From Department of Molecular Medicine and Surgery
Karolinska Institutet, Stockholm, Sweden

STUDIES OF CONGENITAL GENETIC ABERRATIONS BEHIND CHILDHOOD LEUKEMIA

Benedicte Bang

Karolinska Institutet

Stockholm 2023

Cover artwork by Mathilda Bang, 3 years old
All illustrations not indicated with reference produced by Benedicte Bang

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8016-964-6

Studies of congenital genetic aberrations behind childhood leukemia

Thesis for Doctoral Degree (Ph.D.)

By

Benedicte Bang

The thesis will be defended in public at Eva & Georg Klein, Karolinska Institutet Biomedicum, Solnavägen 9, 25 May 2023 at 09.00 AM

Principal Supervisor:
Professor Ann Nordgren
Karolinska Institutet
Department of Molecular Medicine and Surgery

Co-supervisor(s):
Associate Professor Fulya Taylan
Karolinska Institutet
Department of Molecular Medicine and Surgery

Dr Vasilios Zachariadis
Karolinska Institutet
Department of Oncology-Pathology

Professor Magnus Nordenskjöld
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Yenan Bryceson
Karolinska Institutet
Department of Medicine, Huddinge

Professor Arja Harila
Uppsala University
Department of Women's and Children's Health

Opponent:
Associate Professor Marjolijn Jongmans
University Medical Center Utrecht
Department of Department of Pathology
Princess Máxima Center for Pediatric Oncology

Examination Board:
Associate Professor Linda Holmfeldt
Uppsala University
Department of Immunology, Genetics and Pathology

Associate Professor Larry Mansouri
Karolinska Institutet
Department of Molecular Medicine and Surgery

Associate Professor Karin Wadt
Copenhagen University
Department of Clinical Medicine

Till Mathilda, Alicia och Masken ♥

Preface

I have often found myself ending up working with the subjects that puzzle me the most. Genetics is a prime example, being a “pulling my hair to understand” subject during high school years. Dreaming of becoming a medical doctor and curious about what all that fuzz around research was about, I found myself at a job interview in pediatric oncologist and hematologist Professor Jan-Inge Henter’s office. Despite giving the complete wrong answer to his famous question about “how I like to cook, following recipes or going free-style?” I somehow got the position and started as a “sequencing-slave” in the lab of the very same professor who coined that expression, Magnus Nordenskjöld. Under the protective wings of PhD student Marie Meeths, I spent many hours getting to know pipettes, agarose gels, thermocyclers and making sense out of the little rows of rainbow-colored bumps on the computer screen representing genetic sequence. A job much more enjoyable than the title implies, I sequenced genes in a rare disease project on Familial Hemophagocytic Lymphohistiocytosis (FHL). Just the name of that disease qualifies as dauntingly complicated ☺

Toward the end of my medical studies, I was recruited back to the very same lab as a PhD student. A journey that is now coming to an end little short of one decade, a medical doctor’s degree, a marriage (still ongoing!), two amazing daughters and a few grey hairs later. This journey has meant grasping, to me, vastly complicated but also thrillingly smart concepts and methods of genetics, rare disease and cancer. But it has also encompassed a big leap of personal development through both pleasant and challenging experiences. In the spirit of my main supervisor Ann Nordgren these years have been filled with curiosity, excitement, passion, personal engagement, endurance and lots of warmth. For that I owe her immense gratitude.

This thesis represents my gain of knowledge, understanding and scientific results from four years of PhD studies. Here I attempt to present these components through the tale of BCP-ALL etiology. It begins with congenital susceptibility through constitutional predisposing variants (paper I and III) and prenatally initiated pre-leukemic clones (paper II and III), and continues with progression to overt disease through somatic clonal evolution (paper I and III), including the putative environmental drivers of this process (paper IV).

Happy reading!

Stockholm, 21 April 2023

Abstract

Acute lymphoblastic leukemia (ALL) is the most common cancer in childhood, and most frequently (85%) of B-cell precursor type (BCP-ALL). Acquired chromosomal rearrangements or aneuploidies are the recurrent, often prenatal, initiators of BCP-ALL. These aberrations define distinct molecular subtypes that are associated with differences in prognosis and used to guide treatment. Initiating variants are disease driving, but secondary variants are required to drive progression to overt disease. Although constitutional predisposing variants are found in an increasing share of cases (10-18%), BCP-ALL etiology remains largely unknown. Recent studies have suggested that exposure to common infections may modulate progression of BCP-ALL.

The aim of this thesis was to identify, asses and quantify congenital genetic aberrations behind childhood BCP-ALL predisposition and initiation, as well as to characterize subsequent clonal evolution and identify drivers of progression to overt disease. To this end, we performed whole genome sequencing (WGS) to identify constitutional BCP-ALL-predisposing variants. In **paper I**, we reported familial predisposition mediated by a constitutional t(12;14), where haploinsufficiency of the powerful transcription factor ETV6 was suggested to cause predisposition. In **paper III**, monozygotic twins with concordant BCP-ALL shared a constitutional, maternally inherited, novel variant in *NF1,* predicted to be highly damaging. As none of the carriers has any clinical sign of the cancer syndrome neurofibromatosis type 1 (NF1), we classified the variant to be of unknown significance (VUS), but speculated its possible BCP-ALL-predisposing effect.

We developed a sensitive and quantitative method for backtracking BCP-ALL to pre-leukemic clones (**paper II**), applying chip dPCR in combination with WGS to analyze DNA from neonatal dried blood spots. In **paper II,** only one case of BCP-ALL, diagnosed at age 1 month, had detectable copy numbers of genomic breakpoint sequence at birth. Failed detection in the remaining six cases was suggested to be caused by technical and sample related limitations, and less frequently postnatal initiation. In **paper III**, WGS identified a shared somatic complex rearrangement, generating *ETV6-RUNX1*, in the BCP-ALLs of monozygotic twins. Detection at birth by dPCR failed, but identical breakpoint sequences confirmed its prenatal origin. Surprisingly, a shared (prenatal) deletion in *UBA2* was found to precede the complex rearrangement, persisting after several years in remission.

Clonal evolution of concordant BCP-ALLs was characterized in **paper III,** detecting shared and unique overlapping secondary putative driver variants, supporting independent although convergent clonal evolution. In **paper I,** 7-10 secondary putative driver variants, in genes recurrently targeted in childhood ALL, were identified in BCP-ALLs with *ETV6*-mediated predisposition. This further supported that secondary drivers are required for progression, although phylogenetics of somatic events in *ETV6*-predisposed cases remains to be delineated. In **paper IV**, we assessed a Swedish population-based cohort of 1380 BCP-ALL cases and used GARIMAX to demonstrate informative seasonal variation in onset and interpreted peak onset to fall in August. Four explanatory models, related to exposure to common infections as a driver of final progression to overt disease, were suggested. The likelihood of each model depends on still unknown induction time of childhood BCP-ALL.

Together, these studies add to our understanding of; congenital susceptibility to BCP-ALL through constitutional predisposing variants and prenatally initiated pre-leukemic clones, progression to overt disease through somatic clonal evolution and the genetic and environmental drivers of this process.

List of scientific papers

I. Järviaho T*, Bang B*, Zachariadis V, Taylan F, Moilanen J, Möttönen M, Smith CIE, Harila-Saari A, Niinimäki R[#] and Nordgren A[#]. **Predisposition to Childhood Acute Lymphoblastic Leukemia Caused by a Constitutional Translocation Disrupting ETV6.** *Blood Advances* 3, no. 18 (September 24, 2019): 2722–31.

II. Taylan F, Bang B, Ivanov Öfverholm I, Tran A-N, Heyman M, Barbany G, Zachariadis V and Nordgren A. **Somatic Structural Alterations in Childhood Leukemia Can Be Backtracked in Neonatal Dried Blood Spots by Use of Whole-Genome Sequencing and Digital PCR.** *Clinical Chemistry* 65, no. 2 (February 1, 2019): 345–47.

III. Bang B, Eisfeldt J, Barbany G, Harila-Saari A, Heyman M, Zachariadis V, Taylan F[#] and Nordgren A[#]. **A Somatic UBA2 Variant Preceded ETV6-RUNX1 in the Concordant BCP-ALL of Monozygotic Twins.** *Blood Advances* 6, no. 7 (April 4, 2022): 2275–89.

IV. Bychkov G*, Bang B*, Engsner N, Heyman M, Skarin Nordenvall A, Tettamanti G, Ponten E, Jörnsten R, Strannegård C[#] and Nordgren A[#]. **Evidence of seasonal variation of childhood acute lymphoblastic leukemia in Sweden.** *Manuscript, 2023.*

*Shared first authorship
[#]Shared senior authorship

Related scientific papers

V. Järviaho T, Zachariadis V, Tesi B, Chiang S, Bryceson YT, Möttönen M, Niinimäki R, Bang B, Rahikkala E, Taylan F, Uusimaa J, Harila-Saari A, Nordgren A. **Microdeletion of 7p12.1p13, including IKZF1, causes intellectual impairment, overgrowth, and susceptibility to leukaemia.** *British Journal of Haematology* 185, no.2 (July 13, 2018): 354–357.

VI. Meeths M, Chiang SCC, Wood SM, Entesarian M, Schlums H, Bang B, Nordenskjöld E, Björklund C, Jakovljevic G, Jazbec J, Hasle H, Holmqvist B-M, Rajić L, Pfeifer S, Rosthøj S, Sabel M, Salmi TT, Stokland T, Winiarski J, Ljunggren H-G, Fadeel B, Nordenskjöld M, Henter J-I, Bryceson YT. **Familial hemophagocytic lymphohistiocytosis type 3 (FHL3) caused by deep intronic mutation and inversion in UNC13D.** *Blood* 118, no.22 (Nov 24, 2011): 5783–5793.

Contents

List of abbreviations

AAIR	Age-adjusted incidence rate
AD	Autosomal dominant
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
AR	Autosomal recessive
AYA	Adolescents and young adults
BCP-ALL	B-cell precursor acute lymphoblastic leukemia
chr	Chromosome
CNA	Copy number alteration
CNV	copy number variant
DNA	Deoxyribonucleic acid
DS	Down Syndrome
EFS	Event-free survival
FISH	Fluorescence in situ hybridization
G-banding	Giemsa banding (karyotyping)
GWAS	Genome-wide association study
HeH	High hyperdiploidy
HIC	High Income Countries
HR	High risk
iAMP21	Intrachromosomal amplification of chromosome 21
IGV	Integrative Genomics Viewer
IR	Intermediate risk
LMICs	Low- and Middle-Income Countries
LoD	Limit of Detection
LOH	loss of heterozygosity
MPS	Massive Parallel Sequencing
MRD	Minimal residual disease / Measurable residual disease

NOPHO	Nordic Society of Pediatric Hematology and Oncology
OMIM	Online Mendelian Inheritance in Man
OS	Overall survival
PCR	Polymerase chain reaction
> dPCR	digital PCR
> ddPCR	digital droplet PCR
> qPCR	quantitative PCR
> RT-qPCR	real-time quantitative PCR
RNA	Ribonucleic acid
RAG	Recombination-activating gene
SMT	Somatic mutation theory
SNV	Single nucleotide variant
SR	Standard risk
SV	Structural variant
TOFT	Tissue organization field theory
UCB	Umbilical cord blood
VUS	Variant of unknown significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHO	World Health Organization
WT	Wild type

1 Introduction

1.1 Cancer genetics

Cancer, when defined as a phenomenon, is an evil and destructive force that is hard to contain and eradicate. This is reflected in the biological characteristics of cancer in human disease as conceptualized by Hanahan and Weinberg in 2000. Six acquired physiological characteristics were proposed as the "Hallmarks of cancer"[1]: evading apoptosis, self-sufficiency in growth signals, insensitivity to anti-growth signals, limitless replicative potential, activating invasion and metastasis, sustained angiogenesis. Hallmarks have since then been revised and, reflecting major progress made in cancer biology research, the addition of two emerging hallmarks and two enabling characteristics was made in 2011[2]: deregulating cellular energetics, avoiding immune destruction, genome instability and mutation and tumor-promoting inflammation. Another two of each were added in 2022[3]: unlocking phenotypic plasticity, senescent cells, nonmutational epigenetic reprogramming and polymorphic microbiomes (***Figure 1***). These hallmarks are well aligned with a reductionist view of cancer development and progression; the somatic mutation theory (SMT). The foundation of SMT is the view of cancer originating from a single genetic aberration causing the host cell to undergo clonal expansion and acquire additional somatic variants ultimately causing cancer, all facilitated by genetically induced altered characteristics.[4] An alternate and organicist view of cancer biology, opposing SMT, is the tissue organization field theory (TOFT), seeing tissue disorganization as the cause of cancer while genetic variants are considered no more than a consequence of disorganization's abnormal interactions between parenchymal and stromal tissue.[5–8]

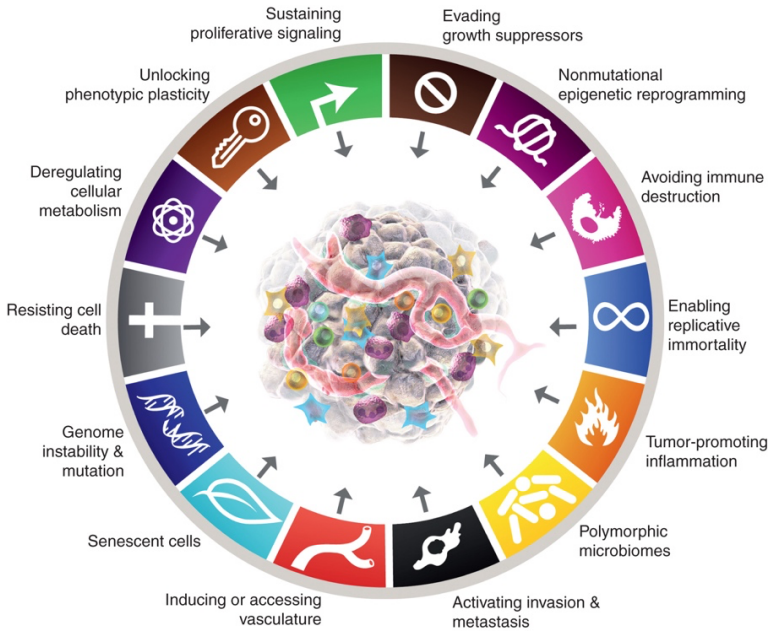


Figure 1. *Hallmarks of Cancer - including originally proposed and prospective new additions. Figure reused in full from the original paper by Hanahan 2022[3] with permission from the publisher.*

The basis of SMT cancer genetics still stands upon the 2-hit hypothesis originally proposed by Nordling in 1953[9] and later confirmed in studies of retinoblastoma and the *RB1* gene by Knudson in 1971[10]. This indirectly led to the definition and discovery of tumor suppressor genes, which upon loss or inactivation by genetic aberrations results in uncontrolled proliferation.[11] In addition, oncogenes were already acknowledged to drive arise from proto-oncogenes and drive cancer following activation by genetic aberrations.[12] Later, DNA-repair genes have also been implicated in cancer initiation and progression.[13] These genes are responsible for a wide repertoire of repair mechanisms ensuring genomic sequence is preserved. More recently, regulatory changes, such as methylation patterns, affecting the genome have demonstrated a role for epigenetics in the development and progression of cancer.[14] Over the last two decades, the emergence and refinement of massive parallel sequencing (MPS) techniques we have seen over the last two decades has skyrocketed our ability to identify and characterize genetic and epigenetic alterations and transcriptional effects in cancer in a time and cost efficient manner.[15–17] Similar advances of high-throughput methods for studying proteomics and metabolomics have enhanced our understanding of functional outcomes. The novel insight following these developments to the genomic landscapes of tumors have demonstrated great interpatient heterogeneity in malignancies of the same tissue of origin. Adding further complexity, each individual tumor displays large intratumor heterogeneity, attributed to a maze of genetically divergent sub-clonal populations. This has pinpointed the vast complexity of cancer on the molecular level.

In human disease, differentially weighted combinations of nature (genetics, epigenetics) and nurture (environmental exposures) lay the foundation of disease etiology, cancer being no exception. Cancer is indeed a genetic disease, according to SMT, though less commonly a heritable one. Nonetheless, approximately 8% of adults with cancer have been found to carry a cancer predisposing constitutional variant, either inherited or de novo.[18] Proportions of constitutional variants varied distinctly by cancer type. In addition to genetic predisposition, cancer typically originates from the combined burden of acquired genetic variants, often arising sequentially in multiple genes as reviewed by Vogelstein et al.[19] These variants accumulating linearly during life are the result of random errors in the replicatory machinery termed “R” and has therefore been come to known as the R-factor in cancer etiology. Not all accumulated variants are of equal importance, the bulk are regarded as passive passengers while a few (2-8) are driver variants providing a selective advantage by promoting growth. The frequencies of somatic pathogenic variants, and with them the lifetime risk of cancer, differ markedly between tissues.[20] This relationship has been proposed to depend on differences in rate of stem cell divisions and thus R-factor.[21] Beyond the hereditary and R-factor, epidemiological studies have over the last centuries identified a plentitude of environmental factors, such as UVB light, radioactive radiation and various chemicals, capable of causing cancer upon prolonged exposure. The molecular mechanism has often been found to be induction of DNA-damage leading to emergence of cancer driving variants. Altogether, these etiological factors are by no means isolated entities, rather closely interlaced.

The hallmarks of cancer have, since they were first proposed in 2000, been praised and widely adopted but also questioned and reevaluated. Naturally by adherers of TOFT, but also by others for their limitations in guiding our understanding of cancers enormous complexity and heterogeneity on multiple functional levels, from molecular/single cell to microenvironmental and all in-between.[22] Thus, hallmarks have been rendered insufficient as a framework for cancer research nurturing calls for and suggestions of a systemic and evolution based alternative in which hallmarks may be incorporated.[22–25]

1.2 Childhood cancer

As reviewed by Pfister et al,[26] childhood cancers differ from adult cancers in many ways; different spectrum of cancer types, embryonic origin in mesoderm and ectoderm as opposed to adult's predominant epithelial origin, occurring in immature cells and often caused by a lone oncogenic fusion, while rates of somatic pathogenic variants are low. The most apparent difference may be that of etiology. While adult cancers typically originate from the combination of carcinogen exposure, the R-factor and a portion of heredity; the etiology of childhood cancer is largely unknown. However, constitutional genetic predisposition has more recently been shown to account for an increasing number of cases, estimated to 10-18%.[27–29] As for adult malignancies, contribution of constitutional predisposition varies by cancer type. Hematological malignancies as a group display less predisposition even though individual types and subtypes may have a more substantial contribution.[30] For all the ways in which they differ from adult cancers and for best possible clinical applicability, childhood cancers have recently been dedicated an independent WHO classification.[26]

Childhood cancer is a rare disease, accounting for 1% of malignancies in humans of all ages.[31] Global incidence of cancer in children and adolescents (0-19 years) has been estimated to 413 000 cases yearly, out of a world population of 2.6 billion in this age group according to the United Nations[32]. Though, it has been approximated that almost half (44%) of childhood cancer cases will die without being diagnosed or treated, mainly due to limited access to adequate care in low and low-middle income countries.[33] About 89% of the world's children (age 0-19 years) live in LMICs, and they account for 95% of the mortality from childhood cancer worldwide.[34,35] Still, despite access to modern treatment strategies and advanced supportive care, cancer is yet the most common disease-related cause of death in children in high income countries (HIC).[26] In response, the World Health Organization (WHO) has set a global survival target of 60% for all children with cancer, with the goal of saving a million more lives by the year 2030.[36]

The average age at diagnosis for cancer in children and adolescents aged 0-19 years is 10 years, more specifically 6 years in children aged 0 to 14, and 17 years in adolescents aged 15 to 19.[37] Although rare, cancer in children is vastly heterogenous, well-illustrated by the 20 types and hundreds of subtypes classified by the WHO to date.[26] Although substantial differences are seen in different parts of the world and for different ethnicities, leukemia is the most common cancer worldwide in 0-19 year old's, followed by brain- and other CNS-tumors and lymphomas.[38] The same pattern is seen in children (0-14 years) while adolescents (15-19 years) have an more lymphomas, followed by epithelial tumors and melanomas, brain and other CNS-tumors and leukemias.

1.3 Childhood B-cell precursor Acute Lymphoblastic Leukemia (BCP-ALL)

Acute lymphoblastic leukemia (ALL) is a malignancy characterized by the rapid clonal expansion of lymphoid blast cells in the bone morrow. Blasts are immature cells in which differentiation has been stalled and proliferation is left unchecked, leading to depression of other hematopoietic linages of the bone marrow. Consequently, patients experience symptoms such as bone pain, easy bruising and increased bleeding tendency in mucous membranes (thrombocytopenia), pallor and fatigue (anemia), susceptibility to infections (leukopenia) at diagnosis.

1.3.1 Descriptive epidemiology of childhood ALL

Childhood ALL is the most common childhood cancer worldwide, accounting for approximately 30% of all pediatric cancers.[39] The disease can be of either T- or B-cell precursor type. The latter (BCP-ALL) is the predominant subtype constituting 85% of all childhood ALL[40], and thus approximately 25% of all childhood cancer.

In Sweden, the incidence of ALL is 4.2 per 100 000 children (<15 years of age)[41], a number comparable to most parts of the world[38]. The general view is that incidence numbers have steadily increased at a low rate of 0.7-1.0% per year[42–44] while some claim they have remained stable over the last decades[45,46]. Global differences in incidence numbers of ALL are notable, with lower incidence rates in LMIC's such as India (1.9 per 100 000 children) and Sub-Saharan countries (<1 per 100 000 children).[45] In a recent international report the age-standardized incidence rate (AAIR, Segi World Standard Population) of leukemias (all types) in 0-14 year old's was lowest in Sub-Saharan Africa (12.5, total average in all regions 46.4).[38] These lower incidence rates may in part be due to underdiagnosis, as described above, and incomplete registration[38], well-illustrated by markedly lower incidence rates across all childhood cancer types in this region.[38] Nonetheless, this does not explain the full extent of these disparities. Interestingly, the significant peak of BCP-ALL onset seen at age 2-5 years in HIC is markedly less pronounced or even non-existing in regions with lower incidence rates.[47] The reason for this is not known, but has been hypothesized to depend on differences in general hygiene conditions and the consequential extent of exposure to common microbes (see section 1.4.2).

Another intriguing observation is that incidence numbers among ethnic groups differ notably in nations such as USA with high grade of ethnic diversity in the population. Measuring AAIRs (2000 U.S. Standard Population), Hispanic children and adolescents aged 0-19 years in the US have been reported to develop ALL more often than non-Hispanic whites (AAIR 44.5 compared to 33.7), while Asians/Pacific Islanders (30.6) and African Americans (18.5) were less commonly affected than non-Hispanic whites.[48] Similar observations were made in a recent study by Steliarova-Focher et al. where all leukemias in children 0-14 years in the US were found with the following distribution of AAIR (Segi World Standard Population) per ethnic group: non-Hispanic White 54.5, Hispanic 65.4, Asian/Pacific Islander 43.9 and African American 31.3. Of note, AAIR of Hispanic 0-14 year old's (65.4, total average all regions 46.4) was the highest in the world.[38] These observations could be explained by differences in both genetic and environmental factors. However, an observed larger increase in incidence of ALL among 0-19 year old Hispanics compared to other ethnicities[48] is more likely of environmental origin.

