From the DEPARTMENT OF MOLECULAR MEDICINE AND SURGERY
Karolinska Institutet, Stockholm, Sweden

IDENTIFICATION AND EVALUATION OF NOVEL PROGNOSTIC GENETIC MARKERS FOR CHILDHOOD ACUTE LYMPHOBLASTIC LEUKEMIA

Ingegerd Ivanov Öfverholm

Stockholm 2017

Cover and book art by Louise Brattström Stolt
Illustrations and figures by Ingegerd Ivanov Öfverholm et al.

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Eprint AB 2017

ISBN 978-91-7676-621-7

IDENTIFICATION AND EVALUATION OF NOVEL PROGNOSTIC GENETIC MARKERS FOR CHILDHOOD ACUTE LYMPHOBLASTIC LEUKEMIA

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Ingegerd Ivanov Öfverholm

Principal Supervisor:
Associate professor Gisela Barbany
Karolinska Institutet
Department of Molecular Medicine and Surgery

Co-supervisor(s):
Professor Magnus Nordenskjöld
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Ann Nordgren
Karolinska Institutet
Department of Molecular Medicine and Surgery

Associate professor Mats Heyman
Karolinska Institutet
Department of Women's and Children's health

Opponent:
Professor Shai Izraeli
Tel Aviv University
Department of Human Molecular Genetics and Biochemistry

Examination Board:
Associate professor Johanna Ungerstedt
Karolinska Institutet
Department of Medicine Huddinge

Associate professor Larry Mansouri
Uppsala Universitet
Department of Immunology, Genetics and Pathology

Associate professor Catharina Lavebratt
Karolinska Institutet
Department of Molecular Medicine and Surgery

It's not what you look at that matters,
it's what you see.

Henry David Thoreau

.

ABSTRACT

Childhood acute lymphoblastic leukemia (ALL) is the most common form of childhood cancer today. Due to advances in risk stratification and treatment, survival rates have increased drastically the last decades. Currently, children with acute leukemia in the Nordic countries are diagnosed and treated according to the NOPHO-2008 treatment protocol. In this protocol, a number of cytogenetic markers are used for risk stratification and guidance of treatment intensity. However, genetic markers associated with high risk are infrequent and relapses occur across all genetic subtypes, including those associated with a favorable outcome. Importantly, over 25% of childhood ALL cases harbor none of the currently used genetic risk markers in their bone marrow cells at diagnosis.

The aim of this thesis was to generate a greater understanding of the genetic landscape in ALL, as well as to identify novel genetic markers of prognostic relevance, with special focus on the group of patients lacking risk-stratifying markers.

In **paper I**, we investigated the frequency and prognostic impact of *IKZF1* deletions in patients diagnosed with B-cell precursor (BCP) ALL in the Stockholm region; *IKZF1* deletions were present in 15% of cases and significantly associated with inferior outcome. These results led to **paper II**, where the cohort was extended to include BCP ALL cases with available *IKZF1* data from other centers in Sweden. This study verified that *IKZF1* deletion was an independent risk factor for decreased survival, and could confirm that the frequency and prognostic effect was most pronounced in patients without risk-stratifying markers.

A high frequency of *IKZF1* deletions could also be detected in **paper III**, where we investigated the genetic copy number landscape of BCP ALL across the different cytogenetic subtypes. This study showed that a majority of cases without risk-stratifying markers harbor deletions with potential prognostic significance, suggesting that a large proportion of this group could be assigned to distinct genetic subtypes.

Intrachromosomal amplification of chromosome 21 (iAMP21) is an intermediate/high-risk subtype for which the biological cause of the high relapse risk is unknown. In **paper IV**, we used an integrated molecular approach to investigate the iAMP21 subtype, and identified significant overexpression of tree potential candidate genes, i.e. *DYRK1A*, *SON* and *CHAF1B,* with leukemia-relevant functions that could represent future targets for therapy.

Together, these studies have identified a number of potential novel prognostic genetic markers that may contribute to the clinical risk-evaluation of children diagnosed with BCP ALL, and to our understanding of the biology behind relapse and poor outcome in this disease.

