasticity is not a flaw but a fundamental condition for evolution and ultimately for life itself. However, when mutations affect key regulators of maturation, growth, death and/or DNA stability and arise in the right cellular context, they may confer selective advantages, promoting uncontrolled cell division.[1]

The word *cancer* is both feared and infamous, evoking anxiety and dread across societies. It is not a single disease, but rather a broad group of disorders characterized by the uncontrolled growth and spread of abnormal cells. As the second leading cause of death worldwide and a major cause of disease related mortality among children, cancer claims over 8 million lives each year, and alarmingly, its incidence is expected to rise by more than 50% in the coming decades.[2,3] Despite many years of research, cancer remains in many ways a black box, a complex and still partially unexplored territory of biology and medicine.[2]

This thesis is dedicated to hematological malignancies and seeks to further explore the genetic landscape, mechanisms of relapse and therapy resistance in a patient cohort with a poor prognosis. The following introduction aims to provide context for these studies.

Cancer cell evolution

Carcinogenesis, the transformation of normal cells into malignant cells, is a complicated, multistep process, for which numerous models of evolution have been proposed.[4] A combination of genetic, epigenetic and phenotypic changes together

with a permissive local microenvironment is likely required and collectively contributes to the *heterogeneity* of tumours.[4-6] Common traits acquired during carcinogenesis have been defined as *hallmarks of cancer*, an expanded framework comprising eight hallmark capabilities and two enabling characteristics, with four additional emerging dimensions proposed in 2022 (**Figure 1**).[6-8]

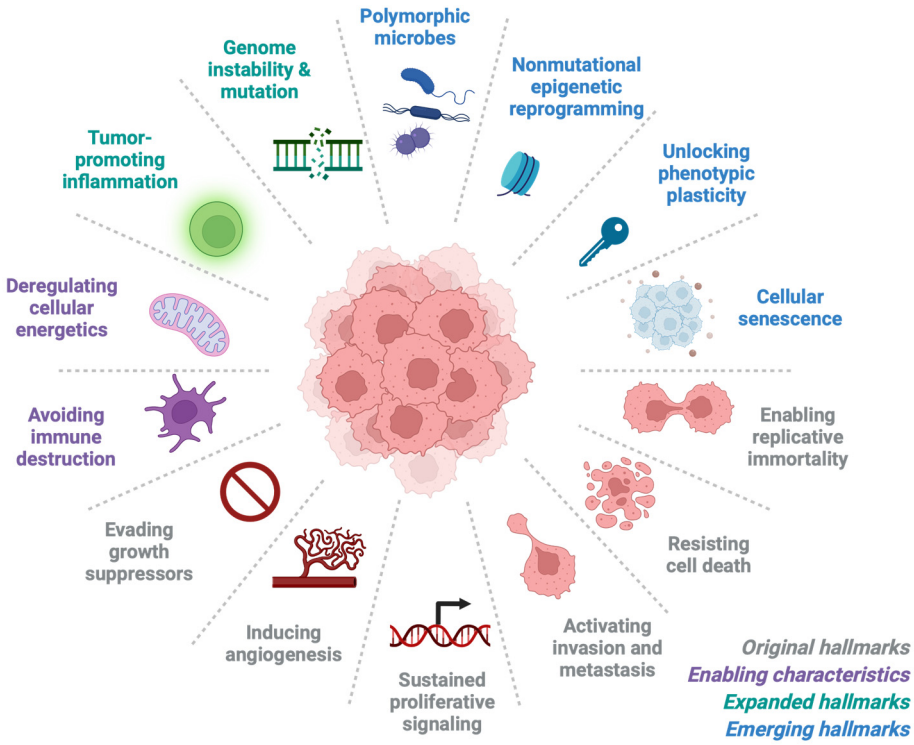


Figure 1: Hallmarks of cancer. The original six hallmarks described in 2000 were expanded with two additional hallmarks and two enabling characteristics in 2011 and further expanded in 2022 to encompass fourteen biological capabilities of cancer. Adapted from Hanahan and Weinberg (2000, 2011) and Hanahan (2022). Illustration created with BioRender.com.

Through the gradual accumulation of genetic alterations, cancer evolves through repeated cycles of mutation, clonal expansion and selection. This process may begin with an initiating *driver* alteration that promotes cell survival or proliferation and/or disrupts normal differentiation or cellular identity, allowing the affected cell to persist and form a microscopic *clone*. As the clone expands, additional genetic alterations accumulate, most of which are harmless *passenger* mutations. Occasionally, however, a cell acquires an additional driver alteration that confers a selective advantage, triggering a new wave of clonal expansion and the emergence of a new *subclone* (**Figure 2**).[9,10]

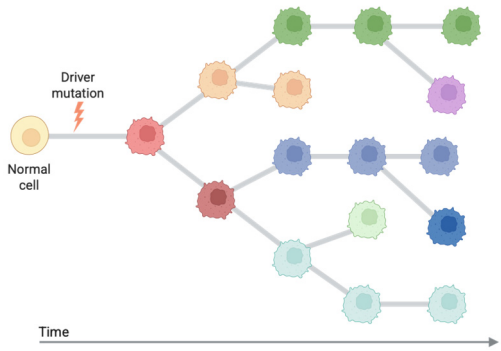


Figure 2: Clonal evolution during tumour development. Illustration of a normal cell acquiring a driver mutation, followed by clonal expansion and the emergence of genetically distinct subclones. Illustration created with BioRender.com.

As clonal evolution continues, subsequent driver alterations can emerge in expanding subclones, further increasing the genetic diversity of the tumour. Some driver genes are recurrently altered across many cancer types (so-called *mountains*, e.g. *KRAS* and *TP53*), whereas many others are altered at lower frequencies or within specific tumour types (so-called *hills*).[10]

Driver genes include activated *oncogenes* and inactivated *tumour suppressor genes*. For oncogenes (e.g. genes regulating growth factors, signal transduction molecules and transcription factors), alteration of a single allele is usually sufficient, whereas tumour suppressor genes (e.g. genes regulating apoptosis, cell growth and proliferation) typically require inactivation of both alleles. Persistent activation of proto-oncogenes or loss-of-function mutations and deletions in tumour suppressor genes can lead to a sustained state of cellular proliferation.[8]

Viewed through the lens of Darwinian evolution, carcinogenesis reflects a process of natural selection occurring at the cell level.[11] Some of these alterations provide a selective advantage, allowing certain cells to outcompete others.[12] This drives the formation of *subclones*, each with different genetic profiles. The tumour's subclonal architecture is continually shaped by internal dynamics such as competition or cooperation between clones, spatial limitations within the tissue and external pressures, including therapy.[13]

DNA structure, genetic alterations and gene regulation

A normal human cell consists of 46 chromosomes, arranged in 22 pairs of *autosomes* and one pair of *sex chromosomes*, XX in females and XY in males. Thus, each cell is *diploid*, meaning that it has two copies of each chromosome. The chromosomes are divided into two arms, the shorter p-arm (*petite*) and the longer q-arm (simply

because q follows p alphabetically), bound together by a *centromere*. Each chromosome contains one long double-stranded deoxyribonucleic acid (DNA) molecule wrapped around proteins known as *histones*. Multiple histones, together with the wrapped DNA, form a fundamental unit called a *nucleosome*. These nucleosomes coil and fold to create a structure called *chromatin*, which is condensed into chromosomes during cell division (**Figure 3**).[5]

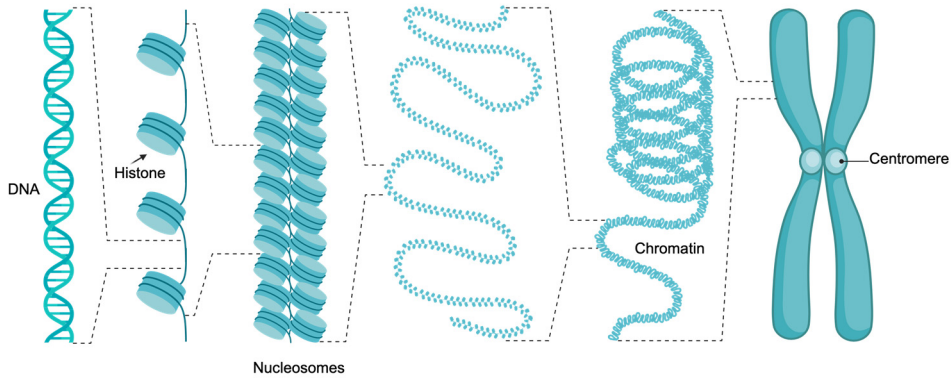


Figure 3: Levels of DNA organisation. Illustration of how DNA is packed into chromosomes. The DNA wraps around histones to form nucleosomes, which then coil into chromatin and finally form the chromosome. Illustration created with BioRender.com.

The DNA molecule, carrying our genetic code in the cell nucleus, is a double-stranded helix consisting of a long sequence of four different nucleotide bases with sugar and phosphate groups. These four bases are divided into *purines*, adenine (A) and guanine (G) and *pyrimidines*, cytosine (C) and thymine (T), with A pairing with T and C pairing with G. Three consecutive nucleotides form a *codon*, which specifies either an amino acid or a translation start or stop signal. Because there are 64 possible codons but only 20 amino acids, multiple codons encode the same amino acid. Amino acids are linked together to form proteins, which are essential for virtually all cellular functions.

DNA contains genes, which consist of exons (coding regions) and introns (non-coding sequences between exons), as well as intergenic regions. Perhaps unexpectedly, exons account for less than 2% of the human genome.[14] Before a cell divides through *mitosis*, its DNA is replicated to ensure that each daughter cell receives a complete copy of the genome. When a cell needs to produce a certain protein, the corresponding region of DNA is transcribed by *RNA polymerase* into single stranded precursor messenger RNA (pre-mRNA). Before it is transported to the cytoplasm, the pre-mRNA undergoes *splicing,* during which the introns are removed and exons are joined together, forming mature messenger RNA (mRNA). The mRNA is then exported to the cytoplasm, where it is *translated* by ribosomes into a chain of amino acids, creating a protein (**Figure 4**).

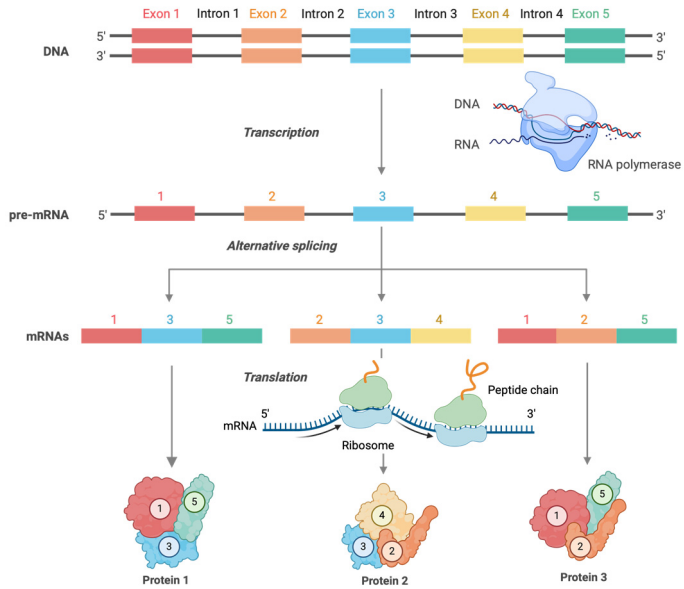


Figure 4: Overview of transcription, alternative splicing and translation. Illustration showing how genetic DNA undergoes transcription by RNA polymerase to produce pre-mRNA. Through alternative splicing, different combinations of exons are joined together, resulting in three mRNA transcripts from a single gene. Finally, these mRNAs undergo translation, where ribosomes decode the mRNA sequences to synthesize three structurally and functionally unique proteins. Illustration created with BioRender.com.

During DNA replication and cell division, genetic alterations may arise because of replication errors, reactive oxygen species or environmental factors such as radiation and chemical exposure.[6] Under normal conditions, several DNA repair mechanisms, including proofreading by DNA polymerases and the mismatch repair system, maintain genomic integrity by correcting most of these errors. However, these repair processes are not completely reliable, allowing some alterations to persist and become permanently incorporated into the genome. Such alterations range from single nucleotide changes to large scale abnormalities affecting chromosome segments or entire chromosomes. The major classes of genomic alterations are described below, from the smallest to the largest.

Single nucleotide changes

A single nucleotide variant (SNV) is a change affecting a single nucleotide and can be either inherited (germline) or acquired during life (somatic). A somatic SNV can arise when a nucleotide on one DNA strand is altered or incorrectly incorporated. If the change escapes DNA repair, the next DNA replication can introduce a complementary base on the opposite strand, establishing a stable altered base pair that is passed on to daughter cells. Germline SNVs present in at least 1% of the population are referred to as single nucleotide polymorphisms (SNPs). While most

SNPs are biologically neutral, some are associated with an increased risk of developing cancer or other diseases.[15]

SNPs and SNVs can occur throughout the genome. If a variant occurs in an exon, it can be either *nonsynonymous* or *synonymous*, where a nonsynonymous mutation causes a change in the amino acid sequence, whereas a synonymous variant does not. Curiously, a synonymous variant can still have functional consequences, as it can affect splicing, mRNA stability or the rate of translation, potentially influencing protein folding and function.[16] Nonsynonymous variants, however, result in a different codon that changes the protein. Nonsynonymous mutations can be divided into *missense* mutations, where amino acids are changed, or *nonsense* mutations, where a premature stop codon is introduced, resulting in a truncated protein. Both missense and nonsense mutations can result in dysfunctional or nonfunctional proteins (**Figure 5**).

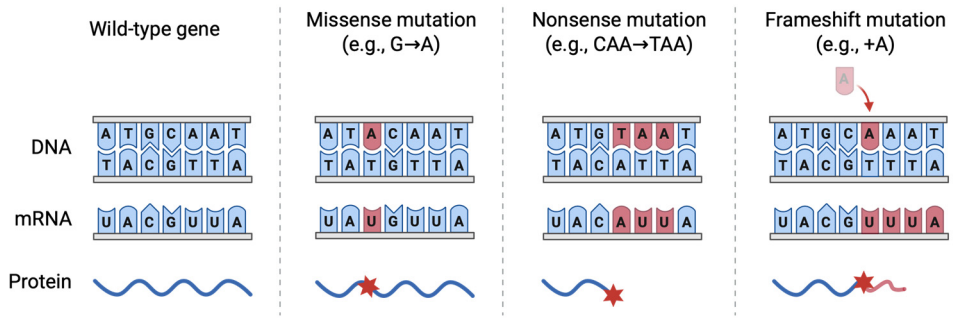


Figure 5: Different SNVs. Illustration of missense, nonsense and frameshift mutations compared to the wildtype sequence on the DNA, mRNA and protein levels. Red indicates the specific nucleotide change or mutation site. For the nonsense and frameshift mutations, the alteration in the genetic code can lead to a truncated protein product. Illustration created with BioRender.com.

If a variant occurs within an intron, the amino acid sequence is usually not directly affected, since introns are removed during RNA splicing. Still, intronic variants can disrupt splice sites or other regulatory elements, resulting in aberrant mRNA transcripts and altered proteins that can contribute to cancer.[17] Although intronic and intergenic regions do not encode proteins, they can contain regulatory elements that control gene expression through *promoters*, *enhancers* and chromatin organisation.[18] Alterations in these regions may therefore disrupt regulatory networks and contribute to cancer development, although their functional effects can be more difficult to identify using conventional approaches.

Insertions and deletions

When one or more nucleotides are added to DNA, this is called an *insertion*, while the removal of one or more nucleotides is called a *deletion*. Together, these changes are called *indels*. Indels occurring in multiples of three nucleotides preserve the

reading frame (*in frame* indels) but result in the insertion or deletion of one or more amino acids, which can alter protein structure or function. In contrast, indels that are not in multiples of three disrupt the reading frame (*frameshift* mutations), typically resulting in an altered amino acid sequence downstream of the mutation and often introducing a premature stop codon, leading to a truncated protein (**Figure 5**). Larger insertions and deletions, typically exceeding 50 base pairs, are generally classified as *structural variants* (SVs).

Structural variants

SVs are larger rearrangements in our DNA. They include deletions, duplications, insertions, inversions and translocations, and they impact the genome in several ways. Firstly, they can directly change gene dosage. Duplications add extra copies of a gene, which can cause the body to produce too much of a certain protein, while deletions remove genetic material, leading to a shortage or complete loss of that protein. Secondly, SVs can physically disrupt and destroy genes by breaking apart the coding sequences, making the genes non-functional. Finally, even if a gene's code remains untouched, SVs can rearrange the surrounding DNA and alter gene regulation. They can accidentally separate a gene from its natural "on/off switches" (enhancers and silencers) or move it next to a foreign switch, causing the gene to be turned on at the wrong time or in the wrong place. Thus, SVs can cause disease without affecting a single exon.[19] Below is a short description of common SV types.

Translocations

A translocation is a structural chromosomal abnormality in which a DNA segment breaks off and reattaches either elsewhere on the same chromosome (*intrachromosomal*) or to a different chromosome (*interchromosomal*). When such rearrangements disrupt two genes, a chimeric *fusion gene* may be generated with an altered gene function. Although interchromosomal translocations are the most common mechanism underlying fusion gene formation, fusion genes can also arise through intrachromosomal rearrangements, including inversions (180° reversal of a chromosome segment), insertions, tandem duplications (an additional copy of a chromosome segment inserted immediately adjacent to the original sequence) and, less frequently, deletions.[5]

Breakpoints commonly occur within introns, partly because introns are much longer than exons. In addition, intronic breakpoints can preserve the coding sequence of the exons, allowing the formation of an *in frame* fusion transcript that encodes a functional *fusion protein* (e.g. *BCR::ABL1*). In contrast, breakpoints within exons can disrupt the reading frame, resulting in a truncated or non-functional fusion protein. In some cases, the disrupted exon is skipped during RNA splicing, allowing fusion to a neighbouring exon and thereby preserving the reading frame.[20,21]

Following two DNA double-strand breaks and subsequent re-ligation, the regulatory elements (e.g. promoters or enhancers) of one gene become positioned adjacent to

another intact gene. Although no fusion protein is formed, the relocated regulatory elements can drive aberrant overexpression of the intact gene, thereby contributing to cancer development (e.g. *IGH::MYC*).[5,22]

Translocations can either be *reciprocal*, where chromosome segments are exchanged between non-homologous chromosomes, or *non-reciprocal*, where only one segment is moved without exchange. Reciprocal translocations are often *balanced,* meaning that no genetic material is lost or gained, while non-reciprocal translocations result in genetic imbalances that can have major biological consequences (**Figure 6**).[5]

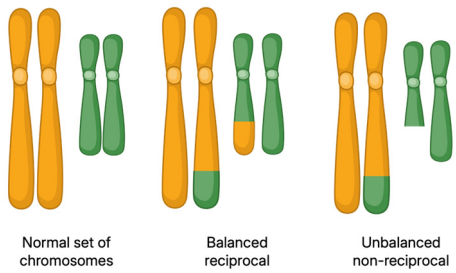


Figure 6: Chromosomal translocations. Examples of balanced reciprocal and unbalanced non-reciprocal translocations. Balanced translocations exchange genetic material without net loss or gain, whereas unbalanced translocations result in gain and/or loss of chromosomal material. Illustration created with BioRender.com.

Fusion genes can also arise at the RNA level through aberrant splicing, where exons from two genes are mistakenly joined together. These are called *spliced fusions* and arise without any direct alteration of the DNA (**Figure 4**).[23]

Chromosomal translocations arise when double-strand breaks are incorrectly repaired, often through non-homologous end joining (NHEJ). Their formation is not random and is influenced by where the breaks occur, the spatial proximity of genomic regions, chromatin structure, cellular stress and the cell type in which the breaks arise. In hematopoietic cells, physiological processes such as V(D)J recombination can create DNA breaks that are misrepaired into translocations.[24]

Copy number changes

Copy number alterations (CNAs) are defined as duplications or deletions of DNA sequences that can affect whole chromosomes or specific chromosome segments. Gains or losses of entire chromosomes are defined as *aneuploidy.* Losses can be *hemizygous* if only one allele is affected or *homozygous* if both copies are lost. If a hemizygous loss results in a dosage that is insufficient to maintain normal function, the gene is considered *haploinsufficient*. *Loss of heterozygosity* (LOH) refers to the loss of one parental allele at a specific genomic *locus*. LOH occurs together with copy number loss or through copy-neutral mechanisms (CN-LOH).

In a healthy cell, aneuploidy is usually not tolerated and results in cell cycle exit or *apoptosis* (although exceptions exist, e.g. trisomies 13, 18 and 21 and sex chromosome aneuploidies). In a cancer cell, the imbalance in gene expression disturbs cellular homeostasis and has been linked to inflammation, proteotoxic and replication stress and genomic instability. However, cancer cells can adapt to these challenges, with aneuploidy acting as a driver of cellular adaptation (**Figure 7**).[25]

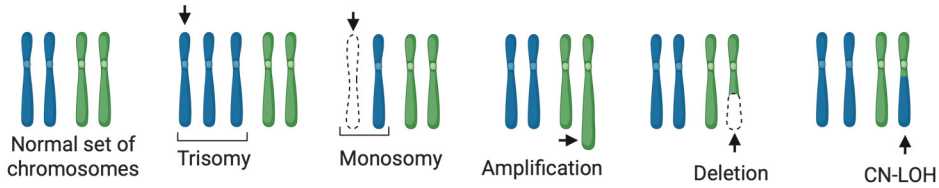


Figure 7: Different CNAs. Schematic illustration of a normal diploid chromosome set and different types of CNAs, including trisomy, monosomy, amplification, deletion and copy-neutral loss of heterozygosity (CN-LOH). Illustration created with BioRender.com.

Epigenetic alterations

Gene expression is tightly regulated by several epigenetic mechanisms. *DNA methylation* involves the addition of methyl groups to DNA and, when occurring at gene promoters, is generally associated with transcriptional repression or silencing. *Histone modifications*, including *acetylation* and *methylation*, alter chromatin structure and regulate DNA accessibility to the transcriptional machinery. Histone acetylation is generally associated with open chromatin and active transcription, whereas specific histone methylation marks are linked to gene repression. *Chromatin remodelling* regulates gene expression by repositioning or restructuring nucleosomes, influencing the accessibility of transcription factors and other DNA-binding proteins. Lastly, *non-coding RNAs* contribute to post-transcriptional regulation by modulating mRNA stability and translation (**Figure 8**).[26]

Enzymes involved in epigenetic modulation are grouped into four categories by function: *writers, erasers, readers* and *remodelers*. Writers add chemical modifications to DNA or histone proteins, while erasers remove these modifications, often having opposite effects on gene activity. Readers recognize these modifications and recruit protein complexes that influence chromatin structure and gene activity. Lastly, remodelers adjust chromatin by moving or removing nucleosomes at key regulatory sites, subsequently altering DNA accessibility.[27]

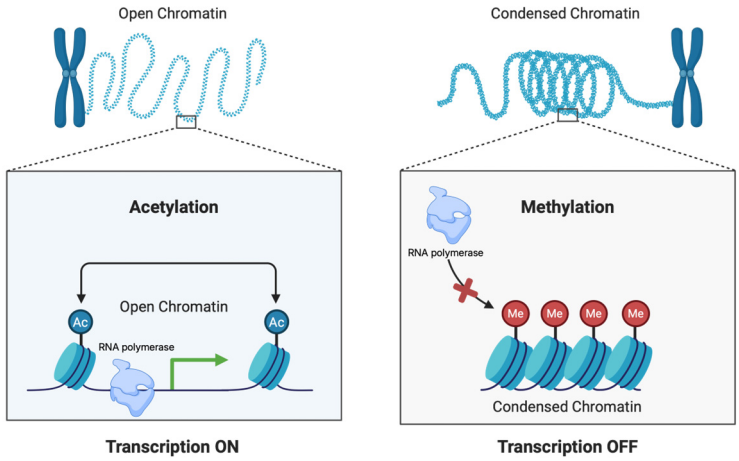


Figure 8: Histone acetylation and DNA methylation regulate chromatin accessibility and transcription. Histone acetylation promotes open chromatin, allowing RNA polymerase to access genes and promote transcription. DNA methylation promotes condensed chromatin, restricting RNA polymerase access and transcription. Illustration created with BioRender.com.

Which regions of the genome are prone to mutations?

While mutations were historically viewed as entirely random events, it is now clear that some regions of the genome are more susceptible to change than others. One of the best-known examples is CpG dinucleotides, where a C nucleotide is located next to a G nucleotide. These sites are frequently methylated as part of normal gene regulation. Methylated cytosines are chemically unstable and can spontaneously change into T, making CpG dinucleotides some of the most prominent mutation hotspots in the human genome. Repetitive DNA sequences, such as microsatellites and segmental duplications, are also highly unstable and can acquire insertions, deletions or larger structural rearrangements during DNA replication. In contrast, other genomic regions accumulate fewer mutations due to more efficient DNA repair or lower susceptibility to replication errors.[28]

Mutation rates are also influenced by chromatin organisation and DNA replication timing. In general, mutations accumulate more frequently in regions of closed chromatin and in late-replicating regions, where DNA repair is less efficient. Thus, the human genome contains both mutational hotspots and relatively protected regions, resulting in an uneven distribution of genetic variation.[29]

Gene regulation

Gene expression is regulated by non-coding DNA sequences known as regulatory elements, which control when, where and to what extent a gene is expressed. *Promoters* are located around the transcription start site and act as docking sites where the transcriptional machinery binds to initiate transcription. *Enhancers*

increase expression by recruiting *transcription factors* and bringing the transcriptional machinery closer to the promoter, often over long genomic distances through chromatin looping.[30] In contrast, *silencers* reduce expression by recruiting proteins that inhibit transcription. Finally, *insulators* act as genomic boundaries that prevent enhancers from activating the wrong genes, thereby ensuring proper gene regulation.[31]

Methods to study cancer

Over the past few decades, genome-wide technologies have transformed our understanding of cancer, shifting the focus from individual gene studies to comprehensive analysis of the entire genome. Early attempts at understanding the molecular basis of cancer were limited by available technologies. Today, high-throughput techniques allow us to analyse massive amounts of data and address fundamental questions about the genetic and epigenetic alterations that drive cancer initiation, progression and treatment response.

The development of cytogenetic techniques in the mid-20th century marked an important moment in cancer research, offering the first concrete insights into the chromosomal abnormalities in various malignancies. A milestone discovery was made in 1960 by Nowell and Hungerford in Philadelphia. They studied bone marrow (BM) samples from chronic myeloid leukemia (CML) patients and identified a small recurrent marker chromosome, which they named the Philadelphia chromosome (**Figure 9**).[32] However, it was not until detailed chromosome banding techniques were introduced in 1970[33] that researchers understood that the Philadelphia chromosome was the result of a translocation between chromosomes 9 and 22.[34] These findings constituted a paradigm shift, as they were the first direct evidence that certain cancers could be driven by specific genetic alterations. In 1985, breakpoint characterization resulted in the detection of the *BCR*::*ABL1* fusion gene,[35,36] which was linked to the oncogenic properties of the Philadelphia chromosome.[37,38] The discovery of this fusion gene paved the way for the development of the tyrosine kinase inhibitor imatinib, which specifically binds to the ATP binding site of the *BCR::ABL1* tyrosine kinase and inhibits its constitutive kinase activity. Imatinib has revolutionized the treatment of CML.[38]

In 1977, Sanger sequencing was unveiled and suddenly each exact nucleotide base of the DNA string was exposed. Sanger sequencing allowed detection of point mutations and small insertions and deletions[39] and has played a critical role in the discovery of important cancer genes, including *TP53*, *KRAS* and *BRCA1*. In the 1980s, fluorescence in situ hybridization (FISH) was developed[40] and is still used together with chromosome banding techniques for the clinical identification of chromosomal abnormalities.

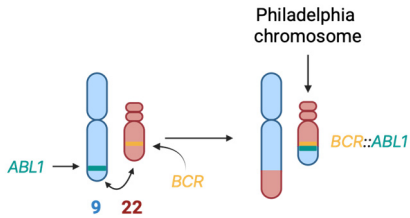


Figure 9: Formation of the Philadelphia chromosome. A schematic illustration of the reciprocal translocation between chromosomes 9 and 22, t(9;22), resulting in the formation of the Philadelphia chromosome and the *BCR::ABL1* fusion gene. Illustration created with BioRender.com.

During the 1990s and early 2000s, microarray technologies enabled the simultaneous analysis of large numbers of genes and genomic loci. Complementary DNA (cDNA) microarrays allowed genome-wide comparisons of gene expression,[41] while SNP arrays enabled high-throughput genotyping and the detection of CNAs and allelic imbalances.[42] These technologies represented an important transition from targeted analyses to genome-wide molecular profiling.

The next major step was the Human Genome Project, which spanned from the 1990s to 2003, during which over 90% of the human genome was sequenced.[14] This was the starting point of a new era and catalysed the development of next-generation sequencing (NGS) technologies, allowing rapid and massively parallel sequencing of both DNA and RNA. These technologies have made it possible to comprehensively characterise cancer genomes by detecting SNVs, SVs and CNAs across large patient cohorts.[10] The aptly named The Cancer Genome Atlas (TCGA) is a major repository of cancer genomic data, including publicly available data from over 20 000 tumour and matched normal samples across 33 cancers.[43] NGS technologies for genomic studies comprise several different methods, each with specific applications and advantages, including whole genome sequencing (WGS), which analyses the entire genome; whole exome sequencing (WES), which targets the protein-coding regions; targeted sequencing, which focuses on predefined genes or genomic regions of interest; and RNA sequencing (RNA-seq), which enables the analysis of gene expression, alternative splicing and fusion transcripts.

At the same time, epigenetic profiling techniques started to emerge, including whole genome bisulfite sequencing (WGBS) for DNA methylation, chromatin immunoprecipitation (ChIP) sequencing for histone modifications and later assay for transposase-accessible chromatin (ATAC) sequencing to measure chromatin accessibility. These techniques revealed that gene regulation can be disrupted at an epigenetic level in cancer cells, contributing to oncogenesis and therapy resistance.[44]

In 2009, the field was again transformed by the development of single cell RNA sequencing.[45] These methods allow sequencing of individual cells, thereby capturing the cellular heterogeneity of both healthy and malignant tissues. Since

then, several single-cell applications have emerged, including single-cell RNA sequencing (scRNA-seq), single-cell DNA sequencing (scDNA-seq) and multimodal approaches that simultaneously profile multiple molecular features within the same cell.

Today, multi-omics platforms combining genomic, proteomic, transcriptomic, epigenomic and spatial data are used to gain a more complete understanding of the complexity of cancer. These advances continue to push the boundaries of our understanding of cancer, while also creating new challenges for future cancer care and the successful translation of these findings into clinical practice.

Bioinformatics

Bioinformatics is an interdisciplinary field combining biology, computer science, mathematics and statistics to analyse and interpret the immense amounts of data generated by sequencing methods. It emerged as a response to the rapid growth of large-scale biological datasets, which required specific expertise for analysis. Bioinformatics is crucial in modern biological and medical research as it facilitates the translation of complex data into patterns, predictions and new biological insights.[46]

Creating and processing sequencing data involves a multistep workflow, from the lab bench to the computer. There are numerous sequencing techniques. Illumina sequencing was used in all studies (**I-III**), and the general steps are described below.

The process begins with extraction of DNA or RNA. During library preparation, the nucleic acids are fragmented, followed by ligation of sequencing adapters, which contain sample specific index barcodes to enable multiplexing. Then, the library is amplified by polymerase chain reaction (PCR) to generate sufficient material for sequencing. The fragments are added to a flow cell, immobilized and amplified into *clusters*, each containing multiple copies of the same fragment. Fluorescently labelled nucleotides are incorporated, and the emitted fluorescence from each cluster is detected during every sequencing cycle. In this workflow, a paired-end sequencing strategy of 2x150 base pairs (bp) is used. This means that each DNA fragment is sequenced from both ends, generating two separate 150 bp reads per fragment and improving the accuracy of downstream mapping.

This information is saved as raw data in binary base call (BCL) files. After sequencing is completed, BCL files are converted to FASTQ format, which contains nucleotide sequences and quality scores for each base, indicating the probability of an incorrect base call. This is followed by *mapping* and *alignment* of the reads to a reference genome (hg19 in **Articles I-III**). The *coverage* refers to how many times

a specific region of the genome has been read, with higher coverage generally making the results more reliable (**Figure 10**).

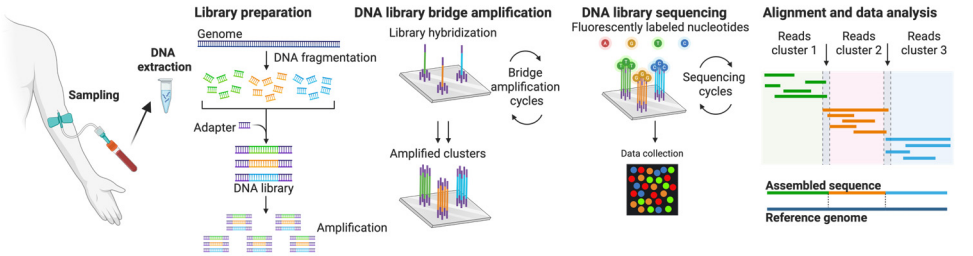


Figure 10: Sequencing workflow. Schematic overview of the sequencing workflow, from sample collection to data analysis. Blood is collected and DNA is extracted, followed by library preparation including DNA fragmentation, adapter ligation and amplification. The DNA library is then subjected to bridge amplification and hybridization to generate amplified clusters, which are sequenced using fluorescently labelled nucleotides. Sequencing reads are aligned to the reference genome to generate assembled sequence data and identify genetic alterations. Illustration created with BioRender.com.

Following alignment to the human reference genome, sequence alignment map (SAM) files are generated, containing information on how each sequencing read maps to the reference. As these files are large, they are typically compressed into binary alignment map (BAM) files for storage and downstream analyses. *Variant calling* is subsequently performed to identify genetic alterations. In cancer genomics, somatic variant calling is ideally carried out by comparing the tumour sequence with matched normal DNA from the same individual, allowing inherited germline variants to be distinguished from acquired somatic mutations. The identified variants are stored in variant call format (VCF) files, followed by filtering steps to reduce false positive calls and improve data quality.

There is a wide array of bioinformatic analysis tools available to support each stage of the pipeline, from alignment to variant interpretation. Each tool has unique strengths and limitations, and the choice of tool often depends on the research question and data type.[47]

Since genomic data are sensitive, all computational analyses and data storage in this thesis were performed in a secure computing environment to ensure patient privacy.

Normal hematopoiesis

The cells in the human body undergo an astonishing turnover, with over 300 billion new cells produced daily. More than 90% of these are blood cells generated through *hematopoiesis* within the BM.[48,49] During childhood, all bones contribute to blood cell production. However, as we age, the hematopoietically active marrow in the

peripheral (long) bones is gradually replaced by fatty, inactive marrow. Consequently, blood cell formation becomes restricted primarily to the central (axial) skeleton, including the vertebrae, pelvis and sternum.[50]

All hematopoietic cells originate from quiescent hematopoietic stem cells (HSCs), which are characterized by their capacity for self-renewal and multipotent differentiation.[51-53] Through proliferation and differentiation, HSCs give rise to mature, specialized blood cells essential for survival: erythrocytes responsible for oxygen transport; thrombocytes involved in blood clotting; granulocytes and monocytes mediating immune defence; B-cells producing antibodies; and T-cells eliminating virus-infected cells.[54,55]

Hematopoietic hierarchy

The classical view of hematopoiesis is built upon a hierarchical structure with the HSC at the top. This model is often depicted as a branching tree with progressively committed progenitors along its branches.[38] During the first differentiation step, HSCs give rise to multipotent progenitors (MPPs), which retain multilineage differentiation potential but have limited self-renewal capacity.[56,57] The cells downstream of the MPPs are divided into two lineages, the common myeloid progenitors (CMPs), which produce myeloid cells and the lymphoid primed multipotent progenitors (LMPPs), which retain the potential to generate both myeloid and lymphoid cell types.[49] CMPs differentiate into granulocyte monocyte progenitors (GMPs), which give rise to monocytes, granulocytes (neutrophils, eosinophils and basophils) and dendritic cells (DCs), and megakaryocyte erythroid progenitors (MEPs) which further differentiate into megakaryocytes (and later platelets) and erythrocytes.[49,58] LMPPs also generate GMPs, as well as common lymphoid progenitors (CLPs), which give rise to natural killer (NK) cells, a subset of DCs, and T- and B-cells (**Figure 11**).[49]

Yet, accumulating evidence shows that this representation might oversimplify a process that is not completely understood. Recent studies have revealed that HSCs actually are heterogeneous and can be primed towards particular lineages. Contrary to the classical model, which predicts a balanced lineage output, HSCs often exhibit biases favouring certain lineages. This lineage bias has been demonstrated by transplanting individual HSCs and following which blood cell types they produced, as well as by genetically marking cells in mice and tracing their descendants during normal hematopoiesis. These studies showed that individual HSCs can contribute more strongly to some lineages than others.[57,59-61] By combining flow cytometry (FC), transcriptomics and functional analyses at single cell resolution, researchers have now proposed an alternative model: *the continuum model* (**Figure 11**). According to this model, hematopoiesis is a dynamic and continuous process, in which HSCs gradually acquire lineage-specific biases without passing through defined progenitor stages.[62]

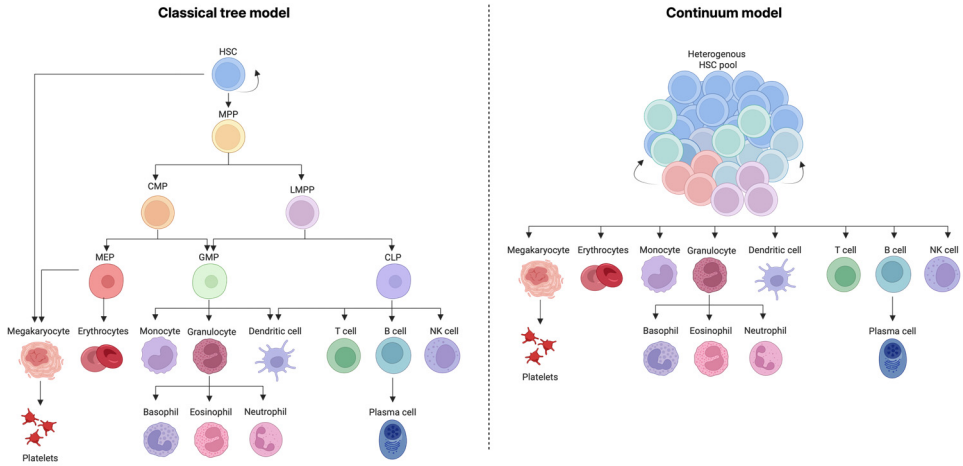


Figure 11: Classical and continuum models of hematopoietic differentiation. Classical stepwise "tree" model of hematopoiesis (left) compared with the continuum model (right), in which heterogeneous HSCs display lineage biases and differentiate along a continuous spectrum. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lymphoid-primed multipotent progenitor; MEP, megakaryocyte-erythroid progenitor; GMP, granulocyte-monocyte progenitor; CLP, common lymphoid progenitor; DC, dendritic cell. Illustration created with BioRender.com.

Regulation of hematopoiesis

Hematopoiesis is a tightly regulated yet highly dynamic process, continuously replenishing short lived blood cells while adapting swiftly to physiological demands such as blood loss or infection.[54,55] It is orchestrated by a combination of extrinsic factors, such as *cytokines* and *growth factors*, and intrinsic factors such as transcription, epigenetic regulation and intracellular signalling pathways.[63] Disruptions in external or internal signals can upset the delicate balance of hematopoiesis, with fatal consequences.

Extrinsic factors

The environment around the HSCs, the *bone marrow niche*, adapts to support the HSC function.[63] HSCs are localized in the trabecular regions of the long bones, adjacent to capillaries and mesenchymal stromal cells.[64,65] The niche is a complex tissue, composed of support structures as well as over 20 different cell types (both hematopoietic and non-hematopoietic) that are closely intertwined.[63,66] These cells release cytokines (secreted molecules that bind to cell surface receptors) that sustain HSC survival and guide migration and lineage fate.[67] Two of the best characterized cytokines with a regulatory role for HSC maintenance are CXCL12 (binding to CXCR4) and stem cell factor (SCF, binding to KIT).[68,69]

Our understanding of the niche is incomplete. However, recent single cell studies have started to decipher the cellular architecture and molecular landscape.[70,71] It is worth noting that the risk of developing hematological disease increases with age, as both the niche and HSCs inevitably age in parallel with us. An increased understanding of these age-related changes and their impact on malignant transformation can help identify treatment targets, guide strategies to restore an aged niche or inform new preventive approaches.[66,68]

Cytokines and growth factors can also be released from other tissues, such as erythropoietin (EPO), which is primarily secreted from the kidneys into the plasma. EPO binds to receptors on erythroid progenitors to sustain red cell production and systemic oxygen homeostasis.[72,73] Further, thrombopoietin (TPO), which is mainly produced in the liver, stimulates the production of megakaryocytes and subsequently platelets,[74] and granulocyte colony stimulating factor (G-CSF), produced by endothelial cells, fibroblasts and macrophages in virtually all organs, supports the survival and proliferation of neutrophil progenitors.[75]

Intrinsic factors

During hematopoietic differentiation, dynamic changes in gene expression are essential for proper lineage commitment and cellular maturation. Disruption of these tightly regulated programmes is a hallmark of hematological malignancies, where genes controlling blood cell development are frequently mutated, thereby contributing to leukemogenesis. Gene expression is regulated by transcription factors that can activate or repress transcription in response to external signals.[1,55] Recent studies have shown that multipotent hematopoietic stem and progenitor cells can simultaneously express low levels of genes associated with multiple lineages, a phenomenon known as *multilineage priming*, which prepares cells for rapid and flexible fate decisions.[56,76] As hematopoietic progenitors commit to a specific lineage, one transcriptional programme becomes dominant while alternative lineage programmes are actively repressed. For example, GATA1 promotes erythroid and megakaryocytic differentiation, whereas SPI1 (PU.1) promotes myeloid development. These transcription factors activate lineage-specific genes while simultaneously repressing each other's activity, thereby ensuring commitment to a single lineage. Similarly, *CEBPA* plays a central role in granulocytic differentiation by activating myeloid gene expression and cooperating with SPI1. Together, these regulatory networks allow hematopoietic stem and progenitor cells to balance alternative cell fates while ensuring lineage commitment.[77-82] In this manner, different transcription factors regulate B-cell, T-cell and myeloid differentiation, including IKZF1, PAX5 and VPREB1 in B-cells[83]; GATA3, HEB and TCF1 in T-cells[84]; and IRF8, RUNX1 and GFI1 in myeloid cells.[85]

Epigenetic regulation is also important for the initiation and continued preservation of transcriptional memory throughout hematopoiesis. The dynamic process of attaching or removing chemical groups on histone proteins and DNA molecules

modulates the three-dimensional organization and accessibility of chromatin. This, in turn, affects gene expression by enabling recruitment of either activating or repressive transcriptional complexes. The importance of epigenetic regulators in hematopoiesis is underscored by the frequent disruption of these genes in hematologic malignancies, including *TET2*, *DNMT1*, *DNMT3A* and *DNMT3B*.[55,86]

Lastly, the activation of specific signalling pathways can influence HSC fate. These pathways can be activated by extrinsic factors binding to receptors on HSCs or niche cells. Examples include Wingless (WNT), which can promote either myeloid or lymphoid signalling in a dose-dependent manner;[87], Smad, which contributes to self-renewal;[88] Hedgehog signalling whose role in normal hematopoiesis remains controversial but which is recognized as a regulator of HSC differentiation and leukemogenesis,[89,90] and RAS/mitogen-activated protein kinase (MAPK) signalling, which transmits cytokine and growth factor signals that regulate proliferation, survival and differentiation.[91] Signals can also be transmitted through direct cell-to-cell interactions, such as NOTCH signalling, which is essential for T-cell lineage commitment.[53,92,93]

Hematological malignancies

Hematological malignancies are among the most common cancers and a major contributor to the global tumour burden. They are broadly classified as leukemia, multiple myeloma, non-Hodgkin lymphoma and Hodgkin lymphoma.[94] Today, the World Health Organization (WHO) and the International Consensus Classification recognize over 50 entities, defined by clinical features and genetic alterations.[95-97]

These malignancies evolve through the accumulation of somatic mutations, generating clonal heterogeneity that shapes disease behaviour and treatment response. Extensive efforts have been devoted to characterising their genomic landscape. However, genomic profiling alone cannot fully explain how genetic alterations influence clonal evolution or treatment response. Addressing these questions requires complementary approaches, such as longitudinal sampling of patients and functional experimental models.