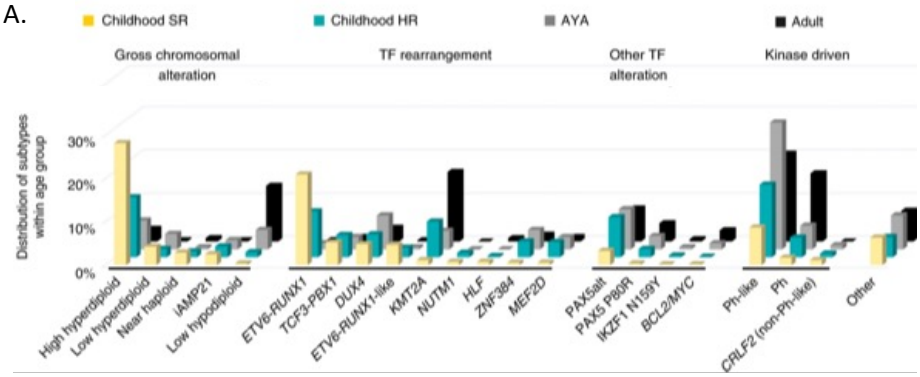
1.3.2 Genetic characterization and subtype classification of childhood BCP-ALL

The clinical applicability of detailing the heterogenous genomic landscape of cancer is sometimes questioned.[22,24,49] Nonetheless, childhood ALL is a disease where genomic characterization of recurrent aberrations has been of great value in clinical practice. BCP-ALL is a genetically heterogenous disease[1-3] in which molecular subtype classification has traditionally been based on recurrent genetic aberrations, predominantly chromosomal translocations or non-random loss or gain of entire chromosomes (aneuploidy) but also other chromosomal aberrations and their associated gene expression profiles.[50,51] Translocations either move strong enhancers or create chimeric oncogenic fusion genes and commonly involve genes coding for transcription factors, epigenetic modifiers, cytokine receptors and tyrosine kinases.[52] These subtype-defining genetic events are also considered

initiating and disease driving. Many have clinical implications to risk stratification and prognosis, although not all are yet used in clinical practice.

Until recently 75-80% of BCP-ALLs were given a molecular diagnosis, leaving a substantial portion of cases without means of risk stratification for treatment.[39] This number has improved substantially in recent years thanks to the application of MPS techniques, not least transcriptome sequencing.[53] Still, in 5% of childhood BCP-ALL cases (10% in adolescents and young adults) a molecular diagnosis cannot be determined, commonly grouped as "B-other". To date, 24 genetically distinct subtypes with clinical implications to prognostics and therapeutics have been characterized.[51,54]. In the recent inaugural WHO Classification of Pediatric Tumors[26], 11 clinically used genetic subtypes of BCP-ALL were recognized (***Figure 2B***).

A.



B.

B-cell precursor lymphoblastic leukemia

B-LBLL with t(9;22)(q34.1;q11.2); BCR::ABL1
B-LBLL with t(v;11q23.3); KMT2A-rearranged
B-LBLL with t(12;21)(p13.2;q22.1); ETV6::RUNX1
B-LBLL with hyperdiploidy, high
B-LBLL with hypodiploidy, near-haploid
B-LBLL with hypodiploidy, low
B-LBLL with hypodiploidy, high
B-LBLL with t(5;14)(q31.1;q32.3); IGH::IL3
B-LBLL with t(1;19)(q23;p13.3); TCF3::PBX1
B-LBLL, BCR::ABL1-like (Philadelphia-like B-ALL)
B-LBLL with iAMP21

Figure 2. *(**A**) Distribution of BCP-ALL subtypes within age groups: childhood (0-15 years), adolescents and young adults (AYA)(16-39 years), adult (40-79 years). Treatment stratification indicated by SR=Standard risk and HR= High risk. Grouped as gross chromosomal alteration, transcription factor (TF) rearrangement, other TF alteration, kinase-driven and others. Adapted from Gu et al.*[55] *with permission from the publisher. (**B**) Summary of genetic subtypes of BCP-ALL defined by the 2022 Inaugural WHO Classification of Pediatric Tumors as reported by Pfister et al.*[26]

Frequency of molecular subtypes differ by age[50] and clinical risk stratification (SR, HR)[51,55] (***Figure 2A***) but also by study i.e. investigated population. In the latter, ethnic differences, as reviewed by Hein et al., likely contribute to variations in frequency of subtypes, adding to the complexity of BCP-ALL's genetic origins. For example, the common *ETV6-RUNX1* subtype is markedly less common among the indigenous people of New Zeeland (Māori), East Asians and Hispanics. With that said, t(12;21)/*ETV6-RUNX1* and High Hyperdiploidy (HeH)[51,56] are generally accepted to be the most

common subtypes, constituting approximately 25 and 30% of cases respectively.[50] Most HeH and t(12;21)/*ETV6-RUNX1* cases present at young age and are the major contributors to the common peak of BCP-ALL onset at 2-5 years of age, however increasingly rare with older age and almost absent in adults.[50] In infants (<1 year of age) *KMT2A*-rearrangements (previously *MLL)* are most common, accounting for 70-75% of cases[57] while t(9;22)/*BCR-ABL1* and *BCR-ABL1*-like are much more common in adolescents and young adults[50,55].

Although diagnostic routines differ between centers, the standard methods used for subtype classification are different combinations of karyotyping with G-banding, fluorescence in-situ hybridization (FISH), reverse transcription PCR and arrays (SNP-array/arrayCGH) or MLPA. The use of WGS in first line diagnostics has been shown to be superior to previous methods[58], but has commonly not yet been implemented in standard diagnostic routines. Nonetheless, all families with a child newly diagnosed with cancer in Sweden are today offered to participate in a project within Genomic Medicine Sweden (GMS), where WGS analysis of the tumor (in addition to standard diagnostic methods) and constitutional material is performed to improve diagnosis, detect targets for treatment and identify actionable predisposing variants. Similar projects are also in place in other countries.

After initiation of BCP-ALL by subtype-defining genetic events, further somatic variants are typically required for progression to overt leukemia (for which supporting epidemiological observations and rationale are covered in section 1.4.1.1 below). A recent study found a median of 4 such putative disease driving variants in every BCP-ALL (including initiating events) although frequencies differed by subtype.[51] These variants were most often CNV's (deletions) and commonly affected B-cell transcription factor genes, such as *ETV6*, *PAX5* and *IKZF1*, epigenetic regulators and Ras signaling. Variants are often recurrent and found in several different subtypes, although more frequently associated with a specific one.[40] An example is the strong association between t(12;21)/*ETV6-RUNX1* leukemia and deletion of the non-translocated *ETV6* allele.[59] These additional somatic hits may have impact on the prognosis and treatment, as exemplified by deletion of *IKZF1*, an independent negative prognostic marker present in 15% of all BCP-ALL[60]. Although childhood BCP-ALL most often does require multiple hits for overt disease the overall rate of pathogenic variants is significantly lower than for solid tumors. This may be explained by the fact that precursor hematopoietic cells, from which leukemias are thought to arise, naturally possess some of the qualities defined as hallmarks of cancer.[44]

1.3.3 Treatment and Survival in childhood BCP-ALL

The overall goal of cancer treatment is to maximize cure while limiting severe toxicities. BCP-ALL is treated with a combination of chemotherapy drugs and corticosteroids, aimed at killing and preventing further growth of leukemic cells. Treatment of childhood BCP-ALL is based on three pillars. First, the induction phase aims to rapidly kill of the bulk mass of leukemic cells and induce remission. Next, the consolidation phase is entered once remission is reached (if not treatment enters intensification) and serves to maintain remission and eliminate any remaining leukemic cells still residing in the body. Finally, the maintenance phase continues over the course of approximately 2 years aiming to maintain remission long-term. Evaluation of Minimal/Measurable Residual Disease (MRD) in bone marrow is used at defined checkpoints during treatment to monitor treatment response and sustained remission. Patients are stratified into three main risk groups to which treatment is tailored: standard (SR), intermediate (IR) and high risk (HR). Risk stratification is based on factors such as age, white blood

cell counts, genetic subtype and other genetic risk markers found at diagnosis, as well as treatment response evaluated by MRD after induction and consolidation phase. Clinical risk factors in ALL are summarized in ***Table 1***.

In many parts of the world larger organizations have been formed to standardize treatment for children, adolescents and young adults with BCP-ALL (and other childhood cancers). In Sweden, children are treated according to protocols from the Nordic Society of Pediatric Hematology and Oncology (NOPHO). NOPHO brings together Sweden, Denmark, Norway, Finland, Iceland, Lithuania and Estonia in the fight against childhood cancers. The latest protocol, ALLTogether implemented since 2019 in Sweden, has been developed by a consortium of peer pediatric oncology organizations in addition to NOPHO; from France, Ireland, UK, Belgium, the Netherlands, Germany and Portugal. This treatment protocol, just as its predecessors, also functions as a clinical trial with several investigative arms to continuously follow up and improve on treatment strategies. In Sweden, all clinical data are registered in a national quality registry, the Swedish Childhood Cancer Registry.

Survival rates in childhood BCP-ALL have improved remarkably over the last few decades. Today, 90% will become long-term survivors, but there is a significant variation among subtypes.[39,61] In Sweden 10-year survival increased from 5% in the early 1970's to 85% in 2010.[41] This dramatic improvement is mainly a result of adjustments of treatment regimens in combination with advances in supportive care. More recent numbers show that the NOPHO ALL-2008 trial resulted in a 5-year EFS of 85% and OS of 91%.[62] These results are comparable to other European and American protocols as reviewed by Inaba et al.[61]

Prognostic factors	Positive	Negative
Patient and clinical characteristics		
- Age at dx	1 to 9 years	<1 or >9 years
- Sex	Female	Male
- Ethnicity	Caucasian, Asian	Hispanic, African American
- Down syndrome	No	Yes
- WBC counts at dx	<50x10*9/L	= or >50x10*9/L
- CNS involvement at dx	CNS 1	CNS 2&3, traumatic tap with blasts
- Testicular involvement	No	Yes
Genetic	High Hyperdiploidy (HeH) *ETV6-RUNX1* *NUMT1* -rearrangement	Hypodiploidy (<44 chrs) *KMT2A* -rearrangement *BCR-ABL1* *BCR-ABL1-like* *TCF3-HLF* *MEF2D* -rearrangement *iAMP21* *BCL2 or MYC rearrangements*
MRD	Negative Continiously decreasing and becoming negative	Positive Increasing and/or persistently positive while monitored

Table 1. *Positive and negative prognostic factors in childhood BCP-ALL. Dx=diagnosis. Table adapted from Inaba and Mullighan 2020*[61] *with permission from the publisher.*

The outcomes of recent treatment protocols, where over-treatment is now one of the failures of current strategies, indicate we have reached ways end in improving outcome through general intensification of chemotherapy. At the same time, under-treatment leading to relapse is still a problem in some cases. Therefore, there is a need for refined molecularly based stratification to sift out individual patients

eligible for de-intensification, and those in need of intensified treatment through targeted therapy, to reduce toxicity but sustain antileukemic effect.

1.4 Etiology of childhood BCP-ALL

Extensive molecular and epidemiological studies of childhood BCP-ALL have indeed identified individual contributing risk factors and generated hypotheses about the causes of this disease. Nonetheless, a fuller understanding of etiology from initiation and through progression to disease onset remains for BCP-ALL. To guide the search for intrinsic and extrinsic etiological factors it has been essential to delineate the timing and characteristics of disease initiation, progression and onset. In this section, the current knowledge and hypotheses regarding these components will be outlined.

1.4.1 Congenital predisposition – Prenatally initiated pre-leukemic clones and constitutional variants

There are two different ways in which an individual may be predisposed to develop BCP-ALL: being a carrier of a pre-leukemic clone initiated during fetal life and carrying a predisposing genetic variant (inherited or *de novo*). Both ways of predisposition are congenital and based on the presence of a genetic aberration, one found only in the pre-leukemic cells (clonal) and the other in every cell of the body (constitutional). Although not required, predisposing genetic variants contribute to initiation of a pre-leukemic clone.

1.4.1.1 Prenatally initiated pre-leukemic clones

Scientists have long strived to elucidate different aspects of BCP-ALL biology. One such aspect is the timing and characteristics of disease initiation. The recurrent translocations and rearrangements characterizing subtypes of BCP-ALL are considered the initiating events of disease, and are thought to establish a pre-leukemic clone that may progress to overt leukemia through the acquisition of additional driver variants. Even prior to the recognition of these molecular events the observation of concordant childhood BCP-ALL in monozygotic twins led to speculations that the timing of leukemia initiation could be prenatal. The theory was that leukemia was initiated in one twin and spread to the other through vascular transfusion in the shared placenta during fetal life.[63,64] Later, Ford *et al.*[65] were the first to demonstrate that concordant BCP-ALL in monozygotic twins shared identical gene fusion sequences, in that report shown for non-constitutional *KMT2A(MLL)*-rearrangements. As fusion sequences are unique to each individual leukemia, this provided molecular evidence of both a common clonal origin and prenatal initiation of *KMT2A*-rearranged leukemia. Subsequently, evidence emerged of molecularly concordant *ETV6-RUNX1* fusion BCP-ALLs in monozygotic twins, supporting prenatal initiation also for this recurrent subtype.[66,67] To date, studies backtracking leukemias to pre-leukemic clones in neonatal samples from both twin and non-twin cases have confirmed prenatal initiation of *KMT2A*-rearrangements, *ETV6-RUNX1* fusion, high hyperdiploidy (HeH), *TCF3-PBX1* fusion generated by t(1;19) and *BCR-ABL1* fusion created by t(9;22), as reviewed by Hein et al.[56]

In addition to the timing of leukemia initiation, studies of BCP-ALL's prenatal origin have also taught us about the nature of different leukemia initiating events. Although initiated in utero, molecularly concordant BCP-ALLs in a twin pair display variable latency (time until diagnosis) depending on their molecular subtype. For example, twins with infant ALL, in which *KMT2A*-rearrangements are predominant, are by definition diagnosed within the first year of life and thus the latency until

diagnosis is short. The concordance rate of infant ALL in monozygotic twins is nearly 100%.[68] These observations collectively suggest that *KMT2A*-rearrangements are strong oncogenic drivers. In further support of this, whole genome sequencing efforts have revealed a low rate of pathogenic somatic variants in BCP-ALLs with *KMT2A-AF4* fusions, implying that *KMT2A*-rearrangements are either alone sufficient to cause leukemia or efficient in causing the additional driving events required for overt leukemia.[69] *ETV6-RUNX1* positive leukemias on the other hand, in which incidence peaks at age 2-5 years[50], have longer latency (median 4 years, range 1-12)[70–73]. This subtype also displays larger differences in latency between concordant cases, as described by Wiemels *et al.* in a monozygotic twin pair diagnosed at age 5 and 14 years respectively, thus separated by 9 years.[67] At 10-15%, concordance rates are also markedly lower in non-infant leukemias.[68] Altogether, these observations suggest that the *ETV6-RUNX1* fusion gene has a weaker oncogenic effect and that additional hits are of greater importance for disease progression. The potential for in utreo transfusion of pre-leukemic clones between twins is assumed to be equal regardless of the initiating genetic aberration. However, the low concordance rates in ETV6-RUNX1-positive leukemias led to suspicions that all pre-leukemic clones may not progress into overt leukemia.

To further investigate this notion, several efforts have been made to elucidate the incidence of *ETV6-RUNX1* chimeric fusion sequences in healthy newborns, so called postnatal screening studies. Mori *et al.*[74] were the first to report an incidence of 1% in umbilical cord blood (UCB) from healthy neonates, a number far exceeding the incidence of the corresponding leukemia. The implication would be that a vast majority of carriers do not develop overt disease. Following this initial report, several studies demonstrated frequency's ranging from 0.01% to 8% using either neonatal blood spots or cord blood samples and various combinations of RNA- and DNA-based techniques.[75–82] Surprisingly, many of the reported frequency's far exceed that of the corresponding leukemia, indicating that the vast majority never progress to overt leukemia. Reports have therefore been accompanied by much debate regarding methodological issues and the true frequency of pre-leukemic clones in healthy neonates at birth. The most recent report addressed some of the previous methodological issues using the DNA-based method GIPFEL, detecting *ETV6-RUNX1* fusions in 5% of analyzed UCB samples from healthy neonates.[83] The above studies further emphasized the importance of secondary genetic events for progression of *ETV6-RUNX1* positive leukemia but also raised questions about potential modulating factors of both intrinsic and extrinsic nature (discussed further in section 1.4.2). Altogether, there is now convincing evidence that pre-leukemic clones are often initiated prenatally and, in the case of *ETV6-RUNX1*-fusion, occur more frequently than the corresponding overt leukemia is observed.

Except for the contribution from constitutional genetic variants (discussed in the following section), the etiological mechanisms behind initiating genetic variants remain to be understod, although some argue they are no more than "developmental accident"[84]. Some epidemiological studies have suggested maternal influenza, varicella, rubella or genital infection may increase risk of the child developing BCP-ALL, yet not all reports support these findings.[85–87] No leukemia-transforming virus has yet been molecularly detected.[84] Other prenatal environmental exposures associated with an increased risk of BCP-ALL are maternal age and exposure to paint, pesticides and large coffee intake during pregnancy, as reviewed by Williams et al.[88] Interestingly, a small number of studies have identified environmental exposures specifically associated to certain molecular subtypes of BCP-ALL. Only one, home paint exposure during pregnancy, was a clear pre-natal exposure and could thus be implicated in disease initiation. Nevertheless, all of the above exposures lack convincing support.

Studies of prenatal origin of BCP-ALL have vastly increased our understanding of the timing and nature of initiating events. However, methods for detection and quantification of initiating genetic variants in prenatal samples have been associated with technical limitations, illustrated by the large discrepancy in rates of *ETV6-RUNX1* found by different methods among healthy neonates (reviewed by Hein et al.).[56] Of the most frequently used methods, nested PCR is sensitive but lacks quantitative ability, while qPCR is a quantitative analysis but has limited sensitivity compared to nested PCR. Both are subject to contamination. Most often RNA from UCB has been used which allows detection of recurrent fusion sequences without designing specific primers for each patient, which is required when analyzing genomic breakpoints in DNA. This is a huge advantage when backtracking or screening more than a few samples. But, to its disadvantage, RNA is utterly instable. The recent application of GIPFEL allowed use of streamlined assays of stable DNA instead of RNA, facilitating use of DNA in backtracking or prenatal-screening studies of larger numbers of samples.[75] However, the quantitative ability and sensitivity of qPCR applied in this study left room for improvements. This motivated our studies in **paper II**, developing a method based on the more sensitive, quantitative less contamination exposed chip dPCR, in a first step combined with WGS-guided design of patient specific assays. Regarding the source of material, RNA or DNA, for backtracking and/or screening studies, both umbilical cord blood (UCB) and dried neonatal blood spots have been used. Coming from bio-banked Guthrie cards, neonatal blood spots are readily available compared to UCB. They are routinely sampled for clinical purposes (newborn screening of metabolic disease), which eludes the need for additional sampling as for UCB. Guthrie cards are also easy and cheap to handle and store. These factors motivated our choice of blood spots as source of DNA. On the downside, DNA/RNA-purification from dried blood spots do not allow for cell sorting prior to purification, which is one way of increasing the likelihood of "catching" traces of pre-leukemic clones at birth.

1.4.1.2 Constitutional variants predisposing to childhood BCP-ALL

The contribution of genetic predisposition to childhood cancer has previously been estimated to 10%, originally derived from a study by Narod *et al.* in 1991.[89] A more recent study identified predisposing variants in 18% of cases, in a prospective cohort of childhood cancers including newly diagnosed (85%) and relapsed/refractory (15%) cases. However, patients were unselected for tumor type, and thus not a representative distribution of diagnoses (n=309)[28]. For children and adolescents with BCP-ALL, pathogenic variants in known cancer predisposition genes have previously been found in 4.4% of a large cohort (subtypes not specified).[27] However, these results likely underestimate the prevalence due to strict inclusion criteria of "known" predisposition genes analyzed and a conservative evaluation of variant pathogenicity.[90] Shortcomings in both clinical recognition (non-cancer symptoms may be subtle or absent) and molecular diagnostics have posed a challenge to find the "true" fraction of cases attributable to genetic predisposition. To date, with the substantial number of familial and sporadic cases of childhood BCP-ALL reported to carry a rare constitutional variant with highly penetrant predisposing effect, there is solid evidence for the contribution of genetic predisposition to BCP-ALL initiation and possibly also progression.[91–93] In addition, GWAS-studies have identified a number of common susceptibility loci associated with BCP-ALL but with low penetrance (reviewed by Rüchel et al).[94]

There is a variety of ways in which one can categorize constitutional genetic predisposition. Predisposing genetic aberrations can either be *de novo* or inherited, although family history is not always evident in the latter case. Most cancer predisposition syndromes described to date are autosomal dominant (AD) with incomplete penetrance, although autosomal recessive (AR) and X-

linked patterns exist. Consequently, variants can be classified according to their effect size i.e., the penetrance of a certain phenotypic expression. A single variant can express several phenotypes, both cancer and non-cancer, for which the effect size may differ. A second variable for variant classification is allele frequency, meaning the frequency of a particular variant in the general population. When searching for novel variants causing rare conditions, such as childhood cancer predisposition, low allele frequency (defined as minor allele frequency (MAF) <1%) is commonly used as one of many criteria for pathogenicity. The assumption is that a rare conditions, by definition, will not be caused by variants common in the general population. Also, within a cancer syndrome, variants affecting the same gene may confer susceptibility to a wide range of cancer types, both adult and pediatric. Thus, even if a specific variant in a known cancer predisposition gene is often associated with a certain type of leukemia, it is not necessarily a common phenotype of other variants in that gene.[29] Based on these three preconditions, this overview will focus on predisposing genes and syndromes where variants are rare but with variable effect size and where BCP-ALL is a recurrent disease phenotype or.

Li Fraumeni syndrome (OMIM#151623) is a well characterized autosomal dominant cancer predisposition syndrome predominantly caused by constitutional loss-of-function variants in the tumor suppressor gene ***TP53***.[95] The transcription factor TP53 has been described both as a *"guardian of the genome"*[96] and *"gatekeeper for growth and division"*[97] for its prominent role in inducing DNA-repair, apoptosis, senescence and cell cycle arrest in cells subjected to serious stressors such as DNA-damage. Li Fraumeni syndrome is associated with an accumulated lifetime risk of cancer of 75% in males and 100% in females.[98,99] The predominant cancer types presenting in this syndrome are breast cancer, brain tumors, sarcomas and adrenocortical carcinoma but other malignancies also occur within the syndrome although less commonly. Somatic variants in *TP53* are common in most sporadic human cancer, including acute myeloid leukemia (AML). In BCP-ALL, somatic *TP53* aberrations are rare.[100] Consequently, Li Fraumeni syndrome was previously not considered associated with ALL, but recent reports of both familial ALL and genomic characterization of hypodiploid BCP-ALL support a role for *TP53* constitutional variants in BCP-ALL predisposition.[101–104] Collectively, these findings stress the importance of detecting constitutional *TP53* pathogenic variants to enable further clinical studies, apply therapeutic considerations, initiate further cancer surveillance and offer genetic counselling to patients and their families.

ETV6, located on chr 12, is a powerful transcriptional repressor essential to both fetal and postnatal hematopoiesis.[105–107] It is a well-known gene in BCP-ALL, where the somatic translocation t(12;21)/*ETV6-RUNX1* is the hallmark of the second most common subtype. In addition, deletions of *ETV6* are common as a secondary event in BCP-ALL, especially in t(12;21)-positive cases (54-77% of cases).[108–110] This suggests *ETV6* is a tumor suppressor gene, supported by functional studies of WT *ETV6* inducing apoptosis and reduced growth rates in BCP-ALL cell lines.[111,112] Given ETV6's role in normal hematopoiesis and BCP-ALL, its involvement in constitutional predisposition is perhaps not surprising. Numerous familial cases with rare constitutional variants in *ETV6*, segregating with susceptibility to BCP-ALL but also other hematological malignancies have been reported (autosomal dominant, incomplete penetrance).[113–120] Also, 35 additional non-familial cases of BCP-ALL with constitutional pathogenic variants in *ETV6* have been reported,[114,121] showing an overrepresentation of HeH subtype[114,122] and association with older age at diagnosis than non-carriers[114]. Reported variants primarily impair DNA-binding and decrease the repressional activity of ETV6.[113,114,122–124] It is also worth noting that second hits to *ETV6* or other leukemia driving genes were required for disease initiation and progression in these patients.[113,122]

PAX5 encodes a B-lymphoid transcription factor essential to early B-cell lineage commitment and further differentiation.[125] *PAX5* is also the most common target of somatic aberrations overall in BCP-ALL (rate 31.7%), consisting mainly of deletions but translocations and SNVs occur.[126,127] An important example is the recurrent somatic unbalanced translocation dic(9;20), invariably targeting *PAX5*, defining a subtype constituting 4.7% of all BCP-ALL.[128] Constitutional pathogenic variants in *PAX5* associated with predisposition to BCP-ALL, AD with incomplete penetrance, have been reported in 3 familial cases.[129,130] All reported leukemias had somatic loss of WT *PAX5*, suggesting BCP-ALL progression required marked reduction of PAX5 function. Several cases also had homozygous deletion of *CDKN2A/2B*, in line with obligate homozygous loss of *CDKN2A* in sporadic cases of dic(9;20) BCP-ALL. Recently, PAX5 (and IKZF1, another B-lymphoid transcription factor) was shown to prevent malignant transformation in B-cells by restricting glucose metabolism and thus energy availability.[131] Pathogenic variants in *PAX5* disrupted this restriction allowing energy levels to drastically increase. This provided one mechanistic insight to leukemogenesis in BCP-ALL conferred by aberrations in *PAX5*, be they constitutional or somatic.