LIST OF SCIENTIFIC PAPERS

I. **Öfverholm I**, Tran AN, Heyman M, Zachariadis V, Nordenskjöld M, Nordgren A, Barbany G. Impact of *IKZF1* deletions and *PAX5* amplifications in pediatric B-cell precursor ALL treated according to NOPHO protocols. *Leukemia* (2013) **27**: 1936–1939

II. Olsson L*****, **Ivanov Öfverholm I***, Norén-Nyström U, Zachariadis V, Nordlund J, Sjögren H, Golovleva I, Nordgren A, Paulsson K, Heyman M, Barbany G, Johansson B. 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. *British Journal of Haematology* (2015) 170: 847–85
***** Equal contribution

III. **Ivanov Öfverholm I**, Tran AN, Olsson L, Zachariadis V, Heyman M, Rudd E, Syk Lundberg E, Nordenskjöld M, Johansson B, Nordgren A, Barbany G. Detailed gene dose analysis reveals recurrent focal gene deletions in pediatric B-cell precursor acute lymphoblastic leukemia. *Leukemia & Lymphoma* (2016) 57: 2161–2170

IV. **Ivanov Öfverholm I**, Zachariadis V, Nordlund J, Taylan F, Marincevic Zuniga Y, Tran AN, Tesi B, Dahlberg J, Saft L, Pokrovskaja K, Grandér D, Nilsson D, Vezzi F, Nordenskjöld M, Lönnerholm G, Heyman M, Nordgren A, Syvänen A-C, Barbany G. Overexpression of the tyrosine kinase gene *DYRK1A* and the chromatin remodeling genes *CHAF1B* and *SON* in the iAMP21 subtype of pediatric acute lymphoblastic leukemia. In Manuscript.

LIST OF RELATED SCIENTIFIC PAPERS

V. Barbany G*, Gauffin F*, **Öfverholm I**, Karlsson H, Thörn I, Arvidson J, Heyman M, Gustafsson B, Nordgren A. The *ETV6/RUNX1* fusion transcript is not detected in RNA isolated from neonatal dried blood spots from children later diagnosed with the corresponding leukemia. *Leukemia & Lymphoma* (2013); 54(12): 2742–2744 * Equal contribution

VI. Zachariadis V, Schoumans J, **Öfverholm I**, Barbany G, Halvardsson E, Forestier E, Johansson B, Nordenskjöld M, Nordgren A. Detecting dic(9;20)(p13.2;p11.2)-positive B-cell precursor acute lymphoblastic leukemia in a clinical setting using fluorescence in situ hybridization. *Leukemia* (2014) 28, 196–198

VII. Nordlund J, Bäcklin C L, Zachariadis V, Cavelier L, Dahlberg J, **Öfverholm I**, Barbany G, Nordgren A, Övernäs E, Abrahamson J, Flaegstad T, Heyman M, Jónsson Ó G, Kanerva J, Larsson R, Palle J, Schmiegelow K, Gustafsson M G, Lönnerholm G, Forestier E and Syvänen A-C. DNA methylation-based subtype predictors for pediatric acute lymphoblastic leukemia. *Clinical Epigenetics* (2015) 7:11

VIII. Tran AN*, Taylan F*, Zachariadis V, Vezzi F, Lötstedt B, **Ivanov Öfverholm I**, Lindstrand A, Nordenskjöld M, Nordgren A, Nilsson D, Barbany G. High resolution detection of structural chromosomal rearrangements in leukemias through mate pair sequencing. *Manuscript submitted to Genes, chromosomes and cancer.* * Equal contribution

IX. Schwab C*, Nebral K*, Chilton L, Leschi C, Waanders E, Boer JM, Zaliova M, Sutton R, **Ivanov Öfverholm I**, Ohki K, Yamashita Y, Groeneveld-Krentz S, Fronova E, Bakkus M, Tchinda J, da Conceição Barbosa T, Fazio G, Mlynarski W, Pastorczak A, Cazzaniga G, Pombo-de-Oliveira M S, Trka J, Kirschner-Schwabe R, Imamura T, Barbany G, Stanulla M, Attarbaschi A, Panzer-Grümayer R, Kuiper R P, den Boer M L, Cavé H, Moorman A V, Harrison C J, Strehl S. Intragenic amplification of *PAX5*: A novel subgroup in B-cell precursor acute lymphoblastic leukemia. *Under review in Blood.* *Equal contribution