Leukemia

Among the hematological malignancies, leukemia accounts for a particularly large proportion of hematological cancer deaths.[94] It is broadly classified as acute or chronic based on cell maturity and the rate of disease progression. In acute leukemia, differentiation is arrested early, resulting in the accumulation of immature *blast*

cells, whereas in chronic leukemia, the arrest occurs later, allowing the malignant cells to retain some degree of maturation and function. Consequently, acute leukemia generally progresses rapidly, whereas chronic leukemia often follows a more indolent course. Leukemia can further be classified as myeloid or lymphoid according to the hematopoietic lineage of the malignant cells. [98,99] Rarely, leukemic blasts exhibit features of both myeloid and lymphoid differentiation and are classified as mixed phenotype acute leukemia (MPAL).[100]

In adults, acute myeloid leukemia (AML) is the most common acute leukemia and its incidence rises with age,[98] while acute lymphoblastic leukemia (ALL) has its peak prevalence between the ages of 2-5 years.[99] Leukemia is the most common childhood malignancy, accounting for 20-30% of childhood cancers.[3] According to the Swedish Childhood Cancer Registry (SBCR; sbcr.se), this corresponds to approximately 100 cases annually in Sweden.

Leukemia development

Leukemia is a heterogeneous disease driven by genomic rearrangements and somatic mutations acquired in hematopoietic stem and progenitor cells.[101,102] The first evidence that leukemia arises from clonal expansion of a single transformed cell came from studies of chromosome X inactivation and CML in the 1960s.[103,104] These findings contributed to the concept of *clonal hematopoiesis*, in which hematopoietic stem or progenitor cells carrying somatic mutations acquire a selective growth advantage and expand over time. The cancer stem cell concept was later experimentally demonstrated in AML in 1994, when a rare population of $CD34^{+}CD38^{-}$ cells was shown to regenerate leukemia following transplantation into immunodeficient mice.[105]

Somatic mutations accumulate throughout life in virtually all human cells. As their number increases with age, they contribute to the development of both malignant and non-malignant diseases.[106] In HSCs, some acquired mutations can confer a competitive advantage, leading to clonal expansion and a progressive reduction in the diversity of the stem cell pool. Clonal fitness is context dependent, with intrinsic and extrinsic selection factors determining the fate of a clone.[107] Clonal hematopoiesis of indeterminate potential (CHIP) refers to the presence of somatic mutations in leukemia-associated driver genes with a variant allele frequency (VAF) of ≥2%, but without overt hematological malignancy. It is rare in people under the age of 50, but present in 10-20% of people over 70 years. CHIP is not malignant per se but confers an increased risk of hematological transformation.[97,108-110]

Leukemia contains different populations of cellular clones. How this heterogeneity is formed has been extensively studied and today there are several proposed models of evolution. One of the earliest models was *linear evolution*, in which clonal evolution was viewed as a linear stepwise process, with successive mutations

leading to the emergence of new, increasingly fit subclones that replace ancestral populations through selective sweeps.[111] However, more recent studies describe clonal evolution as a highly dynamic process that begins before clinical diagnosis, with the clonal landscape continuously reshaped by selective pressures and the accumulation of both genetic and non-genetic changes. While most blast cells arise from one or a few dominant clones, competition and selection during premalignant development can result in a complex and heterogeneous clonal architecture.[112]

This complexity is better captured by the *branching* evolution model, in which clones diverge from a common ancestor and evolve in parallel, with each acquiring different alterations that may confer a selective advantage.[66,113] Unlike solid tumours, leukemia cells are free to move within the blood and BM and are therefore not restricted by space. In branching evolution, different subclones can coexist and evolve in parallel.[9]

Another model is called *punctuated evolution*. In contrast to linear and branching evolution, where genetic alterations accumulate as clones evolve, punctuated evolution proposes that many genomic alterations arise in rapid bursts, often early in tumour development, followed by the stable expansion of one or more dominant clones.[114] Lastly, the *neutral evolution* model proposes that tumours accumulate random passenger mutations, which confer no fitness advantage. This genetic drift generates extensive intratumoral heterogeneity, with limited selection among subclones during subsequent tumour growth[115] (**Figure 12**).

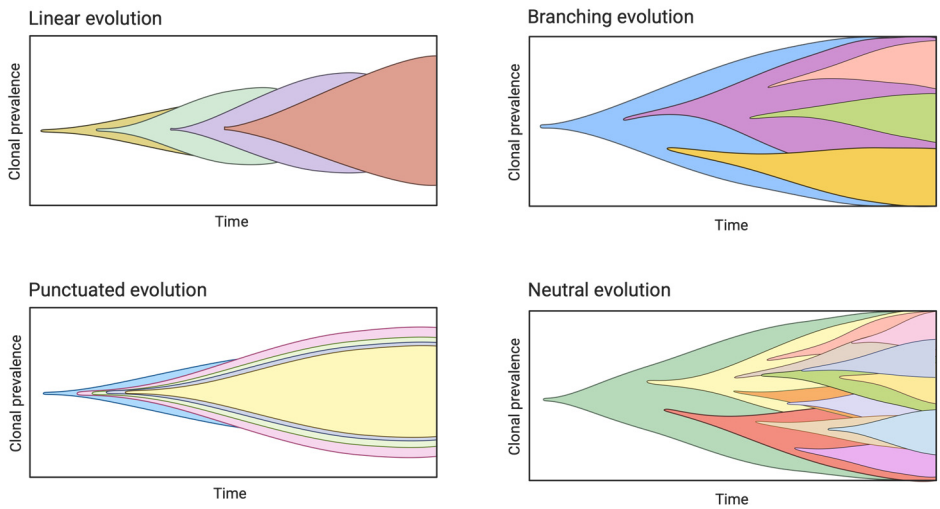


Figure 12: Clonal evolution in leukemia. Illustration of four models of leukemic evolution: linear evolution (sequential selection of dominant clones), branched evolution (parallel outgrowth of multiple subclones), punctuated evolution (bursts of early genetic changes) and neutral evolution (expansion driven by random genetic drift without strong selection). Illustration created with BioRender.com.

Our understanding of clonal evolution has improved with advances in sequencing technologies. Early studies were based on cytogenetics and bulk sequencing, which could only estimate the clonal composition of a sample. More recently, single-cell sequencing has made it possible to study individual cells, providing a much clearer picture of clonal architecture and evolution.

Leukemia can progress slowly or rapidly, either through the gradual accumulation of genetic alterations that confer a selective advantage or via a specific genetic event that gives a strong survival advantage, such as *KMT2A* rearrangements or the *BCR*::*ABL1* fusion. These initiating events are often essential for leukemia maintenance. The extent to which additional genetic alterations contribute varies considerably between leukemia subtypes.[53,116]

Subclones, relapse and therapy mediated clonal selection

During treatment, cytotoxic and targeted therapies impose strong selective pressures on the leukemic population. Resistant or quiescent subclones can survive and later expand, leading to relapse. If the dominant diagnostic clone persists, relapse follows a linear evolutionary pattern and remains genetically similar to the primary disease. Alternatively, relapse can arise from a minor or ancestral subclone that expands after treatment, resulting in branching evolution.[112,117-120]

Relapse might be driven by resistance already present at diagnosis, alterations acquired during treatment or a combination of both. Although many mutations detected at relapse are likely to be passenger events, some may promote treatment resistance or clonal expansion.[112,121] The relative contributions of pre-existing subclonal diversity and treatment-associated mutagenesis to relapse in *KMT2A*-rearranged (*KMT2A*-r) leukemia were investigated in **Article I** and are further discussed in the *Discussion* section.

Prenatal origin of acute leukemia

A prenatal origin has been established for several major subtypes of childhood acute leukemia, including *ETV6*::*RUNX1*-positive, high hyperdiploid and *KMT2A*-r leukemia. Evidence for a prenatal origin includes shared leukemia-associated rearrangements in monozygotic twins, retrospective detection of fusion genes in neonatal Guthrie cards collected at birth and identification of leukemia-associated alterations in fetal hematopoietic samples.[122-125] The frequent prenatal occurrence of chromosomal rearrangements in childhood leukemia indicates that fetal hematopoiesis provides a cellular and developmental context in which these rearrangements can arise.[126] However, prenatal initiation does not necessarily imply that overt leukemia is present at birth. This is particularly evident in *ETV6*::*RUNX1*-positive ALL, in which the fusion can generate a preleukemic clone that persists for several years but requires additional postnatal genetic or epigenetic alterations for

progression to leukemia.[123,127] Twin studies have provided evidence of prolonged latency, including one pair in which the twins developed *ETV6*::*RUNX1*-positive ALL nine years apart but shared the same clone-specific fusion breakpoint.[128]

In contrast, infant (children <1 year) *KMT2A*-r leukemia is characterized by a short latency and a low burden of somatic mutations, indicating that the *KMT2A* fusion acts as a potent initiating driver that requires relatively few cooperating alterations in the permissive context of fetal hematopoiesis.[123,127,129] Experimental mouse models support a strong leukemogenic role for *KMT2A* fusions, although disease development depends on the fusion partner, cooperating alterations and cellular context.[130] Phylogenetic single cell studies further show that the *KMT2A* rearrangement represents an early, truncal event, followed by divergent clonal evolution.[131]

The factors determining whether a preleukemic clone progresses to overt leukemia remain incompletely understood. Rare, high-penetrance germline variants account for only a minority of childhood ALL cases, whereas common low-penetrance susceptibility variants can modify disease risk more broadly. Environmental and immune-mediated factors might also influence progression. According to the delayed infection hypothesis, limited microbial exposure during infancy could predispose some children to dysregulated immune responses when common infections are encountered later in childhood. These immune responses may promote the acquisition or selection of secondary alterations in a pre-existing preleukemic clone. No specific infectious agent has yet been established as a direct cause of leukemia.[53,132]

Pediatric acute lymphoblastic leukemia

ALL can arise from immature lymphoid progenitor cells committed to either the B- or T-cell lineage, with T-cell ALL (T-ALL) accounting for 15% of pediatric and 25% of adult ALL cases.[133,134] Both B-cell ALL (B-ALL) and T-ALL are divided into multiple genetic subtypes, which carry diagnostic and prognostic information.[53,134,135] Thanks to RNA-seq, today over 23 molecular subtypes have been identified in B-ALL and 17 in T-ALL with an age dependent distribution.[116,136,137]

B-ALL

Around 75% of B-ALL cases harbour aneuploidy or recurrent chromosomal rearrangements, which are early initiating events in leukemogenesis and fundamental for clinical diagnostics, disease monitoring and treatment decisions.[138] The genetic abnormalities considered clinically actionable in the pediatric ALLTogether treatment protocol (clinicaltrials.gov: NCT04307576)[139] are described below. The list includes abnormalities relevant to disease classification,

genetic risk stratification or treatment decisions, rather than all recognized molecular ALL subtypes. Overall risk classification also incorporates clinical and treatment-response variables (see the section *The clinic* for further information). *KMT2A*-r leukemia is discussed separately in a dedicated section (**Figure 13**).

Hyperdiploid

Hyperdiploid B-ALL is characterized by chromosomal gains, resulting in more than 46 chromosomes, and is generally associated with a favourable prognosis. *High hyperdiploid*, defined by more than 51 chromosomes, is present in around 30% of pediatric B-ALL cases, but rare in adults (3%). It is associated with a low white blood cell count and rarely involves extramedullary sites.[132,140,141] Large cohort studies have established that chromosomal gains are non-random, with recurrent gains of chromosomes X, 4, 6, 10, 14, 17, 18 and 21[140] and neonatal studies have shown that these gains frequently originate prenatally.[141]

Hyperdiploid B-ALL is thought to originate from a single abnormal cell division in an immature precursor cell, in which the chromosomes fail to segregate properly. Exactly what causes this event and the preferential selection of these specific chromosomes remains unclear.[140,141] One proposed explanation is a gene dosage effect, meaning that the presence of extra copies of certain chromosomes leads to increased expression of genes located on those chromosomes. However, gene dosage alone is unlikely to fully explain the non-random pattern of chromosomal gains, and additional mechanisms are likely to contribute to the development and progression of the disease.[140] Subsequent acquisition of secondary genetic alterations, particularly in the RAS pathway (including *NRAS*, *KRAS*, *FLT3* and *PTPN11*) and epigenetic regulators (*CREBBP*), seems to further promote clonal evolution and progression to leukemia.[140,141]

Hypodiploid

Hypodiploid B-ALL is characterised by the loss of whole chromosomes and comprises two biologically different entities: *low hypodiploid* ALL (32-39 chromosomes) and *near haploid* ALL (24-31 chromosomes). Low hypodiploid ALL is rare, accounting for approximately 1% of childhood ALL and up to 10% of adult cases and is associated with an adverse prognosis. Near haploid ALL is slightly more common in children (approximately 3% of cases) and likewise carries a poor prognosis. Interestingly, like high hyperdiploid ALL, near haploid ALL frequently exhibits dysregulation of the RAS signalling pathway. This most commonly results from deletions affecting negative regulators of RAS signalling, such as *NF1*, rather than activating mutations in *NRAS* or *KRAS*.[142]

Masked hypodiploidy is a phenomenon in which an initially hypodiploid clone undergoes whole genome duplication, resulting in a hyperdiploid chromosome number. Consequently, masked hypodiploidy cytogenetically resembles hyperdiploid ALL and there is a risk of misclassification, resulting in an

underestimation of the number of hypodiploid cases. These cases can be distinguished by SNP arrays or WGS through detection of the characteristic CN-LOH created by whole-genome duplication. Masked hypodiploid cases differ biologically and prognostically and are often associated with adverse genetic features, including germline *TP53* alterations. Accurate distinction is therefore essential for correct risk stratification and treatment decisions.[142,143]

The inferior survival in hypodiploid ALL compared with hyperdiploid likely reflects both the biological consequences of chromosome loss and the specific genes affected. Loss of whole chromosomes reduces the dosage of multiple genes, including tumour suppressor genes.[132,143,144] In particular, low hypodiploid ALL is strongly associated with *TP53* alterations (90%), which are linked to genomic instability and increased resistance to therapy.[143,145,146] In approximately 40% of *TP53* mutated pediatric cases, the alteration is also detected in non-leukemic cells, consistent with a germline origin and an association with Li-Fraumeni syndrome.[143] In contrast, high hyperdiploid ALL is generally highly responsive to chemotherapy, partly because leukemic cells accumulate higher intracellular concentrations of active drug metabolites and are more susceptible to drug-induced apoptosis.[147]

ETV6::RUNX1

ETV6::RUNX1 results from t(12;21)(p13;q22) and is the most common fusion gene in pediatric leukemia (25%) but is rare in adults (<1%).[132] The fusion gene is regarded as an initiating event arising *in utero,* acting as a weak oncogene.[148] *ETV6::RUNX1* encodes a chimeric transcription factor in which the repression domain of *ETV6* is fused to the DNA binding domain of *RUNX1*. This abnormal protein deregulates gene expression in precursor B-cells, represses *RUNX1* target genes, disrupts normal hematopoiesis and alters pathways controlling proliferation, differentiation and cellular homeostasis, thereby promoting leukemogenesis.[149] However, *ETV6*::*RUNX1* alone is insufficient to induce leukemia and additional genetic alterations are required. These vary between patients and may include deletion of the remaining non-rearranged *ETV6* allele, alterations affecting B-cell development (e.g. *PAX5* and *BTG1*) or mutations affecting signalling pathways (e.g. *KRAS* and *NRAS*).[53,148]

ETV6::RUNX1-positive ALL is associated with a favourable prognosis and a low relapse risk across age groups (5%), but accounts for a larger proportion of relapses (20%) due to its high prevalence.[148]

Intrachromosomal amplification of chromosome 21 (iAMP21)

The abnormal iAMP21 chromosome is thought to develop through repeated cycles of chromosome 21 breakage and fusion. This repeated damage creates progressive genomic instability and can eventually lead to *chromothripsis*, in which chromosome 21 is shattered and reassembled in a disorganized way. This results in gains, losses and rearrangements along chromosome 21. Variation in the size and

location of these changes results in different iAMP21 structures.[116,132] Genomic analyses support the view that chromosome 21 instability represents the initiating genetic event in iAMP21 leukemia, whereas recurrent alterations affecting genes such as *IKZF1*, *ETV6*, *CDKN2A*/*B* and *PAX5* are acquired later during disease evolution.[116,150] Activating mutations in the RAS/MEK/ERK signalling pathway are also found in iAMP21, often as subclonal events, consistent with their acquisition during disease progression.[151,152] Recurrent alterations affecting *FLT3* and the tumour suppressor *SH2B3* have also been identified in iAMP21. Both genes are involved in JAK/STAT signalling. *SH2B3* inactivation is often associated with CN-LOH of chromosome 12. These alterations may represent potential therapeutic targets.[152,153]

iAMP21 is rare in adults but present in 2-6% of pediatric ALL.[116,132] iAMP21 is associated with poor outcomes with standard ALL therapy, with higher relapse rates and inferior survival.[154] In ALLTogether, iAMP21 is therefore classified as a high-risk genetic feature and incorporated into risk-adapted treatment.[139]

TCF3::HLF

TCF3::HLF, t(17;19)(q22;p13), is very rare (<1%) but highly aggressive. The fusion encodes an aberrant transcription factor that deregulates genes involved in lymphoid differentiation and cell survival, contributing to chemoresistance and a high risk of early relapse. Despite treatment intensification over the past three decades, outcomes remain poor and long-term survival is limited.[155] *TCF3::HLF* leukemia is frequently accompanied by alterations affecting B-cell differentiation, particularly *PAX5* and *VPREB1*, as well as alterations in the RAS pathway.[156]

Philadelphia-positive and Philadelphia-like ALL, including ABL-class fusions

The translocation t(9;22)(q34;q11.2) gives rise to the Philadelphia chromosome and the *BCR::ABL1* fusion gene (**Figure 9**). *BCR::ABL1* encodes a tyrosine kinase that drives leukemia by activating signals promoting cell growth and survival.[142,157]

BCR::ABL1 is the main driver of CML but is also present in approximately 3-5% of pediatric ALL and 40% of adult ALL. Outcomes have improved since the introduction of tyrosine kinase inhibitors (TKIs, e.g., imatinib and dasatinib) in combination with chemotherapy.[142,158,159] *IKZF1* alterations are common in *BCR::ABL1* leukemia and are associated with poor treatment response and an increased risk of relapse, despite TKIs.[135] Notably, *IKZF1* deletions are absent in *BCR::ABL1* positive CML, suggesting that *IKZF1* loss is specifically involved in lymphoid transformation and the development of *BCR*::*ABL1* positive ALL.[159] These patients are currently treated in a separate trial within the ALLTogether.[139]

Philadelphia-like (Ph-like) ALL is a genetically heterogeneous subtype characterized by kinase-activating alterations and accounts for approximately 15% of high-risk pediatric B-ALL. Among Ph-like ALL, approximately 10-15% harbour

ABL-class fusions involving *ABL1, ABL2, PDGFRB* or *CSF1R*. These patients receive TKI therapy in addition to chemotherapy.[139,160]

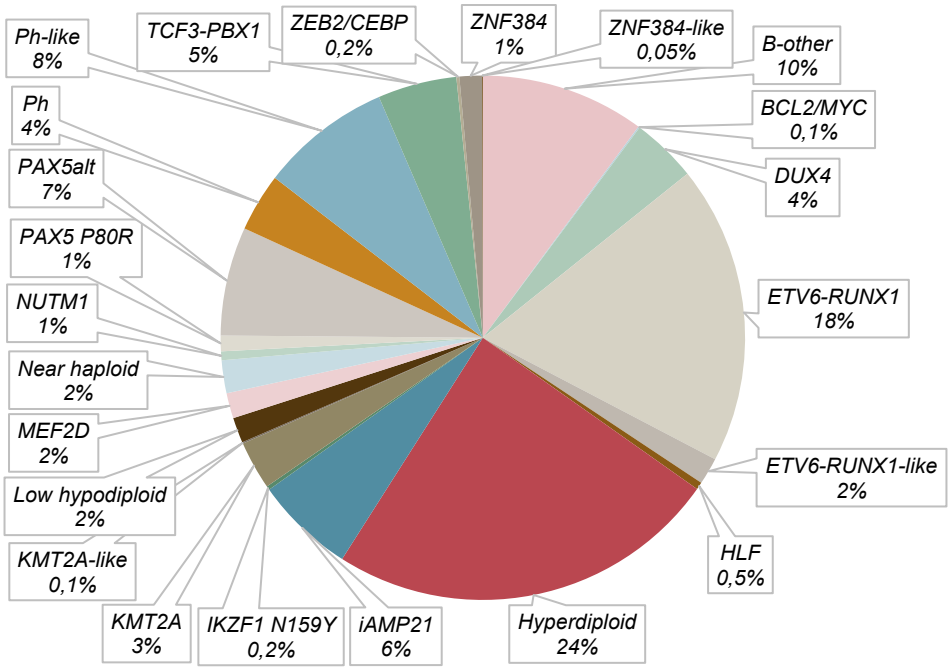


Figure 13: Distribution of 23 molecular categories in a large cohort of pediatric B-ALL (0-18 years). The pie chart includes data from a cohort of 2 176 B-ALL cases reported by Brady et al. (2022). ABL-class fusions are included in the Philadelphia-like (Ph-like) group. Cases with an age >18 years were excluded.

T-ALL

T-ALL is genetically heterogeneous, with recurrent alterations affecting transcriptional regulation (e.g. *TAL1, LMO1/2, TLX1, TLX3, HOXA*), NOTCH signalling (e.g. *NOTCH1* and *FBXW7*), intracellular signalling (e.g. JAK/STAT, RAS and PI3K pathways), cell cycle regulation (e.g. *CDKN2A, CDKN2B, RB1*) and epigenetic regulation (e.g. *DNMT3A, EZH2, SUZ12, PHF6*). Activating *NOTCH1* mutations and deletions of *CDKN2A/CDKN2B* are frequent and occur across several molecular subtypes, whereas dysregulation of transcription factors such as *TAL1*, *TLX1*/3 and *HOXA* defines major T-ALL subgroups.[132,161,162]

Other recurrent rearrangements generate fusion genes, including *NUP214::ABL1, KMT2A::MLLT1* and *PICALM::MLLT10*. Recent profiling has classified childhood T-ALL into 17 molecular subtypes based on their genomic and transcriptional profiles.[132,161,162]

Lastly, early T-cell precursor (ETP) ALL represents an immature form of T-ALL, characterized by low expression of typical T-cell markers and expression of stem

cell or myeloid markers. Genetically, ETP-ALL is enriched for alterations in signalling pathways, hematopoietic transcription factors and epigenetic regulators **(Figure 14)**.[163,164]

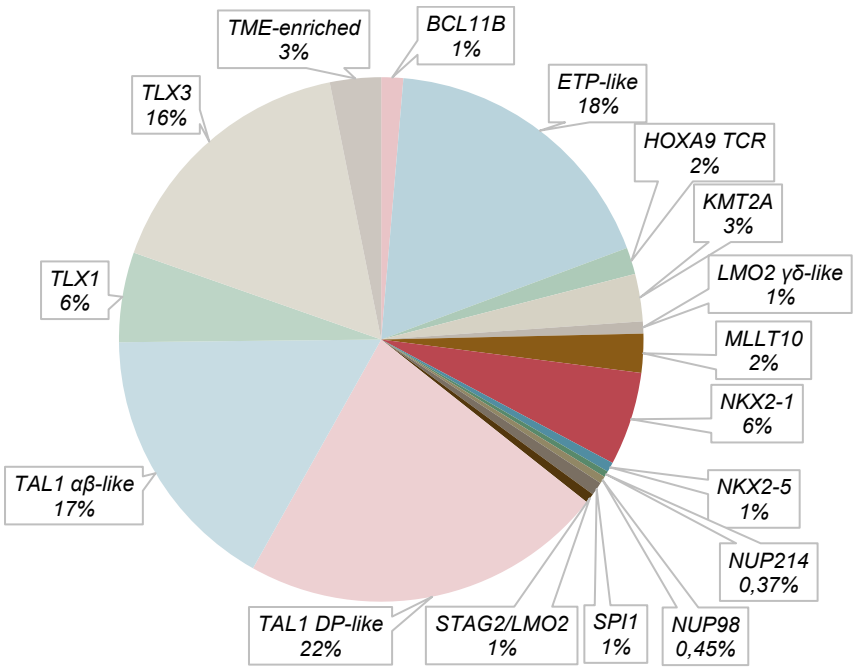


Figure 14: Distribution of 17 molecular categories in a large cohort of childhood (1-18 years) T-ALL. The pie chart includes data from 1 335 T-ALL cases reported by Pölönen et al. (2024).

Pediatric acute myeloid leukemia

AML accounts for a quarter of pediatric leukemia and comprises molecular subtypes defined by recurrent genetic alterations, many of which are important for risk stratification and treatment decisions.[165,166] A quarter of pediatric AML presents with a normal karyotype, making cytogenetic risk classification challenging. The introduction of NGS has enabled identification of recurrent somatic mutation patterns that improve molecular classification for these patients.[166,167] In fact, AML from an adult patient was the first cancer genome to be comprehensively characterized by WGS.[168] In the current protocol for pediatric AML, CHIP-AML22 (clinicaltrials.gov: NCT06262438),[169] the diagnosis and classification are based on the WHO classification of pediatric tumours.[170] The main genetic subtypes are described below (except *KMT2A*-r leukemia, which is discussed separately). In CHIP-AML22, *KMT2A*-r, *CBFA2T3::GLIS2*, *FLT3*^{ITD} and *NPM1* mutations are specifically considered for treatment allocation (**Figure 15**).

AML with NPM1 mutations

NPM1 mutations are among the most frequent genetic alterations in AML, with their incidence increasing with age, occurring in approximately 10% of pediatric cases, 20% of adolescent cases and up to one third of adult AML cases. Nucleophosmin 1 (NPM1) is a nuclear protein involved in several processes including ribosome formation, cell cycle regulation and DNA damage repair. Mutations in *NPM1* create a novel nuclear export signal, causing NPM1 to accumulate in the cytoplasm instead of in the nucleus.[171] Cytoplasmic localisation of mutant NPM1 results in persistent activation of leukemic transcriptional programs, including *HOXA*, *HOXB* and *MEIS1*.[172] Since mutant NPM1 depends on these transcriptional programs several targeted therapeutic strategies, including menin inhibitors, arsenic and all-trans retinoic acid (ATRA), are being explored.[171-174] *NPM1*-mutated AML is generally associated with a favourable prognosis. However, $FLT3^{ITD}$ mutations frequently co-occur and may worsen the clinical outcome.[171]

CBFB::MYH11

CBFB::MYH11 is generated by inv(16)(p13.1q22) or, less commonly, by t(16;16)(p13.1q22) and defines a core binding factor (CBF) AML subtype with a favourable prognosis. The fusion protein disrupts the normal function of the CBF transcription factor complex, which regulates genes required for normal myeloid differentiation, resulting in impaired maturation and accumulation of immature myeloid cells.[171] Additional cooperating mutations, particularly affecting *RAS*, *FLT3* and *KIT* signalling, are usually required for leukemic transformation.[175] *CBFB::MYH11* is present in 10-15% of pediatric AML cases and is less common in adult cases.[171,175,176]

RUNX1::RUNX1T1

RUNX1::RUNX1T1 is generated by t(8;21)(q22;q22) and also defines a CBF AML subtype with a generally favourable prognosis. The fusion protein retains the DNA binding ability of *RUNX1* but replaces its normal activating function with transcriptional repression from *RUNX1T1*. As a result, *RUNX1* target genes required for myeloid differentiation are silenced, blocking normal maturation of myeloid progenitors and leading to accumulation of immature leukemic cells. *RUNX1::RUNX1T1* is one of the most common recurrent chromosomal rearrangements in AML and is more frequent in children than in adults.[171,176]

PML::RARA in acute promyelocytic leukemia

Acute promyelocytic leukemia (APL) is characterized by the fusion gene *PML*::*RARA*, t(15;17)(q24.1;q21.2), and represents a distinct AML entity.[171] The fusion protein blocks normal myeloid differentiation at the promyelocyte stage.[177] APL accounts for approximately 5-10% of pediatric AML cases and 10-15% of adult cases.[178] Unlike most AML subtypes, APL can be treated with targeted

differentiation therapy rather than conventional intensive chemotherapy. ATRA targets the RARA portion of the fusion protein, while arsenic trioxide binds the PML part, leading to degradation of the fusion protein and restoration of differentiation.[177] Although APL has survival rates exceeding 90%, it remains a hematological emergency due to the risk of fatal hemorrhage. ATRA should therefore be initiated immediately upon clinical suspicion, before molecular confirmation. [178]

NUP98 rearrangements

NUP98 rearrangements occur in 4-9% of pediatric and around 2% of adult AML, and over 30 different fusion partners have been discovered.[171,179] *NUP98* fusion proteins dysregulate gene expression through aberrant recruitment of chromatin-modifying complexes, leading to increased HOX expression and impaired differentiation.[171] *NUP98*-rearranged AML is associated with an adverse prognosis, although clinical outcome is influenced by both the fusion partner and cooperating genetic alterations, particularly *FLT3*^{ITD} and *WT1* mutations.[171,179,180]

MECOM rearrangements

MECOM rearrangements are rare in both pediatric and adult leukemia (1-2%) and are associated with a poor prognosis.[171] *MECOM* is a gene locus that includes the *EVI1* transcription factor. In the classical inv(3)/t(3;3) rearrangements, a distal *GATA2* enhancer is repositioned near the *MECOM* locus, leading to aberrant activation of *EVI1* and reduced *GATA2* expression. Increased *EVI1* promotes stem cell self-renewal and impairs normal myeloid differentiation, while reduced *GATA2* disrupts normal hematopoiesis, together contributing to leukemogenesis.[181]

CBFA2T3::GLIS2

CBFA2T3::GLIS2 predominantly affects very young children and usually presents with megakaryoblastic differentiation in the absence of Down syndrome. The fusion is generated by a cytogenetically cryptic inv(16)(p13q24) and is detected in about 3% of AML and 20-30% of pediatric non-Down syndrome acute megakaryoblastic leukemia (AMKL) cases. In contrast to AMKL associated with Down syndrome, this subtype is frequently resistant to chemotherapy and is associated with very poor survival. The resulting chimeric transcription factor dysregulates gene expression, promotes leukemic self-renewal and interferes with normal megakaryocytic differentiation. Recurrent additional driver alterations appear to be uncommon, although RAS pathway mutations occur in a small subset of cases.[182,183]

CEBPA bZIP-mutated AML

This is a recently refined molecular subgroup associated with a favourable prognosis. The bZIP domain is the C-terminal region of the *CEBPA* transcription factor responsible for DNA binding. Mutations affecting this domain impair normal CEBPA transcriptional activity and myeloid differentiation. Co-occurring *WT1* and

GATA2 mutations are relatively common and are each detected in approximately 20% of cases.[184,185]

Additional rare entities

Additional rare subgroups, each present in <2% of cases, will not be further discussed. These include *DEK::NUP214, RBM15::MRTFA, MNX1::ETV6, KAT6A::CREBBP* and *FUS::ERG.*

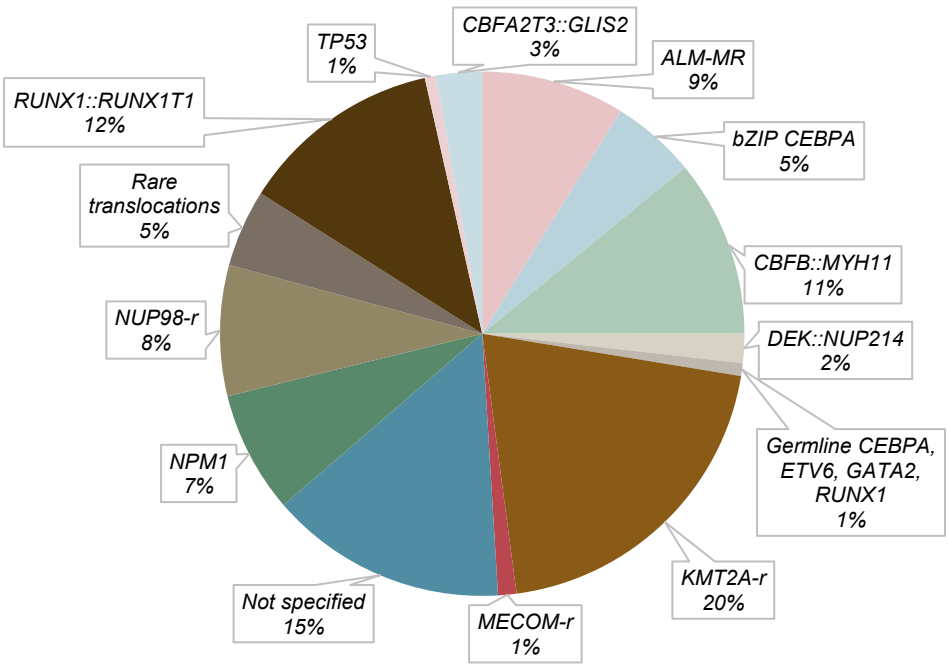


Figure 15: Distribution of molecular categories in a large cohort of pediatric AML. The pie chart includes 881 non-Down syndrome patients aged ≤18 years with diagnostic AML samples from the genomic cohort reported by Umeda et al. (2024). The distribution provides an estimate of molecular subtype frequencies in pediatric AML and does not reflect population-based prevalence. APL was underrepresented in the cohort and was therefore excluded.

Mixed phenotype acute leukemia

MPAL accounts for approximately 2-5% of pediatric leukemia.[142] It is defined by leukemic blasts fulfilling diagnostic criteria for more than one hematopoietic lineage, either as two different blast populations (*bilineage*) or as a single blast population expressing lineage-defining markers of more than one lineage (*biphenotypic*).[97,100,186,187] B-lineage assignment requires either strong CD19 expression together with at least one additional B-cell marker, or weak CD19 with at least two additional markers. T-lineage assignment requires strong cytoplasmic or surface CD3 expression, while myeloid lineage is established by

myeloperoxidase (MPO) expression or clear monocytic differentiation.[187] MPAL most commonly involves B/myeloid and T/myeloid subtypes.[142]

MPAL includes four genetically defined entities characterized by *BCR*::*ABL1* fusions, *KMT2A* rearrangements, *ZNF384* rearrangements or *BCL11B* activation. However, these genetic alterations can also occur in single-lineage leukemia, and the diagnosis of MPAL therefore still requires an appropriate mixed-lineage immunophenotype. Although immunophenotypically heterogeneous, MPAL is thought to arise from early hematopoietic progenitors with multilineage potential. Additional recurrent genetic alterations include mutations in genes such as *WT1*, *RUNX1*, *ETV6* and *CEBPA*, reflecting overlap with both ALL and AML biology.[187] Clinical studies have consistently shown that pediatric patients with MPAL achieve superior outcomes with ALL-directed therapy compared with AML-based or hybrid regimens. Many pediatric MPAL cases harbour genetic features commonly associated with ALL, although the biological basis for the superior response to ALL-directed therapy remains incompletely understood.[142,188]

Secondary genetic alterations in leukemia

There is a complex interplay between genomic subtypes and their secondary alterations, which together influence both leukemogenesis and clinical outcomes in pediatric leukemia. Studies have shown that the number of putative driver gene alterations (4-5 per diagnostic sample) is similar in pediatric and adult leukemia.[116] However, some subtypes, including *KMT2A*-r and aneuploid types, exhibit a lower number (2-3 per diagnostic sample), potentially reflecting a reduced need for secondary driver genes or a younger age at diagnosis.[116,129]

Below is a selection of genes, grouped by pathway, that are frequently mutated or prognostically relevant in pediatric acute leukemia.

Hematopoietic differentiation genes

IKZF1

The transcription factor IKAROS, encoded by the *IKZF1* gene, was first identified in 1992[189] and soon thereafter recognized as playing an important role in leukemia. *IKZF1* alterations include whole-gene or intragenic deletions, the latter generating abnormal *isoforms* (different versions of a protein produced from the same gene). Some of these abnormal isoforms are *dominant negative*, meaning that they interfere with the function of the wildtype IKAROS protein.[190,191] An example is IK6, which is usually generated by deletion of the exons encoding the N-terminal DNA-binding domains. IK6 cannot bind DNA but retains the ability to bind wildtype IKAROS, thereby acting as a dominant negative isoform that interferes with normal IKAROS

function.[190] *IKZF1* alterations can disrupt lymphoid differentiation and are associated with poor prognosis in B-ALL.[190-193]

IKZF1 can also be lost through the formation of an *isochromosome* 7q, or i(7q). An isochromosome results from duplication of one chromosome arm and loss of the other. Thus, i(7q) duplicates the long arm of chromosome 7 while deleting the short arm, 7p, including *IKZF1* (**Figure 16**).[191]

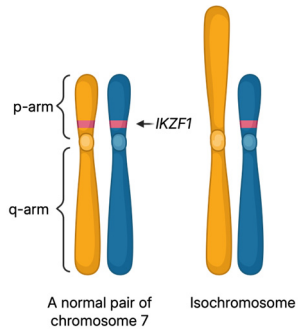


Figure 16: Isochromosome i(7q). A normal pair of chromosome 7 is shown on the left, with *IKZF1* located on the p-arm. To the right, i(7q) is formed through loss of one p-arm and duplication of the q-arm. Illustration created with BioRender.com.

PAX5

PAX5 is a transcription factor and a master regulator of B-cell lineage commitment and differentiation. Genetic alterations affecting *PAX5* include deletions, point mutations and chromosomal rearrangements and are frequently observed in B-ALL.[194] *PAX5* alterations have been associated with an increased risk of relapse and poor prognosis in B-ALL, although their clinical impact varies depending on the alteration type.[193] While *PAX5* alterations often occur as secondary events, specific alterations define distinct B-ALL subtypes, including PAX5alt and PAX5 P80R.[116]

VPREB1

VPREB1 encodes a component of the pre-B-cell receptor, which plays a critical role in early B-cell development. Loss of *VPREB1* impairs pre-B-cell receptor signalling and B-cell differentiation (blocking pro- to pre-B-cell transition) and contributes to leukemogenesis by allowing immature B-cell progenitors to persist and expand.[195]

Cell cycle genes

TP53

TP53 encodes p53, a transcription factor that regulates several essential cellular processes, such as DNA repair, cell cycle arrest, apoptosis, autophagy and metabolism. In response to cellular stress, p53 determines whether a cell should undergo repair or be eliminated, thereby safeguarding genomic integrity and earning

it the title *the guardian of the genome*.[196,197] Loss of p53 function can therefore have devastating consequences.

Over time, research has revealed an increasingly complex role for p53, extending beyond classical tumour suppressive functions to include regulation of the immune microenvironment, stem cell biology and metabolic homeostasis.[197] Given the high frequency of *TP53* mutations across cancers, p53 represents an attractive target for future precision therapies, particularly in leukemia, where *TP53* alterations are associated with inferior prognosis.[96,198,199] However, therapeutic targeting has proven challenging, primarily because the protein lacks well defined druggable binding pockets and there is no clear mechanism for restoring mutant p53 function.[196,200]

The challenge becomes even greater because of the many different *TP53* mutations observed in cancer, most of which are missense mutations in the DNA binding domain. These alterations impair p53 activity through diverse structural and functional mechanisms, making it difficult to develop a single drug capable of broadly restoring function.[200] New approaches are therefore being developed to identify therapeutic vulnerabilities, including prime-editing screens that assess the functional effects of large numbers of *TP53* variants in their normal genomic context.[201]

The clinical impact of a *TP53* alteration depends not only on the presence of a mutation but also on its allelic state. p53 functions as a tetramer, with four protein subunits cooperating to regulate transcription. Like *IKZF1*, monoallelic *TP53* mutations can impair p53 function through dominant negative effects, where mutant p53 interferes with the function of the remaining wildtype protein.[202] In AML, biallelic *TP53* alterations are typically associated with dominant clones, occur independently of other common myeloid gene mutations and are strongly linked to adverse outcomes.[145,203] In contrast, monoallelic mutations frequently coexist with alterations in genes such as *TET2*, *ASXL1* and *DNMT3A*, and are often regarded as later subclonal events with more variable prognostic impact.[145,204]

Exposure to cytotoxic therapies has been associated with a higher frequency of *TP53*-mutant clones. This can reflect selective expansion of pre-existing clones,[117,119] but chemotherapy has also been suggested to directly induce *TP53* mutations.[146,205] This is further discussed in **Articles I** and **II**.

CDKN2A/B

The cyclin-dependent kinase inhibitor genes *CDKN2A* and *CDKN2B* encode key tumour suppressors that control the G1/S checkpoint through the retinoblastoma (RB) and *TP53* pathways. Deletions affecting *CDKN2A/B* are among the most common genetic alterations in childhood ALL, observed in 21% of B-ALL and 50% of T-ALL, and lead to dysregulation of cell cycle control.[206] Homozygous deletions have been associated with an increased risk of relapse in childhood B-ALL.[193]

CDKN2A alterations include CN-LOH, heterozygous and homozygous deletions. In some cases, a large heterozygous deletion is accompanied by a smaller homozygous deletion nested within the larger deleted region, consistent with sequential loss of both *CDKN2A* alleles during leukemic evolution.[206,207]

RAS-pathway genes

The RAS gene family represents the most frequently mutated group of oncogenes in human cancers. RAS proteins function as molecular switches that regulate key signalling pathways controlling cell proliferation and survival, including MAPK and PI3K pathways. Oncogenic RAS mutations lock the protein in an active state, resulting in constitutive downstream signalling and uncontrolled tumour growth.[208] The frequent detection of subclonal RAS pathway mutations at diagnosis indicates that these alterations are secondary events during leukemia evolution rather than initiating lesions.[209] During treatment, some subclonal RAS pathway mutations can persist and expand at relapse, whereas others are lost, reflecting changes in clonal composition under treatment.[210,211] This is further discussed in **Article II**.

In *KMT2A*-r ALL, these activating mutations are frequently subclonal and not mutually exclusive. Instead multiple different RAS-mutated clones often coexist within a single patient.[129,207,212,213] In AML, on the other hand, RAS mutations appear to be mutually exclusive,[214] and a recent study of *KMT2A*-r AML showed that *KRAS* mutations are associated with inferior outcomes.[215] Previous work from our group found that RAS mutations promote clonal evolution and enhance clonal fitness in *KMT2A-r* leukemia through increased proliferative signalling and altered cytokine expression, supporting the founding clone.[216]

KRAS is the most common oncogenic driver in human cancers, accounting for 90% of RAS mutations. Most mutations are gain of function and single base missense changes, with over 90% occurring in just three hotspot codons, G12, G13 and Q61.[217] *NRAS* is also a recurrent cancer driver, particularly in melanoma and leukemia.[218] Although RAS was long considered "undruggable" due to its biochemical properties and lack of suitable drug-binding sites, recent advances have enabled the development of mutant-specific inhibitors such as those targeting $KRAS^{G12C}$, demonstrating that RAS-driven cancers can be targeted.[208,217]

FLT3 encodes a receptor tyrosine kinase. Mutations occur mainly as internal tandem duplications (ITDs), predominantly in AML, or as point mutations in the tyrosine kinase domain (TKD), also found in ALL. Both cause ligand-independent, constitutive activation, driving cell proliferation (**Figure 17**).[219] $FLT3^{ITD}$ mutations are present in approximately 15% of pediatric AML and about twice as frequently in adults. They are enriched in *KMT2A*-r AML and associated with poor prognosis.[220-222] *FLT3* inhibitors, such as quizartinib, are currently implemented for molecularly defined risk groups in CHIP-AML22.[169] Notably, in **Article III**,

characterization of a large *KMT2A*-r leukemia cohort showed that the majority of *FLT3* mutations in AML were TKDs and not ITDs.

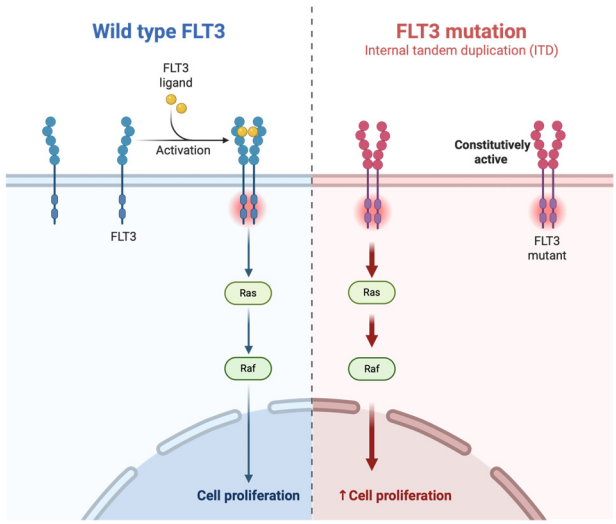


Figure 17: FLT3 signalling. Left: Normal function of the FLT3 tyrosine kinase, which is activated upon ligand binding, triggering the downstream RAS pathway and promoting cell proliferation. Right: The *FLT3*^{ITD} mutation, which results in a constitutively active tyrosine kinase, leading to persistent activation of the RAS pathway and aberrant cell proliferation. Illustration created with BioRender.com.