TYK2 (Tyrosine kinase 2) is part of the Janus kinase (JAK) protein family and JAK/STAT signaling pathway, which has an important role in leukemia upon activation.[100] Although *TYK2* is not commonly somatically mutated in ALL[132,133], other family members of the JAK family (JAK1/2/3) are frequently affected [100,134–136]. However, *TYK2* is involved in translocations of BCR-ABL-like ALL[134,137]. Inactivation of TYK2 has previously been associated with immunodeficiency and autoimmune disease.[132] Driven by the hypothesis that a second primary leukemia in individuals with ALL may be a marker for constitutional predisposition, Waanders *et al.*[132] identified two individuals with constitutional *TYK2* pathogenic variants predisposing to ALL. Functional evaluation confirmed a significant effect on the JAK/STAT pathway, by the strong activation of STAT-signaling through constitutive activation of TYK2.

The incidence of constitutional **Robertsonian translocations** is 1:1000 born children[138,139], most commonly rob(13;14)c accounting for 74%[140]. Constitutional translocation rob(15;21) is very rare (0.5-1%) but of particular interest due to its strong association with intrachromosomal amplification of chr 21 (iAMP21), a high-risk subtype of BCP-ALL. Up to 3.2% of individuals with iAMP21 BCP-ALL have been reported to carry a constitutional translocation rob(15;21)c.[141], giving a 2700-fold risk increase of iAMP21 with constitutional rob(15;21)c.[142] This association is remarkable in its specificity for both rob(15;21)c and iAMP21 BCP-ALL.[143]

Constitutional **ring chromosomes** involving chr 21, r(21)c, have recently been acknowledged to also predispose specifically to iAMP21 BCP-ALL.[143] Incidence of r(21)c in live births is unknown but has been approximated to 1:25 000 conceptions.[140]

SH2B3 coding for SH2B adaptor protein 3 is a negative regulator of both cytokine and growth factor signaling pathways, such as RAS/MAPK[144] and JAK/STAT[145]. Through this negative regulation, SH2B3 plays an important role in normal lymphopoiesis by regulating proliferation[144] as well as controlling self-renewal and dormancy of hematopoietic stem cells (HSC's)[145]. SH2B3 deficiency in mice has been shown to cause splenomegaly, with accumulation of pre-B and immature B-cells, and augmented B-progenitor production in bone marrow, indicating a role for *SH2B3* as a tumor suppressor gene in B-lineage lymphocytes.[146] Somatic deletions of *SH2B3* have been associated with both BCR-ABL1-like[147] and iAMP21[148] BCP-ALL. Collectively, these data implicate *SH2B3* in leukemogenesis. One single case of *SH2B3* mediated constitutional predisposition to BCP-ALL has

been reported to date; a homozygous constitutional loss-of-function variant in *SH2B3*, conferring predisposition though an AR inheritance pattern.[149] Despite limited examples, it was argued *SH2B3* should be considered along with other BCP-ALL predisposition genes and syndromes due to the reported case's strong implication of *SH2B3* in the pathogenicity of BCP-ALL.[90] This is a good example of how rare cases can implicate novel genes in the biology of BCP-ALL and help understanding their pathogenicity.

IKZF encodes the protein IKAROS, a zinc finger transcription factor and a key regulator of lymphoid development. Two central functions of IKZF1 are inhibition of proliferation and promotion of differentiation.[150] Deletion of *IKZF1* is a recurrent event in BCP-ALL (overall rate 15%)[60] and is an independent negative prognostic factor in BCP-ALL[151,152]. Constitutional variants in *IKZF1* predisposing to childhood BCP-ALL have been identified in both familial[153–155] and sporadic[156] cases. In two families, variants segregated with immunodeficiency (loss of serum immunoglobulins and B-cells) and decreased the number of B-cell progenitors in bone marrow.[153] In another family, a 6.7 Mb constitutional deletion on chr 7p spanning *IKZF1* segregated with syndromic features (intellectual impairment, dysmorphic facial features and large head circumferences) and 2/3 of carriers developed childhood BCP-ALL.[155] In addition, 28 constitutional variants in *IKZF1* were identified from screening 5008 children with primarily ALL (4902 BCP-ALL and 106 T-cell ALL) identified.[156] Functional effects on IKAROS transcriptional repression was seen for all but one missense variant. Collectively, these findings support a role for constitutional *IKZF1* aberrations in predisposition to BCP-ALL.

A congenital syndrome is by definition a condition characterized by a combination of associated medical symptoms and/or phenotypic findings. Today, a wide variety of syndromes have been characterized and their genetic origins determined. It is well recognized that many congenital syndromes have an increased risk of pediatric cancer. In these syndromes, presentation of cancers and non-cancerous symptoms vary widely. In the setting of pediatric oncology, this phenotypic variability means that cancer may be the presenting symptom and challenges clinicians to recognize patients with underlying congenital syndromes. A genetic syndrome may negatively affect cancer treatment outcomes by influencing therapeutic response and an increased risk of therapy-related toxicities. Such complications can in turn be actionable, which emphasizes the clinical importance of obtaining an etiological diagnosis. For this purpose, several tools have been developed to aid pediatric oncologists in identifying patients in need of genetic investigation, including the widely applied Jongman's criteria.[89,157,158]

Down syndrome (DS) (OMIM#190685) is the most common chromosomal aberration in humans, occurring in approximately 1 in 700 live born children, caused by extra genetic material from the whole, or a critical part, of chr 21. The syndrome is phenotypically characterized by intellectual disability, dysmorphic facial features, growth retardation, malformations and a well-known 20-fold risk increase of childhood BCP-ALL (termed DS-ALL).[159] DS-ALL constitutes a significant portion (2-3%) of all treated cases[160] and is associated with older age at peak incidence and lower EFS due to increased relapse rate and therapy related mortality. Toxicity reactions to therapy are common and often result in dose reductions, which in part may explain the inferior prognosis.[161,162] Although not discussed in this review, it is noteworthy that DS is also associated with a significantly increased risk of acute myeloid leukemia (ML-DS, a subtype of AML)[159] as well as transient abnormal myelopoiesis in infancy[163]. Up to 40% of DS-ALL have a normal karyotype (apart from constitutional trisomy 21) with standard cytogenetic workup[161], compared to 4.5% in all BCP-ALLs. Unfavorable cytogenetic

markers such as *KMT2A*-rearrangements and BCR-ABL1 fusions are rare in DS-ALL. Low-risk subtypes t(12;21) and HeH are also less common, either of them present in only 15% of DS-ALL compared to ~55% of all BCP-ALL.[161,164–166] Two frequent and commonly co-occurring acquired characteristics in DS-ALL are *CRLF2* overexpression[167] and *JAK2* activation[168,169], suggested to have a synergistic disease promoting role[170]. In a murine model of DS-ALL, *JAK2* pathogenic variant, *CRLF2* overexpression as well as loss of *IKZF1* and *PAX5* were all necessary for inducing BCP-ALL.[171] This highlights the requirement of additional somatic aberrations, in addition to constitutional trisomy 21, for development of DS-ALL. However, the mechanism of trisomy 21-mediated predisposition to BCP-ALL is still unknown.

A number of **additional congenital syndromes** are associated with predisposition to childhood ALL but will not be discussed in detail in this review. These syndromes can be divided into different groups based on the underlying genetic mechanisms. One group is the DNA repair disorders including Ataxia Telangiectasia (*ATM*)[172], Bloom syndrome (*BLM*)[173], Nijmegen breakage syndrome (*NBS1*)[174], Fanconi anemia[175] and constitutional mismatch repair deficiency (*MLH1, MSH2, MSH6, PMS2*)[176]. Moreover, Noonan syndrome (*PTPN11, SOS1, LZTR1, KRAS, RAF1, RIT1, SOS2, MAP2K1, NRAS, BRAF, RRAS2, MRAS, NRAS, RRAS2, RASA2*), Neurofibromatosis type 1 (*NF1*) and cardiofaciocutaneous syndrome (*BRAF, MAP2K1/2, KRAS*) are examples from the group of RASopathies.[177,178] Finally, syndromes caused by genetic aberrations in transcription factors or transcription factor coactivators associated with hematopoietic differentiation, such as Börjeson-Forssman-Lehmann syndrome (*PHF6*)[179], Rubinstein-Taybi syndrome (*CREBBP, EP300*)[180,181], Weaver syndrome (*EZH2*)[182], Sotos syndrome (*NSD1*)[183] and familial platelet disorder with associated myeloid malignancies (*RUNX1*)[184], are conditions where childhood ALL has been reported and predisposition to BCP-ALL is plausible considering the genetic basis of the syndromes.[92]

Further studies of large cohorts of both familial, syndrome associated and sporadic BCP-ALL cases are needed to evaluate prevalence and molecularly characterize predisposition to childhood BCP-ALL.[26,92] In addition, studies of familial and syndrome-associated cases of childhood BCP-ALL will likely continue to aid discovery of novel predisposition genes and molecular pathways as well as improved understanding of predisposition mechanisms, penetrance and associated non-cancer symptoms. This motivated our studies of constitutional predisposing variants in **paper I** and **III**. Ongoing studies applying refined clinical predisposition-criteria and WGS analysis to population-based cohorts of childhood cancer hold promise of such advancements. That said, the greatest benefit of these studies will hopefully be for the individual patients, enabling patient tailored therapeutic strategies for patients with poor treatment response and increased risk of relapse due to constitutional variants.[91]

1.4.2 Progression from pre-leukemic state to disease onset

For pre-leukemic clones to progress into overt leukemia additional "hits" are required, as first suggested in the "two-hit hypothesis" for ALL by Greaves.[185] The discovery that prenatal initiation of pre-leukemic *ETV6-RUNX1*-positive clones is a relatively frequent event, compared to diagnosis of the corresponding leukemia[75], has emphasized the importance of secondary molecular events for progression to overt disease. It has also motivated studies of environmental exposures that may modulate the course of progression. Studying the developmental timing of secondary events was long restricted to distinguishing shared (prenatal) from unique (postnatal) variants in concordant twin leukemias[186,187] or variants with high (early) versus low (late, subclonal) variant allele frequency

(VAF). Single-cell sequencing techniques can now help delineate the phylogenetic tree (temporal order) of clonal evolution in BCP-ALL[188], but do not necessarily explain when driver events occurred during the patient's pre-leukemic life. Understanding this aspect would be of value to couple molecular drivers of progression to extrinsic exposures modulating risk of BCP-ALL.

1.4.2.1 Intrinsic secondary drivers – genomic landscapes of BCP-ALL

The nature of secondary driver variants ("second hits") and their transcriptional consequences have been extensively studied using massive parallel sequencing techniques in large cohorts of childhood BCP-ALL.[51,189–192,192–194] These studies have revealed a generally low burden of "second hit" variants which are mainly subclonal and differ between molecular subtypes.[51,189,191,195,196] Secondary driver variants come in all forms: SNVs, indels, focal deletions, translocations and other rearrangements. For a recent and comprehensive overview Brady et al[51] reported an extensive characterization of the genomic landscape of 2288 childhood and AYA BCP-ALLs. Detecting secondary driver variants at a median of 4 variants (rage 0-19) per case, predominantly focal deletions (median 2, range 0-18), rates varied among molecular subtypes from a mean of ~2 in *KMT2A*-rearranged and hypodiploid subtypes to 7.6 in Ph-like-CRLF2 BCP-ALL. The most frequently targeted genes were *CDKN2A*, *PAX5*, *IKZF1*, *ETV6*, *KRAS* and *NRAS* representing essential biological functions such as cell cycle regulation (growth, DNA-synthesis and repair, division) and cell differentiation. Key pathways affected by secondary driver variants were transcription factors regulating B-cell development (e.g. *PAX5, IKZF1*), epigenetic regulators (e.g. *CREBBP, SETD2*), other transcription factors (e.g. *ETV6, ATF7IP, ERG, RUNX1*) and Ras-singling (e.g. *NRAS, KRAS, NF1*). Within molecular subtypes different pathways were more or less frequently targeted. Regarding clonality, *ETV6*, *IKZF1*, and *PAX5* were often targeted by early events and subclonal later events commonly affected kinase signaling such as *KRAS, NRAS, JAK1/2* and *FLT3*. The above knowledge fueled our analysis of secondary driver variants in **paper I** and **III** as well as delineation of clonal evolution in **paper III** to better understand disease pathogenicity in these specific cases. Despite these significant advances in understanding the genomic basis of BCP-ALL, little is known of potential environmental interactions underlying such aberrations and their pathogenicity.

1.4.2.2 Extrinsic secondary modifiers – the role of infections and trained immunity

The strive to understand the etiology of childhood BCP-ALL is driven by a need for preventive strategies. To that end, environmental risk factors can be modulated to a larger extent than intrinsic genetic/epigenetic ones. Exposure to common microbes and consequential training of our immune system has been a focal point of epidemiological studies in recent years, and well deserved given BCP-ALL is a disease of B-cells which are designed to activate and react upon immune stimulation by infective agents.

In 1988, Greaves suggested a model where delayed exposure to common infections during early childhood (first year of life) deprives our immune system of the natural training its programmed for. Once infections are encountered the immune system reacts adversely, promoting progression of pre-leukemic cells in individuals who are carriers of such.[185] This hypothesis originated in observations linking time and dose-dependent exposure to common microbes to risk of BCP-ALL. The geographical variation in BCP-ALL with markedly lower incidence in regions of lower socioeconomic status, such as sub-Saharan Africa, adds support to Greaves' hypothesis.[45,197] Similarly, historic increases of childhood ALL incidence following improvements in general hygiene

conditions, measured by prevalence of hepatitis A infections in the population, have suggested an association between lower exposure to microbes and an increased risk of developing childhood ALL.[198] Epidemiological studies have also associated early livestock and pet contact, early daycare attendance, vaginal delivery, breastfeeding, having older siblings and early BCG vaccination, all considered proxies for exposure to common microbes, to decreased risk of BCP-ALL (reviewed by Hauer et al.).[199] Based on these associations, early training of the innate immune system and tending of the microbiome have been suggested to protect against progression to BCP-ALL.

While early (within first year of life) exposure to microbes may protect against leukemia development, infectious disease has also been suggested to promote disease progression. Early epidemiological studies by Ward (1917) were the first to recognize the peak of BCP-ALL incidence in 2- to 5-year-olds. He reasoned that this age group was more often affected by common infections than others and therefore suggested infections could be a trigger of BCP-ALL.[200] Exploring the interplay of infectious disease and risk of childhood BCP-ALL has historically generated two main hypotheses. In 1988, Kinlen first formulated the hypothesis of "population mixing" after observing an increased incidence of childhood BCP-ALL in isolated populations after an influx of individuals from urbanized areas of residence. He suggested BCP-ALL was triggered by a common mild infectious agent (not specified), transmitted by the urbanized population, to which the previously isolated inhabitants were immunologically naïve.[201–203] Another "natural experiment" reported in a study from the 2003 SARS outbreak in Hong Kong indicated exposure to common infections is required for onset of BCP-ALL.[204] During the implementation of prolonged (6 months) rigorous infection control measures, researchers could detect a simultaneous drop in diagnoses of childhood BCP-ALL and rapid decline in communicable common infections. The speculative conclusions is that at risk individuals, i.e. carriers of pre-leukemic clones, were protected from acquiring secondary driver variants needed for progression when not exposed to common infection. This suggested that exposure to a common flora of infections was required for progression to overt leukemia from a pre-leukemic state, which has also been supported by recent studies in rodents.[205,206]

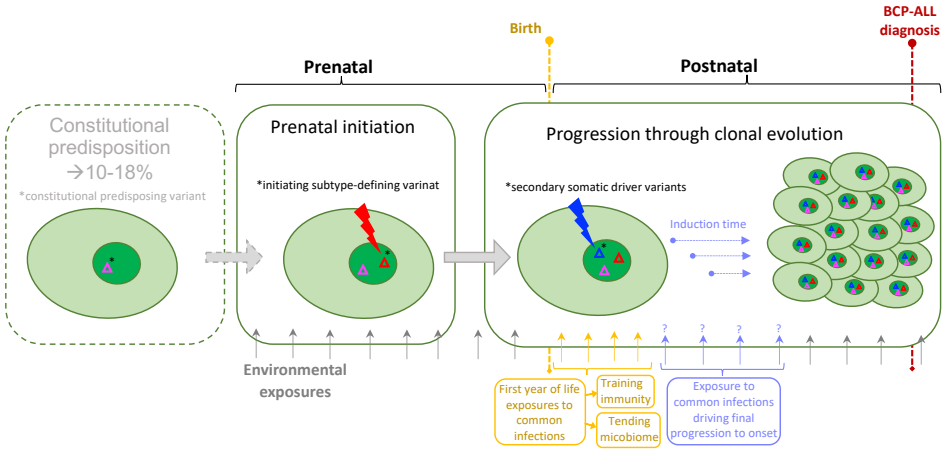


Figure 3. *Schematic overview of genetic drivers and putative environmental drivers or modulators of BCP-ALL initiation and progression. Colored Δ-symbols represent genetic driver variants: constitutional predisposing (pink), initiating (red) and secondary (blue). Lightnings symbolize acquirement of somatic variants.*

The recent Covid-19 pandemic and associated restrictions significantly limited the spread of many other infectious diseases, albeit temporarily. It serves as a real-life pressure test for the dose and time dependent implications of infectious exposure and the association to BCP-ALL progression. Effects of the SARS-CoV-2 virus and Covid-19 pandemic on BCP-ALL incidence is debated. Possible short-term (weeks to months) outcomes are speculated to include both an increasing incidence in response to a novel widespread pathogen (population-mixing hypothesis), and a declining incidence as a consequence of decreased exposure to infectious disease (second-hit hypothesis).[207,208] Due to limitations in sample size, observation time and variable restrictions across countries, reported incidence trends are conflicting and should be interpreted with caution.[209–214] To draw any conclusions about the effects of SARS-CoV-2 infection and the restrictions to prevent it's spread, BCP-ALL incidence will need to be closely monitored in the coming years.

A substantial number of studies have previously assessed seasonal variation in disease onset as a proxy for seasonally variable environmental drivers of disease progression, such as common infections that show strong seasonal patterns in temperate climates. Also, variation in BCP-ALL patient's season of birth has been studied to look for associations to seasonal pathogens that may have influenced in-utero initiation of pre-leukemic clones. Seasonality in onset of childhood ALL has been identified but with onset peaks scattered throughout the year and not supported by all studies. This has been attributed to limitations in reliability of methods and discrepancies in cohort sizes and composition (only children/also adults and only BCP-ALL/also T-ALL). This motivated our studies in **paper IV**, specifically assessing presence of seasonal variation in onset of BCP-ALL in cases aged 0-18 years (children and adolescents) and applying a powerful statistical model for analysis and forecasting of time series data (GARIMAX).

1.4.3 Implications of etiology to prevention

Can childhood BCP-ALL be prevented? This will arguably be the question of our century in childhood leukemia research.[84,199] Therapeutic advances have vastly improved survival of pediatric BCP-ALL patients over the last decades. Nonetheless, 10% of children in HIC, with access to leading healthcare, still die from the disease. Also, the physical and psychological ordeal of living through a cancer diagnosis and treatment for patient, parents and siblings is immense and in itself motivates the need for preventive measures. Possible routes for prevention range from trying to kill off pre-leukemic cells at birth with drug treatment[215], microbiome training, boosting and tending[84,199] and training immunity[199,216] to avoidance of exposure to carcinogens,[216,217] adequate supplement use and a healthy diet during pregnancy[199,216]. The feasibility and putative effects of such measures are worth an extensive discussion and further studies, but are beyond the scope of this thesis.

2 Aims of the thesis

The aim of this thesis was to identify, assess and quantify congenital genetic aberrations behind childhood BCP-ALL predisposition and initiation, as well as to characterize subsequent clonal evolution and identify drivers of progression to overt disease.

The specific aims were to:

⇒ Identify and characterize constitutive genetic aberrations predisposing to BCP-ALL using WGS. (**paper I and III**)

⇒ Develop and apply chip-based dPCR in combination with WGS as a method to detect and quantify genomic breakpoint sequnces of congenital BCP-ALL initiating variants in neonatal blod spots i.e. backtrack pre-leukemic clones. (**paper II and III**)

⇒ Assess and deliniate temporal order (clonal evolution) of somatically acquired predisposing, initiating and secondary driver variants in monozygotic twins' with concordant BCP-ALL. (**paper III**)

⇒ Assess presence of seasonal variation in onset of BCP-ALL in a Swedish population-based cohort by applying a powerful statistical model for analysis and forecasting of time series data (GARIMAX). (**paper IV**)

3 Patients and Methods

3.1 Patients, clinical information and samples

3.1.1 Familial BCP-ALL (paper I)

The family studied in **paper I** was recruited as part of a Nordic collaborative project. Individuals affected by childhood BCP-ALL with a family history of childhood hematological malignancies, raising suspicions of an underlying genetic predisposition, were identified and selected individuals were examined by WGS for shared constitutional genetic aberrations. In **paper I,** 11 family members had been analyzed by karyotyping with G-banding (according to standard procedures) prior to our study. From this analysis nine individuals were identified to carry a shared constitutional t(12;14)(p13.2;q23.1), two carriers were affected by childhood BCP-ALL. DNA was extracted and stored according to standard procedures prior to WGS. (***Figure 4***)

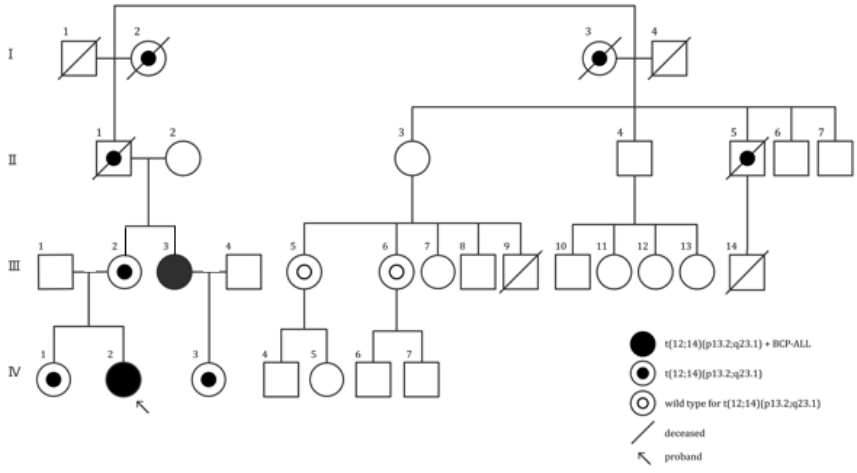


Figure 4. *Pedigree of family with a t(12;14)(p.13.2;q23.1) (**paper I**). Adapted from Järviaho and Bang et al.*[119] *with permission from the publisher.*

3.1.2 BCP-ALL cases for molecular studies of pre-leukemic clones including monozygotic twins with leukemia (paper II, III)

Seven patients affected by BCP-ALL were studied in **paper II**, out of which the pair of female monozygotic twins (patient 8P and 9P) were also the study subjects of **paper III**. Patients' diagnostic information and clinical characteristics are summarized in ***Table 2*** for **paper II** and ***Table 3*** for **paper III**. They had all been diagnosed and treated at the Department of Pediatric Oncology, Astrid Lindgren Children's Hospital with diagnostic cytogenetic workup (including karyotyping with G-banding, FISH and array CGH) performed by Department of Clinical Genetics, both Karolinska University Hospital (Sweden).