CONTENTS

LIST OF ABBREVIATIONS

AKML	Acute megakaryocytic leukemia
ALL	Acute lymphoblastic leukemia
Array-CGH	Array-based comparative genomic hybridization
BCP	B-cell precursor
BM	Bone marrow
CADD	Combined Annotation Dependent Depletion
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemia
CNA	Copy number alteration
CpG	CG dinucleotide
CPM	Counts per million mapped reads
DN	Dominant negative
DS	Down syndrome
EFS	Event-free survival
FISH	Fluorescent *in situ* hybridization
FPKM	Fragments per kilobase of exon per million mapped reads
GRCh	Genome Reference Consortium human assembly
GWAS	Genome-wide association study
HeH	High hyperdiploidy
HR	High-risk
iAMP21	Intrachromosomal amplification of chromosome 21
IGV	Integrative Genomics Viewer
ΔIKZF1	Deletion of *IKZF1*
IR	Intermediate-risk
MLPA	Multiplex ligation-dependent probe amplification
MP-WGS	Mate-pair whole genome sequencing
MRA	Minimal region of amplification
MRD	Minimal residual disease
NOPHO	Nordic Society of Pediatric Hematology and Oncology
OS	Overall Survival

PAR	Pseudoautosomal region
PCR	Polymerase chain reaction
RAG	Recombination-activating gene
RNA-seq	RNA sequencing
RT-PCR	Reverse transcriptase polymerase chain reaction
SNP	Single nucleotide polymorphism
SR	Standard risk
UPD	Uniparental disomy
WBC	White blood cell
WGS	Whole genome sequencing

1 INTRODUCTION

Leukemia is a malignant disease of the bone marrow, characterized by pathogenic growth of immature white blood cells. The term Leukemia, formed by the two Greek words "*leukos*" and "*haima"*, meaning "white" and "blood", was introduced in 1847 by professor Rudolf Virchow, who observed an excess of white or colorless cells in the blood sample of a patient showing symptoms of spleen enlargement, bleeding and fever.[1]

Ever since the discovery of the Philadelphia chromosome in 1960,[2] much attention has been focused on the genetic events underlying leukemic transformation. Today, detection of genetic aberrations in diagnostic bone marrow samples is standard of care for most leukemia types, as it provides important diagnostic, prognostic and treatment-related information.

1.1 THE GENETICS OF CANCER

Cancer occurs when genetic mutations or rearrangements disrupt the normal machinery of the cell, causing potentially unlimited cell proliferation and expansion. Many cancers start with one or a few mutations with growth promoting properties; the rapid cell division rate, often in combination with impaired DNA repair mechanisms, results in new mutations, and different subclones emerge.[3-6] Alternatively, several mutations can appear simultaneously in a few catastrophic genomic events.[7] Malignant disease has traditionally been classified according to tissue type rather than mutational profile; however, new genome-wide DNA and RNA sequencing technologies have dramatically changed the understanding of the cancer genome.[8,9] Tumors across different tissue types can show striking molecular similarities, while tumors of the same tissue type can display completely different mutational spectra, suggesting that the current tissue-specific classification might be inadequate.[8]

The types of genetic events underlying cancer development have traditionally been classified into mutations, copy number alterations (CNAs) and rearrangements/translocations.[5,10] In recent years, the term mutation, which refers to single nucleotide variants (SNVs) or small insertions/deletions (indels), has been widely replaced with the term sequence variant. In the context of cancer, sequence variants cause activation of oncogenes or inactivation of tumor suppressor genes. Similarly, CNAs result in amplifications of oncogenes or, more frequently, deletions of tumor suppressors. Through genomic rearrangements or translocations, parts of the genome are relocated, giving rise to fusions that alter the regulation of genes. The effect of these genetic lesions is executed at the protein level. In order to result in an oncogenic protein, an aberrant gene must first be transcribed into RNA molecules, which subsequently translate into the amino acid chain constituting the protein. Thus, studies of RNA expression generate insights into the functional effect of genomic aberrations.