NF1 encodes neurofibromin, a tumour suppressor that negatively regulates RAS signalling. *NF1* alterations are present in around 4-11% of adult AML and 2% in pediatric AML.[214] Lastly, *PTPN11* encodes SHP2, a protein tyrosine phosphatase that is expressed in most cell types and plays an important role in transmitting signals from cell surface receptors to intracellular signalling pathways.[222] Alterations in *NF1* and *PTPN11* have been associated with a poor prognosis in pediatric AML.[214]

Epigenetic mutations

Mutations in epigenetic regulators are among the most common cooperating events in leukemia. By disrupting the machinery responsible for DNA methylation, histone modification or chromatin organization, these mutations reprogram transcriptional networks and impair normal hematopoietic differentiation. These mutations affect three principal layers of epigenetic control: DNA methylation, histone modifications and higher-order chromatin architecture.

Mutations in genes regulating DNA methylation (e.g. *DNMT3A*, *TET2* and *IDH1/2*) disrupt the balance between DNA methylation and demethylation, resulting in aberrant DNA methylation patterns, altered transcriptional programs and impaired hematopoietic differentiation.[223-225] Similarly, mutations in chromatin-modifying genes (e.g. *ASXL1* and *EZH2*) alter histone modifications, leading to dysregulated

gene expression, enhanced self-renewal and blocked differentiation.[226,227] Mutations affecting the cohesin complex, such as *STAG2*, disrupt the normal organization of chromatin within the nucleus. Consequently, genes may become inappropriately activated or silenced, leading to abnormal transcriptional programs that promote leukemogenesis.[228,229] Epigenetic alterations not only contribute to leukemogenesis but also establish transcriptional states that maintain leukemia stem cell identity and promote therapy resistance, highlighting epigenetic plasticity as a central mechanism underlying disease persistence and relapse.[230]

Assessing the functional impact of cooperating mutations

When multiple somatic alterations co-occur, it can be difficult to determine which are functional drivers and which are passenger mutations. Recurrent alterations are more likely to be functionally relevant, while computational predictions and experimental studies can provide additional evidence.

Mutations in signalling genes, such as *FLT3* or *RAS*, can be tested using the murine Ba/F3 cell line. These cells normally depend on a growth factor (the cytokine IL-3) to survive and if IL-3 is removed, they rapidly die. In this assay, the mutated gene of interest (e.g. *FLT3* or *KRAS*) is first inserted into an expression vector, which acts as a molecular transport vehicle to deliver the genetic material. Once this vector is transduced into the Ba/F3 cells, it forces them to produce the mutated protein. The cells are then cultured in a medium without IL-3. If the cells continue to grow and multiply without this growth factor, this demonstrates growth factor-independent proliferation.[129,216,231] Mutant proteins can also be overexpressed in other relevant cell models to assess their effects on survival, proliferation or cellular transformation. However, failure to induce transformation does not exclude a functional role, as some alterations require cooperating mutations or an appropriate cellular context.

To confirm this autonomy at the molecular level, downstream pathway activation can be assessed. Since signalling receptors such as FLT3 and intracellular mediators such as RAS both converge on the MAPK cascade, activating mutations can result in constitutive pathway activation. This can be assessed by Western blotting for phosphorylated ERK (p-ERK), a key downstream signalling protein in this cascade. A strong p-ERK signal in cells cultured without IL-3 provides evidence of constitutive pathway activation.[129,231]

For other classes of alterations, such as inactivating mutations in tumour suppressor genes or epigenetic regulators, different functional approaches are required. In contrast to oncogenes, tumour suppressors generally require loss of function of both alleles to promote tumorigenesis. Functional validation therefore often relies on gene knockout or knockdown strategies or restoration of the wildtype gene in deficient cells, followed by assessment of proliferation, apoptosis, differentiation or colony-forming capacity. Depending on the biological function of the gene,

transcriptomic profiling or differentiation assays may also be used to evaluate downstream effects.[6,232]

Genetic predisposition

A minority of childhood leukemia cases occur in individuals with an underlying constitutional or germline predisposition. Pathogenic germline variants have been identified in approximately 4% of children with leukemia, although the reported frequency varies depending on the patient population, genes analysed and variant interpretation criteria.[233,234] In adults, pathogenic or likely pathogenic germline variants occur in approximately 14% of AML cases,[235] whereas their frequency in adult ALL remains less well characterized.

Several constitutional syndromes are associated with an increased risk of leukemia. Down syndrome increases the risk of both ALL and AML, particularly AMKL. Down syndrome is also associated with transient abnormal myelopoiesis, a preleukemic condition that resolves spontaneously in most children but can precede myeloid leukemia.[236-238] Constitutional DNA repair disorders, such as Fanconi anemia and Nijmegen breakage syndrome, also confer an increased risk of leukemia. Li-Fraumeni syndrome, caused by germline *TP53* variants, is strongly associated with low-hypodiploid B-ALL.[143]

Although inherited leukemia predisposition is uncommon, germline variants in *ETV6*, *PAX5, GATA3* and *IKZF1* are mainly associated with B-ALL, whereas *RUNX1*, *GATA2* and *CEBPA* variants predominantly predispose to myeloid neoplasms. Such variants are found both in families with multiple affected individuals and in patients without an apparent family history.[136,142,239]

Pediatric versus adult malignancies

The notion that “children are simply small adults” can make any pediatrician, and especially any pediatric oncologist, shiver **(Figure 18)**. One major difference is that cancer is much less common in children than in adults. The types of cancer also differ, with adults mainly developing epithelial cancers (e.g., breast, lung and colorectal cancer), while childhood cancers more often arise from developing tissues and include leukemia, brain tumours, embryonal tumours and sarcomas.[4,240] While cancer in older individuals results from a complex interplay between inherited predisposition, environmental exposures and the gradual accumulation of somatic mutations over time, pediatric cancers more often arise from genetic and epigenetic alterations acquired during fetal development or early childhood.[53,238,241] This difference is reflected in tumour mutational burden. Cancers associated with environmental mutagens, such as melanoma and smoking-associated lung cancer, exhibit the highest mutational burdens, whereas pediatric cancers have among the

lowest mutational burdens.[10,242] The relatively low mutational burden of pediatric cancers is thought to reflect their early onset, leaving less time for somatic mutations to accumulate before malignant transformation. Consequently, pediatric cancers are often driven by a limited number of genetic alterations, frequently involving chromosomal rearrangements and developmental pathways. However, as in adult cancers, most somatic mutations are likely to represent passenger events.[10,240,243]

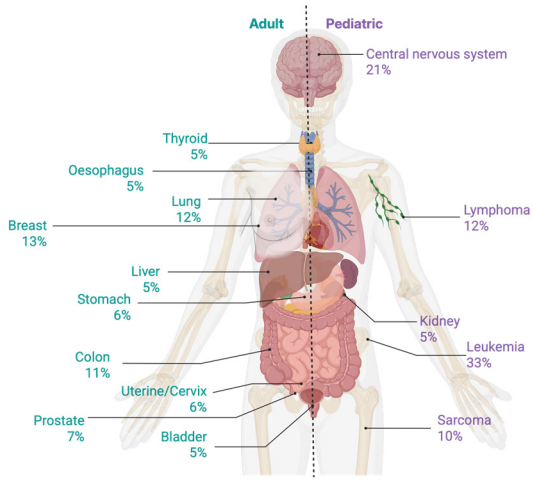


Figure 18: Common organ sites for adult and pediatric cancer. The figure is based on data from Jassim et al. (2023). Only cancer sites accounting for more than 4% of cases were included. Where frequencies differed between adult females and males, the mean of the two values is presented. Illustration created with BioRender.com.

The spectrum of driver genes differs between pediatric and adult cancers. Pediatric malignancies are enriched for alterations affecting transcriptional regulation, epigenetic control, cell-cycle regulation and developmental signalling pathways, with relatively limited overlap with the most frequently mutated driver genes in adult cancers.[244] Consistent with their low mutational burden, many pediatric cancers are driven by structural variants, particularly chromosomal rearrangements that generate oncogenic fusion genes and disrupt normal transcriptional and developmental programmes.[243,245] Since these driver alterations frequently involve transcription factors and epigenetic regulators rather than recurrent kinase mutations that can be directly targeted, the development of effective targeted therapies has been challenging, highlighting the need for therapeutic strategies targeting chromatin regulation and epigenetic mechanisms.[101,245,246]

These age-related differences are also evident in acute leukemia. ALL is the most common leukemia in children, whereas AML is the most common acute leukemia in adults.[247] This age distribution reflects differences in hematopoietic development and leukemic mechanisms: childhood ALL typically arises from lymphoid progenitors during early immune system development and is often initiated by

chromosomal rearrangements that arise during fetal development, whereas adult AML more frequently develops through the gradual accumulation of somatic mutations in hematopoietic cells.[53,248] HSCs also become increasingly biased towards myeloid differentiation with age, while their lymphoid potential decreases.[249]

Lastly, the impact of prolonged survival differs fundamentally between pediatric and elderly patients. An equivalent gain in survival does not have the same impact across age groups. While five additional years of survival in an older individual may represent a substantial proportion of remaining life expectancy, the same gain in a young child falls far short of restoring a normal life span. Consequently, preventing relapse is critical in childhood cancers, where long term disease control is essential to preserve an entire life trajectory.[250]

KMT2A-rearranged leukemia

Prevalence

KMT2A rearrangements occur in leukemia across all age groups and are found in approximately 10% of acute leukemia cases. *KMT2A* rearrangement is the most common genetic abnormality in infants, representing over 70% of infant ALL and 35-50% of infant AML cases.[238,251] In older children, *KMT2A* rearrangements are found in 5-6% of ALL and 20-25% of AML cases, and in adults, in approximately 15% of ALL and 6% of AML cases.[251-254]

Approximately 70% of therapy-related leukemias occurring after treatment with topoisomerase II inhibitors, such as etoposide or doxorubicin, harbour *KMT2A* rearrangements. Topoisomerase II is an enzyme involved in the relaxation of supercoiled DNA during chromatin remodelling. When the enzyme is blocked by these drugs, it forms a complex that causes breaks in both strands of DNA. Repair of these breaks predominantly occurs through the error-prone NHEJ pathway, contributing to genomic alterations involving the *KMT2A* locus.[255] Between 2 and 12% of patients undergoing treatment with topoisomerase II inhibitors are diagnosed with leukemia, of whom a majority develop AML. Therapy-related leukemia is characterized by a short latency and may occur as early as six months after exposure, although it generally develops within 2-4 years. Cleavage sites for topoisomerase II inhibitors are distributed in the susceptible scaffold attachment regions (SARs) of *KMT2A.* Interestingly, breakpoint patterns observed in infant leukemia share similarities with those in therapy related leukemia. Maternal exposure during pregnancy to certain natural substances with topoisomerase II inhibitory activity, such as quercetin found in onions and apples and genistein found in soy products, has been proposed as a possible contributor to *KMT2A*

rearrangements in infant leukemia. However, these exposures are common, whereas infant leukemia is very rare and a causal relationship has not been established.[255]

Although rare outside acute leukemia, *KMT2A* rearrangements have also been identified in B-lymphoblastic lymphoma and in solid tumours, including aggressive thymomas and sarcomas primarily affecting young adults.[256-258]

The normal function of the *KMT2A* gene

The *KMT2A* gene was formerly called *MLL*, an abbreviation for myeloid/lymphoid or mixed-lineage leukemia since *KMT2A* rearrangements can be found in both ALL and AML, with some cases exhibiting ambiguous or mixed-lineage characteristics.[204] Contrary to what this historical nomenclature might suggest, only a minority of MPALs have *KMT2A* rearrangements.[255] *KMT2A* encodes lysine methyltransferase 2A, is located on chromosome 11q23, and was identified and characterized in the early 1990s. It is a relatively large gene with 38 exons spread across a 90 375 base pair region.[259,260]

KMT2A belongs to a group of proteins called *trithorax* proteins, which maintain gene expression patterns through cell divisions and help cells preserve their identity over time. This system works in opposition to the *polycomb* proteins, which mainly silence genes.[259,261] Transgenic mouse models, which are genetically engineered to carry specific alterations, have shown that KMT2A is required both for the generation of HSCs during embryonic development and for the long-term maintenance of HSCs in adulthood.[262,263] One of KMT2A's major targets is the HOX gene cluster, which regulates embryonic development, stem cell renewal and blood cell formation. KMT2A activates these genes by promoting histone H3 lysine 4 (H3K4) methylation, which keeps chromatin in an open and transcriptionally active state.[262,264]

The protein is composed of an N-terminal region responsible for promoter recognition and DNA binding and a C-terminal region involved in chromatin binding and H3K4 methylation.[259] The N-terminal region contains several functional domains, including the menin-binding motif, AT-hooks, CxxC domain, subnuclear localisation domains, repression domain, four plant homeodomain fingers, bromodomain and FYRN domain. The C-terminal region contains the FYRC domain, transactivation domain and SET domain, the latter being responsible for H3K4 methyltransferase activity (**Figure 19**).[265]

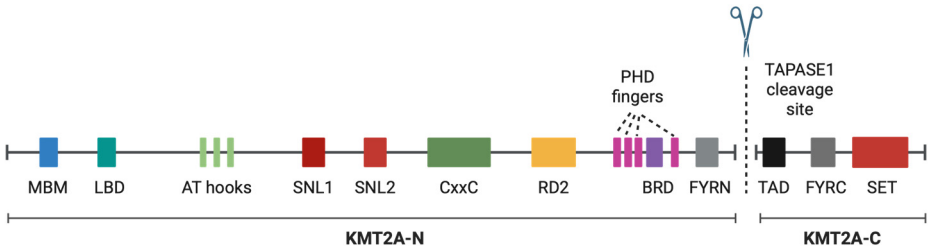


Figure 19: Domain structure of wild type KMT2A. KMT2A is cleaved by the enzyme TASPASE1 into two subunits, with the N-terminal subunit depicted on the left and the C-terminal subunit on the right. KMT2AN, N-terminal KMT2A subunit; KMT2AC, C-terminal KMT2A subunit; MBM, menin-binding motif; LBD, LEDGF binding domain; AT hooks, adenosine-thymidine hooks; CxxC, cysteine-rich zinc finger domain; SNL1/2, subnuclear localisation domains 1 and 2; RD2, repression domain 2; PHD, plant homeodomain; BRD, bromodomain; FYRN, FY-rich N-terminal domain; TAD, transactivation domain; FYRC, FY-rich C-terminal domain; SET, Su(var)3-9, Enhancer-of-zeste and Trithorax domain; H3K4, histone H3 lysine 4. Illustration created with BioRender.com.

Fusion partners in *KMT2A*-r leukemia

Over 130 gene partners have been identified in *KMT2A*-r leukemia, several of which are clinically relevant for risk stratification and disease monitoring.[266] Despite this diversity, a mere six of them account for 85% of all cases (*AFF1, MLLT3, MLLT1, MLLT10, AFDN* and *ELL*). While many *KMT2A* fusions occur across lineages, specific patterns exist. For instance, *KMT2A::MLLT3* is strongly linked to AML, whereas *KMT2A::AFF1* dominates in ALL, particularly in infants and adults.[267] Predicting outcomes is complicated by the fact that the same fusion can carry different prognostic weights across age groups,[268,269] with most *KMT2A* rearrangements being associated with a dismal prognosis[259] (**Figure 20**).

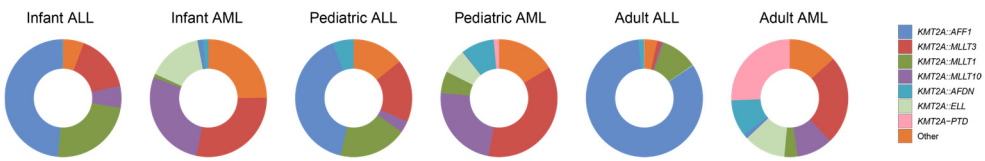


Figure 20: Distribution of *KMT2A* rearrangements across age groups and leukemia subtype. Distribution of the most common *KMT2A* fusion genes across infants, children and adults with ALL or AML. Data are based on 3 267 patients from Meyer et al. (2023).

Over 95% of *KMT2A* breakpoints occur within the major breakpoint cluster region (BCR1), spanning intron 7 to exon 13. The highest concentration is between introns 9 and 11, a region containing repetitive elements and several sites associated with chromatin accessibility and DNA cleavage. Only around 1% of breakpoints occur in the minor breakpoint cluster region (BCR2), between intron 20 and exon 24, most often involving *USP2*, *USP8* or *AFDN* (with the latter occurring in both BCR1 and

BCR2). BCR1 also contains SARs, which attach the DNA to proteins in the nucleus and make this region more prone to breakage (**Figure 21**).[255,259,267,270]

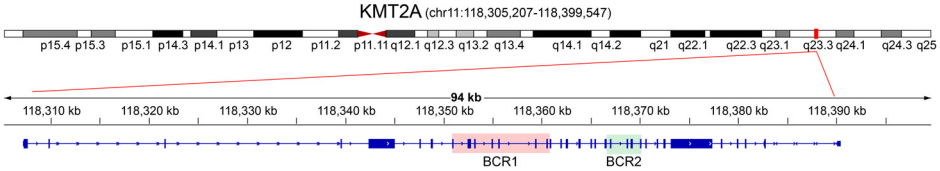


Figure 21: Overview of the *KMT2A* gene. Illustration of the exons and introns in *KMT2A*, with BCR1 and BCR2 marked.

These chromosomal rearrangements disrupt the normal structure of *KMT2A*. The fusion protein retains the N-terminal DNA-binding domains, allowing it to remain associated with its target promoters, including HOX genes, while the C-terminal region, which mediates H3K4 methylation, is lost and replaced by sequences from the fusion partner.[255,259] Consequently, the fusion protein loses its normal epigenetic regulatory function while acquiring aberrant transcriptional activity through interactions mediated by the fusion partner. Together with interaction partners such as MENIN and the polymerase-associated factor complex (PAFc), the fusion proteins remain localized to target genes and maintain abnormal gene activation.[271-273] MENIN is essential for both wildtype and rearranged KMT2A.[272] As a result, *KMT2A* fusion proteins continue to activate stem cell associated transcriptional programs in an abnormal and persistent manner, even though large parts of the protein are lost, promoting leukemic transformation.

KMT2A aberrations drive leukemogenesis through different molecular mechanisms. The most common are fusions with nuclear translocation partners (NTPs), including *AFF1*, *MLLT3* and *MLLT1*. These fusions recruit proteins involved in transcriptional elongation and histone modification, including the super elongation complex (SEC) and DOT1L-associated complexes, to KMT2A target genes. This leads to increased transcriptional elongation and H3K79 methylation, keeping critical hematopoietic progenitor genes such as *HOXA9* and *MEIS1* transcriptionally active.[274-277] Together, these mechanisms activate the transcription factor MYB, which is essential for *KMT2A*-r leukemia.[278] MYB sustains leukemia development by activating oncogenic genes, often leading to prominent MYC-associated transcriptional signatures (**Figure 22**).[279,280] Consequently, these leukemia cells express a gene profile similar to normal progenitor cells, but with an aberrant self-renewal program.[281,282]

A smaller subset of *KMT2A* fusion partners, including *GAS7* and *EPS15*, are cytoplasmic translocation partners (CTPs). Instead of primarily acting through transcriptional activation domains, these proteins can bring several KMT2A fusion proteins together through dimerization or oligomerization. This strengthens their interaction with chromatin and activates leukemia-promoting genes (**Figure 22**).[274-

[277,283,284] *AFDN* also belongs to this group, although its mechanism is more complex. *KMT2A::AFDN* promotes oligomerization but remains dependent on DOT1L-mediated H3K79 methylation.[270]

Both pathways disrupt epigenetic homeostasis and impair hematopoietic differentiation, making components of these complexes potential targets for small molecule inhibition.

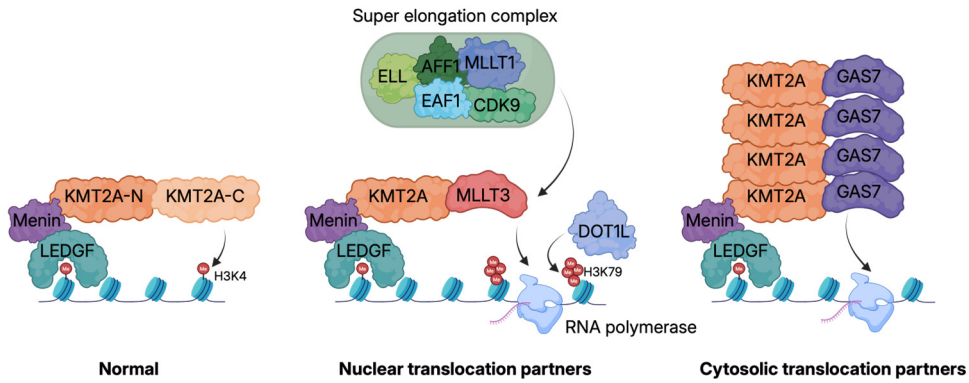


Figure 22: Molecular mechanisms of *KMT2A* fusion proteins. Wildtype *KMT2A* associates with MENIN and LEDGF and regulates transcription through its C-terminal SET domain, which mediates H3K4 methylation. In fusions with nuclear translocation partners, such as *MLLT3*, transcriptional elongation complexes and DOT1L are recruited, promoting RNA polymerase elongation and H3K79 methylation. In contrast, cytoplasmic translocation partners, such as *GAS7*, promote oligomerization of *KMT2A* fusion proteins, resulting in aberrant activation of *KMT2A* target genes through mechanisms that remain incompletely understood. Both mechanisms contribute to persistent activation of stem cell-associated transcriptional programs and leukemic transformation. Illustration created with BioRender.com and adapted from Perner et al. (2025).

Beyond fusion genes, *KMT2A* can also undergo pathogenic remodelling through internal rearrangements. *KMT2A* partial tandem duplications (*KMT2A*-PTD) occur when exons 3-9 or 3-11 are duplicated in tandem within the gene itself.[285] While the precise mechanism underlying *KMT2A*-PTD leukemogenesis remains unclear, it may involve two complementary epigenetic effects. *KMT2A*-PTD activates the HOXA9/MEIS1 oncogenic program while promoting DNA hypermethylation and silencing of tumour suppressor genes.[286,287]

WGS has shown that the majority of *KMT2A* rearrangements are structurally complex and unbalanced, rather than simple reciprocal translocations (**Figure 6**). At the nucleotide level, more than half of these events involve three or more chromosomes and are accompanied by large deletions, insertions or inversions adjacent to the breakpoints. Even translocations that appear cytogenetically balanced harbour focal deletions or insertions at the base-pair level. These rearrangements typically generate an in frame *KMT2A* fusion, while reciprocal products can be heavily disrupted or rarely expressed. This genomic instability is

predominantly driven by error prone NHEJ during the rearrangement process.[129] This is further investigated in **Article III**.

Below is a short description of different characteristics of the six most common fusion genes.

AFF1/AF4/t(4;11)

KMT2A::AFF1 is predominantly associated with ALL and is more common in infants and adults, whereas older children have a more diverse range of fusion partners (**Figure 20**).[267,269] The biological basis for this age-dependent distribution remains incompletely understood but may reflect differences in the developmental cell of origin and the cellular context in which the rearrangement occurs.[274,288]

Like several other *KMT2A* rearrangements, the t(4;11) translocation is often cytogenetically balanced. In these cases, the reciprocal *AFF1*::*KMT2A* fusion is frequently retained and expressed. This was also the case in **Article III**. Notably, the reciprocal fusion does not appear to be a silent bystander but possesses independent oncogenic capacity and is capable of autonomously inducing ALL *in vivo*.[289] Interestingly, expression of the reciprocal *AFF1*::*KMT2A* fusion has also been associated with an improved outcome in infants.[290] Together, these findings highlight the complex biological and clinical role of the reciprocal fusion.

MLLT1/ENL/t(11;19)

KMT2A::MLLT1 is mainly associated with ALL and is the second most common fusion gene in ALL across all age groups.[267] It is associated with a particularly poor prognosis in childhood AML.[291] In about 50% of cases, *MLLT1* can form *spliced fusions*, in which the DNA breakpoint occurs upstream of the *MLLT1* gene, leaving *MLLT1* intact (**Figure 4**). The *KMT2A* and *MLLT1* sequences are not directly fused at the DNA level but are joined during RNA splicing, generating a *KMT2A::MLLT1* fusion transcript without an expressed reciprocal fusion. Other fusion partners rarely form spliced fusions.[23,267,274] The same pattern was also observed in **Article III**.

MLLT3/AF9/t(9;11)

KMT2A::MLLT3 is the most common fusion in AML and the third most common driver fusion in infant and childhood ALL.[267] In CHIP-AML22, this is the only *KMT2A* rearrangement that is not automatically considered high risk, and patients with low measurable residual disease (MRD) are assigned to the standard risk group. The biological basis for its relatively favourable outcome remains unclear.[169] In ALL, *KMT2A::MLLT3* is still considered high risk.[139]

MLLT10/AF10/t(10;11)

KMT2A::MLLT10 is the second most common fusion in AML, is less frequent in ALL and is associated with an adverse prognosis.[267,291,292] Unlike the other fusion

genes, *KMT2A* and *MLLT10* are transcribed in opposite directions and thus an in frame oncogenic transcript cannot be formed by a simple reciprocal translocation. This process requires a more complex rearrangement with at least three chromosomal breaks, which frequently makes the fusion cytogenetically cryptic (**Figure 23**).[293]

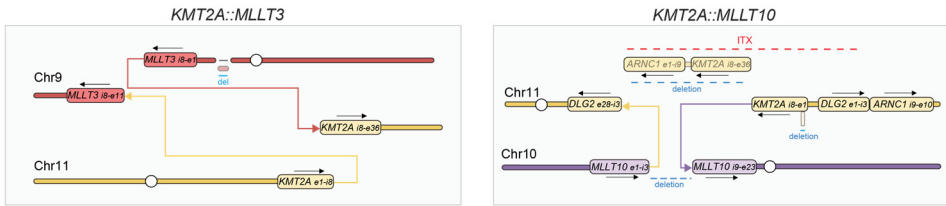


Figure 23: Examples of different *KMT2A* fusions. On the left, a simpler *KMT2A::MLLT3* fusion with a small deletion (del) is shown, and on the right, a more complex *KMT2A::MLLT10* fusion involving a small del, two larger chromosomal deletions (CNA loss) and an intrachromosomal translocation (ITX).

ELL/t(11;19)

KMT2A::ELL is the third most common fusion in infant AML but is rare in ALL. *ELL* and *MLLT1* are both located on chromosome 19, with *ELL* located closer to the centromere.[267]

AFDN/AF6/MLLT4/t(6;11)

KMT2A::AFDN is the third most common fusion in pediatric AML and the fourth most common in pediatric ALL.[267] Remarkably, the small RA1 domain of *AFDN*, comprising approximately 100 amino acids, is sufficient to promote leukemic transformation.[294] *KMT2A::AFDN* is associated with particularly poor outcomes in childhood AML.[291]

The cellular origin of *KMT2A*-r leukemia

Identifying the cellular origin of *KMT2A*-r leukemia remains challenging. Early studies found that introducing fusions like *KMT2A*::*MLLT1* or *KMT2A*::*MLLT3* into murine HSCs and committed myeloid progenitors using retroviral transduction, followed by transplantation into recipient mice, was sufficient to trigger AML.[281,295] *KMT2A* fusion genes induce a transcriptional "self-renewal" program in committed progenitor cells, giving them stem cell-like properties.[281] However, these early models relied on forced fusion gene expression, and transplantation of GMPs expressing endogenous levels of *KMT2A::MLLT3* failed to initiate leukemia.[296]

Later, conditional and inducible mouse models allowing controlled fusion gene expression produced conflicting results. Some studies indicate that HSCs are protected against *KMT2A*-mediated transformation, with downstream progenitors

being susceptible.[297] In contrast, another study showed that *KMT2A*::*MLLT3* can transform both HSCs and GMPs, with more aggressive disease originating from HSCs.[298] These studies model murine AML and should not be generalized to human *KMT2A*-r ALL. Syngeneic mouse models frequently generate myeloid disease, whereas human cord-blood xenograft models have shown that *KMT2A::AFF1* introduced into HSC and progenitor cells can generate B-cell precursor ALL.[299]

In murine *KMT2A*::*MLLT1* AML models, leukemic transformation depends on C/EBPα activity. C/EBPα is a critical transcription factor needed to form GMPs.[82,300] When *CEBPA* is genetically knocked out, the fusions completely lose their ability to initiate leukemia.[301] These findings establish that *KMT2A*-driven transformation depends not only on the fusion itself but also on the transcriptional and differentiation state of the target cell.

The multilineage potential of *KMT2A*-r leukemia is further illustrated by *lineage switch*, most commonly from B-ALL to AML, although the reverse has also been reported.[187] *KMT2A*::*AFF1* is the most frequently reported fusion among cases of lineage switch, although it remains unclear whether this reflects a fusion-specific potential or its high prevalence in *KMT2A*-r ALL.[302] The leukemia retains the same founding genetic alteration, demonstrating a common clonal origin. Although rare, lineage switch is associated with *KMT2A*-r leukemia, particularly following CD19-directed therapies. This phenomenon reflects the remarkable lineage plasticity of *KMT2A*-r leukemia and is thought to result from epigenetic and transcriptional reprogramming rather than the emergence of a new leukemia clone.[303-305]

Together, these studies show that several hematopoietic cell types can support *KMT2A*-mediated transformation, while the precise cellular origin in human disease remains unresolved. Experimental susceptibility to transformation does not necessarily identify the true cell of origin. The resulting phenotype likely depends on both the fusion partner and the developmental and differentiation state of the target cell.[131,305] This concept is further explored in **Article III**.

Gene expression and cellular states in *KMT2A*-r leukemia

When comparing gene expression in *KMT2A*-r ALL and AML to non-*KMT2A*-r leukemia, clear differences are observed, regardless of leukemia type and fusion partner. In *KMT2A*-r leukemia, *HOX* cluster genes, especially *HOXA* genes, and their cofactor *MEIS1* are among the most highly expressed genes. *HOX* genes code for transcription factors that control normal blood cell development and differentiation. *HOXA* and *MEIS1* are normally expressed in stem cells and early progenitor cells, and their expression decreases as cells mature. In *KMT2A*-r leukemia, this downregulation does not occur, resulting in persistent activation of the *HOXA*/*MEIS1* transcriptional program.[255] A similar transcriptional program is observed in other molecular subtypes, including *NPM1*-mutated and *NUP98*-r

AML. In *NPM1*-mutated AML, mutant *NPM1* helps maintain *HOXA* and *MEIS1* expression through a transcriptional machinery that includes the MENIN-KMT2A complex, demonstrating how different initiating alterations can converge on a common oncogenic program.[306] Persistent *HOXA*/*MEIS1* expression maintains an immature, self-renewing cellular state, blocks normal differentiation and supports leukemic proliferation and survival. Although this program is not unique to *KMT2A*-r leukemia, functional studies have shown that *MEIS1* is critical for the proliferation and maintenance of *KMT2A*-r leukemia.[255]

Single-cell approaches have enabled detailed characterization of the developmental states and cellular heterogeneity of *KMT2A*-r leukemia.[131,307,308] *KMT2A*-r B-ALL is characterized by an immature transcriptional state, with a strong early lymphocyte precursor (ELP) program observed in both infants and older children.[131] However, developmental heterogeneity appears to be particularly pronounced in younger patients. Single-cell analyses of *KMT2A*-r ALL from infants and older children showed increased developmental heterogeneity and lineage plasticity in the youngest patients, including more frequent co-expression of lymphoid and myeloid programs.[307] *KMT2A*-r B-ALL can be divided into two transcriptional groups, *KMT2A*-a and *KMT2A*-b.[116] A subsequent study showed that these transcriptional groups could be almost completely separated by their multipotency score (reflecting transcriptional similarity to multipotent hematopoietic progenitors) and classified them as early/multipotent and committed. The early/multipotent group showed greater similarity to early lymphoid and myeloid progenitors, whereas the committed group had a more B-lineage-restricted, pre-B-like state. These states were also associated with the *KMT2A* fusion partner, with *KMT2A::AFF1* being enriched for the early/multipotent state, while *KMT2A::MLLT3* and *KMT2A::MLLT10* were associated with more committed pre-B states. Multipotent states were enriched in subtypes prone to lineage switching, and across pediatric B-ALL, a higher multipotency score was associated with inferior survival.[308] Together, these findings show that developmental heterogeneity is not restricted to infant leukemia and may contribute to the lineage plasticity characteristic of *KMT2A*-r leukemia. This mixed lineage identity can also be reflected by the co-expression of lymphoid and myeloid antigens, potentially creating opportunities for selective immunotherapeutic approaches. This is further discussed in **Article III**.

Single-cell studies of *KMT2A*-r AML have similarly revealed heterogeneity in the leukemia stem cell (LSC) compartment. *KMT2A*-r leukemia can be maintained either by primitive, quiescent LSCs ("*KMT2A*-r early", enriched for *KMT2A::AFDN*) or by more differentiated myeloid progenitors with acquired self-renewal capacity ("*KMT2A*-r committed", enriched for *KMT2A::MLLT3*). Thus, both ALL and AML contain different developmental states despite sharing the same initiating *KMT2A* rearrangement.[307,309]

Single-cell analysis of relapse has identified relapse-initiating cells that are already present at diagnosis in *KMT2A*-r leukemia. These cells are characterized by a

quiescent gene expression program, stem cell like properties, basal activation of glucocorticoid response pathways and reduced cell size, features linked to therapy resistance and disease recurrence.[310] Other single-cell studies have similarly identified subpopulations with reduced glucocorticoid sensitivity and mechanisms of immune evasion from cytotoxic T-cells and NK-cells.[131,307]

Mouse models of *KMT2A*-r leukemia

Mouse models allow us to investigate the genetics and biology of human cancer *in vivo*. Mice are widely used as animal models because of their genetic similarity to humans, size, reliable breeding, short life span and cost efficiency. However, there are important differences between human and murine tumorigenesis, including oncogene and tumour suppressor biology, genetic background, hematopoiesis and the cellular targets of transformation.[311] Several *in vivo* models have been developed to study hematologic malignancies in a physiologically relevant context and to assess leukemic cell proliferation and disease progression (**Figure 24**).

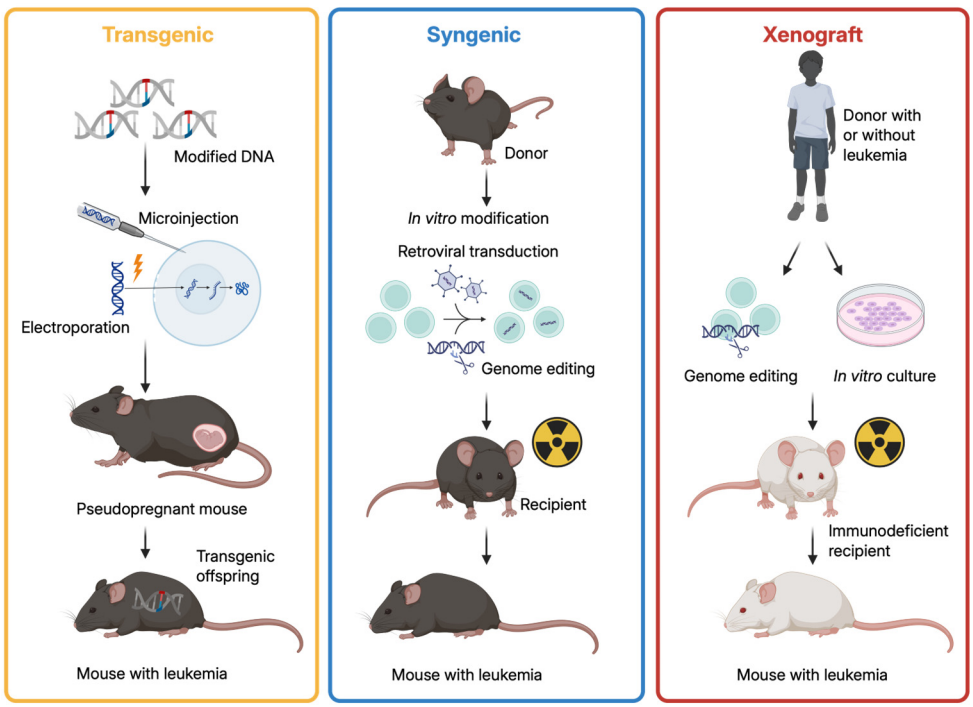


Figure 24: Mouse models used to study leukemia. Transgenic models with genetic alterations introduced directly into the mouse genome; syngeneic models with murine hematopoietic cells genetically modified by lentiviral or retroviral transduction and transplanted into recipient mice; and xenograft models with human hematopoietic cells transplanted into immunodeficient mice. Illustration created with BioRender.com and adapted from Rodriguez-Zabala (2023).

Transgenic models are generated by introducing defined genetic alterations directly into the mouse genome, for example through modification of fertilized eggs or embryonic stem cells. This allows leukemia-associated alterations to be expressed within the murine hematopoietic system and their contribution to leukemogenesis to be studied *in vivo*.[312]

Syngeneic models involve transplantation of murine hematopoietic cells into genetically compatible mice. To model leukemia, the cells can be genetically modified before transplantation, for example by introducing leukemia-associated alterations through retroviral or lentiviral transduction. Syngeneic and transgenic models of *KMT2A*-r leukemia often show a bias towards myeloid leukemia.[312] A syngeneic model was used in **Article III**.

Xenograft models instead involve transplantation of human hematopoietic cells into immunodeficient mice. These may include genetically modified human cells, such as cord-blood cells expressing a *KMT2A* fusion, or primary patient leukemia cells.[312,313] In contrast to murine models, experimental xenograft models of *KMT2A*-r leukemia often show a bias towards ALL or MPAL, while AML can be more difficult to establish. Additional genetic alterations or changes to the mouse microenvironment can, however, promote myeloid leukemia.[312,313]

Model	Advantages	Limitations
Transgenic	Defined genetic alterations in an immunocompetent system; avoids viral overexpression; useful for studying leukemogenesis and cell of origin	Time-consuming to generate; murine biology; can show a myeloid bias
Syngeneic	Immunocompetent; fast and easy to manipulate; useful for studying leukemogenesis, treatment response and immune interactions	Murine disease; viral transduction may cause non-physiological expression; can show a myeloid bias
Xenograft	Genetically modified human cells or human leukemia cells; useful for treatment response and clonal evolution	Requires immunodeficient mice; influenced by the murine BM niche; may show engraftment bias; depending on mouse strain, can be lymphoid-biased and difficult to establish as AML

Table 1: Mouse models used to study *KMT2A*-r leukemia. Overview of commonly used mouse models for *KMT2A*-r leukemia, including transgenic, syngeneic and xenograft models. The table summarizes the main advantages and limitations of each model.

Serial transplantation models can be applied across these model systems to assess leukemia self-renewal capacity. In primary transplantation, genetically modified hematopoietic cells are transplanted into mice to assess their ability to engraft and give rise to leukemia. Cells from the resulting leukemia can then be transplanted into secondary recipients. Successful secondary transplantation supports the

presence of self-renewing, leukemia-propagating cells. A long disease latency can indicate that the introduced alteration is insufficient for leukemic transformation and that additional cooperating events are required. However, latency alone cannot determine whether these events are genetic, epigenetic or related to the cellular environment.[130,313-315] Serial transplantation was used in **Article III**.