Patient ID	Age at diagnosis[a]	Karyotype at diagnosis	FISH results at diagnosis	Flow Cytometry (Percentage of Blasts at diagnosis)
1P	3y 3 mo	46,XX,dic(9;20)(p13.1;q11.2), +21[18]/46,XX[8]	CDKN2A (INK4) (9p21) deletion 86%; three copies of RUNX1 86%	70%
2P	1 mo	46,XY,t(10;11)(q25;q23)	*KMT2A* (11q23) 96%	87%
3P	1 y 8 mo	46,XY,del(12)(p13p13), t(12;21)(p13;q22)	t(12;21)(p13;q22) 99%	90%
4P	3 y	46,XY,?del(12)(p?)[5]/46,XY[5]	t(12;21)(p13;q22) 99%, +21 14%	91%
5P	1 y 1 mo	45,XX,dic(9;20)(p11-13;q11) [8]/46,XX[18]	*CDKN2A (INK4)* (9p21) deletion 76%	65%
8P	3 y 6 mo	46,XX,t(11;12)(q21;p13) [6]/46,XX[19]	t(12;21) 70%	69%
9P	4 y	46,XX,?t(11;12;21)(q23;p13;q22) [5]/47,XX,sl,t(1;12)(p13;p13), +der(1)t(1;12)(p13;p13)[9]/46,XX[3]	t(12;21) 88%	78%

Table 2. *Clinical characteristics of BCP-ALL cases studied in* ***paper II****. Adapted from Taylan et al.*[218] *with permission from the publisher.*

DNA from diagnostic bone marrow (tumor sample) used in **paper II** and **III** were available from the Department of Clinical Genetics. Genomic DNA from time of birth (constitutional/preleukemic sample) was extracted in our research lab from 3 mm punches of neonatal dried blood spots from all seven patients in **paper II** and their adjacent controls (retrieved from the Swedish National PKU Register, Karolinska University Hospital) using QIAmp DNA Micro Kit (Qiagen) following manufacturer's instructions. In **paper III,** five years into clinical remission in both twins genomic DNA from saliva, collected using Oragene DNA saliva collection kit (DNA Genotek, Ontario, Canada), was extracted using prepIT-L2P kit (DNA Genotek, Ontario, Canada) following the manufacturer's instructions.

Clinical characteristics	Twin 1 (Tw1)	Twin 2 (Tw2)
Diagnosis	BCP-ALL	BCP-ALL
Age at diagnosis	3y 4m	3y 10m
Full blood count at diagnosis (peripheral blood):		
-Hb (g/l)	41	69
-WBC (10E9/l)	2,8	2,8
-Platelets (10E9/l)	128	163
Immunophenotype (bone marrow)	CD45dim, CD19+, CD10++, CD20-, TdT+, CD22+, CD cyt79a+, CD38+, HLADR+, CD123dim, CD58+, CD66c-, cytIgM-, no myeloid markers, subpopulation (36%) CD34+	CD45dim/neg, CD19+, CD10++, CD20-, Tdt+, CD22+, CD cyt79a+, CD38+, HLADR+, CD123dim/neg, CD58+, CD66c-, cytIgM-, no myeloid markers, CD34hetero/+, no T-cell markers, CD99++
Blast count at dianosis (bone marrow)	69%	74%
CNS engagement of leukmeia	No	No
Cytogenetics (bone marrow at diagnosis)		
-Karyotyping with G-banding	46,XX,t(11;12)(q21;p13)[6]/46,XX[19]	46,XX,t(11;12)(q23;p13)[5]/47,XX,sl, t(1;12)(p13;p13),+der(1)t(1;12)(p13;p13)[9]/46,XX[3]
-FISH	nuc ish(ETV6x2,RUNX1x3)(ETV6 con RUNX1x1)[145/206],(ETV6x1,RUNX1x3)(ETV6 con RUNX1x1)[12/206]	nuc ish(ETV6x2,RUNX1x3)(ETV6 con RUNX1x1)[98/222],(ETV6x1,RUNX1x3)(ETV6 con RUNX1x1)[98/222]
-ArrayCGH	No detectable copy number changes	No additional finding
-Interpretation	Main clone with t(11;12) and *ETV6-RUNX1* (70%), subclone with additional *delETV6* (6%)	Main clone with t(11;12) and ETV6-RUNX1 (88%), subclone with additional delETV6 and t(1;12) with extra der(1)t(1;12) (44%)
Image cytometric DNA analysis (ICDA)	Diploid (DNA index: 0.5, S-phase 8%)	Diploid (DNA index: 1, S-phase 8%)
Treatment protocol	NOPHO-ALL 2008 standard risk arm	NOPHO-ALL 2008 standard risk arm
SCT	No	No
MRD:		
-day 15	<0,1%	<0,1%
-day 29	<0,1%	<0,01%
-day 79	<0,01%	<0,01%

Table 3. *Detailed clinical data of monozygotic twins with concordant BCP-ALL studied in detail in* ***paper III*** *but also included in* ***paper II****. Adapted from Bang et al.*[219] *with permission form the publisher.*

3.1.3 Swedish population based BCP-ALL cohort (paper IV)

The cohort of 1380 BCP-ALL cases studied in **paper IV** were identified from the Swedish Childhood Cancer Registry (SCCR) (***Figure 5***). Diagnosis of included patients was restricted to age <18 years to not include adult cases. First day of BCP-ALL diagnosis for inclusion was January 1st 1995, from when reliable diagnostic routines for identifying both genetic subtypes of BCP-ALL studied in this paper (HeH and *ETV6-RUNX1*) were in clinical practice. Last day of diagnosis for inclusion was set to December 31st 2017 as this was the most recent year from which we had complete registry data for BCP-ALL cohort. As seasonality of BCP-ALL was studied as a proxy for potential environmental modifiers of disease progression, such as common infections, only cases born and diagnosed in Sweden were included as migration may have altered panorama and timing of exposures. For analysis of HeH (n=444) and *ETV6-RUNX1* (n=272) cases respectively, information about molecular subtype of cases were used to distinguish subgroups.

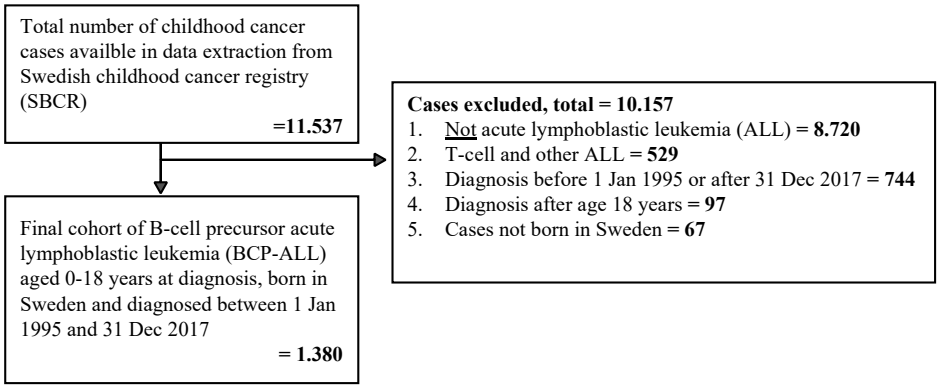


Figure 5. *Flowchart of cohort selection in paper IV.*

3.1.4 Ethical approval and informed consent

All four studies included in the thesis were performed in accordance with the 1975 Declaration of Helsinki, as revised in 2000. Ethical permissions covering **paper I-IV** were granted by the local ethical board at Karolinska Institutet, Stockholm, Sweden. Ethical permission was also granted by the regional ethical committee of Northern Ostrobothnia Hospital District, Oulu; Finland covering **paper I**. Written informed consents were signed by patients or their legal guardians prior to inclusion.

3.2 Methods

3.2.1 Whole Genome Sequencing – constitutional and somatic variant analysis

The ability of whole genome sequencing (WGS) techniques to effectively, and to a nowadays reasonable price, mine the genome for both variants of smaller scale and larger structural aberrations has revolutionized the field of genetics. The application of this and other omics methods in both research and clinical settings have had great impact on our understanding of disease biology and diagnostic abilities. WGS was implemented in **papers I-III** to identify and characterize constitutional genetic aberrations and/or somatic variants in leukemic samples. WGS as opposed to whole exome sequencing (WES) was chosen to facilitate identification of structural variants such as larger insertions/deletions, translocations and inversions.

In **paper I** and for the monozygotic twins part of both **paper II** and **III**, paired-end sequencing was performed on Illumina HiSeq X (Illumina Inc, San Diego, CA, USA) after short-read library preparation from genomic DNA using Illumina TruSeq PCR-free kit with a mean insert size of >350 base pairs. The mean coverage across all samples was 40 X (range 32-60X) in **paper I** and 37X (range 34-39X) in **paper III**. Each sample had on average 500 million unique mapped sequences (range 386-740 million) in **paper I** and average 700 million (range 729-915 million) in **paper III**. Reads were mapped to human reference genome (GRC37/hg19) and variants called using an in-house bioinformatic pipeline developed by Science for Life Laboratory (Stockholm, Sweden) who also performed sequencing. The pipeline used Burrows Wheeler Aligner (version 0.7.12)[220] for alignment to reference genome and Genome Analysis Tool Kit (version 3.3-0-geee94ec)[221] for de-duplicating, recalibrating and cleaning raw alignments. Qualimap (version 2.0)[222] was used to generate quality control information.

For the remaining 5 samples in **paper II**, paired-end sequencing was performed on Illumina HiSeq 2500 (Illumina Inc, San Diego, CA, USA) after mate-pair library preparation from genomic DNA using Nextera Mate Pair Sample Preparation Kit (also Illumina) the effective insert size being ~2 kb, resulting in an average coverage of 5x. CASAVA RTA 1.18 was used to basecall raw sequence reads. Trimmomatic removed adapter sequences[223] before using Burrows Wheeler Aligner (version 0.7.4)[220] to align reads to human reference genome (GRC37/hg19). Sorting and de-duplicating was performed using Picard tools (v2.0.1, publicly available from GATK, Broad Institute).

In **paper I** and **III**, somatic SNP's and indels were called by MuTect2[224] while constitutional variants were identified using Genome Analysis Tool Kit tools and best practices[221]. Functional annotation was done using Variant Effect Predictor (VEP, version 89).[225] Variants were finally explored in Gemini (version 0.20.0)[226] and visualized in Integrative Genomics Viewer (IGV)[227]. Variant analysis in **paper III** also included the use of a Genomics England PanelApp gene panel for "Hematological malignancies cancer susceptibility".[228]

In **papers I**, **II** and **III**, calls from TIDDIT[229], or TIDDIT and CNVnator (V0.3.2)[230] merged by the FindSV pipeline (https://github.com/J35P312/FindSV), were used to identify structural variants, which were then annotated by VEP[225]. In **paper III**, T-cell receptor and immunoglobulin heavy chain (IgH) rearrangements were excluded, the latter instead called using IgCaller.[231] Structural variants were filtered and finally sorted against local structural variant frequency databases. Breakpoint reads were visualized and manually inspected in IGV and UCSC Genome Browser tool BLAT[232] was used to identify the exact coordinates of breakpoints. In **paper III** circus was used to plot complex translocations.[233] Additional loss-of-heterozygosity (LOH) analysis was performed in **paper III** letting GEMINI identify any regions-of-homozygosity to identify any copy number neutral somatic structural variants.

3.2.2 Validation of WGS variants by Sanger sequencing

Sanger sequencing on ABI3730xl DNA Analyzer using BigDye Terminator v3.1(both Applied Biosystems, Foster City, CA, USA) was applied to validate translocation breakpoints and fusion gene transcript sequences (**paper I**, **II** and **III**) as well as to perform segregation analysis (**paper I**). Primers for PCR and sequencing were designed using online software Primer3.[234] Sequence fragments were analyzed on UCSC Genome Browser to identify exact coordinates and sequence compositions at breakpoints.

3.2.3 SNP-array

In **paper I** SNP-arrays (Infinium Omni2.5Exome BeadChip, Illumina) were performed on the two leukemic bone marrow samples of affected translocation carriers (III:3 and IV:2). This was done to further detail the karyotypes of leukemias that had been defined from karyotyping with G-banding and FISH-analysis at diagnosis. To generate log-transformed ratios (logR ratios), probe intensities were normalized against a panel of normal diploid internal controls which were set around 0. Tumor Aberration Prediction Suite[235] and circular binary segmentation[236] were applied in R (R Foundation for Statistical Computing) to call copy number alterations (CANs). After excluding CNAs led than 20 kb or 10 consecutive probes variants were reviewed using IGV, constructing karyotypes based on logR ratios´ genome wide profile.

3.2.4 RT-qPCR expression analysis

In **paper I** real-time quantitative PCR (RT-qPCR) was performed to evaluate RNA expression levels of WT *ETV6* and *RTN1*, and to investigate whether fusion transcripts of the identified constitutional t(12;14)(p13.2;q23.1) were expressed. Total RNA was extracted using standard kits and procedures. Ribosomal RNA *18S* was used as an internal control for each sample. To ensure transcript specific binding, primers were designed spanning exon-exon boundaries of WT transcripts and breakpoints of fusion transcripts, respectively. Samples were run in triplicate using Power SYBRGreen on Quant-Studio 7 Flex Real-Time PCR System (both Thermo Fisher Scientific).

Constitutional expression of WT *ETV6* and *RTN1* was evaluated in the two affected (IV:2 and III:3) and one unaffected (III:2) carrier. The average expression of 6 normal controls (3 sex and age matched to each affected carrier) generated a control value which was set to one (y = 1). Normalizing against *18S* and PCR-efficacy the expression levels were calculated as folds of control $[(E_{target})^{\Delta CPreference(control-sample)}]/[(Ereference)^{\Delta CPreference(control-sample)}]$. Expression of WT *ETV6* and *RTN1* was also evaluated in leukemic bone marrow of the two affected carriers (IV:2 and III:3) and compared to 8 unrelated BCP-ALL cases (3 WT for ETV6, 2 del*ETV6*, 3 t(12;21)+del*ETV6*). Expression levels were calculated as ΔCt (Ct[reference] - Ct[target] normalizing against *18S* (reference).

Fusion transcript expression was evaluated in constitutional plus leukemic samples from the two affected carriers (IV:2 and III:3) and constitutional sample from one unaffected carrier (III:2) with two negative controls. This last analysis was merely intended to investigate presence of fusion transcript and not to be quantitative.

3.2.5 Digital PCR

Digital PCR (dPCR) is a third generation of PCR technique allowing for absolute quantification of target template in a sample. The technique holds both qualitative and quantitative measuring abilities with high specificity and sensitivity. It builds on the principle of partitioning, dividing the PCR-reaction into a very large number of smaller units (pico-scale wells on a chip or water-oil emulsion droplets) with the theoretical goal of isolating individual DNA/RNA molecules in each compartment and running a multitude of PCR-reactions in parallel. The copy number and concentration of target template DNA/RNA in the original sample is calculated based on the fraction of PCR-reactions positive for the target template, giving us an absolute quantification. In addition, Poisson statistics are applied to correct for that fact that, due to random distribution, some compartments will contain more

than one template, even when the density of target template in the total sample is low (in theory <1 per well).[237]

3.2.5.1 Chip-based

In **paper II** and **III**, we applied chip-based dPCR to DNA samples from birth (constitutional/pre-leukemic sample), leukemia diagnosis (tumor sample) and remission (constitutional sample) to a total of 7 individuals with BCP-ALL, two monozygotic twins overlapped both studies (described in section 3.1.1.2).

For each target breakpoint in **paper II**, primers and hydrolysis probes were manually designed using online software Primer3[234], target probes labelled FAM™ in 5' end and MGB (minor groove binder) in 3' end and ordered as custom TaqMan® assays from Applied Biosystems. TaqMan® Copy Number Reference Assay, human, RNase P (Applied Biosystems), detecting the Ribonuclease P RNA component H1 (H1RNA) gene (RPPH1) on chr 14(GRCh37): 20811565l was chosen as an internal control since all samples were diploid and did not harbor any somatic variants at this locus. the internal control was assessed suitable. This internal control was labelled VIC®.

In addition to analysis of the shared leukemia specific structural variant in monozygotic twins (performed in **paper II**), digital PCR assays for two additional shared somatic variants detected by WGS in both leukemias, UBA2 deletion and NSD2 p.E1099K, were performed in **paper III**. Custom TaqMan® assays labelled FAM™ were designed by ThermoFisher Scientific's internal bioinformatics platform. TaqMan® Copy Number Reference Assay (Applied Biosystems) labelled VIC® was again used as internal control.

Prior to analysis, specificity was evaluated by running dPCR assays on control neonatal blood spot DNA from each patient's adjacent Guthrie card in the PKU Registry and NTCs (total 16 chips per assay) detecting number and rate of false positive target template (FAM) signal for each assay. Also, serial dilutions of diagnostic bone marrow DNA from each patient in commercial WT human genomic DNA (Promega) was run with each corresponding assay. The limit of detection (LoD) was then calculted by the QuantStudio®3D AnalysisSuite™ software using rate of false positives for each assay together with data from each assay's dilution series. DigitalPCR amplification and chip analysis was performed as described below.

Genomic DNA roughly equivalent to 2000-7500 cells (15-50 ng), custom (target) and standard (control) TaqMan® assays and 1X QuantStudio 3D Digital PCR Master mix were mixed and loaded into QS3D Digital 20K V2 chips. Amplification was performed in GeneAmp™ PCR System 9700 thermocycler, after which chips were read by QuantStudio®3D reader. Data analysis included PoissonPlus algorithm (version 4.4.10) with a confidence interval of 95% and a desired precision of 10% by QuantStudio®3D AnalysisSuite™ (version 3.1.2-PRC-build-03) to correct for random distribution of template sequence as well as classifying amplification signals as pos/neg for target/template using FAM and VIC channels viewed in 2D scatter plots. Each reaction was performed at least in duplicate. Results were plotted using Microsoft® Excel (version 15.41). Please see publication's supplemental material for more detailed information about methods, PCR-conditions etc.

3.2.5.2 Digital droplet PCR (ddPCR)

In **paper III** we applied digital droplet PCR (ddPCR) to genomic DNA from saliva 5 years into clinical remission of both monozygotic twins to detect and quantify shared *UBA2* deletion. We used the same TaqMan® assays for *UBA2* deletion as in **paper II**. 67 ng of genomic DNA in droplet digital PCR (ddPCR) Supermix for Probes (No dUTP) kit (BioRad, Hercules, CA, USA) were run on Automated Droplet Generator (BioRad, Hercules, CA, USA) to generate droplets and amplified in triplicates for each twin according to manufacturer's instructions. Droplets were read QX200 on the Droplet Reader (BioRad, Hercules, CA, USA) and data analyzed using QuantaSoft Analysis Pro v.1.0. The ratio of positive to total droplets was calculated using Poisson with 95% confidence intervals after adjusting threshold for true positive signal to the signal in control samples.

3.2.6 GARIMAX seasonal variation analysis

In **paper IV** we implemented Bayesian Generalized Autoregressive Integrated Moving Average with external variables model (GARIMAX)[238,239], a powerful statistical model for analysing and forecasting time series data. We used the model for identification of seasonal variation in onset of BCP-ALL in a population based Swedish cohort of 1380 cases, as well as for subgroup analysis of 444 HeH and 272 *ETV6-RUNX1* molecular subtype cases. Before applying the model to our data, BCP-ALL cases were aggregated into quarters based on their time of diagnosis. Three different types of quarter arrangements were created (***Table 4***). The analysis was then performed in two steps answering the following questions; First, is there a seasonal wave/periodicity in onset of BCP-ALL? Second, is there a significant peak in number of cases in any of the examined quarters?

Quarter types	Quarter I	Quarter II	Quarter III	Quarter IV
1st	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
2nd	Feb-Apr	May-Jul	Aug-Oct	Nov-Jan
3rd	Mar-May	Jun-Aug	Sep-Nov	Dec-Feb

Table 4. *Definitions of quarters for all three quarter types to which cases (total cohort, HeH and ETV6-RUNX1 respectively) were aggregated prior to analysis in paper IV.*

In the first step of analysis, harmonic functions were applied to make predictions of all theoretically possible seasonal curves based on the defined quarters, for each of the three types of quarters respectively. Next, our data was run against these predicted curves to identify if any of the predicted curves matched any curve seen in our cohort. In the second step, a seasonal matrix was created for each quarter type. First the quarter with least number of cases was chosen as base quarter. Thereafter, each of the remaining three quarters in the respective quarter types were compared to their designated base quarter to identify any significant peaks in numbers of cases per quarter. Seasonal matrix does not provide information on repeatability of significant peaks while harmonic seasonal waves do not provide information on which seasonal curve is seen in the data. Thus, we considered it essential that both steps of analysis supported a seasonal wave for us to infer presence of informative seasonality.

4 Results and Discussion

4.1 Congenital predisposition to BCP-ALL

4.1.1 Prenatally initiated pre-leukemic clones (paper II, III)

4.1.1.1 Backtracking BCP-ALL to pre-leukemic clones in neonatal blood spots using dPCR (paper II)

In **paper II**, following identification and characterization of genomic breakpoints from WGS (validated by Sanger sequencing), six custom dPCR assays were designed, one for each patient with the exception of monozygotic twins (8P and 9P) who shared assay. Specificity evaluation for each assay showed that number of false positives per 16 analyzed chips (per assay) ranged from 2/16 (4P's *ETV6-RUNX1* assay) to 9/16 (1P's dic(9;20) assay). Each respective assay's dilution series data and false positive rate were used to calculate the limit of detection (LoD). Ranging from 0.02% to 0.08%, these numbers equaled the possibility of detecting one copy of target sequence in 1320 to 5500 copies of reference sequence (***Figure 6***).

As expected, chip based digital PCR readily confirmed and quantified the presence of genomic breakpoint sequences in diagnostic bone marrow samples of all seven patients (***Table 5***). This, together with dilution series provided proof that the method and assays were able to detect the intended targets. In contrast, we failed to detect breakpoint sequences in all but one DNA sample obtained from Guthrie cards. The patient positive at birth was 2P, diagnosed only 1 month after birth with *KMT2A* rearranged BCP-ALL. This likely explained why copy numbers were higher and could be readily detected in this patient. For the remaining six patients, failing to detect genomic breakpoint sequences could be the result of that leukemia was not initiated in utero (except for monozygotic twins who we know share a prenatally initiated clone). Alternatively, breakpoint sequence DNA was not present in the blood spot samples even though present in peripheral blood and/or bone marrow. Blood spots only contain small amounts peripheral blood and we only extracted DNA from a 3 mm punch per patient. Also, poor yield of DNA from purification could affect breakpoint sequence presence in samples.[240] In the case breakpoint sequences were in fact present in samples, sequence abundance below our assays' LoD is another possible explanation for failure of detection.