1.1.1 Genetic analyses of cancer

Traditionally, targeted methods, such as Sanger sequencing, fluorescent *in situ* hybridization (FISH) and real-time reverse transcriptase polymerase chain reaction (RT-PCR) have been used to detect the different kinds of genetic lesions or aberrant RNA expression.[11] However, with the advent of massively parallel sequencing methods, all types of pathogenic lesions can be detected with the same method and in the same run.[9,11] Whole genome sequencing (WGS) enables detection of mutations, small indels, CNAs and rearrangements present in both coding and non-coding genomic regions.[12] With whole exome sequencing, the analysis is narrowed down to the 1,5 % of the genome that consists of coding genomic regions, i.e. the exons.[13] Whole transcriptome sequencing (RNA sequencing) provides a simultaneous overview of gene expression levels, fusion gene transcripts and sequence variants in the transcribed DNA.[14]

1.1.2 Prognostic genetic markers

A prognostic factor is a patient characteristic that can be used to estimate the chance of recovery or the risk for relapse in a disease. In the field of cancer genetics, a prognostic marker refers to a genetic aberration in the tumor DNA that is associated with outcome. Traditionally, genetic cancer research has focused on driver mutations, i.e. highly oncogenic mutations providing strong growth capacities and usually present in a majority of cancer cells in diagnostic samples.[5,6,15] The protein product of the mutated driver gene has been considered the most interesting target for new therapies. However, studies investigating paired diagnostic and relapse samples from the same patients have shown that this might be an oversimplification. A mutation considered as a driver in a diagnostic sample can be absent in the relapse sample, and conversely, the dominating clone in relapse samples can be absent at diagnosis or present in only a small fraction of cells in the diagnostic sample.[16-18] Furthermore, many solid tumors have been shown to harbor different clones in different parts of the tumor, a phenomenon termed tumor heterogeneity.[19] A tissue sample will thus only show the mutational spectrum in a limited part of the tumor, possibly missing the clone with a relapse-prone mutation. Taken together, these factors represent challenges in the interpretation of both established and potential novel prognostic markers.

1.2 LEUKEMIA

In leukemia, pathological and dysfunctional blood cells accumulate and suppress the growth of functional red blood cells, platelets and immune cells, thereby causing the classical clinical symptoms fatigue, bleedings and infections, respectively.

1.2.1 Hematopoiesis and leukemia

All blood cells originate from hematopoietic stem cells in the bone marrow. The hematopoietic stem cells have the ability to divide and differentiate through either the myeloid or lymphoid line in response to signals based on the current needs of the organism.[20] The myeloid line generates the red blood cells, responsible for oxygen transportation, the platelets, important for primary wound healing, and the granulocytes, which constitute the innate immune system. The lymphoid line gives rise to the adaptive immune cells, i.e. natural killer cells, B- and T-lymphocytes. In response to an infectious agent, the lymphoid stem cell differentiates into immature lymphoblasts, which eventually mature into lymphocytes, primed with customized proteins to target a particular infectious antigen.[21] The transformation from hematopoietic stem cell to mature lymphocyte requires several steps of differentiation and maturation followed by rounds of proliferation, as well as genomic recombination.[22,23] In B-cells, the recombination of immunoglobulin genes generates antibodies with affinity for a specific antigen. The different steps of proliferation and differentiation of lymphoblasts are largely regulated by the selective expression of hematopoietic transcription factors. Leukemia arises due to acquired mutations in hematopoietic stem cells or precursor cells, disrupting the tightly regulated growth-differentiation-maturation process and resulting in pathological expansion of immature and dysfunctional cells.[21,24]

In order to supply the body with enough blood cells, the bone marrow has a high baseline cell division rate, producing of over 10^{11} cells per day programmed to leave its primary tissue through the bloodstream. Furthermore, the cells of the adaptive immune system have a dynamic genome, inherently prone to genomic rearrangements.[25] Thus, the normal, healthy immune cells display several cancer-promoting features:[6] a high proliferation rate of immature cells, unstable genomes and the capacity to spread into the bloodstream.