The strong leukemogenic activity of several *KMT2A* fusions has enabled the development of numerous mouse models of *KMT2A*-r leukemia.[316,317] Early mouse models established that *KMT2A::MLLT3* is a potent leukemogenic driver capable of inducing AML *in vivo*. Homozygous expression of the knock-in allele was embryonically lethal, whereas heterozygous mice developed AML, illustrating both its leukemogenic potential and the developmental consequences of constitutive expression.[318]

Mouse models have also shown that disease phenotype depends on the fusion partner and the cellular context in which it is expressed. For example, *KMT2A::MLLT1* can transform T-cell progenitors and induce lineage reassignment, whereas *KMT2A::MLLT3* failed to transform the same population.[319] Furthermore, HSCs were resistant to *KMT2A::MLLT1*-mediated transformation,[297] whereas *KMT2A::MLLT3* expression in HSCs could generate particularly aggressive and invasive AML.[298]

In contrast to *KMT2A*::*MLLT1* and *KMT2A*::*MLLT3*, modelling *KMT2A*::*AFF1* proved more challenging. Early transgenic mouse models mainly developed mature B-cell leukemia rather than the acute leukemia typically observed in patients.[316,320] Later, a refined transgenic model generated both B-ALL and AML and resembled human *KMT2A*-r leukemia.[317] Another xenograft model using human hematopoietic progenitor cells further showed that *KMT2A::AFF1* actively instructs a pro-B leukemic program, while retaining substantial lineage plasticity and the capacity to switch towards a myeloid phenotype.[321]

Several studies have investigated *KMT2A*-r leukemia using syngeneic models with retroviral transduction of hematopoietic cells followed by transplantation into recipient mice. Across different fusion partners, *MLLT1*, *MLLT3, ELL* and *MLLT10* predominantly gave rise to AML after engraftment,[289,322-324] although lymphoid or mixed-lineage phenotypes can occur under specific experimental conditions.[317] In contrast, *KMT2A::AFF1* itself has been difficult to model using this approach. As mentioned previously, expression of the reciprocal *AFF1::KMT2A* fusion in murine hematopoietic progenitors induced pro-B ALL and biphenotypic acute leukemia even in the absence of *KMT2A::AFF1,* while co-expression of both fusion proteins also gave rise to mixed lineage leukemia.[289]

Xenograft models have been used to study the leukemogenic potential of *KMT2A* fusions in primary human hematopoietic cells. In an early study, cord blood-derived hematopoietic stem and progenitor cells were retrovirally transduced with

KMT2A::MLLT1 or *KMT2A::MLLT3* and transplanted into immunodeficient mice. The cells generated myeloid, lymphoid or mixed-lineage acute leukemia, demonstrating that *KMT2A* fusions can transform human progenitor cells while retaining broad lineage potential.[325] These experiments also emphasize an important species-specific effect, as human and murine hematopoietic cells expressing the same fusion can generate different disease phenotypes. A limitation of patient-derived xenograft models is engraftment bias, where only a subset of leukemic subclones efficiently engraft, resulting in a different clonal architecture from that of the primary patient sample. This can select for more engraftment-competent clones and influence the disease phenotype and interpretation of treatment responses.[326]

Mouse models have been instrumental in dissecting the contribution of cooperating mutations to *KMT2A*-r leukemogenesis. Studies from our group have shown that activating signalling mutations shape clonal evolution and increase the fitness of *KMT2A*-r leukemic cells. In particular, $FLT3^{ITD}$, $FLT3^{N676K}$ and $NRAS^{G12D}$ accelerated the onset of *KMT2A*::*MLLT3*-driven leukemia in mice. Even when present only in a subclone, $FLT3^{N676K}$ enhanced disease development. This effect was partly mediated through changes in the leukemic microenvironment, as $FLT3^{N676K}$ expressing *KMT2A::MLLT3* cells secreted macrophage migration inhibitory factor (MIF), which promoted the survival of *KMT2A::MLLT3* leukemia-initiating cells.[216] We have also shown that in a human *KMT2A*::*MLLT3* xenograft model, addition of $FLT3^{N676K}$ shifted leukemia development towards a myeloid phenotype, whereas *KMT2A::MLLT3* alone predominantly gave rise to lymphoid leukemia.[313] Studies of *KMT2A::MLLT3*-driven leukemia have found that these cooperating signalling mutations can also alter the chromatin landscape, transcriptional networks and immune signalling. $NRAS^{G12D}$ and $FLT3^{N676K}$ increased the expression of NKG2D ligands on leukemic cells, making them more sensitive to CAR T-cell-mediated killing.[317,327] Together, these findings show that cooperating mutations can accelerate and modify *KMT2A*-driven leukemogenesis. However, the long latency observed in several transgenic models suggests that additional alterations may be required for full leukemic transformation.[316,317]

Collectively, *in vivo* model systems have provided important insights into *KMT2A*-r leukemogenesis. Transgenic and syngeneic models allow controlled investigation of defined genetic lesions and hematopoietic differentiation, while xenograft models enable functional studies in a human cellular context using primary human cells or patient material. Together, these models establish *KMT2A* fusions as potent leukemogenic drivers, while demonstrating that fusion partner, target cell, developmental context and cooperating alterations all influence the resulting disease phenotype. Nevertheless, all models have inherent limitations in recapitulating the full complexity and heterogeneity of human disease, illustrating the importance of interpreting findings within their experimental context.[130]

The clinic

Symptoms and diagnosis

Symptoms of acute leukemia usually develop over days or weeks. Most symptoms are due to the consequent cytopenias, including weakness from anemia, infections due to neutropenia and bleeding due to thrombocytopenia.[132,136] Hepatosplenomegaly, lymphadenopathy and bone pain are also common at diagnosis. These symptoms are caused by infiltration of leukemic cells into the liver, spleen, lymph nodes and BM, and are generally more frequent in ALL than in AML. Less common manifestations include skin involvement (leukemia cutis), central nervous system (CNS) symptoms (headache, seizures or stroke) and pulmonary symptoms in patients with severe hyperleukocytosis.[287,328] *KMT2A*-r leukemia correlates with high-risk clinical features, especially in infants, such as hyperleukocytosis and CNS involvement, but also leukemia cutis.[132,255]

T-ALL often presents with lymphoma like features, including bulky mediastinal masses and involvement of extramedullary tissues such as the thymus and lymph nodes.[132,328] The proportion of T-ALL among ALL cases increases with age and is nearly twice as high in adults as in children,[329] and is also twice as frequent in males. The reason for this difference is not fully understood, although X-linked tumour suppressor genes such as *KDM6A* could contribute to the male predominance.[330]

Diagnosing leukemia can be challenging since the initial symptoms are often nonspecific. In children presenting with persistent fatigue, pallor, fever, bruising or bone pain, a complete blood count is usually the first investigation and can reveal anemia, thrombocytopenia and abnormal white blood cell counts, raising suspicion of leukemia. This is followed by a differential white blood cell count, which quantifies the different leukocyte populations (neutrophils, lymphocytes, monocytes, eosinophils and basophils) and can detect circulating leukemic blasts.[139,169,287,328] In suspected AML, blast morphology is further assessed, including the proportion of myeloblasts, monoblasts, promonocytes and megakaryoblasts.[169,331]

The diagnosis is based on BM examination, combining cytomorphological, immunophenotypic and genetic analyses. A blast count of ≥20% is generally required.[169,287,328] A spinal fluid sample is analysed for the presence of leukemic blasts, and during the same procedure a skin biopsy can be obtained for germline DNA analysis and identification of constitutional variants. Given that 11% of children with cancer have been found to carry genetic variants associated with cancer predisposition, Sweden has established a national ChiCaP program to identify germline cancer predisposition in children with cancer, including leukemia.[332] Additionally, a chest X-ray is performed to assess the presence of a mediastinal mass.[139,169]

In AML, cytochemical staining can support the diagnosis by identifying myeloid differentiation through MPO positivity and monocytic differentiation through nonspecific esterase positivity.[169]

Immunophenotyping is usually performed by multiparameter FC to determine lineage assignment, maturation stage and leukemia-associated immunophenotype. Representative markers include CD13, CD117 and cytoplasmic MPO in AML; CD19, CD10 and CD34 in B-ALL; and cytoplasmic CD3, CD7 and T-cell receptor (TCR) αβ and TCRγδ in T-ALL.[139,169]

Genetic aberrations are assessed (see sections *Pediatric acute lymphoblastic leukemia* and *Pediatric acute myeloid leukemia*) using karyotyping by G-banding and molecular methods. Karyotyping requires viable leukemic cells undergoing mitosis to obtain metaphase chromosomes. In ALL, this can be challenging because metaphase chromosomes are often of poor quality and highly condensed, making G-band interpretation difficult. FISH, PCR and NGS are therefore used to detect and further identify clinically important abnormalities.[139,169] For *KMT2A*-r, a dual-colour break-apart probe is commonly used for FISH. In normal cells, the red and green probes remain adjacent, generating a fused yellow signal, whereas rearrangement of *KMT2A* separates the probes, producing red and green signals. However, some rearrangements remain cytogenetically cryptic when they arise through insertions, small inversions or other complex structural events. Examples include the novel *KMT2A*::*ACIN1* and the *KMT2A*::*USP2*, which are described in **Article III**, as well as several *NUTM1* rearrangements.[139,169]

Molecular analyses are used to detect recurrent gene mutations such as *NPM1*, *WT1*, *CEBPA* and $FLT3^{ITD}$ in AML, using PCR or NGS.[139,169] Targeted NGS panels are increasingly used in routine diagnostics and specialized centres also perform comprehensive analyses, including RNA-seq, WES and WGS to identify clinically relevant genetic alterations such as ABL-class and *KMT2A* fusion partners.[139]

Measurable residual disease

Unlike solid tumours, leukemia allows repeated sampling of BM and peripheral blood (PB), enabling longitudinal monitoring of leukemic cells and genetic alterations throughout treatment. This is fundamental for prognostic assessment and evaluation of therapy response.[333-336] Highly sensitive assays have been developed to detect MRD, which refers to leukemic cells that persist during or after treatment. MRD is now one of the strongest independent predictors of relapse and survival in both ALL and AML and is routinely incorporated into risk stratification and treatment adaptation.

In ALL, MRD samples are typically collected during induction therapy (day 15) and at the end of induction therapy (day 29), with additional time points depending on the treatment protocol and risk group (**Figure 25)**.[139]

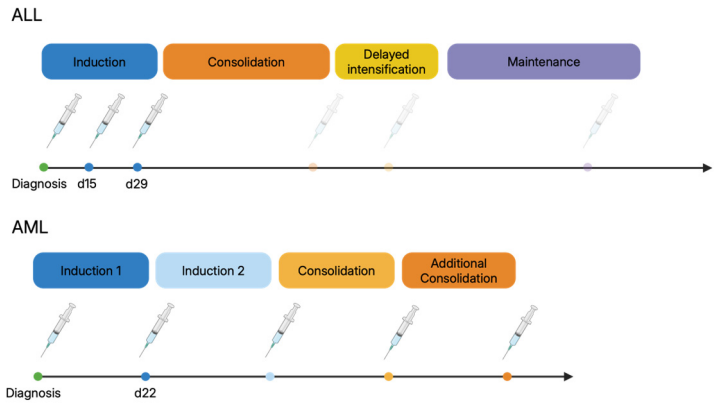


Figure 25: Simplified overview of MRD time points in pediatric ALL and AML. The upper timeline shows MRD assessment time points according to ALLTogether, and the lower timeline according to CHIP-AML22. Transparent symbols indicate assessments performed only in selected patients or treatment pathways. Illustration created with BioRender.com.

In AML, MRD samples are typically collected on day 22 (after induction course 1), at the end of induction therapy (after induction course 2) and after each consolidation course (**Figure 25**).[169]

MRD is commonly detected using two main methods. The choice of method depends on the leukemia subtype and the availability of suitable leukemia-specific molecular markers:

- *PCR:* PCR detects leukemia-specific genetic markers, including fusion transcripts and patient-specific immunoglobulin and TCR rearrangements in ALL, as well as recurrent fusion transcripts or gene mutations in selected AML subtypes, with high sensitivity. However, some PCR-based MRD approaches require patient-specific reagents, which can make it difficult to generate results quickly enough for early treatment decisions.[337-340]
- *Flow cytometry:* FC identifies leukemia cells based on their unique surface markers, which differ from those of normal BM cells. It is less sensitive than PCR but faster and more practical for clinical use.[341,342]

These methods are becoming increasingly sensitive, allowing residual disease to be detected at progressively lower levels. Patients who achieve morphological remission but remain MRD-positive have a substantially higher risk of relapse, with MRD levels after induction therapy being particularly predictive of outcome.[343-345]

Emerging methods, including digital droplet PCR (ddPCR) and NGS, enable highly sensitive detection of leukemia-specific mutations and SVs. [333,335,336,346] Patient-specific genomic fusion breakpoints are also being explored as DNA-based MRD markers, complementing RNA-based fusion transcript detection.[267]

Risk classification in pediatric acute leukemia

ALL

Risk stratification is based on a combination of genetic subtype, clinical features and early treatment response. Standard risk patients have no high-risk genetic features, no CNS involvement and undetectable MRD at the end of induction. Intermediate risk groups are subdivided based on age, genetic subtype and MRD thresholds at defined time points. High risk patients are defined by adverse genetic lesions (*KMT2A*-r, near haploidy, low hypodiploidy, iAMP21, *TCF3::HLF*), poor early MRD response or persistent disease at specific extramedullary sites such as the testes or mediastinum. T-ALL is stratified separately based primarily on MRD response (**Table 2**).[139]

Risk group	Key criteria
SR	No HR genetics/ABL-class; no T-ALL; no CNS3/TLP; MRD d29 undetectable
IR-low	<16 y; no HR genetics; *ETV6::RUNX1* <0.1%, HeH <0.03%, CNA-GR <0.05% at d29; T-ALL: MRD-negative d71
IR-high	CNA-PR: detectable <5% d29; HR genetics below HR criteria; ABL-class: <0.05% d71; CNS3/TLP
HR-chemo	*TCF3::HLF* without HSCT indication; selected patients with ≥5% d29 but <0.05% d71; persistent extramedullary disease
HR-HSCT	B-ALL: ≥0.05% d71 or ≥5% d29 + ≥0.5% d50; T-ALL: protocol-specific persistent MRD criteria

Table 2: Risk groups in pediatric ALL. Simplified overview of risk stratification in ALLTogether[139], based on genetics, age, clinical features and MRD response. Some patient groups, including patients with Down syndrome and ABL-class alterations, follow separate stratification criteria. SR, standard risk; IR-low, intermediate risk-low; IR-high, intermediate risk-high; HR, high risk; HR-chemo, high-risk chemotherapy; HR-HSCT, high-risk hematopoietic stem cell transplantation; HSCT, hematopoietic stem cell transplantation; MRD, measurable residual disease; d29, d50 and d71, treatment days 29, 50 and 71; B-ALL, B-cell acute lymphoblastic leukemia; T-ALL, T-cell acute lymphoblastic leukemia; CNS3, central nervous system status 3; TLP, traumatic lumbar puncture; HeH, high hyperdiploidy; CNA-GR, copy number alteration good risk; CNA-PR, copy number alteration poor risk; ABL-class, rearrangements involving *ABL1, ABL2, CSF1R* or *PDGFRB*; HR genetics, *KMT2A*-r, NH, HoL, iAMP21 or *TCF3::HLF*; NH, near haploidy; HoL, low hypodiploidy.

AML

Risk stratification is based on a combination of genetic aberrations and treatment response. High-risk features include *KMT2A*-r leukemia (except for *KMT2A*::*MLLT3*, which carries a more favourable prognosis), *FLT3*^{ITD} in the absence of a concurrent *NPM1* mutation and *CBFA2T3::GLIS2*. Patients are stratified into three groups: standard risk, high risk and refractory disease. Standard risk patients lack high-risk genetic features and achieve complete remission (CR) after up to two induction courses. High-risk patients include those with adverse genetic lesions and/or delayed treatment response. Refractory disease is defined as ≥5% blasts after two induction courses (**Table 3**).[169]

Risk group	Key criteria
SR	No HR criteria; CR after 2 induction courses
HR	≥15% leukemic cells d22 or 0.1-4.9% after 2 induction courses; *FLT3*^{ITD} and *NPM1* wildtype; *CBFA2T3::GLIS2*; *KMT2A*-r (excluding *KMT2A::MLLT3*) with ≥0.1% after course 1
RD	≥5% leukemic cells after 2 induction courses or persistent extramedullary disease

Table 3: Risk groups in pediatric AML. Simplified overview of risk stratification according to CHIP-AML22[169]. SR, standard risk; HR, high risk; RD, refractory disease; CR, complete remission; d22, day 22; *KMT2A*-r, *KMT2A*-rearranged; *FLT3*^{ITD}, *FLT3* internal tandem duplication.[169]

Prognosis

ALL

The treatment of pediatric ALL stands as a remarkable success story in modern medicine. Only 60 years ago, the prognosis was grim, with most affected children dying from the disease. Today, however, the survival rate exceeds 90%, propelled by a greater understanding of the disease and major advancements in diagnosis, risk stratification, therapy and supportive care, including improved infection management, transfusion support and nutritional care **(Figure 26)**.[3,347] The first major advance came in 1948, when Sidney Farber and colleagues showed that the antifolate chemotherapy drug aminopterin could induce temporary remission.[348] During the 1960s and 1970s, milestones such as combination chemotherapy, multicentre trials and delayed intensification further improved outcomes. In recent decades, cure rates have continued to rise through refined risk stratification, genetic profiling, precise monitoring of MRD, optimized drug combinations and dosing, and improved CNS directed chemotherapy, often replacing irradiation.[349]

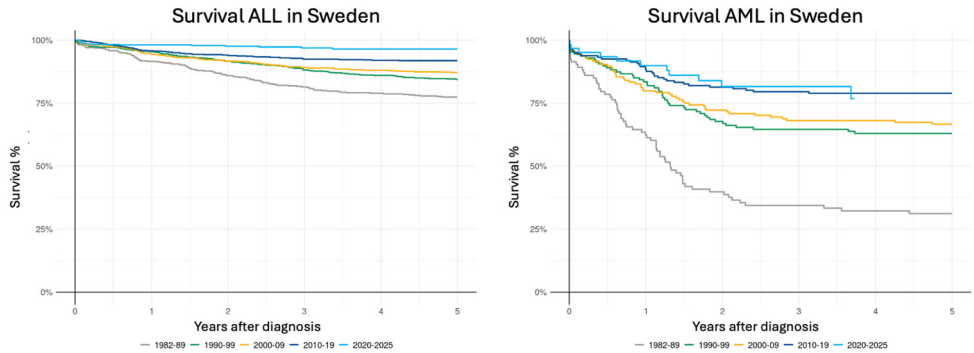


Figure 26: Survival of pediatric ALL and AML in Sweden since 1982. Graphs obtained from the Swedish Childhood Cancer Registry (SBCR; sbcr.se).

Between 1968 and 1988, Swedish pediatric oncology centres gradually established national collaboration in the treatment of childhood ALL, leading to nationwide treatment protocols. This collaboration was subsequently expanded through the Nordic Society of Paediatric Haematology and Oncology (NOPHO), established in the early 1980s, enabling larger Nordic clinical trials. Over the following decades, treatment was continuously refined through successive protocols, including NOPHO ALL1992, NOPHO ALL2000 and NOPHO ALL2008.[3] Today, the NOPHO countries participate in the international ALLTogether study, a collaborative effort involving fourteen European countries.[139]

In 1999, the Interfant consortium was established to improve the treatment of infant ALL. Unlike ALLTogether, the ongoing Interfant-21 protocol (clinicaltrials.gov: NCT05327894) is specifically designed for infants with *KMT2A*-r ALL and combines conventional ALL chemotherapy with AML-like consolidation blocks and blinatumomab as part of first-line therapy.[350] Treatment dosing is adapted to account for the unique developmental pharmacokinetics of infants, as rapid maturation during the first year of life affects drug absorption, distribution, metabolism and elimination.[351]

KMT2A-r ALL is associated with poor outcomes across age groups, particularly in infants diagnosed before 6 months of age.[255,352,353] In the Interfant-06 study, infants with *KMT2A*-r ALL had a 6-year event-free survival of 36%, with the poorest outcomes among high-risk infants.[348] Outcomes remain unfavourable in older children and adults, although prognosis varies with treatment, disease biology and MRD response.[255,353,354] More recently, adding blinatumomab to frontline therapy has markedly improved outcomes in infants with *KMT2A*-r ALL, with early studies reporting overall survival above 90%.[355] New treatment approaches, including blinatumomab, CAR-T-cell therapy and menin inhibitors, are discussed below.

AML

Between 1984 and 2001, pediatric AML treatment in the Nordic countries was gradually developed through three consecutive NOPHO protocols: NOPHO-AML 84, NOPHO-AML 88 and NOPHO-AML 93. These studies formed the basis for a standardized Nordic treatment strategy in childhood AML. Early protocols consisted of multi-agent chemotherapy with cytarabine, anthracyclines and thioguanine, followed by consolidation with high dose cytarabine. Over time, treatment intensity was increased and the intervals between treatment courses were shortened to improve dose intensity and clinical outcomes.[356]

The NOPHO-AML 2004 protocol refined response-adapted treatment by using early BM response to guide risk stratification and treatment intensity, with high remission rates (95%) and improved survival, most notably in patients receiving allogeneic hematopoietic stem cell transplantation (HSCT) in first CR (clinicaltrials.gov: NCT00476541).

The NOPHO-DBH AML 2012 protocol aimed to improve prognosis by introducing MRD based response assessment as the main tool for risk stratification and treatment evaluation, and by intensifying induction therapy compared to previous NOPHO protocols (clinicaltrials.gov: NCT01828489).

Currently, the CHIP-AML22 protocol represents the next step in this evolution. It builds on the NOPHO-DBH AML 2012 but incorporates additional refinement of risk stratification, MRD-guided treatment decisions and targeted therapies such as *FLT3* inhibitors (quizartinib) and the CD33-targeted antibody-drug conjugate gemtuzumab ozogamicin, within an international collaborative trial structure.[169]

Thanks to improved treatment protocols, overall survival is around 70-80% in pediatric AML.[357-360] In pediatric *KMT2A*-r cases, the overall survival is around 58%.[361] Contrary to infant *KMT2A*-r ALL, infants with *KMT2A*-r AML do not have a separate protocol and their prognosis is not inferior to that of older children.[255]

Treatment

Treatment protocols are divided into several sequential therapy phases, each with a specific purpose. Treatment begins with induction therapy (to rapidly eliminate the majority of leukemic cells and achieve CR), followed by consolidation therapy (to eradicate residual leukemic cells and reduce the risk of relapse). In ALL, this is followed by a prolonged maintenance phase (to suppress residual disease and prevent recurrence), whereas maintenance therapy is generally not given in AML. The treatment period is usually shorter but more intensive in AML (around 5-8 months) compared with up to 2 years after remission in ALL.[139,169]

Some leukemic cells evade therapy by entering a quiescent state, making them less susceptible to treatments targeting proliferating cells. The BM microenvironment

can protect residual leukemic cells by providing signals that promote survival and treatment resistance.[230,362] Intrathecal chemotherapy is administered to overcome the blood brain barrier and eradicate leukemic cells within the CNS. While CNS-directed intrathecal therapy is routinely incorporated into ALL treatment protocols as prophylaxis,[139] intrathecal therapy is also included in pediatric AML protocols, although its role in AML is less well established.[169]

To understand how therapies affect the leukemic, as well as the healthy cells, it is essential to understand their pharmacodynamics. Several classes of cytotoxic agents are used in the treatment of leukemia, but they share a common feature: the ability to target rapidly dividing cells by inducing DNA damage, ultimately leading to cell death or cell cycle arrest.[363,364] The fact that DNA damage effectively kills cancer cells can appear paradoxical, given that genomic instability is a hallmark of malignancy (**Figure 1**). However, even cancer cells rely on the integrity of essential genomic regions for survival, making them vulnerable when these are disrupted.[364] The main treatment modalities used in pediatric leukemia are described below.

Antimetabolites

Antimetabolites (e.g. cytarabine, fludarabine, 6-mercaptopurine, 6-thioguanine and methotrexate) are structural analogues of purines, pyrimidines or folates that disrupt DNA synthesis in rapidly dividing cells. They act either by inhibiting enzymes required for nucleotide synthesis or by being incorporated into DNA or RNA as false nucleotides, ultimately leading to cell cycle arrest and apoptosis.[365] Their clinical efficacy is limited by the development of drug resistance through altered transport, metabolism or DNA repair mechanisms.[366]

Antimetabolites are central components of both ALL and AML treatment, although their roles differ between the diseases. In ALL, several antimetabolites, including methotrexate, 6-mercaptopurine and cytarabine, are administered throughout different phases of treatment, including consolidation, delayed intensification and maintenance.[139] In contrast, AML treatment relies primarily on cytarabine, which forms the backbone of induction and consolidation therapy and is commonly combined with anthracyclines during induction.[169]

Glucocorticoids

Glucocorticoids (e.g. prednisolone and dexamethasone) have been a cornerstone of ALL treatment for over 70 years. Glucocorticoids bind to the intracellular glucocorticoid receptor, which translocates to the nucleus and acts as a transcription factor. This alters the expression of genes regulating cell survival and apoptosis, leading to activation of the intrinsic mitochondrial apoptotic pathway in lymphoid leukemic cells. Glucocorticoids are used during induction therapy. In some earlier protocols, steroid response was a strong predictor of outcome and was used for risk stratification. Today, MRD has largely replaced steroid response for treatment stratification. Dexamethasone penetrates the CNS more effectively than

prednisolone but is associated with increased toxicity, infections, osteonecrosis and muscle weakness.[139,367] Interestingly, glucocorticoids have a limited anti-leukemic effect in AML[169] potentially reflecting lineage specific differences in glucocorticoid receptor signalling and apoptotic responses between myeloid and lymphoid leukemic cells.

Topoisomerase II inhibitors, including anthracyclines

Topoisomerase II inhibitors include anthracyclines such as daunorubicin, doxorubicin and idarubicin, as well as etoposide and mitoxantrone. These prevent the repair of DNA breaks during replication, causing DNA damage and ultimately apoptosis of rapidly dividing cells. Anthracyclines also generate reactive oxygen species, which contribute to DNA damage. They are among the most effective drugs in both ALL and AML but their cumulative dose is limited by the risk of cardiotoxicity.[368] Dexrazoxane is therefore used as a cardioprotective agent.[169]

Anthracyclines are a cornerstone of AML therapy and are administered during both induction courses together with cytarabine-based chemotherapy. In contrast, their use in ALL is risk-adapted and restricted to T-ALL or higher-risk B-ALL patients. Due to long-term toxicities, ALLTogether aims to further reduce anthracycline exposure in lower-risk patients.[139,169] Paradoxically, as described previously, topoisomerase II inhibitors can cause therapy-related *KMT2A*-r leukemia.[255]

Alkylating agents

Cyclophosphamide is an alkylating agent that is activated in the liver to phosphoramide mustard. The active metabolite forms DNA crosslinks, preventing DNA replication and transcription, ultimately leading to apoptosis.[369]

Cyclophosphamide is an important component of ALL treatment for children with higher risk during consolidation and delayed intensification[139] but is not part of standard AML induction therapy.[98] Cyclophosphamide-containing regimens are widely used as conditioning prior to HSCT in AML.[169]

Vinca alkaloids

Vinca alkaloids (e.g. vincristine and vinblastine) are microtubule-destabilizing agents. Microtubules are essential for mitosis, where they form the mitotic spindle required for accurate chromosome segregation during cell division. As cancer cells undergo proliferation, they are dependent on intact microtubule dynamics. Vinca alkaloids bind to β-tubulin and suppress microtubule dynamics, resulting in mitotic arrest and subsequent apoptotic cell death.[370]

In ALL, vincristine is administered during induction, consolidation and delayed intensification, and together with glucocorticoids, asparaginase and anthracyclines constitutes one of the cornerstones of therapy.[139] Vinca alkaloids are not part of

standard AML treatment, largely because cytarabine- and anthracycline-based regimens have shown greater clinical efficacy.[169]

Asparaginase

Asparaginase represents a unique class of anti-leukemic agents and is not typically classified within the traditional cytotoxic drug groups. It depletes the circulating amino acid asparagine, thereby inhibiting protein synthesis and inducing cellular stress and apoptosis.[371] Asparaginase is a key component of modern ALL therapy and is administered during induction, but is not part of standard AML therapy.[139,169]

Antibody drug conjugate

Antibody drug conjugates (ADCs) consist of a monoclonal antibody linked to a cytotoxic drug. After binding to a target antigen on the cell surface, the complex is internalized and the cytotoxic payload is released intracellularly, allowing selective targeting of leukemic cells.[372] Gemtuzumab ozogamicin targets CD33-positive AML cells,[169] whereas inotuzumab ozogamicin targets CD22-positive B-ALL cells (**Figure 27**).[139] Both agents deliver calicheamicin, a potent DNA-damaging antibiotic that induces DNA double-strand breaks and triggers apoptotic cell death.[372] Gemtuzumab ozogamicin is being evaluated during induction in CD33-positive AML,[169] while inotuzumab ozogamicin is being evaluated before maintenance in ALL.[139]

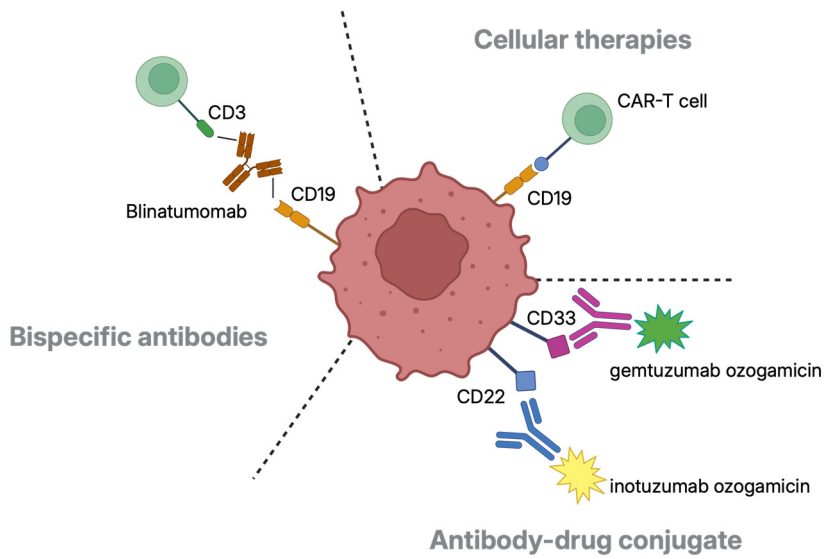


Figure 27: Immunotherapies in pediatric acute leukemia. Illustration of immunotherapeutic approaches used in pediatric ALL and AML. In B-ALL, CD19 can be targeted by the bispecific T-cell engager blinatumomab and by CAR T-cell therapy, while the antibody-drug conjugate inotuzumab ozogamicin targets CD22. In AML, the antibody-drug conjugate gemtuzumab ozogamicin targets CD33. CAR, chimeric antigen receptor. Illustration created with BioRender.com.

Bispecific antibodies

Bispecific antibodies are a form of immunotherapy that simultaneously bind leukemic cells and T-cells (**Figure 27**). Blinatumomab, the first approved bispecific T-cell engager, binds CD19 on B-cell leukemic cells and CD3 on T-cells, bringing the two cell types into close contact and activating T-cell mediated killing of the leukemia cells.[373] Blinatumomab is now incorporated into pediatric ALL treatment protocols, including for infants and children with Down syndrome. In infants with *KMT2A*-r ALL, blinatumomab has improved MRD responses and outcomes and is incorporated into the Interfant-21 protocol.[350,351,355]

Stem cell transplantation

Allogeneic HSCT is used to replace the patient's hematopoietic system with healthy donor stem cells.[139,169] Before transplantation, patients receive conditioning chemotherapy, with or without radiotherapy, to eradicate residual leukemia cells, suppress the immune system and create space in the BM for donor stem cell engraftment. Following infusion, donor hematopoietic stem cells reconstitute normal hematopoiesis, while donor immune cells contribute to disease control through the graft-versus-leukemia effect by recognizing and eliminating residual leukemic cells.[374] HSCT is a highly intensive treatment that places a considerable physical and emotional burden on patients and families, with prolonged hospitalization and isolation, and substantial risks of severe infections, graft-versus-host disease, infertility and other complications. It is therefore reserved for patients in whom the expected benefit clearly outweighs these risks. In childhood ALL, HSCT is used for selected high-risk patients, including those with persistent MRD or high-risk relapse,[139] while in AML it is used for high-risk patients in first remission and for patients achieving remission after relapse.[169]

Chimeric antigen receptor T-cell therapy

CAR-T cell therapy uses the patient's own T-cells, which are genetically modified to express a CAR that recognizes CD19 on leukemic B-cells. After reinfusion, the CAR-T cells specifically recognize and eliminate CD19-positive leukemic cells. In the ALLTogether protocol, CAR-T therapy is offered to selected high-risk patients with persistent MRD after induction and consolidation therapy, with the aim of avoiding intensive chemotherapy and allogeneic HSCT. The most common side effects are cytokine release syndrome, neurotoxicity and B-cell aplasia.[139,375-377] Although CAR-T cell therapy has added a powerful treatment option for B-ALL, its application in AML remains challenging. Unlike CD19 in B-ALL, no AML antigen combines broad expression on leukemic cells with lineage restriction while sparing HSCs. Consequently, CAR-T cells targeting AML-associated antigens can also damage normal hematopoiesis **(Figures 27** and **28**).[378]

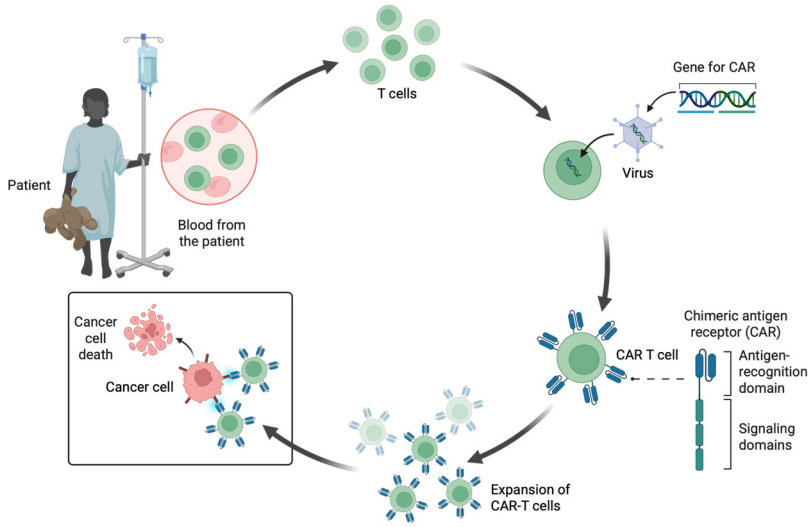


Figure 28: Illustration of CAR-T. T-cells are isolated from the patient's blood and genetically modified using a viral vector to express a chimeric antigen receptor (CAR). The modified T-cells are expanded and infused back into the patient, where they recognize target antigens on cancer cells and kill them. Illustration created with BioRender.com.

Radiotherapy

Radiotherapy has largely been eliminated from frontline ALL treatment, with CNS control achieved through intensive intrathecal and systemic therapy. Its current role is primarily limited to total body irradiation as part of conditioning prior to HSCT.[139] In AML, radiotherapy has no established role in improving survival and is primarily reserved for emergency situations, such as organ compression or other life-threatening manifestations of extramedullary disease.[169]

Novel treatments

Increased understanding of leukemia biology has facilitated the development of novel treatments. Below follows a selection of these, with a focus on *KMT2A*-r leukemia.

Menin inhibitors

Menin inhibitors are a new class of targeted therapies developed for *KMT2A*-r and *NPM1* mutated leukemia. These drugs disrupt the interaction between menin and *KMT2A*, leading to downregulation of leukemogenic genes such as *HOXA* and *MEIS1,* thereby promoting differentiation of leukemic cells. Preclinical studies have shown that menin inhibition activates tumour suppressor pathways and that some forms of resistance may be overcome by combining menin inhibitors with CDK4/6 inhibitors. Menin inhibitors have yielded promising clinical results, leading to the

first regulatory approval of a menin inhibitor for relapsed or refractory acute leukemia with *KMT2A* translocations in the United States.[379,380]

DOT1L inhibitors

DOT1L inhibitors target the histone methyltransferase DOT1L, which is aberrantly recruited by *KMT2A* fusion proteins to leukemogenic target genes. Inhibition reduces H3K79 methylation, suppresses expression of genes such as *HOXA9* and *MEIS1,* thus promoting differentiation, cell cycle arrest and apoptosis in *KMT2A*-r leukemic cells. Clinical evaluation of DOT1L inhibition has demonstrated target inhibition but only modest clinical activity as monotherapy.[381,382]

Gene editing

Gene edited allogeneic CAR-T cell therapy is an emerging treatment strategy for relapsed or refractory T-ALL. In this approach, donor T-cells are base edited to inactivate CD7 and the T-cell receptor. This prevents the CAR-T cells from attacking each other and reduces the risk of graft-versus-host disease. The engineered cells express a CAR targeting CD7, which is present on T-cell leukemic cells. Early clinical studies have shown high remission rates, allowing subsequent allogeneic HSCT, demonstrating the potential of gene-edited CAR-T cells as a treatment option for T-ALL.[383,384]

Treatment sequelae

Decades after the phrase "cure is not enough" was coined,[385] it remains relevant. With survival rates reaching 90%, the focus has increasingly shifted to long term health and quality of life.[349] Many survivors face an elevated risk of late complications such as hormonal, reproductive, pulmonary, renal, auditory, cardiac and neurological issues.[386] No therapy can yet distinguish perfectly between malignant and healthy cells. Greater potency often comes at the expense of increased side effects. The dilemma is especially pronounced in children, where ongoing growth and development make normal tissue more susceptible to unintended harm.[387] New strategies are needed to identify patients at greatest risk of late effects due to genetic, treatment-related or lifestyle factors, while also identifying those who could be safely treated with less intensive therapy.

Relapse

ALL

More than 90% of patients will achieve remission, but approximately 10-15% relapse.[388] In pediatric *KMT2A*-r ALL, the relapse risk is higher, particularly in infants, among whom historical relapse rates have approached 50-60%.[255,351] Older

children with *KMT2A*-r ALL generally have better outcomes than infants, although their prognosis remains inferior to children with *KMT2A*-wildtype ALL.[389]

Relapse is defined according to the ALLTogether protocol as:

- detection of ≥1% blasts, in the BM, confirmed by two independent techniques (FC/molecular MRD/leukemia specific genetic markers) in two repeated samples, 1 to 4 weeks apart

  OR

- ≥25% blasts in the BM by morphology alone after CR has been achieved

Relapses most commonly occur in the BM, which carries the worst prognosis, followed by extramedullary sites in the CNS and testes.[255,343,388]

Prognostic factors at relapse include time to relapse, site of relapse and immunophenotype, with time to relapse being the most important factor.[343] Relapse is classified as very early (within 18 months of diagnosis), early (within 6 months of stopping therapy) or late (more than 6 months after the end of treatment).[139] These factors, together with MRD response, guide risk stratification and treatment according to the IntReALL protocol (ClinicalTrials.gov: NCT07476729). Earlier BM relapses are high risk and usually require induction treatment to achieve a second remission, followed by immunotherapy and HSCT, with CAR T-cell therapy used in selected cases. In contrast, late B-ALL relapses with a good MRD response can be treated without transplantation. IntReALL incorporates blinatumomab and inotuzumab into relapse treatment. After relapse, 5-year overall survival is approximately 32% in children with *KMT2A*-r ALL and 18% in infants, compared with 53% in B-ALL overall and 46% in infants with *KMT2A*-wildtype ALL.[388]

AML

Approximately 30-40% of pediatric AML patients experience relapse and survival after relapse remains poor, at around 20-40%. Relapses are commonly classified as early (within 12 months of diagnosis) or late (>12 months from diagnosis), with better survival reported for patients with late relapse than for those with early relapse (51% versus 25%). Within the early relapse group, outcomes are influenced by relapse timing, with patients relapsing within 6 months having a poorer prognosis than those relapsing between 6 and 12 months after diagnosis (27% versus 37%).[390]

In relapsed pediatric AML, there is no single standardized relapse protocol. Induction chemotherapy, sometimes combined with targeted agents, is used to achieve a second remission, usually followed by HSCT. Treatment depends on previous therapy, disease genetics and response to induction treatment.

Relapse is more comprehensively discussed in **Article I** and in the *Discussion* section.

The present investigation

Aims of the study

The overall aim of this thesis was to improve our understanding of the molecular basis of *KMT2A*-r acute leukemia development and relapse. Highly sensitive sequencing approaches were applied to characterize the genomic landscape of samples collected at diagnosis, during therapy and at relapse, with a focus on clonal evolution, treatment response and mechanisms of disease recurrence. Together, these analyses also defined the genetic and biological heterogeneity of *KMT2A*-r leukemia across disease subtypes and age groups (**Figure A**).

More specifically, the aims of this thesis were:

- To study clonal dynamics and the genetic mechanisms driving relapse in *KMT2A*-r infant and childhood leukemia (**Article I**).
- To characterize the subclonal genetic landscape of pediatric *KMT2A*-r leukemia and investigate the clonal dynamics underlying disease relapse (**Article II**).
- To characterize the genomic and transcriptional landscapes of *KMT2A*-r leukemia from infancy to adulthood and identify biological differences between leukemia lineages, fusion partners and age groups (**Article III**).

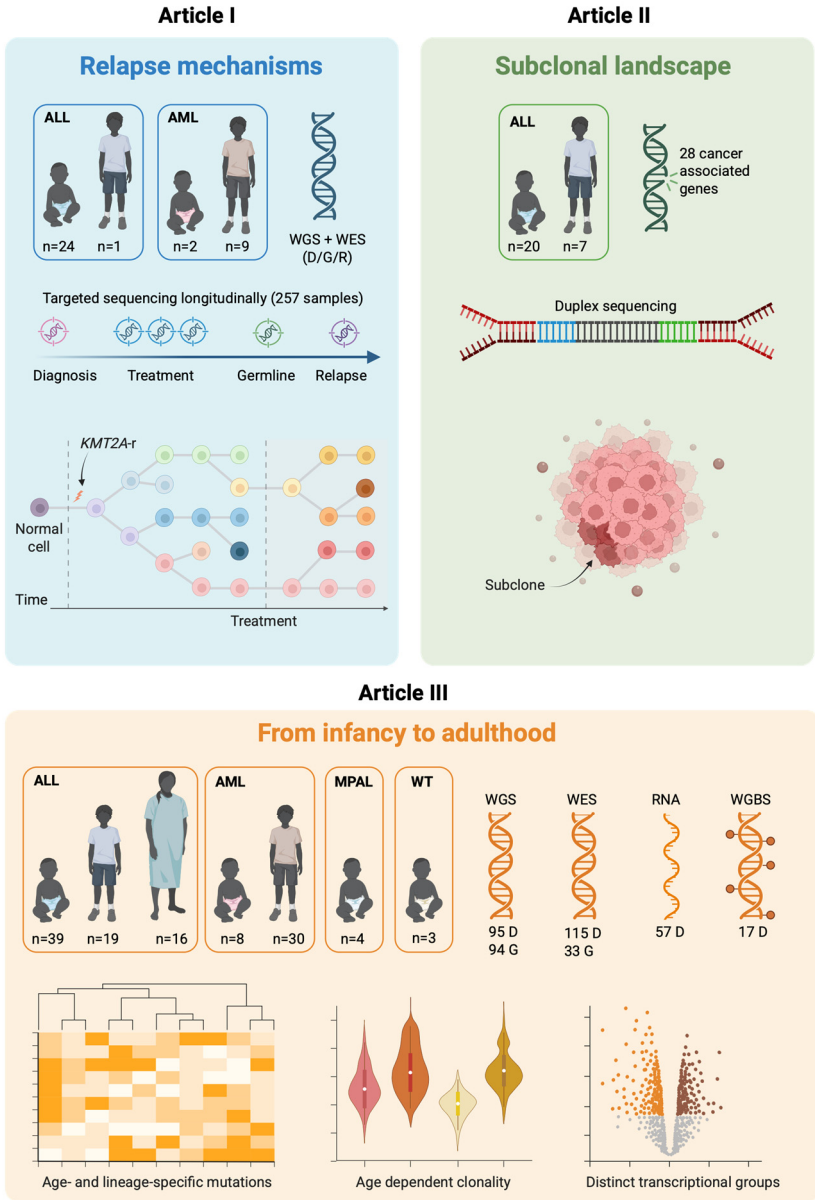


Figure A: Overview of the studies included in this thesis. The three studies investigate *KMT2A*-r leukemia from complementary perspectives, including relapse mechanisms, the subclonal mutational landscape at diagnosis, and genomic and transcriptional characteristics across age groups and disease lineages. The main cohorts, sampling strategies and sequencing approaches used in each study are summarized. ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; MPAL, mixed phenotype acute leukemia; WT, wildtype; *KMT2A*-r, *KMT2A*-rearranged; WGS, whole genome sequencing; WES, whole-exome sequencing; RNA-seq, RNA sequencing; WGBS, whole genome bisulfite sequencing; D, diagnosis; G, germline. Illustration created with BioRender.com.

Materials and methods

Sample collection and processing

Over 140 patients were analysed in this thesis, with full clinical annotation and long-term follow-up. Samples from the Nordic countries were collected at diagnosis, remission or relapse. Most samples were obtained from the Division of Clinical Genetics, Lund University, but given the rarity of this disease, additional samples were provided through key collaborations in Sweden, Denmark and Finland, the NOPHO biobank, and Ronald Stam at the Princess Máxima Center in the Netherlands (Interfant group).

The biological material for these studies included viable cells, Ficoll-isolated cells, cell pellets, TRIzol-preserved samples, fixative-preserved samples, cells from glass slides, as well as DNA and RNA previously extracted as part of routine clinical diagnostics. This provided access to a unique collection of patient samples spanning several decades. Although nucleic acid yield and quality varied depending on sample type and storage conditions, fresh and cryopreserved samples generally yielded high-quality DNA and RNA, whereas archived slides and fixative-preserved material typically provided lower-quality DNA and were not suitable for RNA extraction due to degradation.

A particular challenge was the analysis of serial samples during treatment. During induction therapy, leukemic blast counts rapidly decline, resulting in limited numbers of leukemic and normal cells and consequently low DNA yields. Both BM and PB samples were used, which increased the availability of material for analysis. However, blast counts were generally higher in the BM, making it the preferred source when available.

Polymerase chain reaction

PCR is a fundamental technique used to amplify specific DNA fragments. It allows exponential amplification of a targeted DNA region through repeated cycles of denaturation, primer annealing and extension by a *DNA polymerase*. This results in the generation of large amounts of DNA from very small initial samples, making PCR a key step in many applications, including sequencing-based methods. PCR is rapid, sensitive and cost-effective, but the DNA polymerases can introduce errors during amplification. These errors are generally not detectable by conventional Sanger sequencing but may be relevant when using NGS to identify low-frequency variants, where errors can be mistaken for true mutations **(Figure 29)**.[391]

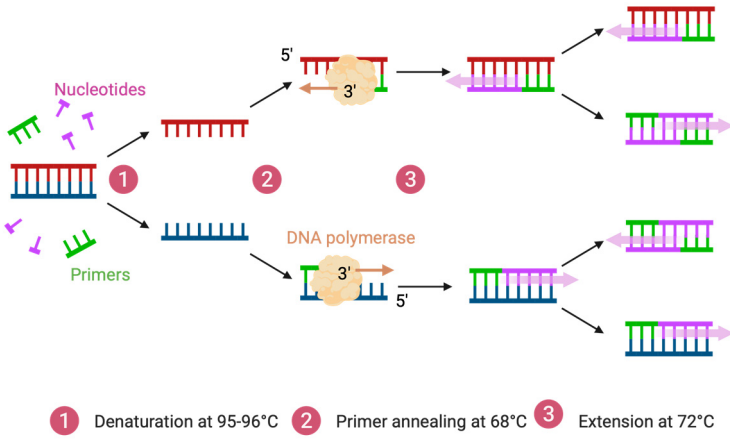


Figure 29: Schematic illustration of PCR, showing the three main steps: denaturation (separation of DNA strands), primer annealing (binding of primers to the template DNA) and extension (DNA synthesis by DNA polymerase). Illustration created with BioRender.com.

Whole genome sequencing

WGS was used in all three articles (**I-III**) to comprehensively characterize the genomic landscape of *KMT2A*-r leukemia. WGS determines the DNA sequence across the entire genome, enabling the identification of both known and novel genetic variants.[392] WGS can identify a broad range of genetic alterations, including exonic and intronic mutations, SVs and CNAs, but is associated with higher costs, increased bioinformatic complexity and lower sensitivity for detecting low frequency variants compared with targeted sequencing approaches (see below).