Patient ID	Age at diagnosis[a]	Karyotype at diagnosis	FISH results at diagnosis	Flow Cytometry (Percentage of Blasts at diagnosis)	dPCR (Target/Total at diagnosis)	dPCR (Target/Total at birth)	Limit of Detection (%)	Average number of false positives identified in control chips	Mean wild type quant in control chips (copies/microliter) [range]
1P	3y 3 mo	46,XX,dic(9;20)(p13.1;q11.2),+21[18]/46,XX[8]	CDKN2A (INK4) (9p21) deletion 86%; three copies of RUNX1 86%	70%	17%	ND[b]	0.08	3	1130 [0-2161]
2P	1 mo	46,XY,t(10;11)(q25;q23)	*KMT2A* (11q23) 96%	87%	26%	3%	0.05	2	988 [0-2206]
3P	1 y 8 mo	46,XY,del(12)(p13p13),t(12;21)(p13;q22)	t(12;21)(p13;q22) 99%	90%	28%	ND	0.02	1	1085 [0-2061]
4P	3 y	46,XY,?del(12)(p?)[5]/46,XY[5]	t(12;21)(p13;q22) 99%, +21 14%	91%	29%	ND	0.02	0	892 [0-2213]
5P	1 y 1 mo	45,XX,dic(9;20)(p11-13;q11)[8]/46,XX[18]	*CDKN2A (INK4)* (9p21) deletion 76%	65%	29%	ND	0.03	1	804 [0-1810]
8P	3 y 6 mo	46,XX,t(11;12)(q21;p13)[6]/46,XX[19]	t(12;21) 70%	69%	26%	ND	0.05	1	596 [0-1862]
9P	4 y	46,XX,?t(11;12;21)(q23;p13;q22)[5]/47,XX,sl,t(1;12)(p13;p13),+der(1)t(1;12)(p13;p13)[9]/46,XX[3]	t(12;21) 88%	78%	29%	ND	0.05	1	596 [0-1862]

Table 5. *Clinical characteristics, detection of breakpoint sequences in DNA samples from diagnosis (bone marrow) and birth (neonatal dried blood spots) using dPCR and results from dilution series determining sensitivity (LoD) and evaluating specificity (rate of false positives) for each custom assay of patients studied in* ***paper II****. Reused from Taylan et al.*[218] *with permission from the publisher.*

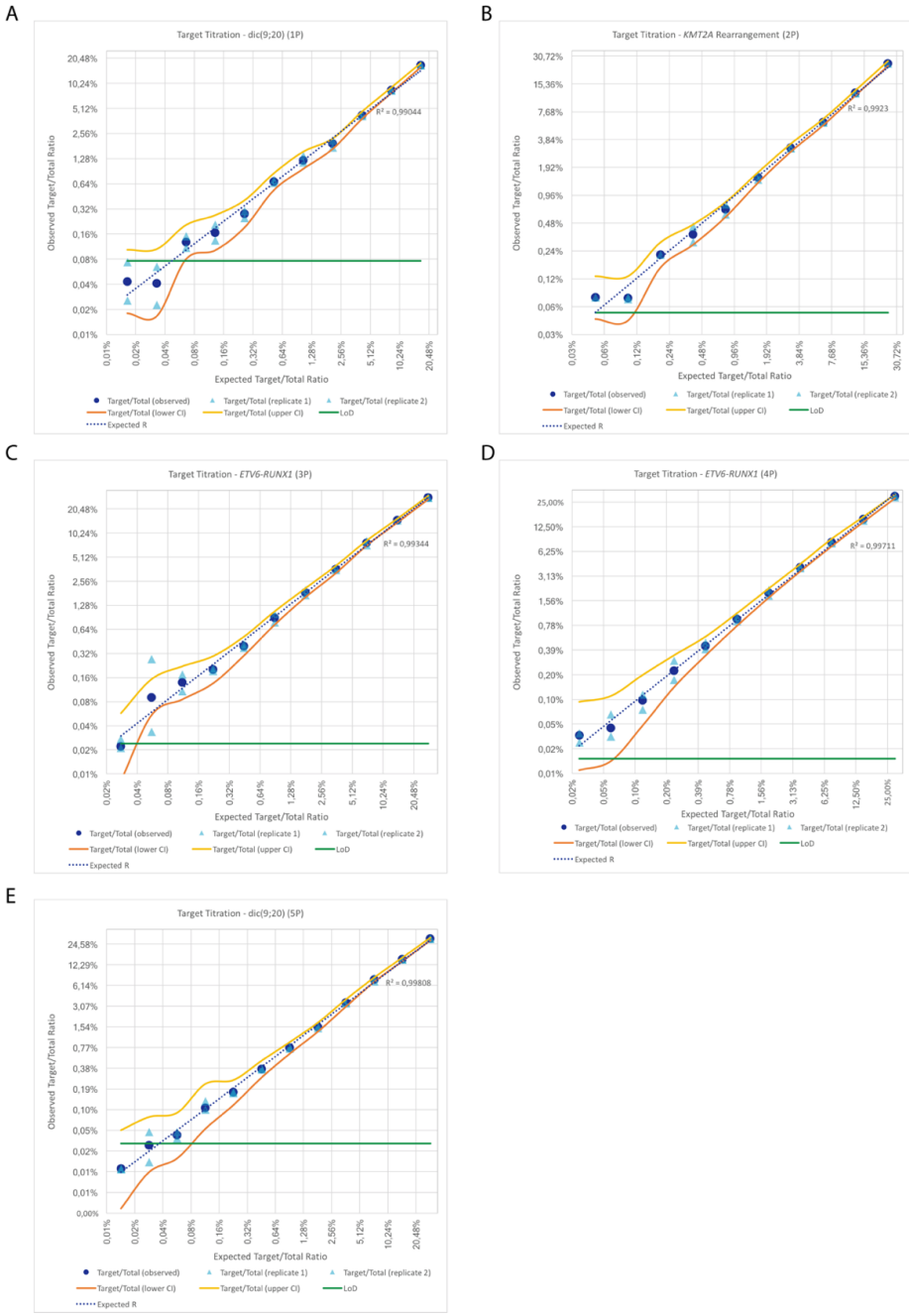


Figure 6. *LoD for each assay was calculated by our software from dilution series data also taking into account assay specific false positive rates. Graphs plots observed against expected target/total ratio for each assay as percentage, indicating how well the assay performs, in turn affecting down-to what level of detection we trust the individual assay. Green line indicates software-calculated LoD. Blue dotted line shows expected and blue dots the observed target/total ratio (converged from two replicates of each dilution) converted to* $\log_2$ *scale. All assays are linear. Orange and yellow lines indicate confidence intervals. Reused from Taylan et al.*[218] *with permission from the publisher.*

Advantages of our approach and method in **paper II** included use of DNA over RNA. RNA is instable and degraded target sequences cannot be detected. The use of RNA has though been advantageous as it facilitates detection of recurrent fusion sequences with streamlined assays, surpassing the obstacle of designing unmanageable numbers of primers for scattered breakpoint clusters required for DNA analysis and allowing for analysis of larger quantities of samples. However, this problem has recently been overcome by the application of genomic inverse PCR for exploration of ligated breakpoints (GIPFEL)[241], which has been applied to detect t(12;21) creating fusion *ETV6-RUNX1* in DNA[75]. Moreover, our use of DNA and customizing assays to the genomic breakpoints identified by WGS overcame previous limitations in backtracking dic(9;20) cases to birth, attributed to the translocation only occasionally giving rise to a fusion transcript[242] and therefore challenging to identify by RNA-based methods. This was affirmed by detection of both dic(9;20)s in leukemic bone marrow samples (likely explanations for failure of detection in birth samples discussed above).

Another advantage of our approach was the sourcing of patient material from dried blood spots as opposed to umbilical cord blood (UCB). Both DNA[243] and RNA[244] may be extracted form neonatal blood spots. Coming from bio-banked Guthrie cards, neonatal blood spots are readily available compared to UCB. They are routinely sampled for clinical purposes (newborn screening of metabolic disease), which eludes the need for additional sampling as for UCB. Guthrie cards are also easy and cheap to handle and store. On the downside, DNA/RNA extraction from dried blood spots do not allow for cell sorting prior to purification, which could increase likelihood of "catching" traces of pre-leukemic clones. Also, small yield of DNA at extraction from blood spots is a drawback.[240] Moreover, the closed system of dPCR chips used in our study diminished previous issues of contamination and added sensitivity compared to nested PCR and qPCR methods. Nonetheless, sensitivity could be further improved by application of digital droplet PCR.

As detailed in the introduction (section 1.3.1.1), there is today solid evidence that pre-leukemic clones of several recurrent BCP-ALL subtypes are initiated in utero. However, previously applied methods are associated with technical limitations affecting reliable detection and quantification of genetic markers of pre-leukemic clones (as reviewed by Hein et al.).[56] This affects both studies backtracking leukemic clones to birth in BCP-ALL patients and those screening for the same in larger cohorts of healthy neonates. Limitations are exemplified by the large discrepancy in rates of *ETV6-RUNX1* fusion found among healthy neonates by different methods. The objective of our studies in **paper II** was to further develop quantitative methods for backtracking pre-leukemic clones to neonatal samples from patients with BCP-ALL. We attempted to do so through the application of chip-based dPCR in combination with WGS to detect and quantify genomic fusion sequences of congenital BCP-ALL initiating variants in neonatal blood spots.

In summary, detection of genomic breakpoint sequences in all tested diagnostic samples validated the ability of all assays to detect their intended target sequence. Moreover, dilution series confirmed the methods ability to accurately quantify target sequence down to the calculated LoD. Finally, failing to detect breakpoint sequences at birth in monozygotic twins (with all certainty) and other two *ETV6-RUNX1* positive cases (most likely) holds methodological over biological explanations, while dic(9;20) cases could be explain by either or. Therefore, we conclude that additional studies to overcome limitations in methods applied in this and previous studies are warranted. Such studies could proposedly include the application of highly sensitive dPCR methods, such as ddPCR, to genomic inverse PCR for exploration of ligated breakpoints (GIPFEL).[241]

4.1.1.2 *Shared pre-leukemic clone in monozygotic twins with concordant BCP-ALL – UBA2 deletion preceding recurrent ETV6-RUNX1 fusion (paper III)*

In **paper III** we studied a pair of monozygotic twins with concordant BCP-ALL of t(12;21)(p13.2;q22.1) *ETV6-RUNX1* subtype (***Table 3*** in Methods). The *ETV6-RUNX1* fusion is currently believed to both initiate pre-leukemic clones in utero and drive disease progression of BCP-ALL.[6] From WGS we were able to determine that the *ETV6-RUNX1* fusions identified at diagnosis in these two cases were in fact a product of a complex chromosomal rearrangement t(11;12;21)(q23;p13;q22) (***Figure 7***). The identical fusion sequences of this complex rearrangement in both leukemias confirmed it was shared and had thus inevitably been initiated in utero. As expected, the complex rearrangement was readily detected, by chip-based dPCR using custom assays, at diagnosis in both twins (Tw1 26% and Tw2 29%). In remission it was expectedly undetectable by dPCR, neither did any reads from WGS support it remained. Although we can confidently say that the rearrangement was present in pre-leukemic clones at birth, given the identical fusion sequences in both leukemias, it was not detected by dPCR in DNA from neonatal dried blood spots. This could be explained by copy numbers below our assays LoD (1 in 1000 copies) or insufficient yield of DNA-purification from dried blood spots. Alternatively, copies of the complex rearrangement were not present in the blood spot sample at all or not in the 3 mm punch we received for analysis.

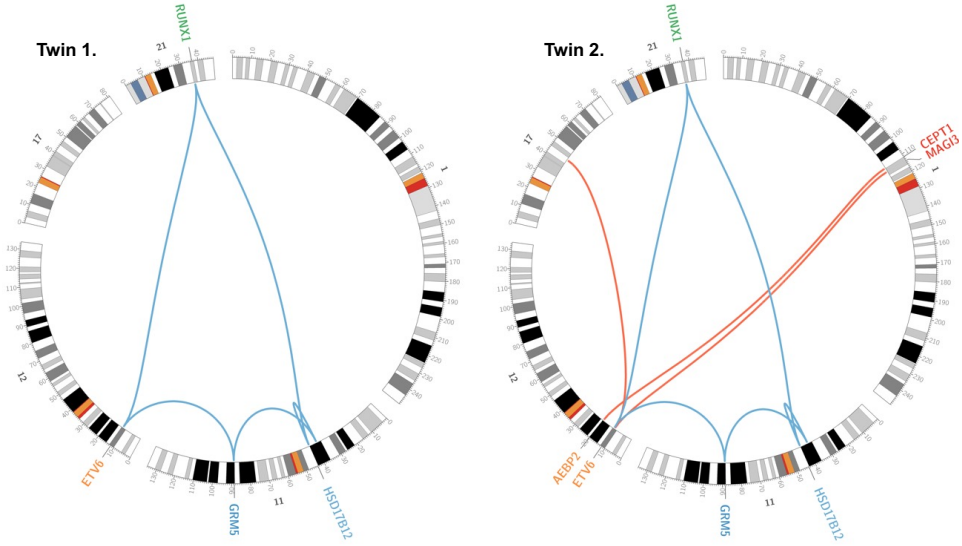


Figure 7. *Circos plots illustrating shared somatic complex rearrangement, t(11;12;21)(q23;p13;q22), involving 2 inversions on chr 11 and giving rise to ETV6-RUNX1 fusion (blue lines). Twin 2's plot also depicts three unique subclonal structural variants, two t(1;12) and one t(11;12;17)(red lines) discussed in section 4.2.1.1. Only involved chromosomes displayed in circus. Colored lines indicate how fusions of translocated chromosomes are arranged. Genes disrupted by or in close proximity of the breakpoints are stated in text. Adapted from Bang et al.*[219] *with permission from the publisher.*

Apart from showing another biological example of prenatal initiation of pre-leukemic clones and spread between twins in utero, our findings added to previous reports of complex rearrangements underlying the classical *ETV6-RUNX1* fusion.[61-65] This demonstrates the strengths of using WGS in leukemia diagnostics enabling detailed characterization of relevant genetic aberrations.[61,63,65] Also, identification of a single identical IGH D-J rearrangement, shared D-J segment (IGHJ4 - IGHD3-22)

but different V segments (IGHV3-69 and IGHV1-73), indicated initiation of the shared pre-leukemic clone took place in between pro- and pre-B cell stages. While analysis of Ig and TCR rearrangements have been used to infer the cell-stage of initiation of pre-leukemic clones in non-twins BCP-ALL cases, this analysis should indeed be more advantageous in twin cases as their shared Ig and TCR rearrangements are not affected by the continued RAG-mediated recombination taking place during clonal evolution.[245,246]

In addition to the complex rearrangement, we identified 58 SNVs/indels shared by the twins leukemias, adding further support of a common clonal origin (***Figure 8***). Particular emphasis was put on a frame shift deletion in *UBA2*, introducing a premature stop in exon 6 out of 17 (NM_005499.2: c.463_470del; NP_005490.1: p.(Thr156Leufs*2)) and with VAF's at diagnosis (WGS) of 40% and 44% in Tw1 and 2 respectively, supporting extensive clonal expansion at diagnosis of the cells carrying this deletion. Our interest for this variant was fueled by *UBA2*'s recent implication as a novel BCP-ALL driver gene[56,75] and reported enrichment *UBA2* variants in *ETV6-RUNX1* positive cases.[76] Further analyses with custom dPCR assays readily detected the *UBA2* deletion at birth (Tw1 3% and Tw2 6%), which was in agreement with it being present in pre-leukemic clones (shared variant). Yet, this was surprising as the shared complex rearrangement we considered being the initiating event had been below limit of detection (LoD) at birth. To our great surprise, dPCR assays also detected the *UBA2* deletion in remission samples (collected 2.5/3 years post diagnosis) of both twins (2,6% in both). Manual inspection of WGS confirmed this finding in Tw1 with 1 out of 45 reads (2,2%) supporting deletion in remission. We further confirmed our results by ddPCR in remission DNA from saliva (lymphocyte content approximately 13,5%[57]), again detecting the *UBA2* deletion at low copy numbers (Tw1 1,14%, Tw2 0,78%). Collectively, our data suggested a clone harboring *UBA2* deletion, but not the complex rearrangement, remained in both twins after several years in remission.

The abundance of *UBA2* deletion at birth was at minimum 30 times higher than that of the complex rearrangement, and the *UBA2* deletion was reliably detected in remission samples from two separate timepoints by both chip-based dPCR and ddPCR. This led us to conclude the *UBA2* deletion most likely gave rise to a subclone of pre-B cells that settled in bone marrow of both twins. Subsequently, the complex rearrangement arose in a del*UBA2* positive cell in one twin, initiating a pre-leukemic clone that spread to the twin sibling over the placenta. We therefore hypothesized that *UBA2* deletion predisposed the occurrence of t(11;12;21)(q23;p13;q22). To the best of my knowledge, *UBA2* has not previously been reported to precede any of the known initiating events associated to BCP-ALL.

A wider elaboration on possible mechanisms by which disruption of *UBA2* may predispose to/promote BCP-ALL can be found in **paper III**. In short, UBA2 protein is essential in SUMOylation[77], a post-translational protein modification process that regulates crucial cellular processes such as gene expression, cell signaling, DNA-damage repair, cell cycle progression and apoptosis[78]. As SUMOylation has been found upregulated in cancer cells[77] while our detected *UBA2* deletion was predicted to result in downregulation, we find no clear mechanistic explanation to the deletions BCP-ALL predisposing or promoting effects. However, as dysregulated SUMOylation contributes to cancer development is complex, for example operating in a context-dependent manner and cross-talking with many other regulatory systems[80] we argue that a role for downregulated SUMOylation in leukemogenesis should not yet be ruled out.

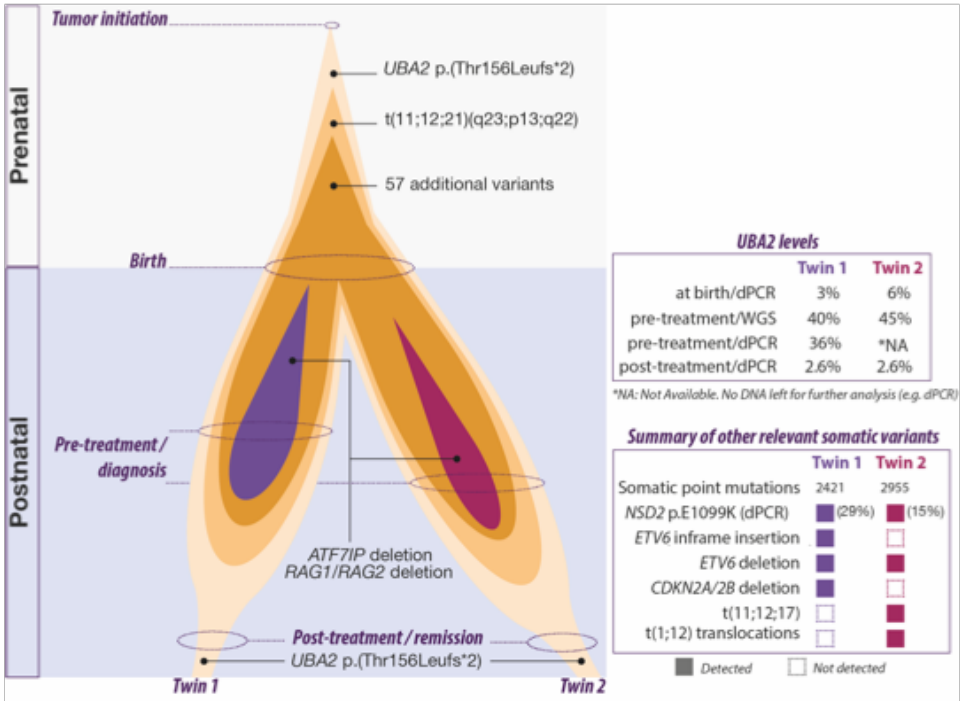


Figure 8. *Overview of clonal evolution in concordant leukemias of monozygotic twins. Beginning with emergence of a subclone harboring UBA2 deletion in which a pre-leukemic clone is then initiated by a shared complex rearrangement generating the ETV6-RUNX1 fusion. Continued by the postnatal addition of secondary putative drivers. Adapted from Bang et al.*[219] *with permission from the publisher.*

Further to this point, BCP-ALL's have been found to harbor somatic *UBA2* variants likely downregulating SUMOylation.[55] Also, it seems highly unlikely a complex rearrangement creating the *ETV6-RUNX1* fusion would develop by pure chance in an *UBA2* deleted B-cell clone. In fact, *ETV6-RUNX1* fusions are thought to arise through illegitimate recombination of multiple simultaneous DSBs by NHEJ[88], a process to which SUMOylation is essential[82], providing an explanation as to how *UBA2* deletion may have promoted the complex rearrangement. As for offering an explanation to how downregulated SUMOylation may promote progression, i.e. secondary driver events in leukemia, we found that the main drivers of secondary genetic events in *ETV6-RUNX1* positive leukemias are RAG endonucleases.[87] Their normal function is to facilitate V(D)J recombination in T- and B-cells,[87] in turn tightly regulated by NHEJ[85,86] to which, again, SUMOylation is essential[82].

4.1.2 Constitutional predisposing variants (paper I, III)

4.1.2.1 Constitutional translocation disrupting transcription factor ETV6 (paper I)

In **paper I** we studied a family with two second degree relatives who developed HeH BCP-ALL; a young female (IV:2) at age 12 years and her maternal aunt (III:3) at age 8 years (***Figure 4*** in Methods). They were treated according to the Nordic Society of Pediatric Haematology and Oncology (NOPHO) high-risk ALL-92[247] (due to myeloid cell surface markers) and NOPHO standard-risk ALL-2008[62] protocol, respectively. HeH karyotypes detected at diagnosis were confirmed and further detailed in our study using SNP-arrays. The SNP-array results are summarized together with clinical

data in ***Table 6***. Complete remission had been sustained for 24 and 5 years respectively at the time of publication. The affected females and 7 additional family members out of those tested (***Figure 4*** in Methods section) had been found carriers of a constitutional t(12;14)(p13.2;q23.1) prior to our study. The pedigree suggested this constitutional translocation segregated with predisposition to childhood leukemia in this family, inherited in an AD fashion with incomplete penetrance.

Patient	Age at dx	Karyotype by SNP-array	WBC*	Blast cells in PB at dx	Blast cells in BM at dx	Treatment protocol	Events	Severe adverse effects
III:3	8	57,XX,t(12;14)(p13.2;q23.1),+X,+X,+6,+10, +der(14)t(12;14)(p13.2;q23.1), +14,+17,+18,+18,+21,+21	2.8	0%	95%	NOPHO-ALL 1992-HR	No	osteonecrosis
IV:2	12	54,XX,t(12;14)(p13.2;q23.1), +X,+4,+6, +der(14)t(12;14)(p13.2;q23.1), +17,+18,+21,+21	4.4	10%	95%	NOPHO-ALL 2008-SR	No	osteonecrosis, vincristine neuropathy, osteoporosis with compression fractures

BCP, B-cell precursor; dx, diagnosis; WBC, white blood cells; PB, peripheral blood; BM, bone marrow;
NOPHO, Nordic Society of Paediatric Haematology and Oncology; SR, standard-risk NOPHO-ALL protocol; HR, high-risk NOPHO-ALL protocol;
**the highest value (10E9/l) at diagnosis*

Table 6. *Clinical data and karyotypes defined by SNP-array (**paper I**). Table adapted from Järviaho and Bang et al.*[119] *with permission from the publisher.*

In our study, we confirmed the constitutional t(12;14)(p13.2;q23.1) in all three analyzed carriers (mother III:2, maternal aunt III:3, and girl IV:2) by WGS analysis and subsequent Sanger sequencing. Identical breakpoints mapped to intron 1 of both *ETV6* (chr 12) and *RTN1* (chr 14) (***Figure 9A***). No other constitutional variants, especially considering genes previously associated with predisposition to hematological malignancies [90,92,248,249], were detected. *ETV6* is closely associated to BCP-ALL through the recurrent somatic t(12;21)/*ETV6-RUNX1* defining the second largest subtype and being a frequent target (22-26% of unselected BCP-ALL cohorts) deletions in clonal evolution.[108,110] ETV6 is also a recurrent target of constitutional variants predisposing to BCP-ALL.[113–116,250,251] However, we found no support for *RTN1* having a role in development of sporadic nor predisposed childhood leukemia. Therefore, we assessed *ETV6* was more likely to be the responsible gene behind suspected leukemia predisposition in our studied family.