For cancer samples, sequencing of a matched normal DNA sample is recommended to distinguish somatic from germline variants. In this thesis, BM or PB obtained during remission were used as matched remission samples. To minimize the risk of residual leukemia contamination, remission samples were selected from time points with sustained deep remission and, whenever possible, distant from relapse. Nevertheless, low-level residual leukemic clones cannot be completely excluded. Alternative sources of constitutional DNA include cultured skin fibroblasts or nail-derived DNA, which carry a lower risk of contamination by leukemic cells. However, these approaches are associated with practical limitations, including invasive sampling, prolonged processing time or lower DNA yield.[393-395] For the workflow, see the section *Bioinformatics* (**Figure 10**).

Whole exome sequencing

WES was also used in all three articles (**I-III**). In contrast to WGS, WES selectively captures and sequences the protein-coding regions of the genome (exons), enabling higher sequencing coverage of coding regions at a lower cost. Library preparation, sequencing and bioinformatics are performed similarly to WGS, except that exonic regions are enriched using hybridization-based capture prior to sequencing.

Since only the exome is sequenced, variants in non-coding regions are generally not detected unless they are located close to the captured regions. Detection of SVs and CNAs is also more limited than with WGS.

Whole genome bisulfite sequencing

In **Article III**, WGBS was used to characterize genome-wide DNA methylation patterns and compare the epigenetic profiles of the transcriptionally defined leukemia subgroups. WGBS determines DNA methylation at single base resolution across the entire genome. DNA is fragmented, sequencing adapters are ligated to the fragments and then the DNA is treated with bisulfite, which converts unmethylated cytosines into uracils while methylated cytosines remain unchanged. During PCR amplification, uracils are read as thymines, allowing methylation status to be inferred from C-T conversions in the sequencing data. The resulting library is sequenced using NGS, generating short DNA reads that reflect genome-wide methylation patterns. These reads are aligned to a bisulfite-converted reference genome and analysed to quantify DNA methylation levels at individual cytosine positions.[396]

RNA sequencing

In **Article III**, RNA-seq was used to characterize gene expression patterns, identify fusion genes and define transcriptional subgroups of *KMT2A*-r leukemia. Total RNA was extracted and mRNA was enriched. The RNA was then converted into complementary DNA (cDNA), which is fragmented and ligated to sequencing adapters. The resulting cDNA library was sequenced using NGS, generating millions of short reads corresponding to expressed transcripts. Following quality control, the reads are aligned to a reference genome to identify differentially expressed genes and detect fusion transcripts.[397]

RNA-seq enables genome-wide analysis of gene expression. Since thousands of genes are evaluated simultaneously, *p*-values were adjusted to account for multiple testing and reduce the risk of false positive findings.[398] Gene Set Enrichment Analysis (GSEA) was used to determine whether predefined groups of functionally related genes are enriched between groups, using gene sets from the Molecular Signatures Database. Beyond differential gene expression, RNA-seq is also

effective for detecting fusion transcripts and alternative splicing events. However, while RNA-seq provides valuable information on transcriptional activity, mRNA abundance does not always reflect protein levels or biological function.[398]

Single cell sequencing

Single cell sequencing enables genomic and transcriptomic analysis at the resolution of individual cells, allowing characterization of cellular and genetic heterogeneity. In **Article III**, scRNA-seq was used to validate the transcriptionally defined subgroups identified by bulk RNA-seq and determine whether the corresponding gene expression programs were maintained at the single-cell level. scRNA-seq was performed using a droplet-based 10x Genomics workflow, in which individual cells were encapsulated in microdroplets together with barcoded beads. This allowed transcripts from each cell to be uniquely labelled before sequencing, providing single-cell transcriptional profiles for identifying different cell types, cell states and rare subpopulations within heterogeneous samples.[45,399]

In **Articles I** and **III**, scDNA-seq was used to define the clonal architecture of leukemia and identify subclonal populations that could not be resolved by bulk sequencing alone. Individual cells were isolated by flow cytometry, after which DNA was whole genome amplified using primary template-directed amplification (PTA). This method generates short DNA fragments that are inefficiently re-amplified, favouring repeated amplification from the original DNA template and thereby reducing amplification bias.[400,401] Depending on the study, the amplified DNA was subsequently analysed by targeted PCR or WES. Despite the high fidelity of PTA, single-cell DNA sequencing remains susceptible to allelic dropout, amplification bias and uneven genome coverage because only two copies of each autosomal locus are available for amplification. Consequently, careful quality control and bioinformatic processing are required for reliable mutation calling, with downstream validation being important for rare *de novo* variants.[401]

Targeted sequencing

Targeted sequencing was used in all three articles (**I-III**) to validate genetic variants identified by WGS/WES, detect low-frequency mutations with higher sensitivity and longitudinally monitor selected variants during disease progression. In contrast to WGS or WES, only selected loci of interest were amplified and sequenced. Target enrichment was performed using multiplex PCR, in which multiple genomic regions were simultaneously amplified using specific primer pairs. Sequencing adapters and sample specific indices were incorporated during library preparation, after which the amplified products were subjected to NGS. For most samples, 5-20 ng of input DNA was used. For cell-poor samples with <3.2 ng of available DNA, PTA was

performed prior to the PCR and library preparation to generate sufficient material for sequencing.

The sensitivity for detecting variants is influenced by the amount and quality of input DNA, as well as sequencing depth. Compared with genome-wide approaches, targeted sequencing reduces sequencing costs and allows greater sequencing depth.

In **Article I**, the analytical sensitivity of the multiplex PCR assay was evaluated using a dilution series in which DNA from the *TP53* mutant Kasumi-1 cell line was serially diluted with DNA from the *TP53* wildtype REH cell line. This established the assay's limit of detection (approximately 0.002 VAF) and reproducibility under controlled conditions.

Duplex sequencing

In **Article II**, duplex sequencing was used to identify ultra-low-frequency cancer-associated mutations at diagnosis and to investigate their contribution to clonal heterogeneity and relapse evolution in *KMT2A*-r ALL. DNA is fragmented and ligated to sequencing adapters containing unique sequences allowing each original DNA molecule to be tracked throughout library preparation and sequencing. Following sequencing, reads originating from the same DNA strand are combined into single-strand consensus sequences and complementary strands are subsequently compared to generate a duplex consensus sequence. As true genetic variants are expected to be present on both strands of the original DNA molecule, whereas most PCR and sequencing errors occur randomly on only one strand, duplex sequencing markedly reduces sequencing errors and enables highly sensitive detection of rare mutations (**Figure 30**).[402] Limitations of duplex sequencing include the requirement for high sequencing depth and relatively large amounts of high-quality input DNA, together with specialized library preparation and bioinformatic analysis, making the method more costly and less suitable for large-scale genome-wide applications. In **Article II**, 250 ng of input DNA was used, providing a theoretical detection limit of approximately 0.005% VAF (corresponding to one mutated cell among ~40,000 cells).

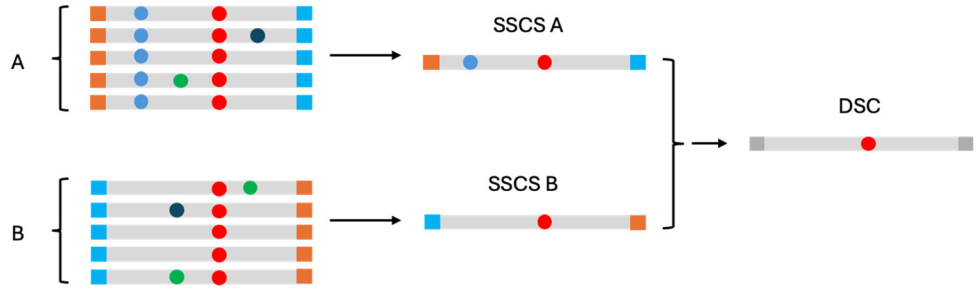


Figure 30: Principle of duplex sequencing. Illustration of duplex sequencing for error-corrected detection of rare genetic variants. DNA fragments are ligated to uniquely tagged adapters and amplified and sequenced. Reads originating from the same DNA strand are combined to generate single-strand consensus sequences (SSCSs), after which complementary strands are compared to generate a duplex consensus sequence (DCS). True variants present on both strands are retained, whereas sequencing or PCR errors present on only one strand are removed.

Digital droplet PCR

ddPCR was used in **Article II** to validate selected low-frequency variants identified by duplex sequencing and to assess the presence of relapse-associated mutations below the detection limit of duplex sequencing. In this study, ddPCR achieved a validated detection limit of 0.0044% VAF or lower when sufficient input DNA (>110 ng) was available. In ddPCR, a DNA sample is partitioned into thousands of nanolitre droplets, with PCR amplification occurring independently within each droplet. Following amplification, droplets are classified as positive or negative based on fluorescence signal, allowing direct counting of target molecules. ddPCR is well suited for highly sensitive detection of predefined low-frequency variants.[403]

CRISPR/Cas9 genome editing

In **Article I**, CRISPR/Cas9 was used to generate *PRPS2* knockout REH cell clones for functional validation of the *PRPS2*^{P320L} variant. CRISPR/Cas9 is a genome editing technology adapted from the bacterial adaptive immune system. The system consists of the Cas9 endonuclease and a guide RNA (gRNA) that directs Cas9 to a complementary DNA sequence, where Cas9 introduces a double-strand DNA break. Although CRISPR/Cas9 enables precise genome editing, editing efficiency varies between cell types, off-target editing can occur and double-strand DNA breaks can introduce unintended alterations.[404] This experiment was performed by collaborators.

Plasmid construction and lentivirus production

Plasmids are circular DNA molecules used as vectors to introduce genes into cells. In **Article I**, plasmids encoding wildtype *PRPS2* or the *PRPS2*^{P320L} variant were generated and packaged into replication-deficient lentiviral particles. These were used to transduce parental and *PRPS2* knockout REH cells, enabling functional comparison of wildtype and mutant *PRPS2*. Lentiviral vectors provide efficient and stable gene delivery through integration into the host genome.[405] This experiment was performed by collaborators.

Western blotting

Western blotting was used to detect specific proteins in biological samples. Proteins are separated by gel electrophoresis, transferred to a membrane and detected using antibodies. In **Article I**, Western blotting was used to confirm expression of wildtype and mutant *PRPS2* following lentiviral transduction of parental and *PRPS2* knockout REH cells. Protein expression was detected using an anti-Flag antibody, while vinculin served as a loading control. In **Article III**, it was used to assess FLT3 signalling in Ba/F3 cells expressing wildtype or mutant *FLT3*, including *FLT3*^{V491L} and *FLT3*^{ITD}, under ligand-stimulated and unstimulated conditions. Although Western blotting enables sensitive detection of proteins, it is dependent on antibody specificity and less suitable for low-abundance proteins.[406] These experiments were performed by collaborators.

Drug sensitivity assessment

Drug sensitivity assessment is used to evaluate how cells respond to pharmacological treatment. Cells are exposed to increasing concentrations of a drug and cell viability is measured after treatment to determine the effect of the compound.[407] In **Article I**, drug sensitivity was assessed using the CellTiter-Glo assay, which quantifies intracellular ATP as a surrogate marker of metabolically active cells. *PRPS2* knockout and parental REH cells expressing wildtype or mutant *PRPS2* were treated with increasing concentrations of 6-mercaptopurine for 72 hours to evaluate the functional effect of the *PRPS2*^{P320L} variant on drug sensitivity. A limitation of cell viability assays is that they provide an indirect measure of cell survival and do not distinguish between reduced proliferation and cell death.[408] This experiment was performed by collaborators.

Mutational signatures

In **Articles I and III**, mutational signature analysis was performed to determine the relative contribution of COSMIC single base substitution (SBS) signatures at

diagnosis and relapse. Based on the dominant SBS signature, each patient was assigned to a primary mutational process, enabling comparison of mutational processes between ALL and AML as well as between diagnosis and relapse.

Mutational signature analysis classifies each single-base substitution according to the type of base change and the identity of the adjacent 5′ and 3′ bases, resulting in 96 possible trinucleotide mutation contexts. The observed mutation profile is then compared with reference signatures to infer the mutational processes that have been active during tumour development. Reliable signature detection requires a sufficient number of somatic mutations, and interpretation can be complicated by overlapping signatures or signatures with unknown biological origins.[242,409,410]

Mouse experiments

In **Article III**, a syngeneic BM transplantation mouse model was used to investigate the leukemogenic potential of the novel *KMT2A::ACIN1* and the rare *KMT2A::USP2* fusion genes *in vivo*. Murine BM cells were transduced with retroviral vectors expressing *KMT2A*::*ACIN1* or *KMT2A*::*USP2*, and *KMT2A*::*MLLT3* served as a positive control. The transduced cells were transplanted into lethally irradiated syngeneic recipient mice, enabling assessment of leukemia initiation and disease latency, followed by secondary transplantation to assess leukemia-propagating capacity. In this model, expression of *KMT2A*::*ACIN1* and *KMT2A*::*MLLT3* in murine BM cells resulted in myeloid leukemia, whereas *KMT2A*::*USP2* showed limited leukemogenic potential.

Statistical considerations

Statistical analyses were performed in all three articles (**I-III**). Owing to the relatively small cohort sizes and the non-normal distribution of several variables, non-parametric statistical tests were primarily applied. Two-group comparisons were performed using the Mann-Whitney U test (Wilcoxon rank-sum test), while comparisons of more than two groups were performed using the Kruskal-Wallis test. Categorical variables were compared using Fisher's exact test. Correlations were assessed using Spearman's rank correlation coefficient. Overall survival was estimated using the Kaplan-Meier method and compared using the Mantel-Cox log-rank test. Cox regression was used to assess associations between recurrent gene mutations and the risk of death or relapse.

All statistical tests were two-sided and $p<0.05$ was considered statistically significant. When multiple statistical tests are performed, the risk of false-positive findings increases. To account for this, p-values were adjusted using the Benjamini-Hochberg method when appropriate. This method controls the false discovery rate,

which represents the expected proportion of false-positive findings among the significant results.

Alternative methodological approaches

Other methods could complement the approaches used in this thesis. Long-read sequencing could improve the characterization of complex structural variants and fusion genes. Larger scRNA-seq studies could offer a more detailed view of transcriptional heterogeneity and help identify rare cell populations that are not resolved by bulk RNA-seq. Serial single-cell analysis could resolve clonal changes during treatment and relapse, while combining genetic and transcriptional information could link specific subclones to their transcriptional profiles. Single-cell chromatin profiling could further reveal how epigenetic states contribute to the transcriptional and lineage heterogeneity of *KMT2A*-r leukemia. Finally, functional studies are needed to determine whether identified genetic alterations contribute directly to the observed phenotypes.

Ethical considerations

Samples were obtained with informed consent from the parents or legal guardians of children, or from the patients themselves if aged 18 years or older, in accordance with the Declaration of Helsinki. All studies were approved by the relevant ethics review authorities in Sweden.

All studies were retrospective and did not affect clinical management or treatment decisions. Only archived material was used, and no additional samples were collected for the studies.

Patient samples were assigned study-specific identification numbers to protect patient identity. Sequencing data contain sensitive genetic information and were therefore stored on secure servers with restricted access in accordance with legal requirements for handling sensitive data.

The studies were restricted to somatic alterations, and germline variants were not investigated. Matched non-leukemic samples were used only to distinguish somatic from inherited variants. This limited the risk of identifying inherited findings with potential implications for patients or their relatives. If a clinically significant constitutional finding were identified incidentally, this would be handled according to established clinical genetics procedures.

Mouse studies were conducted in accordance with Swedish legislation under approved animal ethics permits. Experiments were performed by trained personnel, with predefined humane endpoints to minimise pain and suffering.

Summary of articles

Article I

The genomic landscape of relapsed infant and childhood *KMT2A*-rearranged acute leukemia

Over 90% of infants reach CR, but relapse can occur rapidly, with 90% of relapses occurring within two years of diagnosis and 66% within one year.[352,355] Despite this, the molecular mechanisms driving relapse remain incompletely understood, limiting not only the development of targeted therapies but also efforts to improve risk stratification and identify patients at high risk of relapse.

To gain insight into the mechanisms of relapse and clonal dynamics during treatment, we studied a cohort of 36 patients with relapsed *KMT2A*-r, including 26 infants (24 ALL, 2 AML) and 10 children (1 ALL, 9 AML) with WGS and WES, representing the largest cohort of relapsed pediatric *KMT2A*-r acute leukemia reported to date. In addition, we performed targeted sequencing of 257 samples from 30 cases, including 14 cases from the relapse cohort, during the disease course.

Results in short

- Early relapse ALL (9-36 months after diagnosis) gained alterations in drug-response genes at relapse, most commonly in *IKZF1* and *TP53*, while such alterations were rare in very early relapse ALL (<9 months).
- Mutational signature analysis indicated that chemotherapy was the primary mutational process in early relapse ALL but not in very early relapse ALL or AML.
- In AML, the same mutational processes were active at diagnosis and relapse, suggesting that the cells that gave rise to relapse were likely dormant during therapy and responded differently to therapy than cells in ALL.

Genomic landscape of relapsed KMT2A-r ALL

In ALL, 20 mutated genes affecting five pathways were identified at relapse. These alterations were primarily observed in patients with early relapse (9-36 months after diagnosis, according to Li et al.[205]) and were uncommon in very early relapse (<9 months) or refractory disease. The affected pathways included B-cell maturation (very early 27% vs. early 64%), cell cycle regulation (27% vs. 50%), glucocorticoid receptor signalling (0% vs. 21%), purine metabolism (0% vs. 21%) and kinase/PI3K/RAS signalling (18% vs. 50%). Furthermore, 79% of early relapse ALL cases harboured mutations associated with chemoresistance (*IKZF1, TP53,*

NT5C2, CREBBP, PRPS2, WHSC1 and *NR3C1*), whereas such alterations were rare in very early relapse ALL, where only *IKZF1* alterations were observed (9%).

TP53 and *IKZF1* were the most frequently altered genes (64%) and co-occurred in 56%, suggesting genetic cooperativity. All *TP53* mutations were gained at relapse, and several cases had multiple *TP53*-alterations. In addition, in 67% of cases, alterations affecting purine metabolism (*NT5C2*, *PRPS2*) and glucocorticoid signalling (*CREBBP*, *WHSC1*, *NR3C1*) co-occurred with alterations in *IKZF1* or *TP53*. The $PRPS2^{P320L}$ variant, found at relapse, was novel but knockout of wildtype *PRPS2* together with overexpression of $PRPS2^{P320L}$ revealed that $PRPS2^{P320L}$ had minimal effects on viability in the presence of 6-mercaptopurine.

To validate the frequency of these relapse-associated alterations in *KMT2A*-r infant ALL, we analysed data from three studies[129,205,290] including 80 diagnostic *KMT2A*-r infant ALL cases, of which 18 had a paired relapse sample. We found that 6/12 (50%) of infants with early relapse and none of the 6 infants with very early relapse had such alterations at relapse, aligning with our findings **(Figure 1A)**.

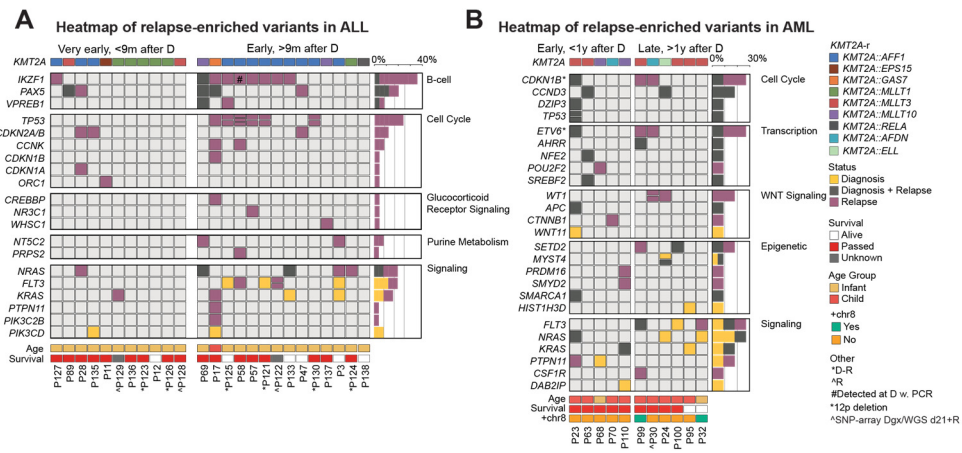


Figure 1A: Heatmap of relapse enriched variants in ALL and AML. A) ALL and B) AML. Alterations in genes within pathways enriched at relapse, with genes in rows and patients in columns. To the right of each heatmap, the bar plot indicates the fraction of patients with alterations in each gene, while the pathways are shown on the left. Yellow indicates diagnosis-specific alterations, grey alterations shared between diagnosis and relapse, and lilac relapse-specific alterations. D-R indicates that no germline sample was available, while R indicates that only a relapse sample was analysed.

Genomic landscape of relapsed KMT2A-r AML

Mutations in five pathways were enriched at AML relapse: cell cycle, transcription, WNT-signalling, epigenetic and signalling, with only six genes recurrently altered, including deletion of 12p affecting *CDKN1B* and *ETV6*, mutations in *CCND3*, *WT1*, *SETD2* and *FLT3*. *WT1*-mutations were gained at late relapse (33%) and detected

in cases with rare *KMT2A*-r (*KMT2A::AFDN*, *KMT2A::ELL*), whereas early relapse harboured mutations in other WNT-signalling pathway genes (40%, *APC*, *CTNNB1*). Alterations in transcription factor genes were enriched at early relapse (60% vs 33%) and included *ETV6* (12p-del), and mutations in *NFE2*, *POU2F2, AHRR* and *SREBF2*.

Signalling mutations were more common at diagnosis than at relapse in AML (73% versus 36%) and often subclonal at relapse. Additionally, diagnostic *KRAS/NRAS* mutations were maintained at early relapse, whereas they were lost at late relapse, which instead gained receptor tyrosine kinase mutations (*FLT3*, *CSF1R*).

Within all enriched relapse pathways, early relapse AML maintained most mutations from diagnosis to relapse (60%), whereas late relapse gained (38%) and lost (29%) more mutations (**Figure 1A**).

Mutational signatures at relapse fingerprint cellular history

To understand which mutational processes were active at relapse, we examined the mutational signatures. In ALL, 9/14 cases had chemotherapy exposure as their primary signature and 36%-86% of their mutations were attributable to chemotherapy exposure, particularly SBS87 which has been correlated with thiopurine treatment. Five cases did not have chemotherapy as their primary signature, of which three had refractory disease. One of the three hypermutated cases at relapse (P11) had a clock-like signature. Cases with mutations affecting purine metabolism (P69, P58 and P3) had the most prominent chemotherapy-induced signatures, whereas very early relapse cases had the least prominent signatures. Combined, early relapse ALL was characterized by mutations that were associated with chemotherapy exposure, consistent with surviving cells accumulating mutations from cytotoxic treatment.

As opposed to ALL, the most common mutational signature at AML relapse was of unknown origin (6/9) with chemotherapy-associated signatures virtually absent. We performed an analysis of relapse-specific mutations only and compared them to the diagnostic profiles, showing that in contrast to ALL, the mutational processes at AML relapse did not differ from those at diagnosis and mirrored the natural acquisition of mutations. This indicates that the relapse originated from an early ancestral cell that survived chemotherapy.

Clonal evolution patterns from diagnosis to relapse

We next investigated the pattern of clonal evolution from diagnosis to relapse. All relapses lost diagnosis specific clones indicative of branching evolution. By comparing the diagnostic and relapse VAFs, we found that in 55% of cases, relapse was seeded by multiple diagnostic clones and in the remaining 45%, relapse arose from a single sweeping clone, 40% of which were detected at diagnosis. Most

patients had additional subclones at relapse, in line with a continuous evolution generating new clones. The lack of detection of the relapse clone at diagnosis does not exclude that it was present but below our resolution (see **Article II**). In ALL, the clonal evolution pattern correlated with time to relapse with multiple diagnostic clones seeding relapse in 80% of very early and 33% of early relapse ALL, whereas in AML the frequency was the same (early, 66% and late, 60%).

One ALL and four AML had multiple BM relapses, allowing us to study how the genetic landscape evolved across consecutive relapses. The mutational burden increased after each relapse and mutations in driver genes were not lost once gained at relapse, suggesting a fitness advantage.

Residual leukemia cells and unique clonal responses

To uncover clonal responses to treatment and study measurable disease at the clinical MRD timepoints, we studied 19 relapse (10 ALL, 9 AML) and 11 remission (7 ALL, 4 AML) cases using patient-specific mutations, including the *KMT2A* rearrangement, by targeted sequencing.

In ALL, at day 15, the first MRD measurement during induction therapy, 13/15 cases had measurable disease, with a similar average VAF in relapse and remission cases, whereas at the end of induction at day 29, relapse cases showed a trend towards higher values (p=0.069). This was mainly driven by very early relapse cases and if no mutations were detected, patients remained in remission.

In AML, around half of cases had measurable disease after the first induction and the average VAF was similar in remission and relapse patients, with two having measurable disease despite being in CR. After the second induction 7/12 cases had measurable disease, including 3 cases in CR by clinical MRD, and 6/7 cases relapsed.

In 4/16 cases with a BM relapse, we detected the relapse clone after induction therapy as long as 25-92 days before relapse. Some diagnostic mutations/SVs were always present and sufficient to detect the relapse. Together, these findings showed that molecular monitoring with patient-specific mutations could readily detect leukemic cells.

Longitudinal sequencing provided insight into clonal evolution during treatment, demonstrating that genetically distinct clones can respond differently to chemotherapy, compete for clonal dominance and persist below the detection limit before re-emerging at relapse (**Figure 1B**).

To further investigate clonal composition and mutational co-occurrence, single cells from two patients were sorted and analysed for selected targets. These results were consistent with the longitudinal sequencing. Together, these findings show that longitudinal genomic profiling can identify relapse-initiating clones months before

overt relapse and provide valuable insight into the evolutionary dynamics underlying treatment failure.

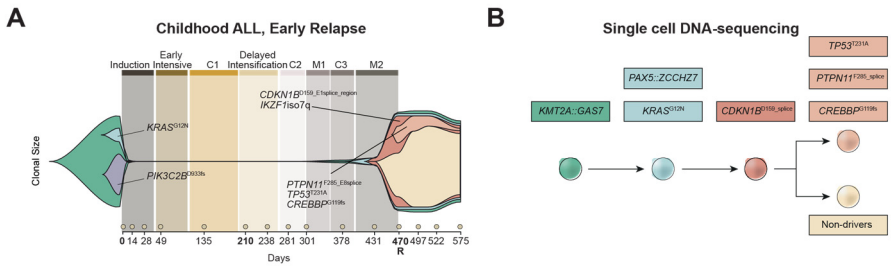


Figure 1B: Longitudinal sequencing of a child with ALL and an early relapse. A) Fish plot showing clonal evolution in a child in whom the *KMT2A*-r was detected at day 378 in 3% of cells, 92 days before relapse and 350 days after CR. The *KRAS*^{G12N}-relapse clone was detected at diagnosis (2%). The diagnostic *PIK3C2B*^{D933fs} was lost at relapse, while *CDKN1B, IKZF1, PTPN11, TP53* and *CREBBP* were gained. B) The order of mutations at relapse as determined by single-cell DNA sequencing. Mutations are gained sequentially with all relapse cells containing *KRAS*^{G12N}, *PAX5::ZCCHZ7*, *CDKN1B*, followed by expansion of cells with *TP53*, *PTPN11* and *CREBBP*-mutations and a clone without additional known driver events.

General conclusions

We demonstrate different relapse mechanisms in *KMT2A*-r ALL and AML. In ALL, early relapse appears to be driven by surviving leukemic cells that acquire chemotherapy-induced mutations, with 79% of cases harbouring relapse-associated alterations in drug-response genes, most commonly *TP53* and *IKZF1*. By contrast, 91% of very early relapse ALL lacked such mutations, in line with inherent resistance. Relapsed AML had the same mutational processes active at diagnosis and relapse and lacked a chemotherapy-exposure signature, indicating relapse from a cell that has remained unaffected by therapy.

Article II

Duplex sequencing uncovers recurrent low-frequency cancer-associated mutations in infant and childhood *KMT2A*-rearranged acute leukemia

Infant ALL has an exceptionally low mutational burden, averaging only 1.3 non-silent somatic mutations in the major leukemia clone per case.[129,411] Despite this paucity of mutations, approximately half of infant *KMT2A*-r ALL cases harbour activating mutations in the PI3K/RAS pathway, suggesting that aberrant signalling represents an important cooperating event.[129,412]

Given this high fraction of subclonal mutations, this study aimed to determine whether low-frequency mutations, especially in RAS signalling genes, are common at diagnosis and whether they could provide a reservoir of genetically diverse clones

for later selection. We targeted 28 leukemia-associated genes using sensitive duplex sequencing in a cohort of 20 infants and 7 children with *KMT2A*-r leukemia. We also analysed 6 paired relapse samples and 5 longitudinal samples.

Results in short

- We identified 87 cancer-associated mutations across the cohort. Only 33 (38%) were present in more than 2% of cells, indicating that conventional NGS would have missed the remaining 62%.
- Expanding clones frequently harboured PI3K/RAS pathway mutations, yet 55% of all signalling mutations resided in non-expanding clones, suggesting that mutation acquisition alone is insufficient for clonal expansion.
- Relapse-associated mutations were already detectable at diagnosis in rare subclones comprising <0.4% of the total cell population, showing that relapse-initiating clones may already be detectable at diagnosis, long before clinical relapse. However, one *TP53* mutation was detected exclusively at relapse, consistent with acquisition during treatment or expansion from a pre-existing undetectable subclone.

Most low-frequency mutations would go undetected by standard NGS

We identified 931 mutations across all investigated bases in the targeted genes. One hypermutated case accounted for 440 mutations and was excluded from subsequent statistical analyses. The remaining cases harboured 491 mutations.

More than 90% of mutations had a VAF below 0.01 and would therefore not be detected by conventional NGS. Thus, duplex sequencing uncovered a substantial burden of low frequency mutations that are missed by standard sequencing.

All cancer-associated mutations target recurrently mutated genes in KMT2A-r ALL

Of the identified mutations, 87 occurred at sites previously associated with leukemia and affected eight of the investigated genes: *KRAS*, *NRAS, FLT3, TP53, PIK3CA, PAX5, PIK3R1* and *PTPN11*. The remaining genes carried no cancer-associated mutations, in line with their rare occurrence in *KMT2A*-r leukemia.

Children harboured more than twice as many cancer-associated mutations as infants (6.0 vs. 2.6 per case); five infants lacked cancer-associated mutations. Only 7% of cancer-associated mutations were present in the dominant leukemic clone, whereas 62% occurred below the detection limit of conventional NGS, highlighting extensive subclonal heterogeneity.

PI3K/RAS mutations were present in larger fractions of cells

Among the 33 cancer-associated mutations present in more than 2% of cells, 31 affected *FLT3* or PI3K/RAS pathway genes, in line with their role as cooperating lesions in *KMT2A*-r leukemia. Notably, 74% of the *KRAS* and *NRAS* mutations affected canonical hotspot codons. When all signalling mutations were considered, 70% of infants and all children harboured at least one PI3K/RAS pathway mutation, often distributed across multiple genetically distinct subclones. However, more than half (55%) of all signalling mutations remained below the detection limit of conventional NGS, showing that many did not undergo substantial clonal expansion despite their oncogenic potential.

Low-frequency drug-resistant clones expand at relapse

To investigate whether low-frequency diagnostic variants could represent relapse-initiating clones, multiplex PCR followed by NGS was performed on six paired diagnosis-relapse samples. In three out of six patients, rare diagnostic subclones carrying $FLT3^{D835H}$ or $NRAS^{G13D/G12S}$ (VAF 0.0003-0.0017) expanded and became dominant at relapse, whereas higher-frequency diagnostic *KRAS* or *NRAS* mutant clones disappeared in two of these patients.

In one patient, a $TP53^{G245V}$ was detected at relapse and remained undetectable at diagnosis despite analysis by both duplex sequencing and ddPCR (detection limit VAF 0.00009), suggesting that it either arose during treatment or was below the detection limit at diagnosis. These findings show that rare diagnostic subclones can survive chemotherapy and subsequently expand at relapse (**Figure 2A**).

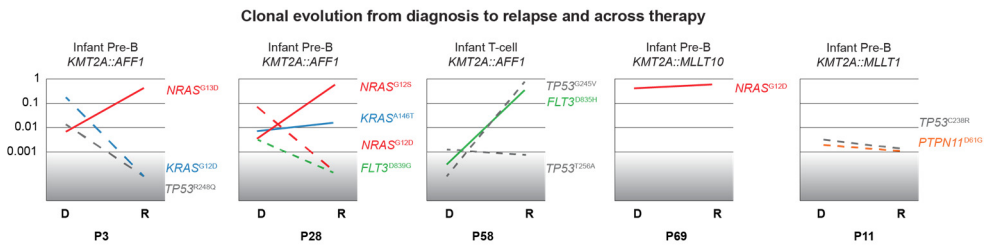


Figure 2A: Paired diagnostic and relapse samples showing the VAF change in cancer associated mutations. Changes in VAF from diagnosis to relapse in five of the six patients with paired relapse samples. Patient 12 had no cancer-associated mutations and is not shown. Red: *NRAS* mutation; Blue: *KRAS* mutation; Green: *FLT3* mutation; Gray: *TP53* mutation; Orange: *PTPN11* mutation. Dashed lines indicate mutations not detected in one of the paired samples. Grey area indicates the detection limit of the targeted sequencing. The scale is log 10 transformed.

Clonal evolution in serial samples reveals clone-specific response kinetics

To investigate how treatment affected low frequency subclones, we used duplex sequencing to track longitudinal samples from a single patient with refractory disease. At diagnosis, the leukemia exhibited at least three distinct subclonal branches off the founding clone (A):

1. The major diagnostic subclone (B) regressed during induction therapy but then started to expand.
2. A minor subclone (D) harbouring an *NRAS*^{G12S} comprised only 0.22% of the major diagnostic clone. It remained quiescent for more than 200 days of treatment. Late in the clinical course, it expanded rapidly, acquired additional mutations and became the dominant relapse clone.
3. A third subclone (C) was present in 56% of the diagnostic cells but was eliminated completely by day 93 of therapy.

Longitudinal analysis of this patient illustrates a classic pattern of branching evolution. It shows that future relapse-driving clones can persist at nearly invisible levels for extended periods under therapeutic pressure **(Figure 2B)**.

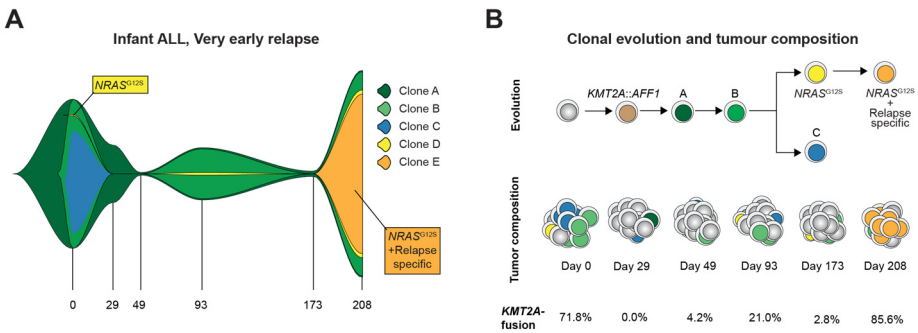


Figure 2B: Longitudinal sequencing with duplex sequencing. A) Fish plot showing the clonal evolution in an infant with a refractory disease. The founding clone "A" is shown in dark green, the major diagnostic clone "B" in light green, clone "C" in blue and clone "D" in yellow. Clone D contained an *NRAS*^{G12S} mutation and expanded in a selective sweep at relapse. By this time, it had acquired additional mutations, defining clone E (orange). B) Schematic view of clonal evolution and tumour composition. The *KMT2A*-r clone in light brown indicates the first event and clone A and B, the first 2 clones before branching. Clone C was lost during treatment, and clone D was seen as a small subclone at diagnosis and through treatment and during the last month before relapse, it expanded and accumulated additional mutations that were only detected at relapse.

General conclusions

Our data show that the subclonal genetic landscape of *KMT2A*-r leukemia is far more heterogeneous than previously assumed. Furthermore, cancer-associated mutations do not always result in clonal expansion. This indicates that microenvironmental factors or clonal competition may influence clonal fate. Finally, some mutations present at relapse can already be detected at diagnosis as rare subclones, persisting through treatment before expanding at relapse. However, the *TP53*^{G245V} was only detected at relapse, illustrating that not all relapse-associated mutations are detectable at diagnosis, either because they are acquired during treatment or remain below the detection limit.

Article III

A genomic atlas of *KMT2A*-rearranged leukemia from infancy to adulthood

The genomes of infants harbour few additional genetic alterations apart from the *KMT2A* rearrangement and little is known about the genomic landscape in adult *KMT2A*-r leukemia.[129] To gain insight into the molecular landscape of *KMT2A-r* leukemia across ages, genetic subtypes and lineages, we characterized the genetic and transcriptional landscapes of 116 *KMT2A*-r cases. This is one of the most comprehensive genomic studies of *KMT2A*-r leukemia to date, including: 39 infant ALL, 19 childhood ALL, 16 adult ALL, 8 infant AML, 30 childhood AML and 4 MPAL (3 infants and 1 child), together with 3 infant ALL *KMT2A*-wildtype cases. Samples were characterized using whole-genome sequencing (WGS; n=96, average 45×), whole-exome sequencing (WES; n=116, average 171×), RNA sequencing (n=57, average 32×), and whole-genome bisulfite sequencing (WGBS; n=17).

Results in short

- Specific genetic alterations were associated with leukemia lineage, *KMT2A* fusion partner and age.
- *FLT3* alterations were enriched in AML with *KMT2A::MLLT3*, whereas B-cell developmental gene alterations (e.g. *PAX5* and *IKZF1*) were enriched in ALL carrying predominantly myeloid-associated *KMT2A* fusion partners, suggesting that cooperating lesions reinforce lineage identity.
- RNA-seq identified distinct transcriptional states in both ALL and AML that extended beyond the *KMT2A* fusion partner and reflected differences in cellular differentiation, underlying biology and clinical outcome.

An age dependent increase in the number of mutations

The median number of SNVs and indels was 222 across all cases and increased with age, ranging from 69 in infant AML to 439 in adult ALL. The adult ALL cases displayed a higher average VAF than infant and childhood ALL/AML, indicating a more clonal disease. Of the 537 non-synonymous mutations identified, 92% were validated by both WGS and WES or by PCR, and 38% were expressed in matched RNA data.

A total of 543 SVs and CNAs were identified, of which all but three were validated by another method, manual inspection or were related to the *KMT2A* rearrangement. The average number was highest in childhood ALL (8.4) and lowest in infant ALL (4.6). Almost two thirds were a direct result of the *KMT2A* rearrangement. Overall, 36% of cases harboured only *KMT2A*-related SVs/CNAs, and this was more common in infants (47-67%), than in children and adults (13-33%). Thus, children and adults had a larger number of *KMT2A*-unrelated SVs/CNAs, suggesting that these lesions are cooperating leukemogenic events.

Mutational signatures revealed that a clock-like signature predominated across subtypes (77%) followed by chemotherapy exposure (13%), with a higher relapse prevalence in the latter group (60% versus 34%). Combined, the mutational burden increased with age, with adult *KMT2A*-r ALL exhibiting a more clonal disease than younger cases, implying differences in their evolution.

Age and subtype dependent mutation spectrum

We identified 31 recurrently altered genes in *KMT2A*-r leukemia including 21 within 4 pathways; PI3K/RAS-signalling (40%), epigenetic (21%), hematopoietic differentiation (17%), cell cycle (11%) and other genes (11%), with *NRAS*, *KRAS*, *FLT3, PAX5, IKZF1*, *CDKN2A/B* and *TP53*, accounting for 62% of the alterations. Fewer genes were recurrently altered in AML (n=8) than in ALL (n=16) (**Figure 3A**).

Including non-recurrent alterations, PI3K/RAS-pathway was the most frequently mutated (46%) and was more common in AML (68%) compared to ALL (35%). Over half of the mutations were subclonal (VAF <0.3) and 34% of cases harboured multiple mutations. In these cases, the combined VAF suggested that the mutations resided in different clones. To investigate the clonal architecture of co-occurring signalling mutations, we performed single-cell WES on an infant ALL case that had previously been analysed by duplex sequencing (**Article II**). Multiple independent *NRAS*, *KRAS* and *FLT3* mutations were identified in distinct clones, revealing extensive branching evolution with at least ten diagnostic clones (**Figure 3A**).

NRAS, *KRAS* and *FLT3* accounted for most signalling mutations, affecting 40% of cases. Activating mutations displayed marked fusion-, lineage- and age-specific patterns. While AML typically harboured single activating driver mutations, particularly *FLT3* in *KMT2A::MLLT3* and *NRAS*/*KRAS* in *KMT2A::MLLT10*, ALL more frequently had multiple co-occurring signalling mutations, with 38% of cases harbouring more than one *NRAS/KRAS/FLT3* mutation compared with only 9% in AML. These findings support that signalling pathway activation is shaped by fusion gene context and that multiple cooperative signalling lesions might be more important for leukemogenesis in ALL than in AML (**Figure 3A**).

Alterations affecting hematopoietic differentiation were detected in 29% of B-ALL cases and were common in childhood (53%) and adult (40%) ALL, but rare in infant ALL (13%) and AML (5%). In ALL, the most frequently affected genes were *PAX5* (14%), *IKZF1* (13%) and *VPREB1* (6%). *PAX5* alterations were enriched in ALL with *KMT2A::MLLT3*, whereas *IKZF1* alterations increased with age and co-occurred with *KMT2A::AFF1*.

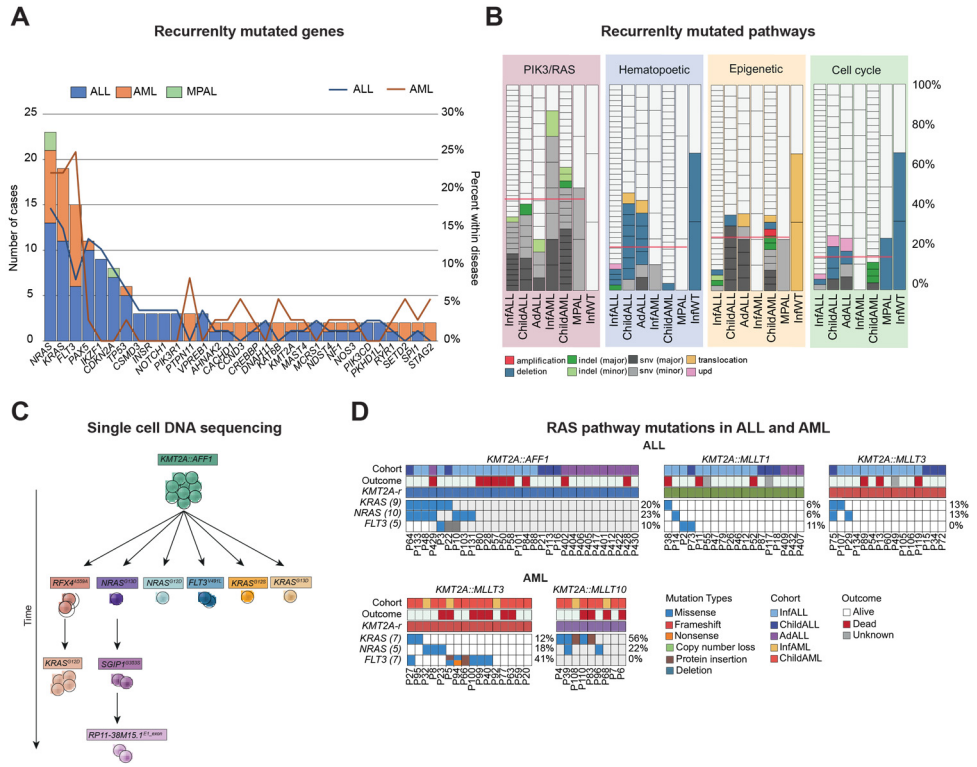


Figure 3A: The genomic landscape of *KMT2A-r* leukemia. A) Bar plot showing the 31 recurrent genes. The left y-axis shows the number of cases with an alteration. Each gene has been coloured based on the leukemia type and the right y-axis shows the fraction of mutations in each leukemia type. B) The fraction of cases with alterations in recurrently mutated pathways is shown, with each tile representing one case within each cohort. The red line indicates the fraction of all *KMT2A*-r cases with an alteration in the respective pathway. C) Single-cell WES of 25 cells in an infant ALL case, with each colour representing a cell with a unique mutational profile. This case had six RAS-pathway mutations in different cells. Due to the limited number of cells analysed, each unique mutational profile was considered a separate clone; thus, a clone could consist of a single cell. D) Heatmap of the most common fusion genes in ALL and AML and the fraction of mutations in *KRAS*, *NRAS* and *FLT3*. At the top, outcome and cohort are indicated.

Cell-cycle alterations (*CDKN2A/B*, *TP53* and *CCND3*) were predominantly observed in childhood and adult ALL and were rare during infancy. *TP53* mutations were largely restricted to *KMT2A*::*AFF1*, whereas *CCND3* mutations were observed exclusively in AML. Mutations affecting epigenetic regulators were rare in infant leukemia but increased with age.

Collectively, these findings demonstrate that the mutational landscape of *KMT2A*-r leukemia is highly dependent on developmental stage, lineage and fusion partner, and suggest that the genetic requirements for leukemogenesis differ fundamentally between ALL and AML (**Figure 3A**).

Cooperating mutations enforcing lineage identity

The observation that ALL with *KMT2A::MLLT3* or *KMT2A::MLLT10* more often harboured alterations in B-cell differentiation genes such as *PAX5*, whereas AML, particularly *KMT2A::MLLT3*, was enriched for activating *FLT3* and RAS alterations, supports a role for cooperating mutations in lineage-specific leukemogenesis.

To explore this, we interrogated the RNA-seq data. Unsupervised PCA identified three transcriptionally distinct groups in ALL that were largely independent of age and only partly explained by the fusion partner. One group displayed an immature stem/progenitor-like program enriched for *HOXA*, *MYC* and proliferation signatures, whereas another exhibited a more differentiated pre-B-cell phenotype with high expression of B-cell markers and frequent *PAX5* alterations. A third, smaller group showed features of multipotent progenitors together with *MYC*, *KRAS*/*MAPK* and oxidative phosphorylation signatures. These transcriptional groups were supported by DNA methylation analyses, reproduced in an independent cohort and associated with differences in relapse frequency and survival, suggesting that they represent biologically distinct subtypes of *KMT2A*-r ALL (**Figure 3B**).

Similarly, AML segregated into three major transcriptional groups. Focusing on the two larger groups, one was characterized by primitive HSC/LSC programs and frequent epigenetic mutations, whereas the other showed more differentiated myeloid, MYC/ERK and metabolic signatures together with frequent *NRAS* mutations. These groups were also validated in an external cohort (**Figure 3B**).

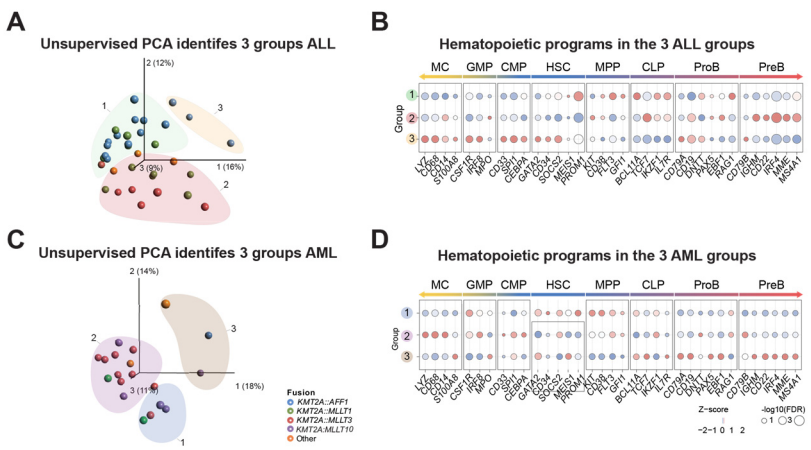


Figure 3B: RNA sequencing reveals distinct transcriptional states. A) Unsupervised principal component analysis identified three groups of B-lineage ALL that were largely independent of age and genetic subtype. B) Hematopoietic transcriptional programs across the three ALL groups. C) Unsupervised principal component analysis identified three groups of AML. D) Hematopoietic transcriptional programs across the three AML groups.

Comparison of fusion gene subtypes showed that ALL with the predominantly myeloid associated fusion partners *MLLT3* and *MLLT10* expressed higher levels of lymphoid differentiation genes and lower levels of myeloid regulators than *KMT2A*::*AFF1* or *KMT2A*::*MLLT1* ALL. These findings suggest that such fusions may require a stronger lymphoid differentiation program, frequently through cooperating B-cell developmental lesions, to develop as ALL.

Collectively, *KMT2A*-r leukemia comprises distinct transcriptional states that extend beyond fusion partner alone. These states are associated with cooperating genetic alterations, differentiation and clinical outcome, highlighting additional biological heterogeneity within *KMT2A*-r leukemia.

Copy number alterations reveal early clonal evolution

Recurrent whole-chromosome gains were uncommon but followed age- and lineage-specific patterns. Trisomy X was the most frequent gain, whereas trisomy 19 was restricted to childhood AML.

Analysis of SNVs within trisomic regions supported an *in utero* origin of infant and childhood ALL and suggested that recurrent copy number gains are acquired early during leukemogenesis. In contrast, adult ALL showed evidence of a later disease origin and acquisition of these chromosomal gains.

Novel KMT2A fusion genes and structural rearrangements

WGS revealed that the formation of the *KMT2A* rearrangements was often more complex than apparent from cytogenetics and occurred in closed loops with chained breakpoints, consistent with formation through a single event. Excluding the *KMT2A* fusion itself, most SVs were predicted loss-of-function events, whereas only a small fraction generated expressed in-frame fusion genes. In addition, half of all *KMT2A*::*MLLT1* cases were formed through splicing, indicating a distinct mechanism of fusion formation. Of note, spliced cases had higher survival than unspliced cases, although numbers were small (100% vs 63%).

WGS and RNA-seq identified *KMT2A* fusions that had not been detected at diagnosis, including the novel *KMT2A::ACIN1* fusion. We also found that reciprocal *KMT2A* fusion transcripts were expressed only in *KMT2A::AFF1* leukemia. Thus, *KMT2A* rearrangements arise through extensive structural complexity that is hidden by conventional cytogenetic analyses. To assess the leukemogenic potential of the novel *KMT2A*::*ACIN1* and the rare *KMT2A*::*USP2* fusions, we performed retroviral transplantation experiments in syngeneic mice. Both fusions induced disease, although with longer latency than *KMT2A*::*MLLT3*. *KMT2A*::*ACIN1* efficiently generated transplantable leukemia, whereas *KMT2A*::*USP2* showed incomplete penetrance and limited serial transplantability, suggesting that additional cooperating genetic events are required for full transformation.

Genomics of non-KMT2A-r infant ALL

The three infant *KMT2A*-wildtype ALL cases lacked PI3K/RAS mutations but harboured recurrent *PAX5* and *CDKN2A/B* alterations together with complex structural rearrangements. One case harboured *ZNF618::NUTM1*, another *TPR::JAK2* consistent with Ph-like ALL, and a third the novel putative *BPIFB1::CNTLN* fusion, highlighting a distinct genetic landscape compared with *KMT2A*-r infant ALL.

General conclusion

Collectively, our findings show that the biological phenotype of *KMT2A*-r leukemia is shaped not by the fusion alone, but also by age, lineage, and cooperating lesions.

Conclusions

The overall aim of this thesis was to further understand the mechanisms underlying the poor prognosis of *KMT2A*-r leukemia and to use this information to improve future diagnostics, follow-up and treatment.

This was addressed through paired analysis of diagnostic and relapse samples to uncover mechanisms of relapse (**I**), sensitive sequencing to characterize the subclonal landscape at diagnosis and relapse (**II**), and comprehensive genomic analysis of a large clinically annotated cohort (**III**). The main conclusions from this thesis and future plans are summarized below:

Article I | The genomic landscape of relapsed infant and childhood *KMT2A*-rearranged acute leukemia

- **Relapse mechanisms differ in *KMT2A*-r AML and ALL.** Very early ALL relapse frequently arose from multiple diagnostic clones, whereas early ALL relapse showed stronger evidence of treatment-associated mutagenesis and acquired resistance. AML relapse showed little evidence of chemotherapy-driven mutagenesis.
- **Chemoresistance in early relapse ALL.** Early relapse ALL frequently acquired alterations associated with treatment resistance, including *IKZF1* and *TP53*, whereas very early relapse was more often associated with pre-existing resistant clones.
- **Tracking resistant clones during treatment.** Longitudinal sequencing uncovered unique clonal responses and allowed tracking of residual leukemia cells and individual clones during treatment, including emerging relapse clones before overt clinical relapse.

Article II | Duplex sequencing uncovers recurrent low-frequency cancer-associated mutations in infant and childhood *KMT2A*-rearranged acute leukemia

- **A vast subclonal genetic landscape.** Duplex sequencing revealed recurrent cancer-associated mutations in low-frequency subclones, many of which were undetectable by conventional sequencing, uncovering previously unrecognized subclonal diversity in *KMT2A*-r ALL.
- **Clonal selection during treatment.** Relapse may result from both selection and expansion of pre-existing subclones and acquisition of new genetic alterations during treatment, highlighting multiple routes to clonal evolution.
- **Mutation acquisition alone does not determine clonal selection.** The limited expansion of many mutation-bearing subclones suggests that cellular state, microenvironmental factors and combinations of cooperating mutations may influence which clones acquire a selective advantage.

Article III | A genomic atlas of *KMT2A*-rearranged leukemia from infancy to adulthood

- **Age-related genetic evolution.** The mutational burden increased with age, with adult cases displaying a more clonal genetic landscape than infants and children, indicating distinct patterns of genetic evolution across age groups.
- **Lineage-specific cooperating alterations.** Specific genetic alterations were associated with leukemia lineage, *KMT2A* fusion partner and age, with *FLT3* alterations enriched in AML and B-cell developmental gene alterations, including *PAX5* and *IKZF1*, enriched in ALL with predominantly myeloid-associated *KMT2A* fusion partners, suggesting that cooperating lesions may reinforce lineage identity.
- **Transcriptional and structural heterogeneity beyond the *KMT2A* fusion.** RNA-seq identified distinct transcriptional states associated with cellular differentiation, underlying biology and clinical outcome, while WGS revealed extensive structural complexity of *KMT2A* rearrangements, including cryptic and novel fusion genes.

Future steps

Across these three articles, we identified different relapse mechanisms, uncovered a vast subclonal landscape, found recurrent genetic alterations associated with age, *KMT2A* fusion partner and lineage, and demonstrated transcriptional heterogeneity in *KMT2A*-r leukemia. We now intend to investigate these findings through a large-

scale scRNA-seq study of *KMT2A*-r ALL and AML including available diagnostic, longitudinal and relapse samples. The study will help determine whether distinct transcriptional states are associated with very early and early relapse (**Article I, III**) and whether specific transcriptional states are selected for during treatment (**Article I, II**). By including probes for specific mutations, we will be able to link specific mutations to transcriptional states and follow their response to treatment. This may also provide further insight into the transcriptional heterogeneity identified in **Article III**.

Patient-derived xenograft models of *KMT2A*-r ALL will be created from patients with very early or early relapse, as well as a patient remaining in remission. Mice will be treated with clinically relevant chemotherapy regimens to model treatment-induced selective pressure. scRNA-seq will be used to investigate how therapy shapes leukemic cell states and clonal composition and WGS to study clonal evolution. The resultant data will also be integrated with scRNA-seq from the patient cohort.

Lastly, functional studies could investigate cooperating alterations, such as *PAX5* and *FLT3*, to determine whether their effects depend on the genetic context of the leukemia. For example, a specific alteration could be studied in different *KMT2A* fusion backgrounds to determine whether its effects vary by fusion partner. Such studies could help clarify how *KMT2A* fusions and cooperating alterations interact to shape the lineage and transcriptional heterogeneity identified in **Article III**.

Discussion

Evolutionary routes to relapse in pediatric acute leukemia: lessons from *KMT2A*-rearranged ALL and AML through a genetic lens

Relapse remains a major challenge in pediatric acute leukemia

The treatment of pediatric acute leukemia has improved substantially over recent decades, particularly in ALL, where most patients today achieve durable remission. However, relapse remains a major cause of treatment failure and leukemia-related death.[413-415] Prognostic factors include timing, age, site of recurrence, disease subtype, immunophenotype and response to reinduction therapy.[388,414,415] Among these factors, time to relapse is one of the strongest predictors of survival, with earlier recurrence associated with poorer outcomes than relapse occurring after a prolonged first remission.[388,415,416]

The clinical behaviour of relapsed leukemia is linked to its underlying biology. Relapse is rarely a simple re-emergence of an unchanged diagnostic population. Instead, treatment imposes a strong selective pressure on a heterogeneous population of leukemic cells and the disease that eventually returns may differ from

the dominant leukemia present at diagnosis.[205,207,209,417-420] Several evolutionary trajectories have been described. Relapse may arise through continued evolution of the dominant diagnostic clone, through selection and expansion of a minor subclone already present at diagnosis, or from an ancestral clone that preceded the dominant diagnostic population. In addition, multiple diagnostic subclones may survive therapy and collectively contribute to a multiclonal relapse. During this evolutionary process, therapy can both select pre-existing resistant populations and favour the emergence of cells carrying newly acquired genetic alterations, some of which confer treatment resistance and promote subsequent expansion at relapse.[205,213,419-421]

KMT2A-r leukemia provides a useful model to study these processes, as it occurs in both lymphoid and myeloid leukemia and is associated with distinct developmental, transcriptional and genetic features.[129,238,255,265,267,274,422] The disease is also notable for its biological heterogeneity despite a shared initiating lesion.

In *KMT2A*-r leukemia, approximately 90% of relapses occur within 2 years of diagnosis and 66% within 1 year.[205,352,355] Clinically, in ALL, relapse is classified as very early (<18 months from diagnosis), early (≥18 months from diagnosis and <6 months after the end of frontline treatment) or late (≥6 months after the end of frontline treatment), whereas in AML, relapse is generally classified as early (<1 year) or late (≥1 year). In **Article I**, we used a different definition for ALL, defining very early relapse as occurring within 9 months and early relapse as occurring after 9 months, based on distinct genomic features between these groups (**Table 4**).[205]

Disease	Relapse category	Definition
ALL - clinical classification	Very early	<18 months from diagnosis
	Early	≥18 months from diagnosis and <6 months after end of frontline treatment
	Late	≥6 months after end of frontline treatment
AML - clinical classification	Early	<1 year from diagnosis
	Late	≥1 year from diagnosis
ALL - **Article I**	Very early	<9 months from diagnosis
	Early	≥9 months from diagnosis

Table 4: Relapse times in ALL and AML. The clinical definitions of relapse in ALL and AML, as well as the definition used in **Article I.**

A central question is not simply why leukemia relapses, but which route to relapse a given leukemia takes. Are relapse-fated cells already present at diagnosis? Are resistance mechanisms generated during treatment? How do cell of origin, cellular state and context influence these trajectories? And can these trajectories be

distinguished early enough to influence treatment? Here, relapse is discussed as an evolutionary process, with particular emphasis on pediatric leukemia and on *KMT2A*-r ALL and AML as models of biologically different routes to relapse.

The diagnostic architecture can pave the way for relapse

The genetic and epigenetic features present at diagnosis can influence which clones contribute to relapse. Pediatric leukemias generally carry fewer somatic mutations than adult malignancies, reflecting their developmental origin and shorter evolutionary history.[240,243,244] This is clearly illustrated in infant *KMT2A*-r ALL, which often develops within the first months of life and is characterized by a remarkably low mutational burden.[129] Prenatal studies support an early origin of childhood leukemia, including *KMT2A*-r disease, showing that initiating fusion events can arise *in utero* before overt leukemia becomes clinically apparent.[122-124,127]

A low total mutation burden with few mutations in driver genes, however, should not be equated with biological homogeneity. Secondary signalling alterations, particularly within the RAS-pathway, are common in *KMT2A*-r leukemia and can occur as subclonal events.[129,216,222,412] This pattern is also evident across the studies presented in **Articles I-III**, with **Article II** providing the highest sensitivity for detecting low-frequency variants. Highly sensitive duplex sequencing uncovered considerably more low-frequency variants than conventional sequencing, including multiple RAS-pathway alterations in several patients at diagnosis.[213] Importantly, these alterations could reside in distinct leukemic clones rather than co-occurring within the same clone, as supported by single-cell sequencing in **Article III** and longitudinal analysis in **Article I**. Together, these findings reveal genetic diversity within the RAS-pathway at diagnosis, with different leukemic clones carrying alternative alterations affecting the same pathway. The recurrence of this pattern is consistent with strong selective pressure for RAS-pathway activation, although it remains unclear which of these pre-existing clones are subsequently favoured to persist and expand at relapse.

The genetic architecture of *KMT2A*-r leukemia also varies with age. In **Article III**, mutational burden increased with age, while adult ALL cases showed a more clonal architecture and infant leukemia a greater proportion of subclonal alterations. One interpretation is that the rapid onset of infant leukemia leaves comparatively little time for secondary clones to undergo extensive selection before diagnosis, whereas longer disease latency in older patients permits additional mutations to arise and expand. This remains a hypothesis, since clonality is also shaped by the strength of the initiating driver, cell of origin, mutation rate and the environment.

The cooperating genetic landscape also differed according to lineage. In **Article III**, *FLT3* alterations were enriched in AML, whereas alterations affecting B-cell developmental regulators, including *PAX5* and *IKZF1*, were predominantly associated with ALL. These observations are consistent with a model in which the

KMT2A fusion drives a shared oncogenic program, while additional alterations interact with developmental context to shape lineage-specific phenotypes.

Thus, even before therapy begins, apparently similar *KMT2A*-r leukemias can differ considerably in the number, identity and cellular distribution of cooperating alterations. The contribution of individual genetic alterations to relapse remains unclear, with relatively few mutations identified at diagnosis being associated with subsequent relapse.[207,209,211] Further, RAS-pathway lesions are known to be lost, maintained or gained at relapse, suggesting that their selective advantage is context-dependent rather than consistently required for relapse.[129,207,212] In line with this, RAS-pathway alterations were not associated with outcome in **Article III**. An alteration that provides a selective advantage during leukemogenesis may therefore not provide the same advantage during treatment or at relapse. Moreover, heterogeneity extends beyond genetics, as epigenetic and transcriptional differences generate distinct cellular states. Together, this diagnostic heterogeneity provides the substrate for subsequent evolution under treatment.

Relapse is an evolutionary process

Paired diagnosis-relapse studies have established that relapse usually remains clonally related to the original leukemia, but that the relationship between diagnostic and relapse populations is often complex.[207,418,419] Oncogenic initiating lesions are generally maintained, whereas secondary alterations are gained or lost. In many cases, the relapse population is more closely related to a minor diagnostic subclone or an ancestral clone than to the dominant population detected at presentation.[207,418] This produces branching evolution in which treatment removes predominant clones, allowing subclones to persist and expand.[207,418] Relapsed disease is generally more difficult to treat, reflecting both pre-existing and acquired resistance mechanisms.[414] The likelihood of remission and cure decreases further with each subsequent relapse.[423]

In **Article I**, we identified three patterns of clonal evolution, all retaining the founding *KMT2A* rearrangement: multiple clones seeding relapse (55%), a single sweeping clone detectable at diagnosis (18%) or a single sweeping clone not detectable at diagnosis (27%). Of note, the latter pattern does not necessarily indicate *de novo* emergence during therapy, as the relapse-founding clone could have been present below the detection limit at diagnosis. These three patterns have also been investigated in a larger cohort of pediatric leukemia relapses, in which relapse was seeded by multiple clones in only 20% of pediatric ALL cases, suggesting that this pattern is enriched in *KMT2A*-r leukemia. Multiple-clone seeding of relapse was associated with very early relapse, consistent with the high frequency of very early relapse in *KMT2A*-r leukemia.[205] This suggests that treatment selects among multiple pre-existing clones rather than driving a selective sweep toward dominance of a single clone. Importantly, very early relapse cases

carried few additional driver mutations beyond the founding *KMT2A* rearrangement, consistent with intrinsic treatment resistance.

Relapse-associated evolution in pediatric ALL repeatedly affects pathways involved in lymphoid development (*IKZF1)*, cell-cycle control (*TP53, CDKN2A/B*), glucocorticoid signalling (*NR3C1, CREBBP*), DNA repair (*MMR*), folate metabolism (*FPGS*) and purine metabolism (*NT5C2, PRPS2)*.[205-207,209,417,419,420,424,425] Several of these alterations were also observed in **Article I,** particularly *IKZF1* and *TP53*, as well as alterations affecting glucocorticoid signalling and purine metabolism. *NT5C2* provides a clear example of treatment-dependent selection. Activating *NT5C2* mutations are enriched in relapsed ALL and confer resistance to thiopurines.[421,426-429] Longitudinal studies have shown that mutant populations can emerge or expand during treatment and experimental models have demonstrated a direct selective advantage in the presence of 6-mercaptopurine (6-MP).[421] The reverse was also seen in **Article I**, where the withdrawal of 6-MP led to negative selection of the *NT5C2* clone.

Treatment can also contribute directly to the generation of genomic diversity. Mutational signature studies have identified relapse-associated patterns consistent with exposure to chemotherapy, raising the possibility that therapy is not only a selective pressure but can also generate mutations from which resistant populations subsequently emerge.[205] This was seen in **Article I**, where a chemotherapy-associated signature was the predominant signature in 89% of early relapse ALL cases, compared with only one of five very early relapse cases. Notably, the only very early relapse case with a predominant chemotherapy-associated signature represented the patient's third relapse and had therefore been exposed to several rounds of treatment. No AML cases had chemotherapy as the predominant signature. Instead, unknown signatures predominated in both very early ALL and AML at relapse. In **Article III**, a signature resembling chemotherapy-associated mutational damage was detected in 13% of diagnostic samples and was associated with a higher relapse rate. Two of these cases were secondary AML and may therefore reflect previous treatment exposure. In treatment-naïve cases, particularly infants, similar signatures may instead reflect overlapping mutational processes or other exposures that generate comparable patterns of DNA damage.[430]

Article II illustrates the difficulty of predicting relapse based only on genetics. Using duplex sequencing and highly sensitive ddPCR, relapse-associated *TP53*, *NRAS* and *KRAS* variants could in some cases be tracked to extremely low frequencies at diagnosis, whereas a *TP53* variant remained undetectable at diagnosis despite sensitivities far exceeding conventional sequencing. While this variant could have been present below the detection limit, previous studies have shown that *TP53* mutations can also arise during therapy, with thiopurine treatment shown to induce *TP53* mutagenesis in the context of mismatch repair deficiency.[146] Together, these findings illustrate both possibilities: relapse-associated variants may pre-exist at

very low frequencies at diagnosis or emerge during treatment. Some diagnostic populations with cancer-associated mutations were not selected at relapse, making it difficult to predict relapse based on this information alone. In **Article I**, AML cases with multiple relapses showed persistence of cancer-associated mutations, with stepwise replacement by fitter clones, potentially due to continued therapeutic pressure. Similarly, in relapsed childhood ALL, RAS-pathway mutations persisted at second relapse in 80% of patients and remained dominant in most cases.[210]

Further, treatment pressure can contribute to differences in relapse patterns between leukemic subtypes. For example, *TCF3*::*PBX1* has a favourable prognosis at diagnosis but, when relapse occurs, it tends to occur early, whereas the opposite pattern is observed for iAMP21. This could reflect differences in treatment intensity, with more intensive primary treatment potentially reducing relapse risk in iAMP21, or differences in relapse timing, clonal evolution, treatment resistance or response to salvage therapy.[142,388,431]

Relapse is not a static event, but a dynamic evolutionary process in which therapy shapes the survival and expansion of leukemic populations. With each round of treatment, resistant clones can be selected, persist and/or acquire additional driver alterations, leading to increasingly therapy-resistant disease.

Time to relapse reflects different evolutionary routes

The strong prognostic effect of relapse timing suggests that time is not merely a clinical variable but could also contain information about the underlying biology.[388,415,416] Earlier relapses may result from rapid expansion of treatment-resistant populations already present at diagnosis, whereas later relapses allow more time for persistent cells to undergo further selection, acquire resistance mechanisms or resume proliferation after treatment.[205] Understanding these biological differences could help identify patients who could benefit from alternative treatment strategies before relapse occurs.

Our findings in *KMT2A*-r leukemia support the existence of distinct evolutionary routes even within the same genetic subtype. In **Article I**, early relapsing *KMT2A*-r ALL frequently acquired alterations in genes associated with treatment response, including *TP53* and *IKZF1*, and showed mutational signatures compatible with chemotherapy exposure. Typically, these relapses were characterized by a clonal sweep, with expansion of a single clone often carrying acquired driver alterations. By contrast, very early ALL relapse had less evidence of acquired genetic evolution and was characterized by multiple diagnostic clones contributing to relapse. These cases appeared more consistent with intrinsic resistance or the rapid expansion of treatment-refractory populations already present at diagnosis. Thus, very early relapse may not simply represent an accelerated version of later relapse, but a biologically distinct evolutionary route.[205]

The relationship between relapse timing and mechanism is unlikely to be absolute. Pre-existing resistant clones can persist for variable periods before expanding, while additional genetic alterations can further shape the relapse clone. Relapse timing therefore reflects the evolutionary history of the leukemia and the selective pressures it has experienced.

Genetics is not enough: cell state, plasticity and the microenvironment

Genetic evolution explains only part of relapse biology. A leukemia cell can change its response to therapy without acquiring new alterations and genetically similar cells can occupy different transcriptional and epigenetic states.[116,308]

Single-cell studies have begun to reveal heterogeneous leukemic populations with features associated with treatment resistance that are difficult to identify through bulk genomic profiling. In infant *KMT2A*-r ALL, relapse-associated populations have been linked to increased cellular plasticity, stem-cell-like transcriptional programs and features associated with treatment resistance.[307] Relapse-initiating populations have also been identified within diagnostic samples, with quiescent or stem-like states, altered glucocorticoid-response programs and other properties that distinguish them from the bulk leukemic population.[310] Single-cell transcriptomic studies have identified a developmentally immature ELP state in infant *KMT2A*-r ALL, particularly in *KMT2A::AFF1*-driven leukemia, which can persist at relapse.[131] These findings show that cellular states associated with treatment resistance vary between leukemia subtypes and can be enriched within specific subpopulations.

RNA-seq in **Article III** further highlighted the importance of cellular state, although the features associated with poor outcome differed by lineage. In ALL, we identified three groups, with G1 displaying a stem-cell-like state and enrichment for *KMT2A::AFF1*. G2 was characterized by a more mature lymphoid state and G3 by a multipotent state with myeloid programs. G2 accounted for 67% of very early relapse cases, all involving *KMT2A::MLLT1*, and partly overlapped with the previously described *KMT2A*-a group.[116] This contrasts with previous studies linking an immature or stem-like transcriptional state to relapse and poor outcome.[307,310] This may reflect biological differences between *KMT2A* fusion partners, as the very early relapses in G2 were driven by *KMT2A::MLLT1*, whereas stem-like relapse biology described in other studies could be more relevant to *KMT2A::AFF1*. Differences in the developmental cell of origin associated with individual fusion partners could also contribute to these transcriptional states. In contrast, the G1 group in AML, which displayed a more stem-cell-like state, had a higher relapse rate than the more mature and myeloid differentiated G2 (60% vs. 28%), although this was not associated with time to relapse. Together, these findings suggest that relapse-associated cellular states differ across *KMT2A*-r leukemia depending on lineage, fusion partner and cellular context. Importantly, a mature bulk phenotype does not exclude the presence of rare stem-like or otherwise

treatment-resistant populations that can preferentially survive therapy and contribute to relapse.

Cellular plasticity is especially relevant to *KMT2A*-r leukemia because lineage identity itself is not fixed. Lineage switching is well documented in *KMT2A*-r disease and can occur under therapeutic pressure, with *KMT2A::AFF1* frequently represented among lineage-switched cases.[302-305] Consistent with this plasticity, G3 in the ALL cohort in **Article III** displayed a multipotent and myeloid transcriptional state and was exclusively composed of *KMT2A::AFF1* cases. This could reflect a cellular state with greater lineage plasticity, although whether this translates into an increased propensity for lineage switching remains unknown.

Rather than being driven by a single mechanism, relapse in infant *KMT2A*-r ALL appears to result from the interplay between genetic evolution and persistent transcriptional and epigenetic programs. While chemotherapy eliminates the majority of leukemic cells, a small population of therapy-resistant cells survives. These cells may retain the *KMT2A*-driven oncogenic transcriptional program and occupy epigenetic and transcriptional states that promote stem cell-like properties and cellular plasticity, allowing them to adapt to therapeutic pressure and ultimately regenerate the leukemia.[351] Altered differentiation therefore represents another potential mechanism of treatment escape. This is important for lineage-directed therapies such as CD19-directed CAR-T cells and blinatumomab, where loss of CD19 expression can allow leukemic cells to escape immune targeting.[302,432,433]

The microenvironment adds another level of selection. Leukemic cells interact continuously with stromal, immune and hematopoietic populations within the BM niche and these interactions can influence survival, quiescence and response to treatment.[63-71] In *KMT2A*-r leukemia, our group has experimentally studied cooperating signalling mutations, showing that subclones can modify the fitness of neighbouring leukemic cells through secreted factors, suggesting that subclonal cooperation contributes to disease behaviour.[216] Similar cooperation between genetically distinct subclones has been demonstrated in solid tumours[434,435] and cytokine-mediated interactions between leukemic and stromal cells have also been described in AML.[436]

These observations argue against a model in which relapse can be understood solely by cataloguing mutations present at diagnosis and recurrence. Genetic alterations, epigenetic state, differentiation, cell-cycle status and environmental interactions collectively determine whether an individual leukemic cell survives treatment. The contribution of each layer is likely to vary between patients and over time.

ALL and AML: a shared initiating lesion, but different routes to relapse

The occurrence of *KMT2A*-r in both ALL and AML provides an opportunity to ask how lineage and treatment context modify evolution from a shared initiating lesion. Although the fusion is central to leukemogenesis in both diseases, the cooperating

genetic landscape, developmental context and therapeutic exposures differ substantially. These differences may shape not only the phenotype at diagnosis, but also the evolutionary route to relapse.

This distinction became more apparent in **Article I**. In *KMT2A*-r ALL, particularly in early relapse, the genetic landscape at recurrence frequently showed evidence of treatment-associated evolution. In contrast, paired *KMT2A*-r AML samples had similar mutational processes at diagnosis and relapse and lacked the clear chemotherapy-associated signatures observed in early relapse ALL. The AML findings were more consistent with the persistence of an early leukemic population during therapy, potentially in a dormant state, followed by its subsequent expansion at relapse. Such persistence could represent a form of non-genetic treatment tolerance, allowing leukemic cells to survive therapy without the marked treatment-associated mutagenesis observed in early relapse ALL (**Figure 4A**).

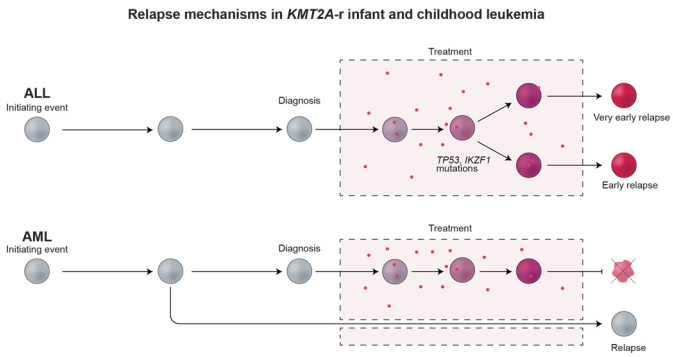


Figure 4A: **Relapse mechanisms in *KMT2A-r* ALL and AML.** Early relapse ALL showed chemotherapy-associated mutagenesis and acquisition of drug-response mutations, particularly in *TP53* and *IKZF1*. In contrast, very early relapse showed little treatment-associated evolution and frequent multiclonal relapse, suggesting inherent resistance. AML similarly lacked chemotherapy-associated mutagenesis, consistent with persistence of potentially dormant leukemic cells through therapy.

ALL and AML are treated with different regimens and differ in cellular hierarchy, disease kinetics and mechanisms of drug response.[139,169,390] ALL treatment includes prolonged maintenance therapy (up to 2 years) to eradicate residual disease and reduce the risk of relapse, whereas AML is treated over a shorter period (5-8 months) with more intensive chemotherapy. ALL treatment relies heavily on glucocorticoids, vincristine and asparaginase, whereas cytarabine forms the backbone of AML chemotherapy.[139,169] Relapse is associated with poor outcome in both diseases, but the prognostic impact of relapse timing differs. In ALL, very early relapse is associated with particularly poor survival and generally requires HSCT in high-risk cases, whereas patients with later relapse are more likely to respond to chemotherapy alone. In AML, the likelihood of achieving durable remission

decreases with each subsequent relapse, although some patients can achieve remission again and proceed to allogeneic HSCT.[390,437] In ALL, there are established relapse protocols incorporating chemotherapy, antibody-based immunotherapy, HSCT and, in selected patients, CAR-T. In AML, relapse treatment is less standardized and is based on intensive reinduction chemotherapy (sometimes with targeted/novel agents) with the aim of proceeding to allogeneic HSCT.

Importantly, these differences in treatment are not only clinically relevant but may also shape relapse evolution. The prolonged and pharmacologically diverse treatment course in ALL creates repeated opportunities for drug-specific selection, whereas persistence of therapy-tolerant or stem-like populations could play a greater role in AML.

From predicting relapse risk to predicting relapse mechanism

Current clinical risk stratification is highly effective at identifying groups of patients with different probabilities of treatment failure, and MRD is among the most powerful predictors of relapse in both ALL and AML.[333-339,346] However, conventional risk assessment gives a relapse probability but less information about how the relapse will occur.

This distinction may become increasingly important. Patients with similar levels of residual leukemia could harbour biologically different populations. In one patient, residual disease could contain a pre-existing drug-resistant clone. In another, the remaining cells could be genetically similar to the diagnostic population but occupy a reversible tolerant state. Some patients with detectable MRD never relapse, whereas relapse can also occur despite undetectable MRD. Another patient could have no detectable resistant population early in therapy but later develop a treatment-selected clone. These scenarios are unlikely to have identical therapeutic implications.

Highly sensitive sequencing offers one approach to resolving this problem. **Article II** identified low frequency subclones that remain invisible to conventional sequencing despite being present at diagnosis, some of which later contributed to relapse. At the same time, simply increasing sequencing depth is unlikely to be sufficient. Rare variants are not necessarily biologically relevant, bulk measurements cannot determine which mutations coexist within the same cell and some relapse mechanisms are not genetic. Thus, increasing analytical sensitivity also increases the challenge of distinguishing detectable heterogeneity from biologically and clinically relevant heterogeneity.

Longitudinal sampling is therefore likely to be more informative than increasingly detailed analysis of a single diagnostic specimen. Serial molecular profiling during treatment can reveal whether resistant populations persist, disappear, emerge or expand. This was also demonstrated in **Article I**.[205] In ALL, molecular disease levels were similar between relapse and remission cases early during induction,

while relapse cases showed a trend towards higher levels at the end of induction, mainly driven by very early relapses. In AML, persistence after the second induction was more strongly associated with subsequent relapse. However, residual leukemic populations were also detected during treatment in patients who did not relapse, indicating that persistence alone does not necessarily imply treatment resistance or impending relapse. Thus, the timing, level and dynamics of residual leukemia can be more informative than its mere detection.

Single-cell and multiomic approaches provide an additional opportunity to connect genotype with cellular state.[307-310,438] Technologies that combine DNA alterations with transcriptomic or chromatin information may help determine whether a minor genetically defined clone also occupies a resistant cell state, whether phenotypically distinct populations share the same ancestry, and whether therapy reshapes differentiation without major genomic change. Longitudinal single-cell profiling of diagnosis and relapse samples could further determine whether rare resistant populations persist through therapy, while combined scRNA and scDNA profiling could directly link genetic subclones to transcriptional states. These approaches are currently more useful for mechanistic discovery than routine clinical decision-making, but they address questions that bulk sequencing cannot resolve.

The clinical challenge is shifting. Predicting who will relapse remains essential, but it may not be sufficient. A more useful long-term goal is to identify which population is surviving treatment, why it survives and which evolutionary route it is beginning to follow. Such information could eventually support intervention before a resistant population becomes the dominant disease and help determine whether different evolutionary routes require different therapeutic strategies. At the same time, a delicate balance must be maintained between intensifying treatment to reduce the risk of relapse and the consequent increase in treatment-related morbidity and mortality.

Together, **Articles I-III** describe different stages of the same evolutionary process. **Article III** defines the genetic and transcriptional heterogeneity present at diagnosis, **Article II** demonstrates that part of this heterogeneity can exist below conventional detection limits, and **Article I** shows how leukemic populations are subsequently selected and reshaped during treatment and relapse.

Conclusions

Relapse is not a single biological event. It is the outcome of an evolutionary process that begins with the cellular and genetic architecture present at diagnosis and continues under the selective pressure imposed by treatment. Pre-existing resistant populations, therapy-associated genetic evolution, persistent cell states and microenvironmental interactions can all contribute, and their relative importance differs between patients.

KMT2A-r leukemia illustrates this diversity particularly clearly. Across our studies, the diagnostic landscape varies with age and lineage, hidden subclonal heterogeneity can be present well below conventional detection limits and the genomic route to relapse differs according to both disease lineage and timing of recurrence. Together, these findings argue against a universal model in which relapse simply reflects expansion of a single resistant diagnostic clone.

The next challenge is to move from retrospective reconstruction of relapse to prospective identification of the processes that precede it. This will require longitudinal sampling and integration of genomic measurements with cellular state, treatment exposure and clinical response. The goal is not merely to detect relapse earlier, but to understand the mechanism of treatment escape while there is still an opportunity to alter its course.

Future perspectives

An important question is whether genomic monitoring can improve relapse prediction beyond current MRD methods. The studies in this thesis show that leukemia evolves during treatment: minor diagnostic clones can disappear, persist or expand, while new alterations can emerge before relapse. However, not all detectable clones are clinically relevant. In **Article I**, genetically detectable clones could persist without subsequent relapse, while in **Article II**, some cancer-associated mutations present at diagnosis were absent at relapse.

For *KMT2A*-r leukemia, this raises the question of what should actually be monitored. Since the *KMT2A* fusion is a founding alteration present across the leukemic population, monitoring the fusion alone may provide a sensitive measure of residual leukemia, but it does not capture changes in the subclonal composition of the disease. The same level of *KMT2A*-positive residual disease could include distinct subclones with different capacities to survive treatment and cause relapse. Conversely, monitoring every diagnostic mutation is unlikely to be informative, as many alterations are lost during treatment.

Future studies should determine which genetic alterations provide prognostic information beyond *KMT2A*-based MRD and whether changes in their clonal abundance can improve relapse prediction. Targeted re-sequencing of recurrently altered genes could also help identify emerging clones associated with treatment resistance or therapeutic vulnerabilities.

As sequencing becomes cheaper and more sensitive, genomic profiling at diagnosis could be used to identify informative patient-specific markers for longitudinal monitoring. The important question is therefore not whether more mutations can be detected, but which alterations are worth following and whether this information can be used to change treatment before clinical relapse.

Clinical impact of this thesis

Together, **Articles I-III** show that *KMT2A*-r leukemia cannot be understood from the presence of the *KMT2A* fusion alone. The disease differs by age, lineage and fusion partner, and evolves during treatment through the selection of genetically distinct subclones. The main contribution of this thesis is therefore to characterize how this heterogeneity relates to treatment response and relapse.

Article I demonstrated that relapse develops through different evolutionary patterns in *KMT2A*-r AML and ALL. Longitudinal sequencing revealed that relapse-associated clones can sometimes be detected during treatment months before overt relapse. The study also found distinct evolutionary patterns in very early and early ALL relapse. These findings support the potential value of longitudinal genomic monitoring as a complement to conventional MRD but also show that detection of a residual clone alone is not sufficient to predict relapse.

Article II revealed that genetic heterogeneity extends below the detection limit of conventional sequencing. Low-frequency cancer-associated subclones were common at diagnosis and followed different trajectories during treatment: some expanded, some remained small and others disappeared. Mutations detected at relapse could either pre-exist at very low levels or arise later during treatment. This highlights that the clinical value of genomic monitoring depends not only on detecting residual leukemia, but also on identifying which clones persist or expand.

Article III showed that *KMT2A*-r leukemia comprises biologically different subgroups. Age, leukemia lineage, fusion partner and cooperating alterations were associated with different genetic and transcriptional profiles. The transcriptional groups were also associated with differences in relapse risk and survival, indicating that classification based primarily on the *KMT2A* fusion partner can overlook clinically relevant heterogeneity. In the future, integrating these additional features may further improve risk stratification.

A potential clinical application of these findings is improved molecular monitoring. Rather than following every mutation detected at diagnosis, future approaches could combine a stable leukemia marker, such as the *KMT2A* rearrangement, with selected alterations that provide additional information about relapse risk, treatment resistance or emerging therapeutic targets. The next step is to determine which markers add clinically useful information beyond conventional MRD and *KMT2A*-based monitoring and to validate them prospectively in independent patient cohorts.

Functional studies are needed to distinguish alterations that actively drive treatment resistance or relapse from those that simply mark an expanding clone. Such studies could identify therapeutic vulnerabilities and help determine whether treatment should be adapted when high-risk clones emerge during therapy.

Limitations of the study

KMT2A-r leukemia is rare and assembling large cohorts therefore takes time and depends on access to well-preserved samples. As treatment protocols continue to evolve, patients included over longer time periods are inevitably treated according to different protocols. This introduces treatment heterogeneity and makes direct comparisons more difficult. In addition, dividing patients according to age, leukemia type and *KMT2A* fusion partner quickly reduces the number of patients in each group. This limits statistical power, makes some findings difficult to generalize and consequently slows their translation into clinical practice. Sample availability also varied between patients, limiting the possibility to perform all genomic and longitudinal analyses across the entire cohort.

Furthermore, many findings are based on retrospective genomic analyses and have not been functionally validated. Although selected alterations and fusion genes were functionally studied, the role of most alterations in treatment resistance, clonal expansion and relapse remains unclear. Further functional studies and validation in independent cohorts are needed to determine their biological and clinical relevance.

AI tools and software for illustrations

Figures 1-12, 16-19, 22, 24, 25, 27-29 and A in this thesis were created using BioRender (www.biorender.com), a tool for creating scientific illustrations.

OpenAI's ChatGPT (GPT-5) was used to improve the clarity and readability of the text. It was only used for linguistic refinement, and it did not contribute to the scientific content, analysis or interpretation of the results. All suggestions were critically reviewed and revised by the author to ensure accuracy and preserve the intended scientific meaning.

Populärvetenskaplig sammanfattning

Livet, trots all sin komplexitet, bygger på en av biologins mest fundamentala processer: delningen av en cell. Från den tidiga embryogenesen till den dagliga förnyelsen av våra vävnader förlitar sig kroppen på ett svindlande antal celldelningar. Varje celldelning kräver en närmast ofattbar precision: omkring sex miljarder baspar och över 20 000 gener ska kopieras. Trots att cellen har flera sofistikerade sätt att förhindra fel är systemen inte helt perfekta. Fel smyger sig in ändå. Många upptäcks och repareras, och av de som blir kvar är de flesta harmlösa. Andra består och kan ackumuleras över tid. Denna genetiska variation ska inte ses som en brist, utan är en förutsättning för evolutionen och därmed för livet självt. Men när förändringar uppstår i nyckelgener som reglerar cellens mognad, tillväxt, död och stabilitet kan de ge cellen en överlevnads- eller tillväxtfördel och i slutändan bidra till okontrollerad celldelning och cancer.

Ordet *cancer* är fruktat världen över och de flesta av oss har blivit berörda. Cancer ska inte ses som en enskild sjukdom, utan som en bred sjukdomsgrupp som karaktäriseras av ohämmad celltillväxt och en förmåga hos onormala celler att överleva och, i många cancerformer, sprida sig i kroppen. Cancer tar flest liv efter hjärtkärlsjukdomar och antalet cancerfall förväntas alarmerande nog att stiga mer än 50 % de kommande decennierna. Trots många år av forskning och enorma medicinska framsteg är det fortfarande mycket i cancerns biologi som vi ännu inte helt förstår.

En typ av cancer är *leukemi*, vilket är den vanligaste cancerformen hos barn. Här uppstår cancern i de blodbildande cellerna i benmärgen. Cellerna förlorar sin normala förmåga att mogna till fungerande blodceller och börjar i stället ansamlas. De kringgår kroppens säkerhetsmekanismer som normalt reglerar celldelning och överlevnad och tränger gradvis undan den friska blodbildningen. Leukemicellerna finns framför allt i benmärgen men kan också spridas ut i blodet. Ansamlingen av dessa cancerceller, *blaster*, leder till brist på normala blodceller som också produceras i benmärgen. Brist på röda blodkroppar leder till *anemi*, med påföljande trötthet och andfåddhet. Brist på fungerande vita blodkroppar ökar risken för infektioner, medan brist på blodplättar ökar risken för blödningar. Vid diagnos kan benmärgen vara helt dominerad av sjuka blastceller och stora mängder blaster kan även finnas i blodet. Det är därifrån som sjukdomen fått sitt namn, leukemi, *vitt blod*.