At the time of the study, this was the 23rd family (total of 88 individuals) reported with a constitutional pathogenic variant in *ETV6*.[113–118,250,251] 16 out of 50 carriers from 11 of these families had been affected by ALL (Supplemental table S4, **paper I**), 15 of which were of B-cell precursor subtype.[113–116,250,251] Inheritance patterns were AD with incomplete penetrance. Also, 38 additional non-familial cases of ALL (35 confirmed B-cell precursor subtype) with constitutional pathogenic variants in *ETV6* had been reported.[114,121] Beyond predisposition to ALL, Thrombocytopenia 5 (THC5) was abundantly associated with constitutional *ETV6* variants, as well as occasional cases of other hematological malignancies (Supplemental table S4, **paper I**). Interestingly, the phenotypic panorama of previously reported ALL cases with constitutional *ETV6*-mediated predisposition is markedly similar to our studied cases with predominance of BCP-ALL (94%) and HeH subtype (80%), older age at diagnosis (median 13.3 vs 6.8 years without risk variants) and leukocyte counts <50x10^9/L at diagnosis.[114,122]

ETV6 is a powerful transcriptional repressor essential to both fetal and post-natal hematopoiesis.[105–107] The predominant functional consequence of previously reported constitutional *ETV6* variants was a dominant negative effect on WT ETV6 protein. Only one previous variant was reported to cause haploinsufficiency of *ETV6* and was associated with mixed phenotype acute leukemia (MPAL) in 1/3 carriers.[117] The t(12;14)(p13.2;q23.1) studied in **paper I** created two fusion genes, *ETV6-RTN1* and *RTN1-ETV6*, with frameshifts downstream of the fusion points introducing premature stop codons (***Figure 9A***). qPCR validated fusion transcript expression in both leukemias and constitutionally in the

three analyzed carriers (III:2, III:3, and IV:2) (***Figure 9B***). However, in silico exploration of the theoretical fusion proteins showed that no functional domains were retained (***Figure 9C***), and no commonly occurring functional domains was created by the frameshift. This rendered any resulting fusion gene products non-functional. Therefore, we anticipated fusion transcripts were most likely targeted for nonsense-mediated decay. Instead, heterozygous loss of *ETV6,* caused by the constitutional translocation, was predicted to result in haploinsufficiency of *ETV6* and its encoded protein.

To confirm haploinsufficiency was the functional consequence of t(12;14)(p13.2;q23.1), we analyzed expression of WT *ETV6* (and *RTN1*) using qPCR. Unfortunately, results were inconclusive as tested carriers (again III:2, III:3, and IV:2) exhibited contradictive patterns in constitutional expression of WT *ETV6*. As expression analyses were performed on total RNA from whole blood in remission, natural variations in composition of cell types in peripheral blood samples, with in turn substantially differing expression levels of *ETV6*[252] could have been concealing differences in expression. In addition, WT *ETV6 e*xpression in leukemic bone marrow did not differ substantially in our two cases compared to sporadic BCP-ALLs with or without genetic aberrations in *ETV6*, thus also producing an inconclusive result. Altogether, because our analyses were inconclusive we suggest haploinsufficiency of *ETV6* may yet be a mechanism by which constitutional pathogenic variants in *ETV6*, such as the t(12;14)(p13.2;q23.1) studied here, causes predisposition to BCP-ALL. Further studies to explore haploinsufficiency as a functional consequence of constitutional ETV6 variants are warranted. Also, characterizing the molecular effects of *ETV6* haploinsufficiency in normal B-cell development would be of value for the understanding of pathogenicity in these cases.

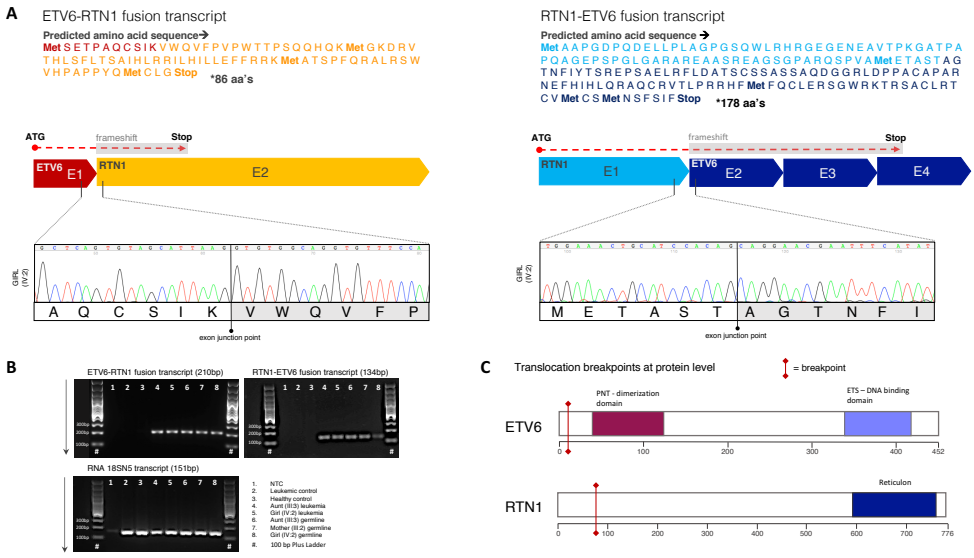


Figure 9. *Chimeric fusion gene transcripts ETV6-RTN1 and RTN1-ETV6. (**A**) In silico prediction and electropherograms of novel exon–exon junctions (**B**) Gel images (2% agarose gel) supporting expression of both fusion transcripts in remission and leukemia. Positive control ribosomal RNA 18S5N is shown in the bottom gel image. 100bp Plus DNA Ladder from GeneRuler (Thermo Fisher Scientific). (**C**) Functional domains are not retained as breakpoints located upstream of functional domains; all sequence downstream subject to frameshift introducing premature stop codons. Adapted from Järviaho and Bang et al.*[119] *with permission from the publisher.*

4.1.2.2 Constitutional NF1 variant (paper III)

Recent insights to the contribution of constitutional predisposition to BCP-ALL etiology, also in non-familial cases of BCP-ALL, motivated our analysis of constitutional variants in the twins studied in **paper II**.[253] This analysis detected a shared heterozygous missense variant in *NF1* (chr17:g.29557883C>A, NM_000267.3:c.3137C>A) NP_001035957.1:p.(Thr1046Lys), a tumor suppressor gene[94] in which pathogenic variants cause susceptibility to a range of malignancies including leukemias,[95,96] constituting the cancer predisposition syndrome neurofibromatosis type 1 (NF1).[53] No somatic second hit variants or LOH was detected in *NF1* in either of the twins' leukemias.

The variant is not previously described in NF1 and both twins had normal psychomotor development, growth parameters within the normal range and no malformations or dysmorphic features associated to NF1.[53] Also the mother, from who the variants was inherited, was healthy. The variant was novel, not found in gnomAD[49], COSMIC[50] nor LOVD[51], and seemingly highly damaging with a CADD[52] (v1.4) c-score of 27.1, also supported by several *in silico* predictions (Supplemental Table 3, **paper III**). Though the variant was novel and predicted highly damaging the absence of any phenotypic expressions obligate for NF1 diagnosis caused us to classify this aberration as a variant of unknown significance (VUS) and dismiss any suspicions of NF1 diagnosis.

While not causing NF1, the detection of this novel highly damaging *NF1* variant in twins with concordant BCP-ALL was still intriguing. We found previous reports of constitutional damaging *NF1* variants in cases with cancer lacking other clinical signs of the syndrome,[96] suggesting cancer predisposition without NF1 could be the case also in our studied twins. In further support of a potential BCP-ALL predisposing effect, *NF1* LOH in malignant cells of NF1 patients is common[56]heterozygous loss of *NF1* is thought result in haploinsifficiency,[95] and. Moreover, significant pairwise co-occurrence of somatic *NF1* and *UBA2* variants has been demonstrated in childhood BCP-ALL,[56] raising the question whether emergence of *UBA2* deletion in the presence of the constitutional *NF1* variant was not a coincidence in our cases? Altogether, we suggest that a cancer predisposing effect of our reported variant is not excluded.

4.2 Progression to disease onset

While previous studies have detected traces of *ETV6-RUNX1* fusions in 1-5% of healthy neonates, indicating presence of pre-leukemic clones at birth,[74,254] comparison to incidence numbers of the corresponding BCP-ALL subtype (approximately 1:10.000 = 0,01% of children aged 0-15 in Europe and United States) indicate that only a fraction progress to overt disease.[74] This illustrates the need for 2nd hits in progression to BCP-ALL[255] and that the course of progression also is subject to intrinsic and/or extrinsic drivers and modifiers. The journey of a pre-leukemic clone from initiation to overt disease is thought to be strongly associated with the dose and timing of a child's exposure to common infectious disease and other microbes.[185,199,201,256–258] To this end, studies of clonal evolution in BCP-ALL, materialized by genetic aberrations, and epidemiological aspects of the disease may both serve as a valuable and accessible recourses to increase our understanding of BCP-ALL etiology.

4.2.1 Clonal evolution (paper I, III)

4.2.1.1 Clonal evolution of concordant BCP-ALL in monozygotic twins (paper III)

In **paper III**, the "first act" in clonal evolution of concordant leukemias is composed of the shared initiating variants delineated in section 4.1.1.1 above. The "second act" was explored by analysis of WGS data for somatic variants unique to either twins' leukemia, uncovering 10 deletions and 2421 SNVs/indels (915 protein coding) in Tw1 and 10 deletions, 2 translocations, one complex rearrangement and 2955 SNVs/indels (1040 protein coding) in Tw2 (Table 2, **paper III**). LOH analysis for shared and unique variants did not reveal any additional deletions, duplications or copy number neutral events.

Among these unique variants, we highlighted six analogous deletions in each twin, meaning they affected the same chromosomal regions although breakpoints differed (***Figure 8*** in section 4.1.1.2). Interestingly, analogous deletions encompassed *ETV6*, *ATF7IP*, and *RAG1/RAG2*, all of which predominantly result from aberrant RAG-mediated recombination[59,87] and are known recurrent targets of secondary events in *ETV6-RUNX1* positive BCP-ALL[92,93] with frequent co-occurrence of *RAG1/RAG2* and *ATF7IP* deletions[56,87]. Also, *NSD2* (*WHSC1*) p.E1099K (COSMIC ID: COSM379334) (Supplemental Table 5, **paper III**), the most common *NSD2* variant in *ETV6-RUNX1* positive BCP-ALL[58], was uncovered in both twins and most likely independently acquired. This *NSD2* variant also frequently co-occurs with variants in *UBA2*.[56]. Two additional variants were highlighted in Tw1's leukemia, a deletion in the recurrent second-hit region *CDKN2A/CDKN2B*[59] and an in-frame insertion in *ETV6*.

In conclusion, the presence of the genetic variants described above are in line with the notion that additional somatic events are required for progression to overt BCP-ALL of *ETV6-RUNX1* positive pre-leukemic clones.[32,33] Also, abundance of CNV's in each leukemia was well within range of reported average, 3.5 (range 0-14), in pediatric ALL, supporting gross genomic instability was not a common trait. Altogether, the analogous deletions and other secondary putative driver variants identified, supported the independent albeit convergent clonal evolution of the twins' leukemias, orchestrated by the selective pressure of shared genetic and environmental drivers.

4.2.1.2 Clonal evolution of HeH BCP-ALL in predisposed t(12;14) carriers (paper I)

Analysis of somatic variants, reflecting clonal evolution, in BCP-ALL's of leukemia predisposed t(12;14)(p13.2;q23.1) carriers revealed 10 and 7 somatic aberrations of interest respectively for IV:2 and III:3 (Supplemental Table 3, **paper I**). Although no somatic second hits to *RTN1* or *ETV6* were found, 13 of the 17 genes that were implicated (in two of them also the very same variant) were previously reported targets of somatic second hits in childhood ALLs (Pediatric Cancer Data Portal, St. Jude Children's Research Hospital [2015-2018]; https://pecan.stjude.cloud/ home, accessed 20 July 2018) (Supplemental Table 3, **paper I**). None of these variants were shared in the leukemias indicating divergent clonal evolution despite common predisposing variant. Also of note, der(14)t(12;14)(p13.2;q23.1) was among the non-randomly gained chromosomes in both leukemias, likely explained by chr14 being frequently gained in HeH BCP-ALL.[259] SNP arrays did not reveal any other somatic CNAs of genes relevant to leukemogenesis.

High hyperdiploidy has been shown to be initiated in utero, establishing this aberrations role as a probable leukemia initiating event.[260–264] These ploidy states were thus most probably the initiating

events also in our cases. In addition, for individuals carrying leukemia predisposing constitutional variants in *ETV6*[114,251] as well as pre-leukemic clones initiated by either *ETV6-RUNX1* fusion or HeH state[113,122,126,187,255,265], additional somatic variants (2nd hits) are required for progression to overt BCP-ALL. This is made evident in our cases by the above-described somatic variants, however the phylogenetics of somatic events required for overt leukemia in cases with constitutional pathogenic variants in *ETV6* remains to be delineated.

4.2.2 Seasonal variation in onset of BCP-ALL (paper IV)

In **paper IV**, we implemented Bayesian Generalized Autoregressive Integrated Moving Average with external variables model (GARIMAX), a powerful statistical model for analysing and forecasting time series data, and detected seasonal variation of BCP-ALL onset in a Swedish population-based cohort of 1380 cases. Informative seasonal variation with peaks in two different quarters, Jun-Aug and Jul-Sep, was identified. Subgroup analysis of genetic subtypes HeH and *ETV6-RUNX1* did not show informative seasonal variation, likely due to limitations in sample size of each subtype. We also intended to asses seasonal variation by date of birth in the same cohort. However, as additional exclusion of cases that had not yet reached age 18 years at the time of our analysis was required, the cohort became too small for detection of an informative seasonal wave by GARIMAX.

As described in more detail in methods-section our cases were aggregated into quarters (of three different types) to optimize chances of detecting informative seasonal variation by increasing numbers of cases in each analyzed time-frame. Aggregation to quarters also helps to account for any year-to-year differences in seasonal variation of a putative environmental driver of BCP-ALL progression, e.g. peak incidence month of common infections varying slightly each year. The GARIMAX analysis was composed of two parts. The first identifying presence of periodicity, using harmonic functions to predict all theoretical variants of seasonal curves and subsequently comparing against curves of the analyzed cohort to identify matches. A significant sinus and cosine wave was identified for the first and third quarter type, respectively, indicating non-random periodicity. The second part identified informative increases of case numbers in specified quarters. Within each quarter type, a base quarter (lowest number of cases) was chosen and the three remaining quarters were compared to the base for identification of informative increases. Informative increases were identified in quarters Jul-Sep and Jun-Aug for first and third quarter type, respectively (Table 4, **paper IV**).

Together, data shows informative seasonal variation in onset of BCP-ALL with peak quarters Jul-Sep and Jun-Aug respectively. The two months overlapping between these two peak quarters were July and August. Manually examining absolute numbers of cases diagnosed in these respective months made evident that August had the highest number with 138 cases (10% of total) while July had 110 (8%) (***Table 7***). June and September, included in either of the respective informative peak quarters but not overlapping between them, had 126 (9%) and 125 (9%) cases respectively. Absolute number in the remaining 8 months, not included in informative peak quarters, ranged from 91 cases in December to 125 in April. Our collective interpretation was that August is the peak month of BCP-ALL onset. However, it is essential to point out that the results with statistical significance generated by GARIMAX model merely informs us of the peak quarters, not specifying individual months. Descriptive data on distribution of age at diagnosis (Supplemental figure S7 in **paper IV**) and genetic subtype (***Table 7***), respectively, did not differ for cases diagnosed in August, nor any other months included in the peak quarters, compared to months not included in the peak quarters and the whole year aggregate.

Time-periods	BCP-ALL	BCP-ALL % of total	HeH	HeH % of cases per time-period	*ETV6-RUNX1*	*RUNX1* % of cases per time-period
Jan	108	8%	37	34%	25	23%
Feb	104	8%	35	34%	19	18%
Mar	120	9%	28	23%	29	24%
Apr	125	9%	48	38%	21	17%
May	100	7%	24	24%	16	16%
Jun	126	9%	47	37%	20	16%
Jul	110	8%	36	33%	18	16%
Aug	**138**	**10%**	**48**	**35%**	**27**	**20%**
Sep	125	9%	40	32%	22	18%
Oct	119	9%	31	26%	31	26%
Nov	114	8%	40	35%	23	20%
Dec	91	7%	30	33%	21	23%
Whole year	**1380**	100%	**444**	**32%**	**272**	**20%**

Table 7. *Descriptive data. Table displays distribution of cases, total and per investigated subtype, diagnosed in each respective months and total for the whole year (equaling total studied cohort 1380 cases). % of total to illustrate how onset is distributed throughout the year. Distribution of the two respective subtypes in % per month (column 5 and 7 row 1-12) for comparison to the whole year (column 5 and 7 last row). Adapted form manuscript of paper IV.*

Exposure to common microbes and the consequential training of our immune system, as a modulator of progression in BCP-ALL, has been extensively studied and debated (section 1.3.2). Epidemiological studies support that early (within first year of life) exposure to microbes protects against BCP-ALL development[199,266], while exposure to infections later in life appears to be required for the same[201–206]. That exposure to infections appears required for development of BCP-ALL has motivated studies of seasonal variation in onset of childhood ALL. Seasonal variation in onset is here used as a proxy for seasonally variable common infections acting as environmental modulators of disease progression in ALL. After reviewing the literature (Table 1, **paper IV**) we can conclude that seasonal variation in ALL onset has been identified in some studies, but was not supported by all. Also, in studies where peaks in onset were detected these were scattered throughout the year. A number of factors could explain these divergent and inconclusive results, such as differences in methods applied, cohort size as well as cohort characteristics such as age at diagnosis (only children or also adults) and disease entities studied (T- and B- or only B-cell ALL). One should also consider that seasonal patterns and panorama of common infections differ by region or climate zone. In addition, susceptibility to develop BCP-ALL, conferred by constitutional genetic variation, has been shown to vary with ethnicity. Also, the protective effect of early (first year of life) microbial exposures are likely to vary with differences in general hygiene conditions between HIC and LMIC. Both these underlying factors could theoretically affect the extent to which progression of BCP-ALL may be modulated by exposure to seasonally variable common infections. This would in turn influence the likelihood of detecting seasonal variation in onset in different regions of the world.

A main strength of our study was the application of a powerful modern statistical method (GARIMAX) increasing reliability of results. In addition, we restricted our cohort to only include BCP-ALL cases in children and adolescents (0-18 years), excluding T-ALL and adult cases. This ensured results were representative of a theoretically etiologically distinct group of patients. The importance of such a distinction is emphasized by previous identification of epidemiologic risk factors specifically associated with BCP-ALL and not T-ALL, and with particular diagnostic age-groups of ALL (reviewed by Williams et al).[88] Although studies are scarce, some heterogeneity in environmental exposures associated to molecular subtypes of BCP-ALL has also been observed[88].

This encouraged our subgroup-analyses of the two most frequent molecular subtypes of BCP-ALL, HeH and *ETV6-RUNX1*. Unfortunately, GARIMAX failed to produce informative results in either of these subgroups due to the limited numbers of cases available for analysis in each group (HeH=444 and *ETV6-RUNX1*=272). To this end, pooling larger datasets of cases to create sufficient numbers for analysis also in subgroups would be of value.

For BCP-ALL the induction time, i.e. the time from exposure to a final driver of disease progression until clinically overt disease, remains to be elucidated. This limits our ability to associate a seasonal peak in onset of BCP-ALL, i.e. final progression to overt leukemia, with exposure to a seasonally variable common infection as a potential driver of that final progression. Therefore, we can only speculate about the environmental etiology behind our observed August onset peak. Assuming a "longer" induction time of 6-9 months, the August onset peak could be associated with the peak of common infections during winter months in Sweden; either the accumulated load of several infections during this period, or one to a few infections being more potent drivers of disease progression. The latter could also be applied to a "shorter" induction time of 1-2 months, associating the August peak to one or a few scarce common infections peaking during summer months, again being potent drivers of progression. Lastly, we speculate that a short but marked decrease in spread of common infections during Swedish summer holidays (mid-June to mid-August), could postpone a portion of cases from onset in July and instead accumulating in August. This model assumes an even "shorter" induction time of a 2-4 weeks. Sweden is well known for its prolonged summer vacation when the vast majority of children, both school- and pre-school-, and their parents are on summer vacation for 4-8 weeks. Which of these models are more likely again depend on the unknown variable of induction time. The possibility of patient's/parent's delay in seeking care for their children during summer vacation, upon leukemia associated symptoms, as an explanation to our observed August onset peak was considered. Although we cannot confidently exclude this explanation, we do consider it less likely as symptoms of BCP-ALL at diagnosis are acute and progress quickly. Also, Swedish parents are in general liberal when it comes to seeking healthcare for their children, and face no generalized economic or accessibility obstacles to do so.

In conclusion we detected informative seasonal variation in onset of BCP-ALL, in a Swedish population-based cohort, with peak incidence of onset in August. We suggest four speculative explanatory models for the August onset peak, relating exposure to common infections to progression to overt disease. The likelihood of each model depends on the induction time of childhood BCP-ALL, which remains unknown. Our study holds advantages to previous reports in applying a powerful statistical model and studying a theoretically etiologically distinct (in regard to immunophenotype and age at diagnosis) cohort of BCP-ALL cases. Nonetheless, expanding the size of the cohort would further strengthen support for seasonal variation in BCP-ALL and enable subgroup-analysis of molecular subtypes. This could be facilitated by including cases from other Nordic countries, with corresponding panorama and seasonal patterns of common infections.

5 Conclusions

This thesis represents the tale of BCP-ALL etiology. It begins with congenital susceptibility through constitutional predisposing variants and prenatally initiated pre-leukemic clones, and continues with progression to overt disease through somatic clonal evolution, including the intrinsic (genetic) and extrinsic (environmental) drivers of this process. Its constituent papers each contribute to a chapter of this tale, starting with the addition of a constitutional structural variant in *ETV6* predisposing to BCP-ALL, conferring susceptibility through insufficient (monoallelic) expression of *ETV6*. Next, it includes the development of a methodological strategy for backtracking somatic variants of childhood BCP-ALL to neonatal samples, inferring in utero initiation of pre-leukemic clone. Predisposition, initiation of a common pre-leukemic clone and clonal evolution of concordant leukemias in a monozygotic twin pair was genetically characterized. This revealed another shared prenatally acquired but non-leukemic clone, with deletion in a BCP-ALL associated gene, in which leukemia had been initiated. The deletion suggestively predisposed this clone to leukemia initiation. Last but not least, informative seasonal variation in onset of BCP-ALL, with a peak interpreted to fall in August, was detected in a Swedish population-based BCP-ALL cohort aged 0-18 years at diagnosis.

6 Future perspective

Studies of constitutional genetic predisposition will continue to be key in uncovering different aspects of BCP-ALL etiology. Today, we have limited knowledge about the prevalence, penetrance, prognosis and pathogenesis in leukemia with constitutional predisposition. This results in many uncertainties regarding medical care and benefits from pre-symptomatic testing for this group of patients. Despite the growing evidence for the contributions of constitutional predisposition to BCP-ALL, we are in need of further studies to understand these aspects of childhood BCP-ALL predisposition.

Further studies backtracking BCP-ALL to time of birth are important to understand the frequency of prenatal leukemia initiation. Screening larger cohorts of healthy newborns for the known recurrent initiating and canonical somatic variants of BCP-ALL will add insights to their frequency and pathogenicity. I imagine that understanding how and why these variants emerge to initiate pre-leukemic clones will be instrumental to developing primary preventive actions against initiation of BCP-ALL in the future.

Delineating clonal evolution in BCP-ALL in a subtype specific-manner together with pinpointing the intrinsic (genetic, epigenetic) and extrinsic (environmental) drivers of this process will expand our understanding of how and why pre-leukemic clones progress to onset of BCP-ALL. Such insights hold the potential to guide secondary preventive measures in children at risk of BCP-ALL, i.e. carriers of pre-leukemic clones, which could motivate screening of newborns at large scale to detect those at risk.