Beroende på vilken blodbildande cellinje som har omvandlats till leukemiceller brukar man traditionellt dela in akut leukemi i akut lymfatisk leukemi (ALL) och akut myeloid leukemi (AML), där ALL är vanligast hos barn och AML är vanligare hos vuxna.

Leukemi hos barn var länge ett mysterium. Symptomen var ospecifika men allvarliga. Utan effektiv behandling återstod ofta bara veckor eller månader innan barnet avled av infektion, blödning eller svår benmärgssvikt. Läkare, forskare och föräldrar stod maktlösa. Diagnosen var i praktiken en dödsdom. Men så på 1940-talet kom äntligen ett oväntat genombrott. Patologen Sidney Farber började experimentera med ett ämne som blockerade folsyrans funktion, folsyra-antagonisten aminopterin. För första gången kunde man se hur blastantalet sjönk. Barn gick i remission. Alla höll andan. Tyvärr var glädjen tillfällig. Blasterna började åter dyka upp i blodet och kunde påvisas i blodprover. Men det hade väckts en gnista av hopp.

Under decennierna som följde växte förståelsen för att leukemi drivs av genetiska förändringar i leukemicellerna. Det kan handla om små genetiska fel på basparsnivå, mutationer, men också större förändringar på kromosomnivå, såsom kopietals-förändringar och translokationer. Nya läkemedel introducerades och kombinerades. Man fann att leukemicellerna inte bara fanns i benmärg och blod, utan även kunde passera blodhjärnbarriären och gömma sig i det centrala nervsystemet. Där skyddades de från cytostatika som inte kunde nå dem. Behandlingen förbättrades steg för steg. Bättre understödjande vård, inklusive antibiotika och transfusionsstöd, kombinerades med behandling riktad mot det centrala nervsystemet, intensiv kombinationsbehandling och långvarig underhållsbehandling. Allt fler barn kunde därmed stanna i remission. I början av 1980-talet växte det nordiska samarbetet fram som 1984 formaliserades som Nordic Society for Paediatric Haematology and Oncology (NOPHO), med målet att genom gemensamma behandlingsprotokoll och populationsbaserade data förbättra vården för barn med blodcancer. Resultaten är häpnadsväckande. Idag överlever mer än 90 % av barn med ALL.

Trots detta finns det fortfarande undergrupper av akut leukemi som är betydligt svårare att behandla och där risken för återfall är hög. En av dessa är *KMT2A*-rearrangerad (*KMT2A*-r) leukemi. Det innebär att kromosom 11 har brutits i en gen som heter *KMT2A* och att den brutna genen har kopplats samman med en annan gen, ofta från en annan kromosom. Resultatet blir en så kallad *fusionsgen*. Förändringen finns i leukemicellerna, men inte i kroppens övriga celler, och när leukemicellerna försvinner vid behandling försvinner också förändringen. Denna typ av genetisk förändring kallas ett *kromosomalt rearrangemang* och kan till exempel uppstå genom en *translokation*, där delar av olika kromosomer byter plats.

KMT2A-rearrangemang återfinns i upp till omkring 10 % av akuta leukemier, men är särskilt vanliga hos spädbarn. Hos spädbarn med ALL bär ungefär tre av fyra

patienter på ett *KMT2A*-rearrangemang. Trots att många patienter kan nå remission är risken för återfall hög, framför allt hos de allra yngsta barnen. Det finns starka belägg för att det genetiska felet i många fall uppkommer redan under fosterlivet. *KMT2A*-rearrangemang har kunnat spåras tillbaka till prover tagna vid födseln och identiska förändringar hos enäggstvillingar har gett ytterligare stöd för ett prenatalt ursprung. Vad som gör att translokationen uppstår är inte klarlagt.

Denna avhandling är ett försök att kasta ytterligare ljus över hur denna komplexa leukemi uppstår, förändras och undkommer behandling. Med hjälp av olika tekniker har vi studerat det genetiska landskapet vid diagnos, under behandling och vid återfall.

I **Artikel I** såg vi att leukemin kan överleva behandlingen på olika sätt. Hos patienter som fick återfall något senare såg vi att leukemicellerna hade förvärvat mutationer som gjorde dem mer motståndskraftiga mot behandlingen. Hos dem som fick återfall mycket tidigt verkade sjukdomen i stället bestå av flera små grupper av leukemiceller som tillsammans kunde överleva behandlingen. Resultaten visar att återfall kan uppstå genom olika mekanismer och att tidpunkten för återfallet kan ge ledtrådar om hur leukemin har utvecklats och undkommit behandlingen. Genom att följa flera patienter under hela sjukdomsförloppet kunde vi dessutom se hur leukemin förändrades över tid och identifiera genetiska förändringar som kan bidra till återfall och som också kan användas som markörer för att följa små mängder kvarvarande leukemiceller under behandlingen.

I **Artikel II** använde vi mycket känsliga sekvenseringstekniker för att upptäcka mutationer i en så liten andel av leukemicellerna att de inte kan ses med vanliga sekvenseringsmetoder. På så sätt visade vi att leukemins genetiska variation är betydligt större än vad den först verkar vara. Genom att följa dessa små cellkloner över tid kunde vi se att vissa överlevde behandlingen och senare växte och blev den dominerande gruppen av leukemiceller vid återfall. Det visar hur en nästan osynlig del av leukemin vid diagnos senare kan få avgörande betydelse för sjukdomens utveckling. Vi såg också en mutation vid återfall som inte kunde upptäckas vid diagnos och som därför antingen hade tillkommit under behandlingen eller funnits i så få celler att den inte gick att upptäcka.

I **Artikel III** studerade vi en av de hittills största genomiskt kartlagda kohorterna av *KMT2A*-r leukemi. Vi undersökte 116 patienter i olika åldrar, med olika leukemityper (ALL och AML) och med olika *KMT2A*-fusionspartner. Vi såg att leukemin förändrades genetiskt med åldern och att vuxna patienter hade fler genetiska förändringar än spädbarn och barn. Vi såg också att olika genetiska förändringar hängde samman med vilken typ av leukemi och vilken *KMT2A*-translokation patienterna hade. Dessutom fann vi många olika förändringar både i enskilda gener och på kromosomnivå, inklusive flera tidigare okända förändringar. När vi studerade vilka gener som var aktiva såg vi dessutom olika biologiska

mönster som speglade hur mogna leukemicellerna var och som var kopplade till patienternas utfall. Sammantaget visar resultaten att *KMT2A*-r leukemi inte är en enda sjukdom, utan en genetiskt heterogen grupp vars biologiska egenskaper skiljer sig beroende på bland annat patientens ålder, leukemityp och fusionspartner.

KMT2A-r leukemi fortsätter att vara en svårbehandlad form av leukemi, där återfall fortfarande är en stor utmaning. Förhoppningsvis kan denna avhandling bidra till en djupare förståelse av hur leukemin utvecklas och överlever behandling och på sikt bidra till bättre sätt att följa sjukdomen, bedöma risken för återfall och utveckla mer effektiva behandlingar. Målet är inte längre bara att få leukemin att försvinna, utan att förstå hur vi kan se till att den aldrig kommer tillbaka.

Acknowledgements

This thesis is finally finished. Getting here has been challenging, sometimes frustrating and certainly time-consuming, but above all, far more memorable than I could ever have hoped for. The fact that these years turned into such a positive part of my life is thanks to the people around me.

First of all, a big thank you to my supervisor **Anna Hagström**. You have taught me an incredible amount and always managed to answer my questions within an hour, regardless of time or day. Our discussions have reached far beyond science, covering the future, choices to be made, and how to make the most of them. Your knowledge seems endless, and I am constantly impressed by how easily you move between the smallest details and the bigger picture. Thanks to you, these years have passed in the blink of an eye.

My co-supervisor **Thoas Fioretos**, despite your extremely busy schedule, you have always set aside time for me to talk about research, career choices or future plans. I admire how you think big without losing sight of the patient, and how you manage to have so many things going on at once. Most importantly, you have opened doors for me that I would never have been able to open on my own. For that, I am truly grateful.

My co-supervisor **Anders Castor**, thank you for bringing a clinical perspective to our manuscript discussions and raising questions that we researchers sometimes overlook. I wish we had had more opportunities to work together.

Mattias Pilheden, you have been invaluable throughout this thesis. Thank you for your contagious enthusiasm and for always making time for me when I need your help. We have discussed everything from science and careers to relationships, people, life and complete nonsense, and I cannot remember ever leaving one of our conversations without a few new thoughts spinning around in my head. You have a rare ability to make me question things I thought I had already figured out. Thank you for all the discussions, all the laughs, and for being my oasis of ventilation in return.

Helena Sturesson, you are the linchpin of the lab and the person who knows where absolutely everything is. You had the patience to take me on when I had never even held a pipette before. You taught me the lab from scratch, and I could not have wished for a better start. Since then, people have come and gone, but you have

remained the person to turn to when something needs to be found, fixed, ordered, remembered or done properly. Thank you for all your help over the years.

Ton Falques, thank you for helping me tie up the loose ends with the final paper. You always take the time to ask how things are going, and somehow you never forget my birthday. Your eye for detail has rescued many manuscript figures, and your skills behind the camera are equally impressive. I am very grateful for all your help, especially during this final stretch.

Khrystyna Hafiichuk, thank you for being so positive and for helping me with all the strange requests I keep throwing at you. You are our little office jukebox, a devoted Lego builder and, somewhat involuntarily, a plant collector. Your little notes always make me smile. I wish you the best of luck with your PhD journey. I am completely convinced that you will rock it.

Mari Lahnalampi, thank you for bringing your Finnish cheerfulness to the group and for being the first to suggest doing something outside the lab. I have enjoyed getting to know you and hope we will find more time to spend together.

Anna Bergman, you left us (me!) far too soon for Stockholm and the lab was not the same without you. You are a fantastic colleague and an even better person, and you made even the most ordinary working days so much more fun. I am very happy our paths crossed, and moving to Stockholm will not get rid of me that easily.

Varsha Singh, thank you for being so kind and cheerful. I really enjoyed your curiosity and willingness to learn new things. It was great having you in the lab, and I wish you and your family all the best in the US.

Qirui Zhang, thank you for always bringing a smile to the lab and for sharing your computational expertise. It was great having you around.

Minjun Yang, thank you for stepping into the bioinformatics void left by Mattias. No matter how busy you are, you always make me feel welcome to drop by your office with questions, big or small, and find the time to help. Thank you for your patience, for untangling my computational problems, and for explaining things until they actually make sense.

I want to thank **The Division of Clinical Genetics**. I am so proud to be part of this team and thankful to all the people I have shared these years with. First, **Felix Mitelman** and **Nils Mandahl**, thank you for your never-ending curiosity and enthusiasm for science, and for always stopping to say hello and get an update in the corridors. And to all PIs: **Marcus Järås, Kajsa Paulsson, David Gisselsson, Karolin Hansén Nord, Bertil Johansson, Fredrik Mertens, Axel Hyrenius-Wittsten** and **Niklas Landberg,** thank you for carrying that spirit forward and for continuing to fill the division with great people. An extra shoutout to **Anette Welin**,

for always being so helpful and for keeping the administrative and financial side of things running smoothly.

And finally, to all the people who have filled the corridors, offices, labs and lunchroom over these years: **Gustav, Jakob H, Tina, Josephine, Katrin, Marianne, Ram, Somadri, Charlotte, Efe, Eleanor, Larissa, Andrea, Stina, Rebeqa, Jenny N, Karim, Linda M, Valeria, Jan, Saskia, Bahar, Caroline, Jenny K, Linda HM, Kaname, Natalie, Subhayan, Jakob HS, Hannah, Carl S, Carl H, Vendela, Gabriel, Mia, Sofia, Rasmus, Maria, Giulia, Anders W, Henrik, Pablo, Helena Å, Kristina, Ylva, Christina, Ludvig, Petr, Lydia, Romain, Adél, Adrian, Ahmad, Julian, Solène, Fanny, Noelia, Hanna** and **Carl N**. Thank you for all the discussions, lunches, after-works, laughs and everything in between. You are what has made this such a fun place to work.

Niklas Landberg, thank you for taking the time to answer my questions. I always feel less stressed after talking to you, perhaps because you have at least as many balls in the air as I do (usually more) and still seem remarkably unfazed by it. **Axel Hyrenius-Wittsten,** thank you for paving the way for those of us who came after you in the Hagström group, and for all the useful tips on how to survive a move across the Atlantic. **Marcus Järås,** thank you for giving me your recommendations to several groups in Boston. I am so grateful for this, and I learnt a lot from meeting all those people thanks to you. **Kajsa Paulsson,** thank you for making my half-time seminar a positive experience, even though you asked the trickiest questions. You are one of those rare people that one immediately likes. **Bertil Johansson**, thank you for always taking the time to stop and talk to me in the corridor and ask what my current future plans are.

Vendela Rissler, for always knocking on my office door when it's time for lunch and fika. Even though I always come embarrassingly late because I am stuck in analysis, you never give up. **Marianne Rissler,** for always remembering not only my birthday, but also my oh so important name day. You ask me questions in a way that makes me feel special and I always grow an inch after speaking with you. **Carl Sandén,** for making all conferences and lunch breaks so much more fun, and for always saying "YES" to all of Hanna's and my sometimes quite questionable ideas.

Fanny Asbaghi, for making the office atmosphere much livelier and more unpredictable. You are a bundle of energy, and I am very glad Umeå lent you to us, and I hope we can convince you to stay. **Saskia Sydow,** for being my solid German rock and for showing me how to navigate the chaos when dissertation deadlines start closing in. **Charlotte Ragnarsson**, thank you for answering my clinical questions and for taking the time to truly give me your best answer. You make me want to consider working as a pediatrician just so I can continue sharing an office with you. **Karim Saba**, for your loyal service to the Party Pixies and for making sure I never

fall behind on the latest Valeria news. **Bahar Rastegar,** for being such a warm and vibrant person. I really miss having you around the lunch table.

To our brilliant technical staff, **Tina, Caroline, Katrin, Jenny N, Linda M, Andrea, Stina, Linda HM, Mia, Marianne** and **Helena S.** You keep everything running, everything in its place, and the rest of us in order. On top of that, you make the lab a very nice place to work.

Valeria Difillippo, in my family we call you the Ferrari, because you talk incredibly fast, look smashing and fulfil just about every Italian stereotype. I am so happy I got to know you. Thank you for all the laughs, talks and incredible food over the years. Moving back to Italy does not mean you get to escape me, and you know it.

Natalie Andersson, you are a born researcher. I am so thankful for all the help and all the laughs we have shared over the years. From medical school to Clinical Genetics, I am very happy that our paths crossed twice. Convincing you to do a PhD was clearly one of the better ideas Hanna and I have had.

Noelia Puente Moncada, every conference with you has also turned into a culinary experiment. Thank you for all the trips and laughs, and for never saying no to trying something slightly questionable with me. I still think leaving Sweden was a poor decision, but there is plenty of time to reconsider.

Now, moving outside the research bubble, to all of you who did not have so much to do with this thesis, but a great deal to do with everything else.

My fantastic friends. **Annika Kranefeld** was für Abenteuer wir schon zusammen erlebt haben! Du bist ein unglaublich toller Mensch und jemand, auf den ich mich wirklich immer verlassen kann. Eine wahre Freundin, und ich bin so froh, dich zu haben. **Alma Dahlberg,** thank you for always answering, always agreeing with me and always being at least as stressed as I am. There is something oddly reassuring about that. **Karin Markendahl,** we found each other in a millisecond at Sweden's smallest amusement park. You always have an opinion, are never afraid to voice it and go your own way without much fuss. I am very glad that our paths crossed. **Johanna Forsvall,** someone to hold on to when the wind picks up, and whose expressions and dance moves I have secretly started to copy. Still not quite sure they look as cool on me. **Margit Vilhelmsson,** I admire how you always follow your own path and go wholeheartedly for what you want, something I truly look up to. **Lovisa Terling,** always positive, always with an opinion and always there. **Anna Székely,** for luring me into Studentafton and opening a whole new world to me, for your honesty and conversations that always leave me with something new to ponder. **Nina Alian** and **Malin Jakobsson**, for making Studentafton such a special time, and for keeping me up to date on the world through our sparkling wine club discussion. **Lovisa Tengnäs,** you bring such good energy wherever you go and are one of those people I always want around. **Ebba Nord** and **Emelie Edvardsson**,

for finding our way back to each other and making it feel as though no time had passed at all. **Fabienne Adriaanse,** for always being so positive, helpful and understanding, for welcoming me into your home, and for making my time in Utrecht unforgettable. My dearest Karlskrona friends, **Kristina Pålsson, Kajsa Witthuhn, Sara Strand, Therese Werkström** and **Ellen Thörneborn**, you were the first group of friends where I truly felt at home and completely myself. We have shared so many unforgettable moments over the years, and I am so grateful that we still have each other.

Moa, Klara, Heléne and **Björn Eklund** and **Frida, William, Catarina** and **Göran Wallin** aka *Barnen i Bullerbyn*. Thank you for all the laughs, chaos and countless hours we spent together growing up. At some point you stopped being family friends and became family.

My *Simrishamn family*: **Hans** and **Birgitta Andersson** for always making me feel that I have a place in your hearts, just as you have in mine. **Anders Hassel Jönsson, Marie, Elvira** and **Inez Hassel.** We have shared both the very best and the very hardest of times, and you have become family in every sense of the word. I am also very happy that this family has branched out to include **Joachim Knutsson**, **Gustaf Sedin** and **Maria Ericsson**. Thank you for always being there. **Marie**, you are deeply missed and will always be part of us.

Lisa and **Charles Walther** for convincing me that a PhD was one of the most fun things you could do and that Clinical Genetics was the place to do it. Charles, I remember sceptically thinking you must have lived a rather boring life. Turns out, you were right and I was wrong.

Hanna Thorsson, I am so incredibly happy that we ended up together at Clinical Genetics. Knowing you has changed me in more ways than I think you realise. Being your friend is not always easy, and I mean that as the greatest compliment. You challenge me, expect things from me and make me grow. You have made me braver, and with you by my side, I sometimes feel almost invincible. Few people make me laugh as much as you do. Pink champagne from Steins forever! **André Schoug,** you are simply so easy to be around: calm, reassuring, but also unexpectedly sly and mischievous. I have to agree with Hanna: you and I are the same person. And thank you both for **Harry Schoug,** of whom I am a very proud godmother. This officially makes you all fall into the family section.

Jan and **Eva Bergfeldt**, thank you for being like an extra pair of grandparents, for always cheering me on, and for creating the fantastic cover of this thesis. I cannot think of two better people to have put the finishing touch on all these years of work.

Sören Ahlgren, I could not have asked for a cooler grandparent. With your motorcycles, earring, and thirst for life (and wine), you are one of the people I look up to most. I have always admired your calm and logical way of approaching things,

and your knack for finding solutions where the rest of us see dead ends. You have made us question things, take less obvious routes, and look at the world from new angles. I can only hope to grow older with the same curiosity and appetite for life as you.

Stefanie Ahlgren, you are one strong-minded lady, and I have often asked myself, "What would mormor say?" when faced with a tricky question. I do wonder what happened to all those opinionated genes, and how on earth none of them made it to me. You have been my proofreader for as long as I can remember, but I decided to spare you this one.

Jan Johansson, it still strikes me every day that research and science would have suited you so well. Instead, you chose a clinical path, which of course is very lucky for all heartbroken Karlskronites. You never give up on a scientific problem, and I still get chills thinking of when you dismantled my physics homework down to its very atoms, only to patiently piece it all back together again until it finally made sense.

Madelaine Ahlgren, thank you for always being there, for the big things and the very small things (that somehow become big). You have an endless capacity to listen and to really understand whatever is in front of you. You never just give an answer, you think, turn things around, and make everything easier to understand and handle. You are incredibly competent, yet remarkably humble. You have a beautiful mind, and I am so happy that you let me share it.

Wilhelm Ahlgren, you are the most mindful minion, and I could not have asked for a better younger brother. You are my best friend, and although you have chosen an entirely different path, you never fail to give me brilliant advice on life, work and relationships. I am convinced that you are destined for something big, but to me, you know that you will always be little.

Erik Rask, my puzzle piece. No one can dive into a quest with quite the same ferocity as you, whether it is tracking down the optimal paper type for our wedding stationery, finding the perfect boat for the hypothetical jetty in our garden in Lund, or going through my Kappa with almost unreasonable thoroughness. I trust you completely, because no matter what comes our way, you somehow make everything work out, from fixing a broken dishwasher and piloting me to Bornholm to building our wonderful patio and navigating the tedious process of our US visa applications. You also make us embarrassingly late to everything, but somehow always charm everyone into forgiving us. Fortunately, you have developed an ECG device, so at least we can keep track of what all these adventures are doing to my heart rate. You challenge my mind every day, and I am so proud to call you my husband and so happy to have you by my side. A life with you will be nothing short of an adventure, and I look forward to every part of it.

References

1. Alberts, B.H., R.; Johnson, A.; Morgan, D.; Raff, M.; Roberts, K. & Walter, P. *Molecular Biology of the Cell Seventh Edition*, (W.W. Norton & Company, New York, 2022).
2. Consortium, I.T.P.-C.A.o.W.G. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
3. Bjork-Eriksson, T. *et al.* Mortality Among Pediatric Patients With Acute Lymphoblastic Leukemia in Sweden From 1988 to 2017. *JAMA Netw Open* **5**, e2243857 (2022).
4. Jassim, A., Rahrmann, E.P., Simons, B.D. & Gilbertson, R.J. Cancers make their own luck: theories of cancer origins. *Nat Rev Cancer* **23**, 710–724 (2023).
5. Heim, S.M.F. *Cancer Cytogenetics Fourth Edition*, (Wiley-Blackwell, 2009).
6. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31–46 (2022).
7. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–74 (2011).
8. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
9. Ferrando, A.A. & Lopez-Otin, C. Clonal evolution in leukemia. *Nat Med* **23**, 1135–1145 (2017).
10. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–58 (2013).
11. Darwin, C. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*, (John Murray, London, 1859).
12. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029–1041 e21 (2017).
13. Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **194**, 23–8 (1976).
14. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
15. Deng, N., Zhou, H., Fan, H. & Yuan, Y. Single nucleotide polymorphisms and cancer susceptibility. *Oncotarget* **8**, 110635–110649 (2017).
16. Supek, F., Minana, B., Valcarcel, J., Gabaldon, T. & Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324–1335 (2014).
17. Jung, H., Lee, K.S. & Choi, J.K. Comprehensive characterisation of intronic mis-splicing mutations in human cancers. *Oncogene* **40**, 1347–1361 (2021).

18. Elliott, K. & Larsson, E. Non-coding driver mutations in human cancer. *Nat Rev Cancer* **21**, 500–509 (2021).
19. Collins, R.L. & Talkowski, M.E. Diversity and consequences of structural variation in the human genome. *Nat Rev Genet* **26**, 443–462 (2025).
20. Liu, S.V., Nagasaka, M., Atz, J., Solca, F. & Mullauer, L. Oncogenic gene fusions in cancer: from biology to therapy. *Signal Transduct Target Ther* **10**, 111 (2025).
21. Mertens, F., Johansson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* **15**, 371–81 (2015).
22. Annala, M.J., Parker, B.C., Zhang, W. & Nykter, M. Fusion genes and their discovery using high throughput sequencing. *Cancer Lett* **340**, 192–200 (2013).
23. Meyer, C. *et al.* Spliced MLL fusions: a novel mechanism to generate functional chimeric MLL-MLLT1 transcripts in t(11;19)(q23;p13.3) leukemia. *Leukemia* **21**, 588–90 (2007).
24. Mani, R.S. & Chinnaiyan, A.M. Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nat Rev Genet* **11**, 819–29 (2010).
25. Chunduri, N.K. & Storchova, Z. The diverse consequences of aneuploidy. *Nat Cell Biol* **21**, 54–62 (2019).
26. Lee, J.E. & Kim, M.Y. Cancer epigenetics: Past, present and future. *Semin Cancer Biol* **83**, 4–14 (2022).
27. Dai, W. *et al.* Epigenetics-targeted drugs: current paradigms and future challenges. *Signal Transduct Target Ther* **9**, 332 (2024).
28. Shendure, J. & Akey, J.M. The origins, determinants, and consequences of human mutations. *Science* **349**, 1478–83 (2015).
29. Schuster-Bockler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–7 (2012).
30. Yang, J.H. & Hansen, A.S. Enhancer selectivity in space and time: from enhancer-promoter interactions to promoter activation. *Nat Rev Mol Cell Biol* **25**, 574–591 (2024).
31. Pang, B., van Weerd, J.H., Hamoen, F.L. & Snyder, M.P. Identification of non-coding silencer elements and their regulation of gene expression. *Nat Rev Mol Cell Biol* **24**, 383–395 (2023).
32. Nowell, P.C. The minute chromosome (Phl) in chronic granulocytic leukemia. *Blut* **8**, 65–6 (1962).
33. Caspersson, T., Hulten, M., Lindsten, J. & Zech, L. Distinction between extra G-like chromosomes by quinacrine mustard fluorescence analysis. *Exp Cell Res* **63**, 240–3 (1970).
34. Rowley, J.D. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–3 (1973).
35. Heisterkamp, N., Stam, K., Groffen, J., de Klein, A. & Grosveld, G. Structural organization of the bcr gene and its role in the Ph' translocation. *Nature* **315**, 758–61 (1985).

36. Shtivelman, E., Lifshitz, B., Gale, R.P. & Canaani, E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* **315**, 550–4 (1985).
37. Salesse, S. & Verfaillie, C.M. Mechanisms underlying abnormal trafficking and expansion of malignant progenitors in CML: BCR/ABL-induced defects in integrin function in CML. *Oncogene* **21**, 8605–11 (2002).
38. Melo, J.V. & Deininger, M.W. Biology of chronic myelogenous leukemia--signaling pathways of initiation and transformation. *Hematol Oncol Clin North Am* **18**, 545–68, vii–viii (2004).
39. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463–7 (1977).
40. Langer-Safer, P.R., Levine, M. & Ward, D.C. Immunological method for mapping genes on Drosophila polytene chromosomes. *Proc Natl Acad Sci U S A* **79**, 4381–5 (1982).
41. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–70 (1995).
42. Matsuzaki, H. *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* **1**, 109–11 (2004).
43. Cancer Genome Atlas Research, N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113–20 (2013).
44. Corces, M.R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**(2018).
45. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377–82 (2009).
46. Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K. & Mardis, E.R. The next-generation sequencing revolution and its impact on genomics. *Cell* **155**, 27–38 (2013).
47. Pevsner, J. *Bioinformatics and Functional Genomics Third Edition*, (Wiley Blackwell, Chichester, 2015).
48. Sender, R. & Milo, R. The distribution of cellular turnover in the human body. *Nat Med* **27**, 45–48 (2021).
49. Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016).
50. J. Hodler, C.L.Z.G.K.v.S. *Musculoskeletal Diseases 2009-2012*, (Springer, New York, 2009).
51. Till, J.E. & Mc, C.E. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res* **14**, 213–22 (1961).
52. Till, J.E. & McCulloch, E.A. Hemopoietic stem cell differentiation. *Biochim Biophys Acta* **605**, 431–59 (1980).
53. Greaves, M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat Rev Cancer* **18**, 471–484 (2018).
54. Takizawa, H., Boettcher, S. & Manz, M.G. Demand-adapted regulation of early hematopoiesis in infection and inflammation. *Blood* **119**, 2991–3002 (2012).

55. Cullen, S.M., Mayle, A., Rossi, L. & Goodell, M.A. Hematopoietic stem cell development: an epigenetic journey. *Curr Top Dev Biol* **107**, 39–75 (2014).
56. Laurenti, E. & Gottgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).
57. Haas, S., Trumpp, A. & Milsom, M.D. Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell* **22**, 627–638 (2018).
58. Pietras, E.M. *et al.* Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* **17**, 35–46 (2015).
59. Carrelha, J. *et al.* Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature* **554**, 106–111 (2018).
60. Rodriguez-Fraticelli, A.E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).
61. Sanjuan-Pla, A. *et al.* Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–6 (2013).
62. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271–281 (2017).
63. Hofmann, J. & Kokkaliaris, K.D. Bone marrow niches for hematopoietic stem cells: life span dynamics and adaptation to acute stress. *Blood* **144**, 21–34 (2024).
64. Morrison, S.J. & Scadden, D.T. The bone marrow niche for haematopoietic stem cells. *Nature* **505**, 327–34 (2014).
65. Wilson, A. & Trumpp, A. Bone-marrow haematopoietic-stem-cell niches. *Nat Rev Immunol* **6**, 93–106 (2006).
66. Mian, S.A., Ngo, S. & Bonnet, D. Is the bone marrow microenvironment the hidden catalyst in malignant haematopoiesis? *Leukemia* **39**, 1589–1592 (2025).
67. Rieger, M.A., Hoppe, P.S., Smejkal, B.M., Eitelhuber, A.C. & Schroeder, T. Hematopoietic cytokines can instruct lineage choice. *Science* **325**, 217–8 (2009).
68. Pinho, S. & Frenette, P.S. Haematopoietic stem cell activity and interactions with the niche. *Nat Rev Mol Cell Biol* **20**, 303–320 (2019).
69. Takeishi, S. *et al.* Haematopoietic stem cell number is not solely defined by niche availability. *Nature* **646**, 687–696 (2025).
70. Tikhonova, A.N. *et al.* The bone marrow microenvironment at single-cell resolution. *Nature* **569**, 222–228 (2019).
71. Bandyopadhyay, S. *et al.* Mapping the cellular biogeography of human bone marrow niches using single-cell transcriptomics and proteomic imaging. *Cell* **187**, 3120–3140 e29 (2024).
72. Kragesteen, B.K. *et al.* The transcriptional and regulatory identity of erythropoietin producing cells. *Nat Med* **29**, 1191–1200 (2023).
73. Grover, A. *et al.* Erythropoietin guides multipotent hematopoietic progenitor cells toward an erythroid fate. *J Exp Med* **211**, 181–8 (2014).
74. Qian, H. *et al.* Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. *Cell Stem Cell* **1**, 671–84 (2007).

75. Kaushansky, K. Lineage-specific hematopoietic growth factors. *N Engl J Med* **354**, 2034–45 (2006).
76. Laurenti, E. *et al.* The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat Immunol* **14**, 756–63 (2013).
77. Enver, T., Pera, M., Peterson, C. & Andrews, P.W. Stem cell states, fates, and the rules of attraction. *Cell Stem Cell* **4**, 387–97 (2009).
78. Rekhtman, N., Radparvar, F., Evans, T. & Skoultchi, A.I. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev* **13**, 1398–411 (1999).
79. Hoppe, P.S. *et al.* Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature* **535**, 299–302 (2016).
80. Burda, P., Laslo, P. & Stopka, T. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia* **24**, 1249–57 (2010).
81. Pundhir, S. *et al.* Enhancer and Transcription Factor Dynamics during Myeloid Differentiation Reveal an Early Differentiation Block in Cebpa null Progenitors. *Cell Rep* **23**, 2744–2757 (2018).
82. Zhang, P. *et al.* Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. *Immunity* **21**, 853–63 (2004).
83. Lee, R.D. *et al.* Single-cell analysis identifies dynamic gene expression networks that govern B cell development and transformation. *Nat Commun* **12**, 6843 (2021).
84. Hosokawa, H. & Rothenberg, E.V. How transcription factors drive choice of the T cell fate. *Nat Rev Immunol* **21**, 162–176 (2021).
85. Rosenbauer, F. & Tenen, D.G. Transcription factors in myeloid development: balancing differentiation with transformation. *Nat Rev Immunol* **7**, 105–17 (2007).
86. Zhao, A., Zhou, H., Yang, J., Li, M. & Niu, T. Epigenetic regulation in hematopoiesis and its implications in the targeted therapy of hematologic malignancies. *Signal Transduct Target Ther* **8**, 71 (2023).
87. Luis, T.C. *et al.* Canonical wnt signaling regulates hematopoiesis in a dosage-dependent fashion. *Cell Stem Cell* **9**, 345–56 (2011).
88. Blank, U. & Karlsson, S. The role of Smad signaling in hematopoiesis and translational hematology. *Leukemia* **25**, 1379–88 (2011).
89. Mar, B.G., Amakye, D., Aifantis, I. & Buonamici, S. The controversial role of the Hedgehog pathway in normal and malignant hematopoiesis. *Leukemia* **25**, 1665–73 (2011).
90. Gao, J. *et al.* Hedgehog signaling is dispensable for adult hematopoietic stem cell function. *Cell Stem Cell* **4**, 548–58 (2009).
91. Alawieh, D., Cysique-Foinlan, L., Willekens, C. & Renneville, A. RAS mutations in myeloid malignancies: revisiting old questions with novel insights and therapeutic perspectives. *Blood Cancer J* **14**, 72 (2024).

92. Luis, T.C., Killmann, N.M. & Staal, F.J. Signal transduction pathways regulating hematopoietic stem cell biology: introduction to a series of Spotlight Reviews. *Leukemia* **26**, 86–90 (2012).
93. Blank, U., Karlsson, G. & Karlsson, S. Signaling pathways governing stem-cell fate. *Blood* **111**, 492–503 (2008).
94. Zhang, N. *et al.* Global burden of hematologic malignancies and evolution patterns over the past 30 years. *Blood Cancer J* **13**, 82 (2023).
95. Arber, D.A. *et al.* International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood* **140**, 1200–1228 (2022).
96. Alaggio, R. *et al.* The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia* **36**, 1720–1748 (2022).
97. Khoury, J.D. *et al.* The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia* **36**, 1703–1719 (2022).
98. Short, N.J., Rytting, M.E. & Cortes, J.E. Acute myeloid leukaemia. *Lancet* **392**, 593–606 (2018).
99. Pui, C.H., Robison, L.L. & Look, A.T. Acute lymphoblastic leukaemia. *Lancet* **371**, 1030–43 (2008).
100. Alexander, T.B. *et al.* The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature* **562**, 373–379 (2018).
101. Lambo, S. *et al.* A longitudinal single-cell atlas of treatment response in pediatric AML. *Cancer Cell* **41**, 2117–2135 e12 (2023).
102. Chen, Y., Li, J., Xu, L., Gaman, M.A. & Zou, Z. The genesis and evolution of acute myeloid leukemia stem cells in the microenvironment: From biology to therapeutic targeting. *Cell Death Discov* **8**, 397 (2022).
103. Beutler, E., Yeh, M. & Fairbanks, V.F. The normal human female as a mosaic of X-chromosome activity: studies using the gene for C-6-PD-deficiency as a marker. *Proc Natl Acad Sci U S A* **48**, 9–16 (1962).
104. Fialkow, P.J., Gartler, S.M. & Yoshida, A. Clonal origin of chronic myelocytic leukemia in man. *Proc Natl Acad Sci U S A* **58**, 1468–71 (1967).
105. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645–8 (1994).
106. Kakiuchi, N. & Ogawa, S. Clonal expansion in non-cancer tissues. *Nat Rev Cancer* **21**, 239–256 (2021).
107. Weeks, L.D. & Ebert, B.L. Causes and consequences of clonal hematopoiesis. *Blood* **142**, 2235–2246 (2023).
108. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371**, 2488–98 (2014).
109. McKerrell, T. *et al.* Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep* **10**, 1239–45 (2015).

110. Singh, J. *et al.* Clonal hematopoiesis of indeterminate potential as a prognostic factor: a systematic review and meta-analysis. *Blood Adv* **8**, 3771–3784 (2024).
111. Kreso, A. & Dick, J.E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**, 275–91 (2014).
112. Stauber, J., Greally, J.M. & Steidl, U. Preleukemic and leukemic evolution at the stem cell level. *Blood* **137**, 1013–1018 (2021).
113. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892 (2012).
114. Zhang, K.Y. *et al.* Tumor evolution: signaling pathways, molecular mechanisms and therapeutic targets. *Signal Transduct Target Ther* **11**(2026).
115. Williams, M.J., Werner, B., Barnes, C.P., Graham, T.A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nat Genet* **48**, 238–244 (2016).
116. Brady, S.W. *et al.* The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat Genet* **54**, 1376–1389 (2022).
117. Wong, T.N. *et al.* Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* **518**, 552–555 (2015).
118. Wong, T.N. *et al.* Cellular stressors contribute to the expansion of hematopoietic clones of varying leukemic potential. *Nat Commun* **9**, 455 (2018).
119. Bolton, K.L. *et al.* Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat Genet* **52**, 1219–1226 (2020).
120. Mossner, M. *et al.* Mutational hierarchies in myelodysplastic syndromes dynamically adapt and evolve upon therapy response and failure. *Blood* **128**, 1246–59 (2016).
121. Rapaport, F. *et al.* Genomic and evolutionary portraits of disease relapse in acute myeloid leukemia. *Leukemia* **35**, 2688–2692 (2021).
122. Gale, K.B. *et al.* Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc Natl Acad Sci U S A* **94**, 13950–4 (1997).
123. Wiemels, J.L. *et al.* Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* **354**, 1499–503 (1999).
124. Calderon, A.S. *et al.* Ontogeny-specific induction of the KMT2A::AFF1-fusion drives development of a distinct CD24 positive pre-leukemic state. *Leukemia* **39**, 2099–2111 (2025).
125. Ford, A.M. *et al.* In utero rearrangements in the trithorax-related oncogene in infant leukaemias. *Nature* **363**, 358–60 (1993).
126. Greaves, M.F. & Wiemels, J. Origins of chromosome translocations in childhood leukaemia. *Nat Rev Cancer* **3**, 639–49 (2003).
127. Hein, D., Borkhardt, A. & Fischer, U. Insights into the prenatal origin of childhood acute lymphoblastic leukemia. *Cancer Metastasis Rev* **39**, 161–171 (2020).
128. Ford, A.M., Colman, S. & Greaves, M. Covert pre-leukaemic clones in healthy co-twins of patients with childhood acute lymphoblastic leukaemia. *Leukemia* **37**, 47–52 (2023).

129. Andersson, A.K. *et al.* The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet* **47**, 330–7 (2015).
130. Milne, T.A. Mouse models of MLL leukemia: recapitulating the human disease. *Blood* **129**, 2217–2223 (2017).
131. Khabirova, E. *et al.* Single-cell transcriptomics reveals a distinct developmental state of KMT2A-rearranged infant B-cell acute lymphoblastic leukemia. *Nat Med* **28**, 743–751 (2022).
132. Pagliaro, L. *et al.* Acute lymphoblastic leukaemia. *Nat Rev Dis Primers* **10**, 41 (2024).
133. Iacobucci, I., Kimura, S. & Mullighan, C.G. Biologic and Therapeutic Implications of Genomic Alterations in Acute Lymphoblastic Leukemia. *J Clin Med* **10**(2021).
134. Iacobucci, I. & Mullighan, C.G. Genetic Basis of Acute Lymphoblastic Leukemia. *J Clin Oncol* **35**, 975–983 (2017).
135. Roberts, K.G. & Mullighan, C.G. The Biology of B-Progenitor Acute Lymphoblastic Leukemia. *Cold Spring Harb Perspect Med* **10**(2020).
136. Kantarjian, H., Pui, C.H. & Jabbour, E. Acute lymphocytic leukaemia. *Lancet* **406**, 950–962 (2025).
137. Polonen, P. *et al.* The genomic basis of childhood T-lineage acute lymphoblastic leukaemia. *Nature* **632**, 1082–1091 (2024).
138. Mullighan, C.G. The molecular genetic makeup of acute lymphoblastic leukemia. *Hematology Am Soc Hematol Educ Program* **2012**, 389–96 (2012).
139. Heyman, M. *et al.* A Treatment Study Protocol for Participants 0-45 Years With Acute Lymphoblastic Leukaemia. (2020).
140. Paulsson, K. *et al.* The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Nat Genet* **47**, 672–6 (2015).
141. Haas, O.A. & Borkhardt, A. Hyperdiploidy: the longest known, most prevalent, and most enigmatic form of acute lymphoblastic leukemia in children. *Leukemia* **36**, 2769–2783 (2022).
142. Inaba, H. & Mullighan, C.G. Pediatric acute lymphoblastic leukemia. *Haematologica* **105**, 2524–2539 (2020).
143. Holmfeldt, L. *et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat Genet* **45**, 242–52 (2013).
144. Pui, C.H. *et al.* Outcome of Children With Hypodiploid Acute Lymphoblastic Leukemia: A Retrospective Multinational Study. *J Clin Oncol* **37**, 770–779 (2019).
145. Bernard, E. *et al.* Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nat Med* **26**, 1549–1556 (2020).
146. Yang, F. *et al.* Chemotherapy and mismatch repair deficiency cooperate to fuel TP53 mutagenesis and ALL relapse. *Nat Cancer* **2**, 819–834 (2021).
147. Lee, S.H.R. *et al.* Prognostic and Pharmacotypic Heterogeneity of Hyperdiploidy in Childhood ALL. *J Clin Oncol* **41**, 5422–5432 (2023).

148. Oksa, L. *et al.* Genomic determinants of therapy response in ETV6::RUNX1 leukemia. *Leukemia* **39**, 2125–2139 (2025).
149. Linka, Y. *et al.* The impact of TEL-AML1 (ETV6-RUNX1) expression in precursor B cells and implications for leukaemia using three different genome-wide screening methods. *Blood Cancer J* **3**, e151 (2013).
150. Rand, V. *et al.* Genomic characterization implicates iAMP21 as a likely primary genetic event in childhood B-cell precursor acute lymphoblastic leukemia. *Blood* **117**, 6848–55 (2011).
151. Ryan, S.L. *et al.* The role of the RAS pathway in iAMP21-ALL. *Leukemia* **30**, 1824–31 (2016).
152. Hormann, F.M. *et al.* Secondary lesions and sensitivity to signaling inhibitors in iAMP21 acute lymphoblastic leukemia. *Hemasphere* **9**, e70069 (2025).
153. Sinclair, P.B. *et al.* SH2B3 inactivation through CN-LOH 12q is uniquely associated with B-cell precursor ALL with iAMP21 or other chromosome 21 gain. *Leukemia* **33**, 1881–1894 (2019).
154. Harrison, C.J. Blood Spotlight on iAMP21 acute lymphoblastic leukemia (ALL), a high-risk pediatric disease. *Blood* **125**, 1383–6 (2015).
155. Tasian, S.K. TCF3::HLF acute lymphoblastic leukemia: still challenging to cure thirty years later. *Haematologica* **108**, 1713–1714 (2023).
156. Fischer, U. *et al.* Genomics and drug profiling of fatal TCF3-HLF-positive acute lymphoblastic leukemia identifies recurrent mutation patterns and therapeutic options. *Nat Genet* **47**, 1020–1029 (2015).
157. Kim, J.C. *et al.* Transcriptomic classes of BCR-ABL1 lymphoblastic leukemia. *Nat Genet* **55**, 1186–1197 (2023).
158. Hunger, S.P. *et al.* Dasatinib with intensive chemotherapy in de novo paediatric Philadelphia chromosome-positive acute lymphoblastic leukaemia (CA180-372/COG AALL1122): a single-arm, multicentre, phase 2 trial. *Lancet Haematol* **10**, e510–e520 (2023).
159. Mullighan, C.G. *et al.* BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453**, 110–4 (2008).
160. Tran, T.H. & Hunger, S.P. ABL-class fusion positive acute lymphoblastic leukemia: can targeting ABL cure ALL? *Haematologica* **105**, 1754–1757 (2020).
161. Girardi, T., Vicente, C., Cools, J. & De Keersmaecker, K. The genetics and molecular biology of T-ALL. *Blood* **129**, 1113–1123 (2017).
162. Liu, Y. *et al.* The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat Genet* **49**, 1211–1218 (2017).
163. Coustan-Smith, E. *et al.* Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol* **10**, 147–56 (2009).
164. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–63 (2012).
165. Zwaan, C.M. *et al.* Collaborative Efforts Driving Progress in Pediatric Acute Myeloid Leukemia. *J Clin Oncol* **33**, 2949–62 (2015).

166. Mo, Q. *et al.* Integrative molecular subtypes of acute myeloid leukemia. *Blood Cancer J* **13**, 71 (2023).
167. Tyner, J.W. *et al.* Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**, 526–531 (2018).
168. Ley, T.J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
169. Kaspers, G.J.L. *et al.* CHIP-AML22/Master: An Open Label Complex Clinical Trial in Newly Diagnosed Pediatric de Novo AML Patients. (2023).
170. Pfister, S.M. *et al.* A Summary of the Inaugural WHO Classification of Pediatric Tumors: Transitioning from the Optical into the Molecular Era. *Cancer Discov* **12**, 331–355 (2022).
171. Conneely, S.E. & Rau, R.E. The genomics of acute myeloid leukemia in children. *Cancer Metastasis Rev* **39**, 189–209 (2020).
172. Brunetti, L. *et al.* Mutant NPM1 Maintains the Leukemic State through HOX Expression. *Cancer Cell* **34**, 499–512 e9 (2018).
173. Kuhn, M.W. *et al.* Targeting Chromatin Regulators Inhibits Leukemogenic Gene Expression in NPM1 Mutant Leukemia. *Cancer Discov* **6**, 1166–1181 (2016).
174. Martelli, M.P. *et al.* Arsenic trioxide and all-trans retinoic acid target NPM1 mutant oncoprotein levels and induce apoptosis in NPM1-mutated AML cells. *Blood* **125**, 3455–65 (2015).
175. Opatz, S. *et al.* The clinical mutatome of core binding factor leukemia. *Leukemia* **34**, 1553–1562 (2020).
176. Umeda, M. *et al.* A new genomic framework to categorize pediatric acute myeloid leukemia. *Nat Genet* **56**, 281–293 (2024).
177. de The, H. Differentiation therapy revisited. *Nat Rev Cancer* **18**, 117–127 (2018).
178. Yilmaz, M., Kantarjian, H. & Ravandi, F. Acute promyelocytic leukemia current treatment algorithms. *Blood Cancer J* **11**, 123 (2021).
179. Bertrums, E.J.M. *et al.* Comprehensive molecular and clinical characterization of NUP98 fusions in pediatric acute myeloid leukemia. *Haematologica* **108**, 2044–2058 (2023).
180. Bolouri, H. *et al.* The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* **24**, 103–112 (2018).
181. Groschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369–381 (2014).
182. Umeda, M., Liu, Y.C., Karol, S.E. & Klco, J.M. Novel classification system and high-risk categories of pediatric acute myeloid leukemia. *Haematologica* **110**, 1962–1973 (2025).
183. Gonzales, F. *et al.* Identifying targeted therapies for CBFA2T3::GLIS2 acute myeloid leukemia. *Leukemia* **40**, 383–396 (2026).
184. De la Torre, E.P. *et al.* Validation of mutated CEBPA bZIP as a distinct prognosis entity in acute myeloid leukemia: a study by the Spanish PETHEMA registry. *Haematologica* **109**, 2682–2687 (2024).

185. Tarlock, K. *et al.* CEBPA-bZip mutations are associated with favorable prognosis in de novo AML: a report from the Children's Oncology Group. *Blood* **138**, 1137–1147 (2021).
186. Bene, M.C. Biphenotypic, bilineal, ambiguous or mixed lineage: strange leukemias! *Haematologica* **94**, 891–3 (2009).
187. Weinberg, O.K. *et al.* The International Consensus Classification of acute leukemias of ambiguous lineage. *Blood* **141**, 2275–2277 (2023).
188. Hrusak, O. *et al.* International cooperative study identifies treatment strategy in childhood ambiguous lineage leukemia. *Blood* **132**, 264–276 (2018).
189. Georgopoulos, K., Moore, D.D. & Derfler, B. Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. *Science* **258**, 808–12 (1992).
190. Mullighan, C.G. *et al.* Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med* **360**, 470–80 (2009).
191. Olsson, L. & Johansson, B. Ikaros and leukaemia. *Br J Haematol* **169**, 479–91 (2015).
192. Den Boer, M.L. *et al.* A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol* **10**, 125–34 (2009).
193. Chang, T.C. *et al.* Genomic Determinants of Outcome in Acute Lymphoblastic Leukemia. *J Clin Oncol* **42**, 3491–3503 (2024).
194. Gu, Z. *et al.* PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet* **51**, 296–307 (2019).
195. Mangum, D.S. *et al.* VPREB1 deletions occur independent of lambda light chain rearrangement in childhood acute lymphoblastic leukemia. *Leukemia* **28**, 216–20 (2014).
196. Sabapathy, K. & Lane, D.P. Therapeutic targeting of p53: all mutants are equal, but some mutants are more equal than others. *Nat Rev Clin Oncol* **15**, 13–30 (2018).
197. Wang, H., Guo, M., Wei, H. & Chen, Y. Targeting p53 pathways: mechanisms, structures, and advances in therapy. *Signal Transduct Target Ther* **8**, 92 (2023).
198. Hof, J. *et al.* Mutations and deletions of the TP53 gene predict nonresponse to treatment and poor outcome in first relapse of childhood acute lymphoblastic leukemia. *J Clin Oncol* **29**, 3185–93 (2011).
199. Stengel, A. *et al.* TP53 mutations occur in 15.7% of ALL and are associated with MYC-rearrangement, low hypodiploidy, and a poor prognosis. *Blood* **124**, 251–8 (2014).
200. Loh, S.N. Arsenic and an Old Place: Rescuing p53 Mutants in Cancer. *Cancer Cell* **39**, 140–142 (2021).
201. Gould, S.I. *et al.* High-throughput evaluation of genetic variants with prime editing sensor libraries. *Nat Biotechnol* **43**, 1648–1662 (2025).
202. Willis, A., Jung, E.J., Wakefield, T. & Chen, X. Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene* **23**, 2330–8 (2004).

203. Grob, T. *et al.* Molecular characterization of mutant TP53 acute myeloid leukemia and high-risk myelodysplastic syndrome. *Blood* **139**, 2347–2354 (2022).
204. Daver, N.G. *et al.* TP53-Mutated Myelodysplastic Syndrome and Acute Myeloid Leukemia: Biology, Current Therapy, and Future Directions. *Cancer Discov* **12**, 2516–2529 (2022).
205. Li, B. *et al.* Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. *Blood* **135**, 41–55 (2020).
206. Sulong, S. *et al.* A comprehensive analysis of the CDKN2A gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. *Blood* **113**, 100–7 (2009).
207. Ma, X. *et al.* Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nat Commun* **6**, 6604 (2015).
208. Moore, A.R., Rosenberg, S.C., McCormick, F. & Malek, S. RAS-targeted therapies: is the undruggable drugged? *Nat Rev Drug Discov* **19**, 533–552 (2020).
209. Waanders, E. *et al.* Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. *Blood Cancer Discov* **1**, 96–111 (2020).
210. Irving, J. *et al.* Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. *Blood* **124**, 3420–30 (2014).
211. Antic, Z. *et al.* Multiclonal complexity of pediatric acute lymphoblastic leukemia and the prognostic relevance of subclonal mutations. *Haematologica* **106**, 3046–3055 (2021).
212. Oshima, K. *et al.* Mutational landscape, clonal evolution patterns, and role of RAS mutations in relapsed acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **113**, 11306–11311 (2016).
213. Pilheden, M. *et al.* Duplex Sequencing Uncovers Recurrent Low-frequency Cancer-associated Mutations in Infant and Childhood KMT2A-rearranged Acute Leukemia. *Hemasphere* **6**, e785 (2022).
214. Kaburagi, T. *et al.* Clinical significance of RAS pathway alterations in pediatric acute myeloid leukemia. *Haematologica* **107**, 583–592 (2022).
215. Shoji, K. *et al.* Age-specific mutation profiles and their prognostic implications in pediatric KMT2A-rearranged acute myeloid leukemia. *Haematologica* **111**, 1235–1245 (2026).
216. Hyrenius-Wittsten, A. *et al.* De novo activating mutations drive clonal evolution and enhance clonal fitness in KMT2A-rearranged leukemia. *Nat Commun* **9**, 1770 (2018).
217. Riedl, J.M. *et al.* Emerging landscape of KRAS inhibitors in cancer treatment. *Cancer Cell* **44**, 471–497 (2026).
218. Kong, G. *et al.* The ability of endogenous Nras oncogenes to initiate leukemia is codon-dependent. *Leukemia* **30**, 1935–8 (2016).
219. Mecklenbrauck, R. *et al.* Prognostic impact of myelodysplasia-related gene mutations in FLT3-ITD-mutated acute myeloid leukemia. *Leukemia* **40**, 622–629 (2026).

220. Meshinchi, S. *et al.* Prevalence and prognostic significance of Flt3 internal tandem duplication in pediatric acute myeloid leukemia. *Blood* **97**, 89–94 (2001).
221. Qiu, K.Y. *et al.* Poor outcome of pediatric patients with acute myeloid leukemia harboring high FLT3/ITD allelic ratios. *Nat Commun* **13**, 3679 (2022).
222. Esposito, M.T. The Impact of PI3-kinase/RAS Pathway Cooperating Mutations in the Evolution of KMT2A-rearranged Leukemia. *Hemasphere* **3**, e195 (2019).
223. Ley, T.J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424–33 (2010).
224. Tulstrup, M. *et al.* TET2 mutations are associated with hypermethylation at key regulatory enhancers in normal and malignant hematopoiesis. *Nat Commun* **12**, 6061 (2021).
225. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–8 (2012).
226. Greenblatt, S.M. & Nimer, S.D. Chromatin modifiers and the promise of epigenetic therapy in acute leukemia. *Leukemia* **28**, 1396–406 (2014).
227. Wang, L. *et al.* Epigenetic targeted therapy of stabilized BAP1 in ASXL1 gain-of-function mutated leukemia. *Nat Cancer* **2**, 515–526 (2021).
228. Fischer, A. *et al.* STAG2 mutations reshape the cohesin-structured spatial chromatin architecture to drive gene regulation in acute myeloid leukemia. *Cell Rep* **43**, 114498 (2024).
229. Scott, J.S. *et al.* Emerging roles of cohesin-STAG2 in cancer. *Oncogene* **44**, 277–287 (2025).
230. Takao, S. *et al.* Epigenetic mechanisms controlling human leukemia stem cells and therapy resistance. *Nat Commun* **16**, 3196 (2025).
231. Gilliland, D.G. & Griffin, J.D. The roles of FLT3 in hematopoiesis and leukemia. *Blood* **100**, 1532–42 (2002).
232. Katti, A., Diaz, B.J., Caragine, C.M., Sanjana, N.E. & Dow, L.E. CRISPR in cancer biology and therapy. *Nat Rev Cancer* **22**, 259–279 (2022).
233. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med* **373**, 2336–2346 (2015).
234. Maese, L.D. *et al.* Update on Recommendations for Surveillance for Children with Predisposition to Hematopoietic Malignancy. *Clin Cancer Res* **30**, 4286–4295 (2024).
235. Yang, F. *et al.* Identification and prioritization of myeloid malignancy germline variants in a large cohort of adult patients with AML. *Blood* **139**, 1208–1221 (2022).
236. Hasle, H., Clemmensen, I.H. & Mikkelsen, M. Risks of leukaemia and solid tumours in individuals with Down's syndrome. *Lancet* **355**, 165–9 (2000).
237. Baruchel, A. *et al.* Down syndrome and leukemia: from basic mechanisms to clinical advances. *Haematologica* **108**, 2570–2581 (2023).
238. Biondi, A., Cimino, G., Pieters, R. & Pui, C.H. Biological and therapeutic aspects of infant leukemia. *Blood* **96**, 24–33 (2000).

239. Rio-Machin, A. *et al.* The complex genetic landscape of familial MDS and AML reveals pathogenic germline variants. *Nat Commun* **11**, 1044 (2020).
240. Chen, X., Yang, W., Roberts, C.W.M. & Zhang, J. Developmental origins shape the paediatric cancer genome. *Nat Rev Cancer* **24**, 382–398 (2024).
241. Stoltze, U.K. *et al.* The evolutionary impact of childhood cancer on the human gene pool. *Nat Commun* **15**, 1881 (2024).
242. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–21 (2013).
243. Grobner, S.N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
244. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).
245. Downing, J.R. *et al.* The Pediatric Cancer Genome Project. *Nat Genet* **44**, 619–22 (2012).
246. Dharia, N.V. *et al.* A first-generation pediatric cancer dependency map. *Nat Genet* **53**, 529–538 (2021).
247. Rowe, R.G. *et al.* The developmental stage of the hematopoietic niche regulates lineage in MLL-rearranged leukemia. *J Exp Med* **216**, 527–538 (2019).
248. Jaiswal, S. & Ebert, B.L. Clonal hematopoiesis in human aging and disease. *Science* **366**(2019).
249. Pang, W.W. *et al.* Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc Natl Acad Sci U S A* **108**, 20012–7 (2011).
250. Kattner, P. *et al.* Compare and contrast: pediatric cancer versus adult malignancies. *Cancer Metastasis Rev* **38**, 673–682 (2019).
251. Muntean, A.G. & Hess, J.L. The pathogenesis of mixed-lineage leukemia. *Annu Rev Pathol* **7**, 283–301 (2012).
252. Cancer Genome Atlas Research, N. *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059–74 (2013).
253. Roberts, K.G. *et al.* High Frequency and Poor Outcome of Philadelphia Chromosome-Like Acute Lymphoblastic Leukemia in Adults. *J Clin Oncol* **35**, 394–401 (2017).
254. Creutzig, U. *et al.* Diagnosis and management of acute myeloid leukemia in children and adolescents: recommendations from an international expert panel. *Blood* **120**, 3187–205 (2012).
255. Winters, A.C. & Bernt, K.M. MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front Pediatr* **5**, 4 (2017).
256. Iaccarino, I. *et al.* KMT2A-rearranged B-lymphoblastic lymphomas are skewed towards a more mature developmental stage. *Leukemia* (2026).
257. Massoth, L.R. *et al.* Pan-Cancer Landscape Analysis Reveals Recurrent KMT2A-MAML2 Gene Fusion in Aggressive Histologic Subtypes of Thymoma. *JCO Precis Oncol* **4**(2020).

258. Massoth, L.R. *et al.* Pan-sarcoma genomic analysis of KMT2A rearrangements reveals distinct subtypes defined by YAP1-KMT2A-YAP1 and VIM-KMT2A fusions. *Mod Pathol* **33**, 2307–2317 (2020).
259. Meyer, C., Larghero, P., Lopes, B.A. & Marschalek, R. The KMT2A/MLL consensus gene structure: a comprehensive update for research and diagnostic implications. *Leukemia* **38**, 1403–1406 (2024).
260. Janssens, D.H. *et al.* Automated CUT&Tag profiling of chromatin heterogeneity in mixed-lineage leukemia. *Nat Genet* **53**, 1586–1596 (2021).
261. Sneppen, K. & Ringrose, L. Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent. *Nat Commun* **10**, 2133 (2019).
262. Jude, C.D. *et al.* Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell Stem Cell* **1**, 324–37 (2007).
263. McMahon, K.A. *et al.* Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal. *Cell Stem Cell* **1**, 338–45 (2007).
264. Ernst, P., Mabon, M., Davidson, A.J., Zon, L.I. & Korsmeyer, S.J. An Mll-dependent Hox program drives hematopoietic progenitor expansion. *Curr Biol* **14**, 2063–9 (2004).
265. Forgione, M.O., McClure, B.J., Eadie, L.N., Yeung, D.T. & White, D.L. KMT2A rearranged acute lymphoblastic leukaemia: Unravelling the genomic complexity and heterogeneity of this high-risk disease. *Cancer Lett* **469**, 410–418 (2020).
266. Zhang, R. *et al.* Outcomes of acute myeloid leukemia with KMT2A (MLL) rearrangement: a multicenter study of TROPHY group. *Blood Cancer J* **15**, 84 (2025).
267. Meyer, C. *et al.* The KMT2A recombinome of acute leukemias in 2023. *Leukemia* **37**, 988–1005 (2023).
268. Szczepanski, T., Harrison, C.J. & van Dongen, J.J. Genetic aberrations in paediatric acute leukaemias and implications for management of patients. *Lancet Oncol* **11**, 880–9 (2010).
269. Hartmann, A.M. *et al.* Intra-subtype heterogeneity shapes treatment response in KMT2A-rearranged ALL across all age groups. *Hemasphere* **10**, e70324 (2026).
270. Deshpande, A.J. *et al.* Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l. *Blood* **121**, 2533–41 (2013).
271. Muntean, A.G. *et al.* The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. *Cancer Cell* **17**, 609–21 (2010).
272. Yokoyama, A. *et al.* The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**, 207–18 (2005).
273. Caslini, C. *et al.* Interaction of MLL amino terminal sequences with menin is required for transformation. *Cancer Res* **67**, 7275–83 (2007).
274. Meyer, C. *et al.* The MLL recombinome of acute leukemias in 2017. *Leukemia* **32**, 273–284 (2018).
275. Krivtsov, A.V. & Armstrong, S.A. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* **7**, 823–33 (2007).

276. Krivtsov, A.V., Hoshii, T. & Armstrong, S.A. Mixed-Lineage Leukemia Fusions and Chromatin in Leukemia. *Cold Spring Harb Perspect Med* **7**(2017).
277. Deshpande, A.J., Bradner, J. & Armstrong, S.A. Chromatin modifications as therapeutic targets in MLL-rearranged leukemia. *Trends Immunol* **33**, 563–70 (2012).
278. Hess, J.L. *et al.* c-Myb is an essential downstream target for homeobox-mediated transformation of hematopoietic cells. *Blood* **108**, 297–304 (2006).
279. Ramsay, R.G. & Gonda, T.J. MYB function in normal and cancer cells. *Nat Rev Cancer* **8**, 523–34 (2008).
280. Kim, J. *et al.* A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* **143**, 313–24 (2010).
281. Krivtsov, A.V. *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818–22 (2006).
282. Krivtsov, A.V. *et al.* Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia* **27**, 852–60 (2013).
283. So, C.W., Lin, M., Ayton, P.M., Chen, E.H. & Cleary, M.L. Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias. *Cancer Cell* **4**, 99–110 (2003).
284. Slany, R.K. The molecular biology of mixed lineage leukemia. *Haematologica* **94**, 984–93 (2009).
285. Sun, Q.Y. *et al.* Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD). *Leukemia* **31**, 1–10 (2017).
286. Whitman, S.P. *et al.* DNA hypermethylation and epigenetic silencing of the tumor suppressor gene, SLC5A8, in acute myeloid leukemia with the MLL partial tandem duplication. *Blood* **112**, 2013–6 (2008).
287. Hoffman R, B.E., Silberstein LE et al. Hematology: Basic Principles and Practice. *Elsevier Inc.* (2017).
288. Rice, S. *et al.* A human fetal liver-derived infant MLL-AF4 acute lymphoblastic leukemia model reveals a distinct fetal gene expression program. *Nat Commun* **12**, 6905 (2021).
289. Bursen, A. *et al.* The AF4.MLL fusion protein is capable of inducing ALL in mice without requirement of MLL.AF4. *Blood* **115**, 3570–9 (2010).
290. Agraz-Doblas, A. *et al.* Unraveling the cellular origin and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis. *Haematologica* **104**, 1176–1188 (2019).
291. van Weelderen, R.E. *et al.* Measurable Residual Disease and Fusion Partner Independently Predict Survival and Relapse Risk in Childhood KMT2A-Rearranged Acute Myeloid Leukemia: A Study by the International Berlin-Frankfurt-Munster Study Group. *J Clin Oncol* **41**, 2963–2974 (2023).
292. Abla, O. *et al.* Structural variants involving MLLT10 fusion are associated with adverse outcomes in pediatric acute myeloid leukemia. *Blood Adv* **8**, 2005–2017 (2024).
293. Van Limbergen, H. *et al.* Molecular cytogenetic analysis of 10;11 rearrangements in acute myeloid leukemia. *Leukemia* **16**, 344–51 (2002).

294. Liedtke, M., Ayton, P.M., Somervaille, T.C., Smith, K.S. & Cleary, M.L. Self-association mediated by the Ras association 1 domain of AF6 activates the oncogenic potential of MLL-AF6. *Blood* **116**, 63–70 (2010).
295. Cozzio, A. *et al.* Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes Dev* **17**, 3029–35 (2003).
296. Chen, W. *et al.* Malignant transformation initiated by Mll-AF9: gene dosage and critical target cells. *Cancer Cell* **13**, 432–40 (2008).
297. Ugale, A. *et al.* Hematopoietic stem cells are intrinsically protected against MLL-ENL-mediated transformation. *Cell Rep* **9**, 1246–55 (2014).
298. Stavropoulou, V. *et al.* MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome. *Cancer Cell* **30**, 43–58 (2016).
299. Pan, F. *et al.* Genome editing-induced t(4;11) chromosomal translocations model B cell precursor acute lymphoblastic leukemias with KMT2A-AFF1 fusion. *J Clin Invest* **134**(2024).
300. Ye, M. *et al.* Hematopoietic Differentiation Is Required for Initiation of Acute Myeloid Leukemia. *Cell Stem Cell* **17**, 611–23 (2015).
301. Ohlsson, E. *et al.* Initiation of MLL-rearranged AML is dependent on C/EBPalpha. *J Exp Med* **211**, 5–13 (2014).
302. Zhou, T. *et al.* Genetics and pathologic landscape of lineage switch of acute leukemia during therapy. *Blood Cancer J* **14**, 19 (2024).
303. Kovach, A.E. *et al.* Plasticity under pressure: biology and detection of lineage switch in acute leukemia. *Leukemia* (2026).
304. Tirtakusuma, R. *et al.* Epigenetic regulator genes direct lineage switching in MLL/AF4 leukemia. *Blood* **140**, 1875–1890 (2022).
305. Iacobucci, I. & Mullighan, C.G. KMT2A-rearranged leukemia: the shapeshifter. *Blood* **140**, 1833–1835 (2022).
306. Uckelmann, H.J. *et al.* Mutant NPM1 Directly Regulates Oncogenic Transcription in Acute Myeloid Leukemia. *Cancer Discov* **13**, 746–765 (2023).
307. Chen, C. *et al.* Single-cell multiomics reveals increased plasticity, resistant populations, and stem-cell-like blasts in KMT2A-rearranged leukemia. *Blood* **139**, 2198–2211 (2022).
308. Iacobucci, I. *et al.* Multipotent lineage potential in B cell acute lymphoblastic leukemia is associated with distinct cellular origins and clinical features. *Nat Cancer* **6**, 1242–1262 (2025).
309. Zeng, A.G.X. *et al.* Single-cell Transcriptional Atlas of Human Hematopoiesis Reveals Genetic and Hierarchy-Based Determinants of Aberrant AML Differentiation. *Blood Cancer Discov* **6**, 307–324 (2025).
310. Candelli, T. *et al.* Identification and characterization of relapse-initiating cells in MLL-rearranged infant ALL by single-cell transcriptomics. *Leukemia* **36**, 58–67 (2022).
311. Rangarajan, A. & Weinberg, R.A. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat Rev Cancer* **3**, 952–9 (2003).

312. Bubnova, X., Hope, L. & Wheadon, H. Evaluation of current mouse in vivo models and advanced in vitro models for leukaemia research. *Blood Rev* **74**, 101318 (2025).
313. Hyrenius-Wittsten, A. *et al.* FLT3(N676K) drives acute myeloid leukemia in a xenograft model of KMT2A-MLLT3 leukemogenesis. *Leukemia* **33**, 2310–2314 (2019).
314. Yun, H. *et al.* Mutational synergy during leukemia induction remodels chromatin accessibility, histone modifications and three-dimensional DNA topology to alter gene expression. *Nat Genet* **53**, 1443–1455 (2021).
315. Welch, J.S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–78 (2012).
316. Chen, W. *et al.* A murine Mll-AF4 knock-in model results in lymphoid and myeloid deregulation and hematologic malignancy. *Blood* **108**, 669–77 (2006).
317. Krivtsov, A.V. *et al.* H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell* **14**, 355–68 (2008).
318. Corral, J. *et al.* An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell* **85**, 853–61 (1996).
319. Drynan, L.F. *et al.* Mll fusions generated by Cre-loxP-mediated de novo translocations can induce lineage reassignment in tumorigenesis. *EMBO J* **24**, 3136–46 (2005).
320. Metzler, M. *et al.* A conditional model of MLL-AF4 B-cell tumourigenesis using invertor technology. *Oncogene* **25**, 3093–103 (2006).
321. Lin, S. *et al.* Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia. *Cancer Cell* **30**, 737–749 (2016).
322. Lavau, C., Luo, R.T., Du, C. & Thirman, M.J. Retrovirus-mediated gene transfer of MLL-ELL transforms primary myeloid progenitors and causes acute myeloid leukemias in mice. *Proc Natl Acad Sci U S A* **97**, 10984–9 (2000).
323. Somervaille, T.C. & Cleary, M.L. Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. *Cancer Cell* **10**, 257–68 (2006).
324. DiMartino, J.F. *et al.* The AF10 leucine zipper is required for leukemic transformation of myeloid progenitors by MLL-AF10. *Blood* **99**, 3780–5 (2002).
325. Barabe, F., Kennedy, J.A., Hope, K.J. & Dick, J.E. Modeling the initiation and progression of human acute leukemia in mice. *Science* **316**, 600–4 (2007).
326. Kawashima, N. *et al.* Comparison of clonal architecture between primary and immunodeficient mouse-engrafted acute myeloid leukemia cells. *Nat Commun* **13**, 1624 (2022).
327. Zhang, Q. *et al.* Activating mutations remodel the chromatin accessibility landscape to drive distinct regulatory networks in KMT2A-rearranged acute leukemia. *Hemasphere* **8**, e70006 (2024).
328. Terwilliger, T. & Abdul-Hay, M. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J* **7**, e577 (2017).

329. Lyu, A., Nam, S.H., Humphrey, R.S., Horton, T.M. & Ehrlich, L.I.R. Cells and signals of the leukemic microenvironment that support progression of T-cell acute lymphoblastic leukemia (T-ALL). *Exp Mol Med* **56**, 2337–2347 (2024).
330. Van der Meulen, J. *et al.* The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Blood* **125**, 13–21 (2015).
331. Dohner, H. *et al.* Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood* **140**, 1345–1377 (2022).
332. Taylan, F. *et al.* The Nordic ChiCaP Network: An interdisciplinary approach to childhood cancer predisposition. *Rare* **3**, 100111–100111 (2025).
333. Kim, T. *et al.* Next-generation sequencing-based posttransplant monitoring of acute myeloid leukemia identifies patients at high risk of relapse. *Blood* **132**, 1604–1613 (2018).
334. Schultz, K.R. *et al.* Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG). *Blood* **109**, 926–35 (2007).
335. Van der Velden, V.H. *et al.* Prognostic significance of minimal residual disease in infants with acute lymphoblastic leukemia treated within the Interfant-99 protocol. *Leukemia* **23**, 1073–9 (2009).
336. Burmeister, T. *et al.* Monitoring minimal residual disease by quantification of genomic chromosomal breakpoint sequences in acute leukemias with MLL aberrations. *Leukemia* **20**, 451–7 (2006).
337. Zhou, J. *et al.* Quantitative analysis of minimal residual disease predicts relapse in children with B-lineage acute lymphoblastic leukemia in DFCI ALL Consortium Protocol 95-01. *Blood* **110**, 1607–11 (2007).
338. Szczepanski, T. *et al.* Comparative analysis of Ig and TCR gene rearrangements at diagnosis and at relapse of childhood precursor-B-ALL provides improved strategies for selection of stable PCR targets for monitoring of minimal residual disease. *Blood* **99**, 2315–23 (2002).
339. van der Velden, V.H. *et al.* Immunoglobulin kappa deleting element rearrangements in precursor-B acute lymphoblastic leukemia are stable targets for detection of minimal residual disease by real-time quantitative PCR. *Leukemia* **16**, 928–36 (2002).
340. Bruggemann, M. *et al.* Rearranged T-cell receptor beta genes represent powerful targets for quantification of minimal residual disease in childhood and adult T-cell acute lymphoblastic leukemia. *Leukemia* **18**, 709–19 (2004).
341. Chen, J.S. *et al.* Identification of novel markers for monitoring minimal residual disease in acute lymphoblastic leukemia. *Blood* **97**, 2115–20 (2001).
342. Lucio, P. *et al.* BIOMED-I concerted action report: flow cytometric immunophenotyping of precursor B-ALL with standardized triple-stainings. BIOMED-1 Concerted Action Investigation of Minimal Residual Disease in Acute Leukemia: International Standardization and Clinical Evaluation. *Leukemia* **15**, 1185–92 (2001).

343. H., P.C. *Childhood Leukemias Third Edition*, (Cambridge University Press, Cambridge, 2012).
344. Borowitz, M.J. *et al.* Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood* **111**, 5477–85 (2008).
345. Brodersen, L.E. *et al.* Morphologic remission status is limited compared to DeltaN flow cytometry: a Children's Oncology Group AAML0531 report. *Blood Adv* **4**, 5050–5061 (2020).
346. Schuurhuis, G.J. *et al.* Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood* **131**, 1275–1291 (2018).
347. Malard, F. & Mohty, M. Acute lymphoblastic leukaemia. *Lancet* **395**, 1146–1162 (2020).
348. Farber, S. & Diamond, L.K. Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid. *N Engl J Med* **238**, 787–93 (1948).
349. Winther, J.F. & Schmiegelow, K. How safe is a standard-risk child with ALL? *Lancet Oncol* **15**, 782–3 (2014).
350. Interfant-21 Treatment Protocol for Infants Under 1 Year With KMT2A-rearranged ALL or Mixed Phenotype Acute Leukemia (Interfant-21). (2022).
351. Stutterheim, J., Van der Sluis, I.M., Vrenken, K.S., Pieters, R. & Stam, R.W. KMT2A-rearranged acute lymphoblastic leukemia in infants: current progress and challenges. *Haematologica* **110**, 1951–1961 (2025).
352. Pieters, R. *et al.* Outcome of Infants Younger Than 1 Year With Acute Lymphoblastic Leukemia Treated With the Interfant-06 Protocol: Results From an International Phase III Randomized Study. *J Clin Oncol* **37**, 2246–2256 (2019).
353. Brown, P., Pieters, R. & Biondi, A. How I treat infant leukemia. *Blood* **133**, 205–214 (2019).
354. Kim, R. *et al.* Genetic alterations and MRD refine risk assessment for KMT2A-rearranged B-cell precursor ALL in adults: a GRAALL study. *Blood* **142**, 1806–1817 (2023).
355. van der Sluis, I.M. *et al.* Blinatumomab Added to Chemotherapy in Infant Lymphoblastic Leukemia. *N Engl J Med* **388**, 1572–1581 (2023).
356. Lie, S.O. *et al.* Long-term results in children with AML: NOPHO-AML Study Group--report of three consecutive trials. *Leukemia* **19**, 2090–100 (2005).
357. Pession, A. *et al.* Results of the AIEOP AML 2002/01 multicenter prospective trial for the treatment of children with acute myeloid leukemia. *Blood* **122**, 170–8 (2013).
358. Creutzig, U. *et al.* Randomized trial comparing liposomal daunorubicin with idarubicin as induction for pediatric acute myeloid leukemia: results from Study AML-BFM 2004. *Blood* **122**, 37–43 (2013).
359. Rubnitz, J.E. *et al.* Clofarabine Can Replace Anthracyclines and Etoposide in Remission Induction Therapy for Childhood Acute Myeloid Leukemia: The AML08 Multicenter, Randomized Phase III Trial. *J Clin Oncol* **37**, 2072–2081 (2019).

360. Rubnitz, J.E. *et al.* Minimal residual disease-directed therapy for childhood acute myeloid leukaemia: results of the AML02 multicentre trial. *Lancet Oncol* **11**, 543–52 (2010).
361. Pollard, J.A. *et al.* Gemtuzumab Ozogamicin Improves Event-Free Survival and Reduces Relapse in Pediatric KMT2A-Rearranged AML: Results From the Phase III Children's Oncology Group Trial AAML0531. *J Clin Oncol* **39**, 3149–3160 (2021).
362. Aguirre-Ghiso, J.A. Models, mechanisms and clinical evidence for cancer dormancy. *Nat Rev Cancer* **7**, 834–46 (2007).
363. Juthani, R., Punatar, S. & Mittra, I. New light on chemotherapy toxicity and its prevention. *BJC Rep* **2**, 41 (2024).
364. Nesic, K., Parker, P., Swisher, E.M. & Krais, J.J. DNA repair and the contribution to chemotherapy resistance. *Genome Med* **17**, 62 (2025).
365. Parker, W.B. Enzymology of purine and pyrimidine antimetabolites used in the treatment of cancer. *Chem Rev* **109**, 2880–93 (2009).
366. Jordheim, L.P., Durantel, D., Zoulim, F. & Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat Rev Drug Discov* **12**, 447–64 (2013).
367. Inaba, H. & Pui, C.H. Glucocorticoid use in acute lymphoblastic leukaemia. *Lancet Oncol* **11**, 1096–106 (2010).
368. Minotti, G., Menna, P., Salvatorelli, E., Cairo, G. & Gianni, L. Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacol Rev* **56**, 185–229 (2004).
369. Emadi, A., Jones, R.J. & Brodsky, R.A. Cyclophosphamide and cancer: golden anniversary. *Nat Rev Clin Oncol* **6**, 638–47 (2009).
370. Jordan, M.A. & Wilson, L. Microtubules as a target for anticancer drugs. *Nat Rev Cancer* **4**, 253–65 (2004).
371. Lee, J.K., Kamran, H. & Lee, K.Y. L-asparaginase induces IP3R-mediated ER Ca(2+) release by targeting micro-OR1 and PAR2 and kills acute lymphoblastic leukemia cells. *Cell Death Discov* **10**, 366 (2024).
372. Ricart, A.D. Antibody-drug conjugates of calicheamicin derivative: gemtuzumab ozogamicin and inotuzumab ozogamicin. *Clin Cancer Res* **17**, 6417–27 (2011).
373. Goebeler, M.E. & Bargou, R.C. T cell-engaging therapies - BiTEs and beyond. *Nat Rev Clin Oncol* **17**, 418–434 (2020).
374. Blazar, B.R., Hill, G.R. & Murphy, W.J. Dissecting the biology of allogeneic HSCT to enhance the GvT effect whilst minimizing GvHD. *Nat Rev Clin Oncol* **17**, 475–492 (2020).
375. Park, J.H. *et al.* Long-Term Follow-up of CD19 CAR Therapy in Acute Lymphoblastic Leukemia. *N Engl J Med* **378**, 449–459 (2018).
376. Maude, S.L. *et al.* Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *N Engl J Med* **378**, 439–448 (2018).
377. Maude, S.L. *et al.* Chimeric antigen receptor T cells for sustained remissions in leukemia. *N Engl J Med* **371**, 1507–17 (2014).

378. Zugasti, I. *et al.* CAR-T cell therapy for cancer: current challenges and future directions. *Signal Transduct Target Ther* **10**, 210 (2025).
379. Soto-Feliciano, Y.M. *et al.* A Molecular Switch between Mammalian MLL Complexes Dictates Response to Menin-MLL Inhibition. *Cancer Discov* **13**, 146–169 (2023).
380. Lynch, E.J. *et al.* Revumenib for Relapsed or Refractory Acute Leukemia With a KMT2A Translocation. *Ann Pharmacother* **59**, 1108–1118 (2025).
381. Yi, Y. & Ge, S. Targeting the histone H3 lysine 79 methyltransferase DOT1L in MLL-rearranged leukemias. *J Hematol Oncol* **15**, 35 (2022).
382. Stein, E.M. *et al.* The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood* **131**, 2661–2669 (2018).
383. Chiesa, R. *et al.* Base-Edited CAR7 T Cells for Relapsed T-Cell Acute Lymphoblastic Leukemia. *N Engl J Med* **389**, 899–910 (2023).
384. Chiesa, R. *et al.* Universal Base-Edited CAR7 T Cells for T-Cell Acute Lymphoblastic Leukemia. *N Engl J Med* **394**, 152–165 (2026).
385. D'Angio, G.J. Pediatric cancer in perspective: cure is not enough. *Cancer* **35**, 866–70 (1975).
386. Robison, L.L. & Hudson, M.M. Survivors of childhood and adolescent cancer: life-long risks and responsibilities. *Nat Rev Cancer* **14**, 61–70 (2014).
387. Laetsch, T.W., DuBois, S.G., Bender, J.G., Macy, M.E. & Moreno, L. Opportunities and Challenges in Drug Development for Pediatric Cancers. *Cancer Discov* **11**, 545–559 (2021).
388. Rheingold, S.R. *et al.* Determinants of survival after first relapse of acute lymphoblastic leukemia: a Children's Oncology Group study. *Leukemia* **38**, 2382–2394 (2024).
389. Attarbaschi, A. *et al.* Outcomes of Childhood Noninfant Acute Lymphoblastic Leukemia With 11q23/KMT2A Rearrangements in a Modern Therapy Era: A Retrospective International Study. *J Clin Oncol* **41**, 1404–1422 (2023).
390. Egan, G. & Tasian, S.K. Relapsed pediatric acute myeloid leukaemia: state-of-the-art in 2023. *Haematologica* **108**, 2275–2288 (2023).
391. Saiki, R.K. *et al.* Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–4 (1985).
392. Bocher, O., Willer, C.J. & Zeggini, E. Unravelling the genetic architecture of human complex traits through whole genome sequencing. *Nat Commun* **14**, 3520 (2023).
393. Godley, L.A. Prioritization of patients for germline testing based on tumor profiling of hematopoietic malignancies. *Front Oncol* **13**, 1084736 (2023).
394. Krystel-Whittemore, M. *et al.* Cell-free DNA from nail clippings as source of normal control for genomic studies in hematologic malignancies. *Haematologica* **109**, 3269–3281 (2024).
395. Bentley, D.R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–9 (2008).

396. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–22 (2009).
397. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).
398. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–50 (2005).
399. Zheng, G.X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017).
400. Gonzalez-Pena, V. *et al.* Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc Natl Acad Sci U S A* **118**(2021).
401. Zhang, C.Z. *et al.* Calibrating genomic and allelic coverage bias in single-cell sequencing. *Nat Commun* **6**, 6822 (2015).
402. Kennedy, S.R. *et al.* Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat Protoc* **9**, 2586–606 (2014).
403. Hindson, B.J. *et al.* High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Analytical Chemistry* **83**, 8604–8610 (2011).
404. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–21 (2012).
405. Tiscornia, G., Singer, O. & Verma, I.M. Production and purification of lentiviral vectors. *Nat Protoc* **1**, 241–5 (2006).
406. Towbin, H., Staehelin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A* **76**, 4350–4 (1979).
407. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J Immunol Methods* **65**, 55–63 (1983).
408. Crouch, S.P., Kozlowski, R., Slater, K.J. & Fletcher, J. The use of ATP bioluminescence as a measure of cell proliferation and cytotoxicity. *J Immunol Methods* **160**, 81–8 (1993).
409. Koh, G., Degasperi, A., Zou, X., Momen, S. & Nik-Zainal, S. Mutational signatures: emerging concepts, caveats and clinical applications. *Nat Rev Cancer* **21**, 619–637 (2021).
410. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
411. Chang, V.Y., Basso, G., Sakamoto, K.M. & Nelson, S.F. Identification of somatic and germline mutations using whole exome sequencing of congenital acute lymphoblastic leukemia. *BMC Cancer* **13**, 55 (2013).
412. Driessen, E.M. *et al.* Frequencies and prognostic impact of RAS mutations in MLL-rearranged acute lymphoblastic leukemia in infants. *Haematologica* **98**, 937–44 (2013).
413. Adomako, J., Jimenez-Camacho, K.E., Correa-Lara, M.V.M., Nunez-Enriquez, J.C. & Schnoor, M. Acute lymphoblastic leukemia relapse: biomarkers, hopes, and challenges. *Trends Mol Med* **31**, 1032–1046 (2025).

414. Bhojwani, D. & Pui, C.H. Relapsed childhood acute lymphoblastic leukaemia. *Lancet Oncol* **14**, e205–17 (2013).
415. Nguyen, K. *et al.* Factors influencing survival after relapse from acute lymphoblastic leukemia: a Children's Oncology Group study. *Leukemia* **22**, 2142–50 (2008).
416. Bailey, L.C., Lange, B.J., Rheingold, S.R. & Bunin, N.J. Bone-marrow relapse in paediatric acute lymphoblastic leukaemia. *Lancet Oncology* **9**, 873–883 (2008).
417. Hogan, L.E. *et al.* Integrated genomic analysis of relapsed childhood acute lymphoblastic leukemia reveals therapeutic strategies. *Blood* **118**, 5218–26 (2011).
418. Mullighan, C.G. *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322**, 1377–80 (2008).
419. Yang, J.J. *et al.* Genome-wide copy number profiling reveals molecular evolution from diagnosis to relapse in childhood acute lymphoblastic leukemia. *Blood* **112**, 4178–83 (2008).
420. Ahlgren, L. *et al.* The genomic landscape of relapsed infant and childhood KMT2A-rearranged acute leukemia. *Nat Commun* **16**, 8964 (2025).
421. Tzoneva, G. *et al.* Clonal evolution mechanisms in NT5C2 mutant-relapsed acute lymphoblastic leukaemia. *Nature* **553**, 511–514 (2018).
422. Armstrong, S.A. *et al.* MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30**, 41–7 (2002).
423. Ko, R.H. *et al.* Outcome of patients treated for relapsed or refractory acute lymphoblastic leukemia: a Therapeutic Advances in Childhood Leukemia Consortium study. *J Clin Oncol* **28**, 648–54 (2010).
424. Yu, S.L. *et al.* FPGS relapse-specific mutations in relapsed childhood acute lymphoblastic leukemia. *Sci Rep* **10**, 12074 (2020).
425. Kuster, L. *et al.* ETV6/RUNX1-positive relapses evolve from an ancestral clone and frequently acquire deletions of genes implicated in glucocorticoid signaling. *Blood* **117**, 2658–67 (2011).
426. Tzoneva, G. *et al.* Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. *Nat Med* **19**, 368–71 (2013).
427. Dieck, C.L. & Ferrando, A. Genetics and mechanisms of NT5C2-driven chemotherapy resistance in relapsed ALL. *Blood* **133**, 2263–2268 (2019).
428. Meyer, J.A. *et al.* Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. *Nat Genet* **45**, 290–4 (2013).
429. Barz, M.J. *et al.* Subclonal NT5C2 mutations are associated with poor outcomes after relapse of pediatric acute lymphoblastic leukemia. *Blood* **135**, 921–933 (2020).
430. Levatic, J., Salvadores, M., Fuster-Tormo, F. & Supek, F. Mutational signatures are markers of drug sensitivity of cancer cells. *Nat Commun* **13**, 2926 (2022).
431. Jeha, S. *et al.* Improved CNS Control of Childhood Acute Lymphoblastic Leukemia Without Cranial Irradiation: St Jude Total Therapy Study 16. *J Clin Oncol* **37**, 3377–3391 (2019).
432. Coorens, T.H.H. *et al.* Clonal origin of KMT2A wild-type lineage-switch leukemia following CAR-T cell and blinatumomab therapy. *Nat Cancer* **4**, 1095–1101 (2023).

433. Bataller, A. *et al.* Myeloid lineage switch in KMT2A-rearranged acute lymphoblastic leukemia treated with lymphoid lineagedirected therapies. *Haematologica* **109**, 293–297 (2024).
434. Hobor, S. *et al.* TGFalpha and amphiregulin paracrine network promotes resistance to EGFR blockade in colorectal cancer cells. *Clin Cancer Res* **20**, 6429–38 (2014).
435. Cleary, A.S., Leonard, T.L., Gestl, S.A. & Gunther, E.J. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature* **508**, 113–7 (2014).
436. Abdul-Aziz, A.M. *et al.* MIF-Induced Stromal PKCbeta/IL8 Is Essential in Human Acute Myeloid Leukemia. *Cancer Res* **77**, 303–311 (2017).
437. Kaspers, G.J. *et al.* Improved outcome in pediatric relapsed acute myeloid leukemia: results of a randomized trial on liposomal daunorubicin by the International BFM Study Group. *J Clin Oncol* **31**, 599–607 (2013).
438. Mumme, H.L. *et al.* Identification of leukemia-enriched signature through the development of a comprehensive pediatric single-cell atlas. *Nat Commun* **16**, 4114 (2025).



As a doctor, one of the greatest challenges is finding the zebra among a herd of horses.

Amid the noise of runny noses and back pain, it means noticing what stands out and identifying the patient who does not quite fit the pattern.

Sometimes, the difference can be seen in the patient. Sometimes, the clue lies even deeper, hidden in the genomic code.

Printed by Media-Tryck, Lund 2026 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Division of clinical genetics
Department of Laboratory Medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2026:145
ISBN 978-91-8021-944-0
ISSN 1652-8220