Further survival improvements in patients with BCP-ALL are unlikely to come from intensification of chemotherapy. Therefore, efforts to further detail and understand this etiologic tale of BCP-ALL are pivotal to future improvements in treatment and, hopefully, the application of preventive measures. In addition to novel targeted drugs and individually tailored treatments, guided by precision medicine, the future is prevention, prevention, prevention!

7 Acknowledgements

First and foremost, a big thanks to the patients and families who have made the studies in this thesis possible. Thank you also to all the treating physicians who have recognized the uniqueness of each case, brought them to research and contributed with their clinical expertise.

To my main supervisor **Ann Nordgren** for your endless support and belief in me. For bringing curiosity, excitement, passion, personal engagement, endurance and lots of warmth to my time in your team. For that I am immensely grateful!

My co-supervisor and closest college **Fulya Taylan** ✿ who has become a dear friend. For your calm, humor and way of making everything feel easier. For sharing your deep knowledge in genetics, for your patience in explaining complicated matters and for listening and sharing achievements and rough patches. You're the best!

My co-supervisor, friend and soon to be college in pediatrics ☺ **Vasilios Zachariadis**. For being a gold mine of knowledge, insight and strong opinion.

To my co-supervisor **Magnus Nordenskjöld** for being the first to give me a lab-roof over my head and later motivating me to indulge in a PhD.

To my co-supervisors **Yenan Brycecon** for always being supportive, and **Arja Harila** for sharing experience in pediatric oncology.

Old and new friends at the lab: **Maria** and **Ellika** for allot of fun and laughter. **Raquel** for being such a happy spirit. **Jesper** for being a person of bottomless kindness, patience and positivism. **Sofia Frisk** for being a good friend and parental-leave companion. **Anna H, Anna L, Cecilia A, Katja E**, **Nina J, Samina A, Bianca T, Alice C, Dominyka B, Jakob S**, **Carolina MG**, **Marlene E**, **Mansoureh S, Sailan W**, **Isabel TP, Angelica DV, K-J dG, Wolfgang H, Josephine W, Ingegerd IÖ, Emelie P, Ida N, Alexandra W.**

People from Clinical Genetics that have helped and inspired me: **Gisela B, Ann Nhi T, Britt-Marie A, Malin K, Kicki L, Erik I, Erik B, Giedre G, Daniel N, Johanna L, Helena M, Agne L.**

Also to **Jan Inge-Henter** for giving me my first job in research and **Marie Meeths** for being a great teacher and friend.

Collaborators in research: **Gleb Bychkov, Niklas Engsner**, **Claes Strannegård** for enjoyable collaborations. **Yunxia Lu** for welcoming me at UCI.

I am also grateful for the financial support of **Barncancerfonden**, **Mary Béves Stiftelse för Barncancerforskning, Sällsyntafonden, Svenska Sällskapet för Medicinsk Forskning** and **Karolinska Institutet.**

Thank you for educating and supporting me, giving me courage and making this thesis possible!

Last but not least, to my family: My loves **Mathilda**, **Alicia** and **Erik.** My parents **Madeleine** and **Peter**. My brothers **Jacob** and **Oscar**. For all your love and support, thank you

8 References

1. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
2. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
3. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery* **12**, 31–46 (2022).
4. Vaux, D. L. In defense of the somatic mutation theory of cancer. *BioEssays* **33**, 341–343 (2011).
5. Sonnenschein, C. & Soto, A. M. The aging of the 2000 and 2011 Hallmarks of Cancer reviews: A critique. *J Biosci* **38**, 651–663 (2013).
6. Sonnenschein, C., Soto, A. M., Rangarajan, A. & Kulkarni, P. Competing views on cancer. *J Biosci* **39**, 281–302 (2014).
7. Sonnenschein, C. & Soto, A. M. Why is it that despite signed capitulations, the war on cancer is still on? *Organisms. Journal of Biological Sciences* **1**, 45–52 (2017).
8. Sonnenschein, C. & Soto, A. M. Over a century of cancer research: Inconvenient truths and promising leads. *PLOS Biology* **18**, e3000670 (2020).
9. Nordling, C. O. A New Theory on the Cancer-inducing Mechanism. *Br J Cancer* **7**, 68–72 (1953).
10. Knudson, A. G. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820–823 (1971).
11. Weinberg, R. A. Tumor suppressor genes. *Science* **254**, 1138–1146 (1991).
12. Cantley, L. *et al.* Oncogenes and signal transduction. *Cell* **64**, 281–302 (1991).
13. Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair and mutagenesis. *Environ Mol Mutagen* **58**, 235–263 (2017).
14. Jones, P. A. & Baylin, S. B. The Epigenomics of Cancer. *Cell* **128**, 683–692 (2007).
15. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339**, 1546–1558 (2013).
16. Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nat Rev Genet* **13**, 795–806 (2012).
17. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
18. Huang, K. *et al.* Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* **173**, 355-370.e14 (2018).
19. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339**, 1546–1558 (2013).
20. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians* **72**, 7–33 (2022).
21. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
22. Horne, S. D., Pollick, S. A. & Heng, H. H. Q. Evolutionary mechanism unifies the hallmarks of cancer. *International Journal of Cancer* **136**, 2012–2021 (2015).
23. Boddy, A. M. The need for evolutionary theory in cancer research. *Eur J Epidemiol* (2022) doi:10.1007/s10654-022-00936-8.
24. Saxena, K., Subbalakshmi, A. R., Kulkarni, P. & Jolly, M. K. Cancer: More than a geneticist's Pandora's box. *J Biosci* **47**, 21 (2022).

25. Weinberg, R. A. Coming Full Circle—From Endless Complexity to Simplicity and Back Again. *Cell* **157**, 267–271 (2014).

26. Pfister, S. M. *et al.* A Summary of the Inaugural WHO Classification of Pediatric Tumors: Transitioning from the Optical into the Molecular Era. *Cancer Discovery* **12**, 331–355 (2022).

27. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *New England Journal of Medicine* **373**, 2336–2346 (2015).

28. Newman, S. *et al.* Genomes for Kids: The Scope of Pathogenic Mutations in Pediatric Cancer Revealed by Comprehensive DNA and RNA Sequencing. *Cancer Discovery* **11**, 3008–3027 (2021).

29. Rossini, L. *et al.* Diagnostic Strategies and Algorithms for Investigating Cancer Predisposition Syndromes in Children Presenting with Malignancy. *Cancers* **14**, 3741 (2022).

30. Nakano, Y., Rabinowicz, R. & Malkin, D. Genetic predisposition to cancers in children and adolescents. *Current Opinion in Pediatrics* **35**, 55 (2023).

31. Schüz, J. & Roman, E. Childhood cancer: A global perspective. *Cancer Epidemiology* **71**, 101878 (2021).

32. United Nations. *World Population Prostects 2022, Online Edition: Population by select age groups - both sexes*. https://population.un.org/wpp/Download/Standard/Population/ (2022).

33. Atun, R. *et al.* Sustainable care for children with cancer: a Lancet Oncology Commission. *The Lancet Oncology* **21**, e185–e224 (2020).

34. Magrath, I. *et al.* Paediatric cancer in low-income and middle-income countries. *The Lancet Oncology* **14**, e104–e116 (2013).

35. Pritchard-Jones, K. *et al.* Sustaining innovation and improvement in the treatment of childhood cancer: lessons from high-income countries. *The Lancet Oncology* **14**, e95–e103 (2013).

36. Global initiative for childhood cancer. https://www.who.int/publications/m/item/global-initiative-for-childhood-cancer.

37. Howlander, N. *et al.* SEER Cancer Statistics Review, 1975-2018, National Cancer Institute. *SEER* https://seer.cancer.gov/csr/1975_2018/index.html.

38. Steliarova-Foucher, E. *et al.* International incidence of childhood cancer, 2001–10: a population-based registry study. *The Lancet Oncology* **18**, 719–731 (2017).

39. Pui, C.-H., Carroll, W. L., Meshinchi, S. & Arceci, R. J. Biology, Risk Stratification, and Therapy of Pediatric Acute Leukemias: An Update. *JCO* **29**, 551–565 (2011).

40. Bhojwani, D., Yang, J. J. & Pui, C.-H. Biology of Childhood Acute Lymphoblastic Leukemia. *Pediatric Clinics of North America* **62**, 47–60 (2015).

41. Gustafsson, G., Kogner, P., & Heyman, M. *Childhood Cancer Incidence and Survival in Sweden 1984-2010 - Report 2013*. (2014).

42. Linabery, A. M. & Ross, J. A. Trends in childhood cancer incidence in the U.S. (1992–2004). *Cancer* **112**, 416–432 (2008).

43. Kaatsch, P. & Mergenthaler, A. Incidence, time trends and regional variation of childhood leukaemia in Germany and Europe. *Radiat Prot Dosimetry* **132**, 107–113 (2008).

44. Wiemels, J. Perspectives on the causes of childhood leukemia. *Chemico-Biological Interactions* **196**, 59–67 (2012).

45. Howard, S. C. *et al.* Childhood cancer epidemiology in low-income countries. *Cancer* **112**, 461–472 (2008).

46. Belson, M., Kingsley, B. & Holmes, A. Risk Factors for Acute Leukemia in Children: A Review. *Environmental Health Perspectives* **115**, 138 (2007).

47. Greaves, M. F. & Alexander, F. E. An infectious etiology for common acute lymphoblastic leukemia in childhood? *Leukemia* **7**, 349–360 (1993).

48. Giddings, B. M., Whitehead, T. P., Metayer, C. & Miller, M. D. Childhood leukemia incidence in California: High and rising in the Hispanic population. *Cancer* **122**, 2867–2875 (2016).

49. Magalhães, J. P. de. Every gene can (and possibly will) be associated with cancer. *Trends in Genetics* **38**, 216–217 (2022).

50. Moorman, Anthony V. The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. *Blood Reviews* **26**, 123–135 (2012).

51. Brady, S. W. *et al.* The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat Genet* **54**, 1376–1389 (2022).

52. Iacobucci, I. & Mullighan, C. G. Genetic Basis of Acute Lymphoblastic Leukemia. *J Clin Oncol* **35**, 975–983 (2017).

53. Mullighan, C. G. How advanced are we in targeting novel subtypes of ALL? *Best Practice & Research Clinical Haematology* **32**, 101095 (2019).

54. Jeha, S. *et al.* Clinical Significance of Novel Subtypes of Acute Lymphoblastic Leukemia in the Context of Minimal Residual Disease–Directed Therapy. *Blood Cancer Discovery* **2**, 326–337 (2021).

55. Gu, Z. *et al.* PAX5-driven Subtypes of B-progenitor Acute Lymphoblastic Leukemia. *Nature genetics* **51**, 296 (2019).

56. Hein, D., Borkhardt, A. & Fischer, U. Insights into the prenatal origin of childhood acute lymphoblastic leukemia. *Cancer Metastasis Rev* **39**, 161–171 (2020).

57. Meyer, C. *et al.* The MLL recombinome of acute leukemias in 2017. *Leukemia* **32**, 273 (2018).

58. Ryan, S. L. *et al.* Whole genome sequencing provides comprehensive genetic testing in childhood B-cell acute lymphoblastic leukaemia. *Leukemia* **37**, 518–528 (2023).

59. Raynaud, S. *et al.* The 12;21 translocation involving TEL and deletion of the other TEL allele: two frequently associated alterations found in childhood acute lymphoblastic leukemia. *Blood* **87**, 2891–2899 (1996).

60. Olsson, L. & Johansson, B. Ikaros and leukaemia. *Br J Haematol* **169**, 479–491 (2015).

61. Inaba, H. & Mullighan, C. G. Pediatric acute lymphoblastic leukemia. *Haematologica* **105**, 2524–2539 (2020).

62. Toft, N. *et al.* Results of NOPHO ALL2008 treatment for patients aged 1–45 years with acute lymphoblastic leukemia. *Leukemia* **32**, 606–615 (2018).

63. Wolman, I. J. Parallel responses to chemotherapy in identical twin infants with concordant leukemia. *The Journal of Pediatrics* **60**, 91–95 (1962).

64. Clarkson, B. & Boyse, E. Originally published as Volume 1, Issue 7701POSSIBLE EXPLANATION OF THE HIGH CONCORDANCE FOR ACUTE LEUKÆMIA IN MONOZYGOTIC TWINS. *The Lancet* **297**, 699–701 (1971).

65. Ford, A. M. *et al.* In utero rearrangements in the trithorax-related oncogene in infant leukaemias. *Nature* **363**, 358 (1993).

66. Ford, A. M. *et al.* Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *PNAS* **95**, 4584–4588 (1998).

67. Wiemels, J. L., Ford, A. M., Wering, E. R. V., Postma, A. & Greaves, M. Protracted and Variable Latency of Acute Lymphoblastic Leukemia AfterTEL-AML1 Gene Fusion In Utero. *Blood* **94**, 1057–1062 (1999).

68. Greaves, M. F., Maia, A. T., Wiemels, J. L. & Ford, A. M. Leukemia in twins: lessons in natural history. *Blood* **102**, 2321–2333 (2003).

69. Dobbins, S. E. *et al.* The silent mutational landscape of infant MLL-AF4 pro-B acute lymphoblastic leukemia. *Genes Chromosomes Cancer* **52**, 954–960 (2013).

70. Golub, T. R., Barker, G. F., Stegmaier, K. & Gilliland, D. G. Involvement of the TEL gene in hematologic malignancy by diverse molecular genetic mechanisms. *Curr Top Microbiol Immunol* **211**, 279–288 (1996).

71. McLean, T. W. *et al.* TEL/AML-1 dimerizes and is associated with a favorable outcome in childhood acute lymphoblastic leukemia. *Blood* **88**, 4252–4258 (1996).

72. Harbott, J., Viehmann, S., Borkhardt, A., Henze, G. & Lampert, F. Incidence of TEL/AML1 fusion gene analyzed consecutively in children with acute lymphoblastic leukemia in relapse. *Blood* **90**, 4933–4937 (1997).

73. Borkhardt, A. *et al.* Incidence and Clinical Relevance of TEL/AML1 Fusion Genes in Children With Acute Lymphoblastic Leukemia Enrolled in the German and Italian Multicenter Therapy Trials. *Blood* **90**, 571–577 (1997).

74. Mori, H. *et al.* Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A* **99**, 8242–8247 (2002).

75. Schäfer, D. *et al.* Five percent of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening. *Blood* **131**, 821–826 (2018).

76. Lausten-Thomsen, U. *et al.* Prevalence of t(12;21)[ETV6-RUNX1]–positive cells in healthy neonates. *Blood* **117**, 186–189 (2011).

77. Lausten-Thomsen, U. *et al.* ETV6-RUNX1 transcript is not frequent in early human haematopoiesis. *European Journal of Haematology* **81**, 161–162 (2008).

78. Eguchi-Ishimae, M. *et al.* Breakage and fusion of the TEL (ETV6) gene in immature B lymphocytes induced by apoptogenic signals. *Blood* **97**, 737–743 (2001).

79. Zuna, J. *et al.* ETV6/RUNX1 (TEL/AML1) is a frequent prenatal first hit in childhood leukemia. *Blood* **117**, 368–369 (2011).

80. Škorvaga, M. *et al.* Incidence of Common Preleukemic Gene Fusions in Umbilical Cord Blood in Slovak Population. *PLOS ONE* **9**, e91116 (2014).

81. Ornelles, D. A., Gooding, L. R. & Garnett-Benson, C. Neonatal Infection with Species C Adenoviruses Confirmed in Viable Cord Blood Lymphocytes. *PLOS ONE* **10**, e0119256 (2015).

82. Kosik, P. *et al.* Low numbers of pre-leukemic fusion genes are frequently present in umbilical cord blood without affecting DNA damage response. *Oncotarget* **8**, 35824–35834 (2017).

83. Schäfer, D. *et al.* 5% of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening. *Blood* blood-2017-09-808402 (2018) doi:10.1182/blood-2017-09-808402.

84. Greaves, M., Cazzaniga, V. & Ford, A. Can we prevent childhood Leukaemia? *Leukemia* **35**, 1258–1264 (2021).

85. He, J.-R. *et al.* Maternal Infection in Pregnancy and Childhood Leukemia: A Systematic Review and Meta-analysis. *The Journal of Pediatrics* **217**, 98-109.e8 (2020).

86. Naumburg, E., Bellocco, R., Cnattingius, S., Jonzon, A. & Ekbom, A. Perinatal exposure to infection and risk of childhood leukemia. *Med. Pediatr. Oncol.* **38**, 391–397 (2002).

87. Nyari, T. A., Dickinson, H. O. & Parker, L. Childhood cancer in relation to infections in the community during pregnancy and around the time of birth. *International Journal of Cancer* **104**, 772–777 (2003).

88. Williams, L. A., Yang, J. J., Hirsch, B. A., Marcotte, E. L. & Spector, L. G. Is There Etiologic Heterogeneity between Subtypes of Childhood Acute Lymphoblastic Leukemia? A Review of Variation in Risk by Subtype. *Cancer Epidemiol Biomarkers Prev* **28**, 846–856 (2019).

89. Narod, S. A., Stiller, C. & Lenoir, G. M. An estimate of the heritable fraction of childhood cancer. *British Journal of Cancer* **63**, 993 (1991).

90. Bannon, S. A., Foglesong, J. & DiNardo, C. D. Germline Mutations Associated with Leukemia in Childhood: New Discoveries and Emerging Phenotypes. *Curr Genet Med Rep* 1–7 (2017) doi:10.1007/s40142-017-0118-5.

91. Moriyama, T., Relling, M. V. & Yang, J. J. Inherited genetic variation in childhood acute lymphoblastic leukemia. *Blood* **125**, 3988–3995 (2015).

92. Kratz, C. P., Stanulla, M. & Cavé, H. Genetic predisposition to acute lymphoblastic leukemia: Overview on behalf of the I-BFM ALL Host Genetic Variation Working Group. *European Journal of Medical Genetics* **59**, 111–115 (2016).

93. Ripperger, T. *et al.* Childhood cancer predisposition syndromes—A concise review and recommendations by the Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology. *Am J Med Genet* n/a-n/a (2017) doi:10.1002/ajmg.a.38142.

94. Rüchel, N. *et al.* In Utero Development and Immunosurveillance of B Cell Acute Lymphoblastic Leukemia. *Curr. Treat. Options in Oncol.* **23**, 543–561 (2022).

95. Bachinski, L. L. *et al.* Genetic Mapping of a Third Li-Fraumeni Syndrome Predisposition Locus to Human Chromosome 1q23. *Cancer Res* **65**, 427–431 (2005).

96. Lee, J. M. & Bernstein, A. p53 mutations increase resistance to ionizing radiation. *Proc Natl Acad Sci U S A* **90**, 5742–5746 (1993).

97. Levine, A. J. p53, the Cellular Gatekeeper for Growth and Division. *Cell* **88**, 323–331 (1997).

98. Guha, T. & Malkin, D. Inherited TP53 Mutations and the Li–Fraumeni Syndrome. *Cold Spring Harb Perspect Med* a026187 (2017) doi:10.1101/cshperspect.a026187.

99. Kamihara, J., Rana, H. Q. & Garber, J. E. Germline TP53 Mutations and the Changing Landscape of Li–Fraumeni Syndrome. *Human Mutation* **35**, 654–662 (2014).

100. Zhang, J. *et al.* Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood* **118**, 3080–3087 (2011).

101. Holmfeldt, L. *et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat Genet* **45**, 242–252 (2013).

102. Powell, B. C. *et al.* Identification of TP53 as an acute lymphocytic leukemia susceptibility gene through exome sequencing. *Pediatr. Blood Cancer* **60**, E1–E3 (2013).

103. Comeaux, E. Q. & Mullighan, C. G. TP53 Mutations in Hypodiploid Acute Lymphoblastic Leukemia. *Cold Spring Harb Perspect Med* **7**, a026286 (2017).

104. Qian, M. *et al.* TP53 Germline Variations Influence the Predisposition and Prognosis of B-Cell Acute Lymphoblastic Leukemia in Children. *JCO* JCO.2017.75.5215 (2018) doi:10.1200/JCO.2017.75.5215.

105. Lopez, R. G. *et al.* TEL Is a Sequence-specific Transcriptional Repressor. *J. Biol. Chem.* **274**, 30132–30138 (1999).

106. Wang, L. C. *et al.* The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev.* **12**, 2392–2402 (1998).

107. Hock, H. *et al.* Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev* **18**, 2336–2341 (2004).

108. Schwab, C. J. *et al.* Genes commonly deleted in childhood B-cell precursor acute lymphoblastic leukemia: association with cytogenetics and clinical features. *Haematologica* **98**, 1081 (2013).

109. Stams, W. a. G. *et al.* Incidence of additional genetic changes in the TEL and AML1 genes in DCOG and COALL-treated t(12;21)-positive pediatric ALL, and their relation with drug sensitivity and clinical outcome. *Leukemia* **20**, 410–416 (2006).

110. Cavé, H. *et al.* ETV6 is the target of chromosome 12p deletions in t(12;21) childhood acute lymphocytic leukemia. *Leukemia* **11**, 1459–1464 (1997).

111. Irvin, B. J. *et al.* TEL, a Putative Tumor Suppressor, Induces Apoptosis and Represses Transcription of Bcl-XL. *J. Biol. Chem.* **278**, 46378–46386 (2003).

112. Boily, G & Sinnett, D. Functional Characterization of ETV6, a Candidate Tumor Suppressor Gene Associated with Childhood Acute Lymphoblastic Leukaemia. in (2002).

113. Topka, S. *et al.* Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. *PLOS Genet* **11**, e1005262 (2015).

114. Moriyama, T. *et al.* Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *The Lancet Oncology* **16**, 1659–1666 (2015).

115. Duployez, N. *et al.* Detection of a new heterozygous germline ETV6 mutation in a case with hyperdiploid acute lymphoblastic leukemia (ALL). *Eur J Haematol* n/a-n/a doi:10.1111/ejh.12981.

116. Melazzini, F. *et al.* Clinical and pathogenetic features of ETV6 related thrombocytopenia with predisposition to acute lymphoblastic leukemia. *Haematologica* haematol.2016.147496 (2016) doi:10.3324/haematol.2016.147496.

117. Dirse, V. *et al.* ETV6 and NOTCH1 germline variants in adult acute leukemia. *Leukemia & Lymphoma* 1–3 (2017) doi:10.1080/10428194.2017.1359742.

118. Poggi, M. *et al.* Germline variants in ETV6 underlie reduced platelet formation, platelet dysfunction and increased levels of circulating CD34+ progenitors. *Haematologica* **102**, 282–294 (2017).

119. Järviaho, T. *et al.* Predisposition to childhood acute lymphoblastic leukemia caused by a constitutional translocation disrupting ETV6. *Blood Adv* **3**, 2722–2731 (2019).

120. Rampersaud, E. *et al.* Germline deletion of ETV6 in familial acute lymphoblastic leukemia. *Blood Adv* **3**, 1039–1046 (2019).

121. Junk, S. V. *et al.* TP53, ETV6 and RUNX1 Germline Variants in Patients Developing Secondary Neoplasms after Treatment for Childhood Acute Lymphoblastic Leukemia. *Blood* **130**, 884–884 (2017).

122. Nishii, R. *et al.* Comprehensive Functional Characterization of Germline ETV6 Variants Associated with Inherited Predisposition to Acute Lymphoblastic Leukemia in Children. *Blood* **128**, 1085–1085 (2016).

123. Zhang, M. Y. *et al.* Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat Genet* **47**, 180–185 (2015).

124. Noetzli, L. *et al.* Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet* **47**, 535–538 (2015).

125. Nutt, S. L., Heavey, B., Rolink, A. G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**, 556 (1999).

126. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).

127. Mullighan, C. G. & Downing, J. R. Global Genomic Characterization of Acute Lymphoblastic Leukemia. *Seminars in Hematology* **46**, 3–15 (2009).

128. Zachariadis, V. *et al.* The frequency and prognostic impact of dic(9;20)(p13.2;q11.2) in childhood B-cell precursor acute lymphoblastic leukemia: results from the NOPHO ALL-2000 trial. *Leukemia* **25**, 622 (2011).

129. Auer, F. *et al.* Inherited susceptibility to pre B-ALL caused by germline transmission of PAX5 c.547G>A. *Leukemia* **28**, 1136–1138 (2014).

130. Shah, S. *et al.* A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nat Genet* **45**, 1226–1231 (2013).

131. Chan, L. N. *et al.* Metabolic gatekeeper function of B-lymphoid transcription factors. *Nature* **542**, 479–483 (2017).

132. Waanders, E. *et al.* Germline activating TYK2 mutations in pediatric patients with two primary acute lymphoblastic leukemia occurrences. *Leukemia* **31**, 821–828 (2017).

133. Atak, Z. K. *et al.* High Accuracy Mutation Detection in Leukemia on a Selected Panel of Cancer Genes. *PLOS ONE* **7**, e38463 (2012).

134. Roberts, K. G. *et al.* Genetic Alterations Activating Kinase and Cytokine Receptor Signaling in High-Risk Acute Lymphoblastic Leukemia. *Cancer Cell* **22**, 153–166 (2012).

135. Harvey, R. C. *et al.* Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. *Blood* **115**, 5312–5321 (2010).

136. Mullighan, C. G. *et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. *PNAS* **106**, 9414–9418 (2009).

137. Roberts, K. G. *et al.* Targetable Kinase-Activating Lesions in Ph-like Acute Lymphoblastic Leukemia. *New England Journal of Medicine* **371**, 1005–1015 (2014).

138. Hamerton, J. L., Canning, N., Ray, M. & Smith, S. A cytogenetic survey of 14,069 newborn infants. *Clinical Genetics* **8**, 223–243 (1975).

139. Jacobs, P. A., Browne, C., Gregson, N., Joyce, C. & White, H. Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. *Journal of Medical Genetics* **29**, 103–108 (1992).

140. Gardner, R. J. Mk., Sutherland, G. R. & Shaffer, L. G. *Chromosome Abnormalities and Genetic Counseling*. (Oxford University Press, USA, 2011).

141. Harrison, C. J. *et al.* An international study of intrachromosomal amplification of chromosome 21 (iAMP21): cytogenetic characterization and outcome. *Leukemia* **28**, 1015 (2014).

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

143. Harrison, C. J. & Schwab, C. Constitutional abnormalities of chromosome 21 predispose to iAMP21-acute lymphoblastic leukaemia. *European Journal of Medical Genetics* **59**, 162–165 (2016).

144. Velazquez, L. *et al.* Cytokine Signaling and Hematopoietic Homeostasis Are Disrupted in Lnk-deficient Mice. *Journal of Experimental Medicine* **195**, 1599–1611 (2002).

145. Bersenev, A., Wu, C., Balcerek, J. & Tong, W. Lnk controls mouse hematopoietic stem cell self-renewal and quiescence through direct interactions with JAK2. *J Clin Invest* **118**, 2832–2844 (2008).

146. Takaki, S. *et al.* Control of B Cell Production by the Adaptor Protein Lnk: Definition of a Conserved Family of Signal-Modulating Proteins. *Immunity* **13**, 599–609 (2000).

147. Mullighan, C. Genomic Characterization of Childhood Acute Lymphoblastic Leukemia. *Seminars in Hematology* **50**, 314–324 (2013).

148. Öfverholm, I. I. *et al.* Detailed gene dose analysis reveals recurrent focal gene deletions in pediatric B-cell precursor acute lymphoblastic leukemia. *Leukemia & Lymphoma* **57**, 2161–2170 (2016).

149. Perez-Garcia, A. *et al.* Genetic loss of SH2B3 in acute lymphoblastic leukemia. *Blood* **122**, 2425–2432 (2013).

150. Koipally, J. & Georgopoulos, K. Ikaros Interactions with CtBP Reveal a Repression Mechanism That Is Independent of Histone Deacetylase Activity. *J. Biol. Chem.* **275**, 19594–19602 (2000).

151. Dörge, P. *et al.* IKZF1 deletion is an independent predictor of outcome in pediatric acute lymphoblastic leukemia treated according to the ALL-BFM 2000 protocol. *Haematologica* **98**, 428–432 (2013).

152. Olsson, L. *et al.* The clinical impact of IKZF1 deletions in paediatric B-cell precursor acute lymphoblastic leukaemia is independent of minimal residual disease stratification in Nordic Society for Paediatric Haematology and Oncology treatment protocols used between 1992 and 2013. *Br J Haematol* **170**, 847–858 (2015).

153. Kuehn, H. S. *et al.* Loss of B Cells in Patients with Heterozygous Mutations in IKAROS. *New England Journal of Medicine* **374**, 1032–1043 (2016).

154. Mendoza-Londono, R., Kashork, C. D., Shaffer, L. G., Krance, R. & Plon, S. E. Acute lymphoblastic leukemia in a patient with Greig cephalopolysyndactyly and interstitial deletion of chromosome 7 del(7)(p11.2 p14) involving the GLI3 and ZNFN1A1 genes. *Genes Chromosom. Cancer* **42**, 82–86 (2005).

155. Järviaho, T. *et al.* Microdeletion of 7p12.1p13, including IKZF1, causes intellectual impairment, overgrowth, and susceptibility to leukaemia. *Br. J. Haematol.* **185**, 354–357 (2019).

156. Churchman, M. L. *et al.* Germline Genetic Variation in IKZF1 and Predisposition to Childhood Acute Lymphoblastic Leukemia. *Blood* **128**, LBA-2-LBA-2 (2016).

157. Knapke, S., Nagarajan, R., Correll, J., Kent, D. & Burns, K. Hereditary cancer risk assessment in a pediatric oncology follow-up clinic. *Pediatr. Blood Cancer* **58**, 85–89 (2012).

158. Jongmans, M. Recognition of genetic predisposition in pediatric cancer patients: An easy-to-use selection tool. *European Journal of Medical Genetics* **59**, 116–125 (2016).

159. Hasle, H., Clemmensen, I. H. & Mikkelsen, M. Risks of leukaemia and solid tumours in individuals with Down's syndrome. *The Lancet* **355**, 165–169 (2000).

160. Whitlock, J. A. Down syndrome and acute lymphoblastic leukaemia. *British Journal of Haematology* **135**, 595–602 (2006).

161. Buitenkamp, T. D. *et al.* Acute lymphoblastic leukemia in children with Down syndrome: a retrospective analysis from the Ponte di Legno study group. *Blood* **123**, 70–77 (2014).

162. Bohnstedt, C. *et al.* Physicians compliance during maintenance therapy in children with Down syndrome and acute lymphoblastic leukemia. *Leukemia* **27**, 866–870 (2013).

163. Zipursky, A. Transient leukaemia – a benign form of leukaemia in newborn infants with trisomy 21. *British Journal of Haematology* **120**, 930–938 (2003).

164. Zeller, B. *et al.* Acute leukaemia in children with Down syndrome: a population-based Nordic study. *British Journal of Haematology* **128**, 797–804 (2005).

165. Arico, M. *et al.* Acute lymphoblastic leukemia and Down syndrome: presenting features and treatment outcome in the experience of the Italian Association of Pediatric Hematology and Oncology (AIEOP). *Cancer* **113**, 515–521 (2008).

166. Zwaan, C. M. *et al.* Different drug sensitivity profiles of acute myeloid and lymphoblastic leukemia and normal peripheral blood mononuclear cells in children with and without Down syndrome. *Blood* **99**, 245–251 (2002).

167. Izraeli, S. Similar yet different. *Blood* **116**, 1019–1020 (2010).

168. Bercovich, D. *et al.* Mutations of JAK2 in acute lymphoblastic leukaemias associated with Down's syndrome. *The Lancet* **372**, 1484–1492 (2008).

169. Gaikwad, A. *et al.* Prevalence and clinical correlates of JAK2 mutations in Down syndrome acute lymphoblastic leukaemia. *British Journal of Haematology* **144**, 930–932 (2009).

170. Lee, P., Bhansali, R., Izraeli, S., Hijiya, N. & Crispino, J. D. The biology, pathogenesis and clinical aspects of acute lymphoblastic leukemia in children with Down syndrome. *Leukemia* **30**, 1816–1823 (2016).

171. Lane, A. A. *et al.* Triplication of a 21q22 region contributes to B cell transformation through HMGN1 overexpression and loss of histone H3 Lys27 trimethylation. *Nature Genetics* **46**, 618–623 (2014).

172. Suarez, F. *et al.* Incidence, Presentation, and Prognosis of Malignancies in Ataxia-Telangiectasia: A Report From the French National Registry of Primary Immune Deficiencies. *JCO* **33**, 202–208 (2015).

173. Adams, M. *et al.* Acute myeloid leukaemia after treatment for acute lymphoblastic leukaemia in girl with Bloom syndrome. *J Genet Syndr Gene Ther* **4**, (2013).

174. Pasic, S., Vujic, D., Fiorini, M. & Notarangelo, L. D. T-cell lymphoblastic leukemia/lymphoma in Nijmegen breakage syndrome. *Haematologica* **89**, ECR27–ECR27 (2004).

175. Spinella, J.-F. *et al.* Whole-exome sequencing of a rare case of familial childhood acute lymphoblastic leukemia reveals putative predisposing mutations in Fanconi anemia genes. *BMC Cancer* **15**, (2015).

176. Wimmer, K. & Kratz, C. P. Constitutional mismatch repair-deficiency syndrome. *Haematologica* **95**, 699–701 (2010).

177. Cavé, H. *et al.* Acute lymphoblastic leukemia in the context of RASopathies. *European Journal of Medical Genetics* doi:10.1016/j.ejmg.2016.01.003.

178. Kratz, C. P., Rapisuwon, S., Reed, H., Hasle, H. & Rosenberg, P. S. Cancer in Noonan, Costello, cardiofaciocutaneous and LEOPARD syndromes. *Am. J. Med. Genet.* **157**, 83–89 (2011).

179. Chao, M. M. *et al.* T-cell acute lymphoblastic leukemia in association with Börjeson–Forssman–Lehmann syndrome due to a mutation in PHF6. *Pediatr. Blood Cancer* **55**, 722–724 (2010).

180. Siraganian, P. A., Rubinstein, J. H. & Miller, R. W. Keloids and neoplasms in the Rubinstein-Taybi syndrome. *Med. Pediatr. Oncol.* **17**, 485–491 (1989).

181. Jonas, D., Heilbron, D., & Ablin, A. Rubinstein-Taybi syndrome and acute leukemia. *The Journal of Pediatrics* **92**, 851–852 (1978).

182. Basel-Vanagaite, L. Acute lymphoblastic leukemia in Weaver syndrome. *Am. J. Med. Genet.* **152A**, 383–386 (2010).

183. Tatton-Brown, K. & Rahman, N. The NSD1 and EZH2 Overgrowth Genes, Similarities and Differences. *Am. J. Med. Genet.* **163**, 86–91 (2013).

184. Schlegelberger, B. & Heller, P. G. RUNX1 deficiency (familial platelet disorder with predisposition to myeloid leukemia, FPDMM). *Seminars in Hematology* **54**, 75–80 (2017).

185. Greaves, M. F. Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia* **2**, 120–125 (1988).

186. Ma, Y. *et al.* Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia. *PNAS* **110**, 7429–7433 (2013).

187. Bateman, C. M. *et al.* Evolutionary trajectories of hyperdiploid ALL in monozygotic twins. *Leukemia* **29**, 58 (2015).

188. Gawad, C., Koh, W. & Quake, S. R. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *PNAS* **111**, 17947–17952 (2014).

189. Lindqvist Carl Mårten *et al.* The Mutational Landscape in Pediatric Acute Lymphoblastic Leukemia Deciphered by Whole Genome Sequencing. *Human Mutation* **36**, 118–128 (2014).

190. Loh, M. L. *et al.* Whole Exome Sequencing of Pediatric Acute Lymphoblastic Leukemia Patients Identify Mutations in 11 Pathways: A Report from the Children's Oncology Group. *Blood* **128**, 455–455 (2016).

191. Ding, L.-W. *et al.* Mutational Landscape of Pediatric Acute Lymphoblastic Leukemia. *Cancer Research* **77**, 390–400 (2017).

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

193. Mullighan, C. G. The genomic landscape of acute lymphoblastic leukemia in children and young adults. *Hematology* **2014**, 174–180 (2014).

194. Roberts, K. G. *et al.* The Genomic Landscape of Childhood Acute Lymphoblastic Leukemia. *Blood* **134**, 649–649 (2019).

195. Lindqvist, C. M. *et al.* Deep targeted sequencing in pediatric acute lymphoblastic leukemia unveils distinct mutational patterns between genetic subtypes and novel relapse-associated genes. *Oncotarget* **5**, (2016).

196. Roberts, K. G. *et al.* The Genomic Landscape of Childhood Acute Lymphoblastic Leukemia. *Blood* **134**, 649 (2019).

197. Miranda-Filho, A. *et al.* Epidemiological patterns of leukaemia in 184 countries: a population-based study. *The Lancet Haematology* **5**, e14–e24 (2018).

198. Smith, M. A. *et al.* Evidence that Childhood Acute Lymphoblastic Leukemia is Associated with an Infectious Agent Linked to Hygiene Conditions. *Cancer Causes Control* **9**, 285–298 (1998).

199. Hauer, J., Fischer, U. & Borkhardt, A. Toward prevention of childhood ALL by early-life immune training. *Blood* **138**, 1412–1428 (2021).

200. Ward: The infective theory of acute leukaemia - Google Scholar. https://scholar-google-com.proxy.kib.ki.se/scholar_lookup?title=The%20infective%20theory%20of%20acute%20leukaemia&journal=Br.%20J.%20Child%20Dis.&volume=14&pages=10-20&publication_year=1917&author=Ward%2CG.

201. Kinlen, L. EVIDENCE FOR AN INFECTIVE CAUSE OF CHILDHOOD LEUKAEMIA: COMPARISON OF A SCOTTISH NEW TOWN WITH NUCLEAR REPROCESSING SITES IN BRITAIN. *The Lancet* **332**, 1323–1327 (1988).

202. Kinlen, L. J., Clarke, K. & Hudson, C. Evidence from population mixing in British New Towns 1946-85 of an infective basis for childhood leukaemia. *The Lancet* **336**, 577–582 (1990).

203. Kinlen, L. & Balkwill, A. Infective cause of childhood leukaemia and wartime population mixing in Orkney and Shetland, UK. *The Lancet* **357**, 858 (2001).

204. Li, C. K. *et al.* Impact of SARS on development of childhood acute lymphoblastic leukaemia. *Leukemia* **21**, 1353–1356 (2007).

205. Rodríguez-Hernández, G. *et al.* Infection Exposure Promotes ETV6-RUNX1 Precursor B-cell Leukemia via Impaired H3K4 Demethylases. *Cancer Res* **77**, 4365–4377 (2017).

206. Martín-Lorenzo, A. *et al.* Infection Exposure Is a Causal Factor in B-cell Precursor Acute Lymphoblastic Leukemia as a Result of Pax5-Inherited Susceptibility. *Cancer Discovery* **5**, 1328–1343 (2015).

207. Taub, J. W., Ge, Y. & Xavier, A. C. COVID-19 and childhood acute lymphoblastic leukemia. *Pediatr Blood Cancer* **67**, e28400 (2020).

208. Greaves, M. COVID-19 and childhood acute lymphoblastic leukemia. *Pediatric Blood & Cancer* **67**, e28481 (2020).

209. Jarvis, K. B., Lind, A., LeBlanc, M. & Ruud, E. Observed reduction in the diagnosis of acute lymphoblastic leukaemia in children during the COVID-19 pandemic. *Acta Paediatrica* **110**, 596–597 (2021).

210. Schüz, J., Borkhardt, A., Bouaoun, L. & Erdmann, F. The impact of the COVID-19 pandemic on the future incidence of acute lymphoblastic leukaemia in children: Projections for Germany under a COVID-19 related scenario. *Int J Cancer* **151**, 153–155 (2022).

211. Ferrari, A. *et al.* Children with cancer in the time of COVID-19: An 8-week report from the six pediatric onco-hematology centers in Lombardia, Italy. *Pediatric Blood & Cancer* **67**, e28410 (2020).

212. Erdmann, F., Spix, C., Schrappe, M., Borkhardt, A. & Schüz, J. Temporal changes of the incidence of childhood cancer in Germany during the COVID-19 pandemic: Updated analyses from the German Childhood Cancer Registry. *The Lancet Regional Health – Europe* **17**, (2022).

213. Silver, L. J. *et al.* New pediatric leukemia/lymphoma diagnoses during the COVID-19 pandemic: A New York perspective. *Pediatr Blood Cancer* **70**, e29879 (2023).

214. Pelland-Marcotte, M.-C. *et al.* Incidence of childhood cancer in Canada during the COVID-19 pandemic. *CMAJ* **193**, E1798–E1806 (2021).

215. Nichols, K. E. & Sánchez-García, I. Towards the prevention of childhood leukemia. *Oncoscience* **9**, 17–19 (2022).

216. Whitehead, T. P., Metayer, C., Wiemels, J. L., Singer, A. W. & Miller, M. D. Childhood Leukemia and Primary Prevention. *Current Problems in Pediatric and Adolescent Health Care* **46**, 317–352 (2016).

217. Metayer, C., Dahl, G., Wiemels, J. & Miller, M. Childhood Leukemia: A Preventable Disease. *Pediatrics* **138**, S45–S55 (2016).

218. Taylan, F. *et al.* Somatic Structural Alterations in Childhood Leukemia Can Be Backtracked in Neonatal Dried Blood Spots by Use of Whole-Genome Sequencing and Digital PCR. *Clin Chem* **65**, 345–347 (2019).

219. Bang, B. *et al.* A somatic UBA2 variant preceded ETV6-RUNX1 in the concordant BCP-ALL of monozygotic twins. *Blood Advances* **6**, 2275–2289 (2022).

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

221. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).

222. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).

223. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

224. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213–219 (2013).

225. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).

226. Paila, U., Chapman, B. A., Kirchner, R. & Quinlan, A. R. GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations. *PLOS Computational Biology* **9**, e1003153 (2013).

227. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotech* **29**, 24–26 (2011).

228. Martin, A. R. *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet* **51**, 1560–1565 (2019).

229. Eisfeldt, J., Vezzi, F., Olason, P., Nilsson, D. & Lindstrand, A. TIDDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. *F1000Res* **6**, 664 (2017).

230. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* **21**, 974–984 (2011).

231. Nadeu, F. *et al.* IgCaller for reconstructing immunoglobulin gene rearrangements and oncogenic translocations from whole-genome sequencing in lymphoid neoplasms. *Nature Communications* **11**, 3390 (2020).

232. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **12**, 656–664 (2002).

233. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645 (2009).

234. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115 (2012).

235. Rasmussen, M. *et al.* Allele-specific copy number analysis of tumor samples with aneuploidy and tumor heterogeneity. *Genome Biology* **12**, R108 (2011).

236. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657–663 (2007).

237. Lin, J., Su, G., Su, W. & Zhou, C. [Progress in digital PCR technology and application]. *Sheng Wu Gong Cheng Xue Bao* **33**, 170–177 (2017).

238. Briët, O. J. T., Amerasinghe, P. H. & Vounatsou, P. Generalized Seasonal Autoregressive Integrated Moving Average Models for Count Data with Application to Malaria Time Series with Low Case Numbers. *PLOS ONE* **8**, e65761 (2013).

239. Benjamin, M. A., Rigby, R. A. & Stasinopoulos, D. M. Generalized Autoregressive Moving Average Models. *Journal of the American Statistical Association* **98**, 214–223 (2003).

240. Rajatileka, S. *et al.* Isolation of human genomic DNA for genetic analysis from premature neonates: a comparison between newborn dried blood spots, whole blood and umbilical cord tissue. *BMC Genet* **14**, 105 (2013).

241. Fueller, E. *et al.* Genomic Inverse PCR for Exploration of Ligated Breakpoints (GIPFEL), a New Method to Detect Translocations in Leukemia. *PLOS ONE* **9**, e104419 (2014).

242. Forestier, E. *et al.* Clinical and cytogenetic features of pediatric dic(9;20)(p13.2;q11.2)-positive B-cell precursor acute lymphoblastic leukemias: A nordic series of 24 cases and review of the literature. *Genes, Chromosomes and Cancer* **47**, 149–158 (2008).

243. Sok, P. *et al.* Utilization of archived neonatal dried blood spots for genome-wide genotyping. *PLOS ONE* **15**, e0229352 (2020).

244. Gauffin, F., Nordgren, A., Barbany, G., Gustafsson, B. & Karlsson, H. Quantitation of RNA decay in dried blood spots during 20 years of storage. *Clinical Chemistry and Laboratory Medicine* **47**, 1467–1469 (2009).

245. Alpar, D. *et al.* Clonal origins of ETV6-RUNX1+ acute lymphoblastic leukemia: studies in monozygotic twins. *Leukemia* **29**, 839–846 (2015).

246. Gawad, C. *et al.* Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. *Blood* **120**, 4407–4417 (2012).

247. Schmiegelow, K. *et al.* Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia* **24**, 345–354 (2010).

248. Furutani, E. & Shimamura, A. Germline Genetic Predisposition to Hematologic Malignancy. *JCO* JCO.2016.70.8644 (2017) doi:10.1200/JCO.2016.70.8644.

249. Feurstein, S. & Godley, L. A. Germline ETV6 mutations and predisposition to hematological malignancies. *Int J Hematol* **106**, 189–195 (2017).

250. Noetzli, L. *et al.* Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet* **47**, 535–538 (2015).

251. Zhang, M. Y. *et al.* Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat Genet* **47**, 180–185 (2015).

252. Watkins, N. A. *et al.* A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood* **113**, e1–e9 (2009).

253. Bloom, M., Maciaszek, J. L., Clark, M. E., Pui, C.-H. & Nichols, K. E. Recent advances in genetic predisposition to pediatric acute lymphoblastic leukemia. *Expert Review of Hematology* **13**, 55–70 (2020).

254. Schäfer, D. *et al.* Five percent of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening. *Blood* **131**, 821–826 (2018).

255. Greaves, M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nature Reviews Cancer* 1 (2018) doi:10.1038/s41568-018-0015-6.

256. Greaves, M., Cazzaniga, V. & Ford, A. Can we prevent childhood Leukaemia? *Leukemia* **35**, 1258–1264 (2021).

257. McNally, R. J. Q. & Eden, T. O. B. An infectious aetiology for childhood acute leukaemia: a review of the evidence. *British Journal of Haematology* **127**, 243–263 (2004).

258. Greaves, M. The 'delayed infection' (aka 'hygiene') hypothesis for childhood leukaemia. in *The Hygiene Hypothesis and Darwinian Medicine* (ed. Rook, G. A. W.) 239–255 (Birkhäuser, 2009). doi:10.1007/978-3-7643-8903-1_13.

259. Paulsson, K. *et al.* The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Nat Genet* **47**, 672–676 (2015).

260. Yagi, T. *et al.* Detection of clonotypic IGH and TCR rearrangements in the neonatal blood spots of infants and children with B-cell precursor acute lymphoblastic leukemia. *Blood* **96**, 264–268 (2000).

261. Maia, A. T. *et al.* Prenatal origin of hyperdiploid acute lymphoblastic leukemia in identical twins. *Leukemia* **17**, 2202 (2003).

262. Taub, J. W. *et al.* High frequency of leukemic clones in newborn screening blood samples of children with B-precursor acute lymphoblastic leukemia. *Blood* **99**, 2992–2996 (2002).

263. Gruhn, B. *et al.* Prenatal origin of childhood acute lymphoblastic leukemia, association with birth weight and hyperdiploidy. *Leukemia* **22**, 1692–1697 (2008).

264. Panzer-Grümayer, E. R. *et al.* Nondisjunction of chromosomes leading to hyperdiploid childhood B-cell precursor acute lymphoblastic leukemia is an early event during leukemogenesis. *Blood* **100**, 347–349 (2002).

265. Sun, C., Chang, L. & Zhu, X. Pathogenesis of ETV6/RUNX1-positive childhood acute lymphoblastic leukemia and mechanisms underlying its relapse. *Oncotarget* **8**, 35445–35459 (2017).

266. Greaves, M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat Rev Cancer* **18**, 471–484 (2018).