1.2.2 Classification of leukemia

The term leukemia comprises several subtypes, with different but often overlapping clinical symptoms. Subtype classification is based on which hematopoietic line is affected, i.e. myeloid or lymphoid, as well as disease progression, i.e. acute or chronic. Morphologically, acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML), are characterized by rapidly expanding immature progenitor cells, while their chronic equivalents show a gradual accumulation of more mature cells. The different leukemia subtypes are further characterized by different cytogenetic alterations that provide information on disease progress, prognosis and suitable treatment. In chronic myeloid leukemia (CML), all malignant cells harbor the t(9;22) translocation, i.e. the Philadelphia chromosome, which results in overexpression of the tyrosine kinase gene *ABL1*, and CML patients can be treated successfully with at tyrosine kinase inhibitor.[26] In AML, mutations in the *CREBBP*, *FLT3* or *NPM1* genes provide important prognostic information.[27] In chronic lymphocytic leukemia

(CLL), deletion of 11q2, including the *ATM* gene, and deletion of 17p13, are associated with poor survival.[28] ALL has a high number of risk-associated genetic markers, with a total of six markers associated with adverse prognosis and relapse risk.[29]

1.3 CHILDHOOD ACUTE LYMPHOBLASTIC LEUKEMIA

While relatively uncommon in adults, ALL is the most frequent cancer type in childhood worldwide, accounting for approximately one third of all childhood cancers. In Sweden, the incidence is 3.6/100 000 children per year, with an incidence peak age of 2-5.[30] Survival rates for childhood ALL have increased dramatically over the last four decades, mainly due to intensified treatment and refined risk stratification, but also improved supportive care.[29,31,32]

1.3.1 Etiology

A small fraction of childhood ALL cases have congenital genetic syndromes, such as Down syndrome (DS), Noonan syndrome or Fanconi anemia, that predisposes for leukemia development.[33,34] Also, healthy carriers of a balanced Robertsonian translocation rob(15;21)c have a 2 700-fold increased risk of developing a specific subtype of ALL.[35] However, the overwhelming majority of leukemia cases are caused by acquired mutations in the bone marrow cells and the etiology of leukemia in children without predisposing conditions is still largely unknown; the only proven risk factor for leukemia development is exposure to ionizing radiation.[36]

The average childhood ALL case harbor few pathogenic mutations relative to adult cancers,[8] reflecting a time-dependent accumulation of mutations in adult cancers, but also that few mutations are needed to cause childhood ALL.

Retrospective genetic analyses of Guthrie cards have shown that some of the leukemia-associated translocations, including the t(12;21) *ETV6/RUNX1* fusion, can be present already before birth in children who later develop leukemia, but also in children who do not develop this disease.[37,38] This observation has generated the “two-hit” hypothesis of childhood leukemia, arguing that leukemia develops only after a second mutation occurs in a cell harboring a predisposing mutation (Figure 1). A few causes for susceptibly to a second hit have been suggested, including an aberrant immune response to infections early in life,[39] or a low exposure to infectious agents during the first year, followed by over-response to common infections later in childhood,[40,41] however, none of these hypotheses has been proven and different studies have shown conflicting results.

Even though the majority of leukemia cases are caused by acquired mutations, familial clustering of childhood ALL cases does occur, suggesting that constitutional, predisposing factors exist. Recent research indicate that heritable predisposition is a much more important cause of childhood leukemia than previously thought. For example, inherited mutations in the

tumor suppressor gene *TP53* are associated with increased risk for a particular high-risk leukemia subtype characterized by a hypodiploid karyotype.[42] Recently, several other germline variants predisposing to childhood ALL have been reported.[34] Genome wide association studies (GWAS) have shown association between increased risk of developing childhood leukemia and constitutional allele variants in genes involved in lymphocyte development and cell cycle regulation, including *IKZF1, PAX5, CDKN2A/B, CEBPE, ARID5B* and *SH2B3* (Figure 1).[43-46] GWAS have also revealed several inherited variants that affect the sensitivity to cytotoxic treatment, as well as risk variants for drug toxicities. Still, the knowledge of predisposing variants is too low to justify screening of healthy children in the current clinical setting.[47]

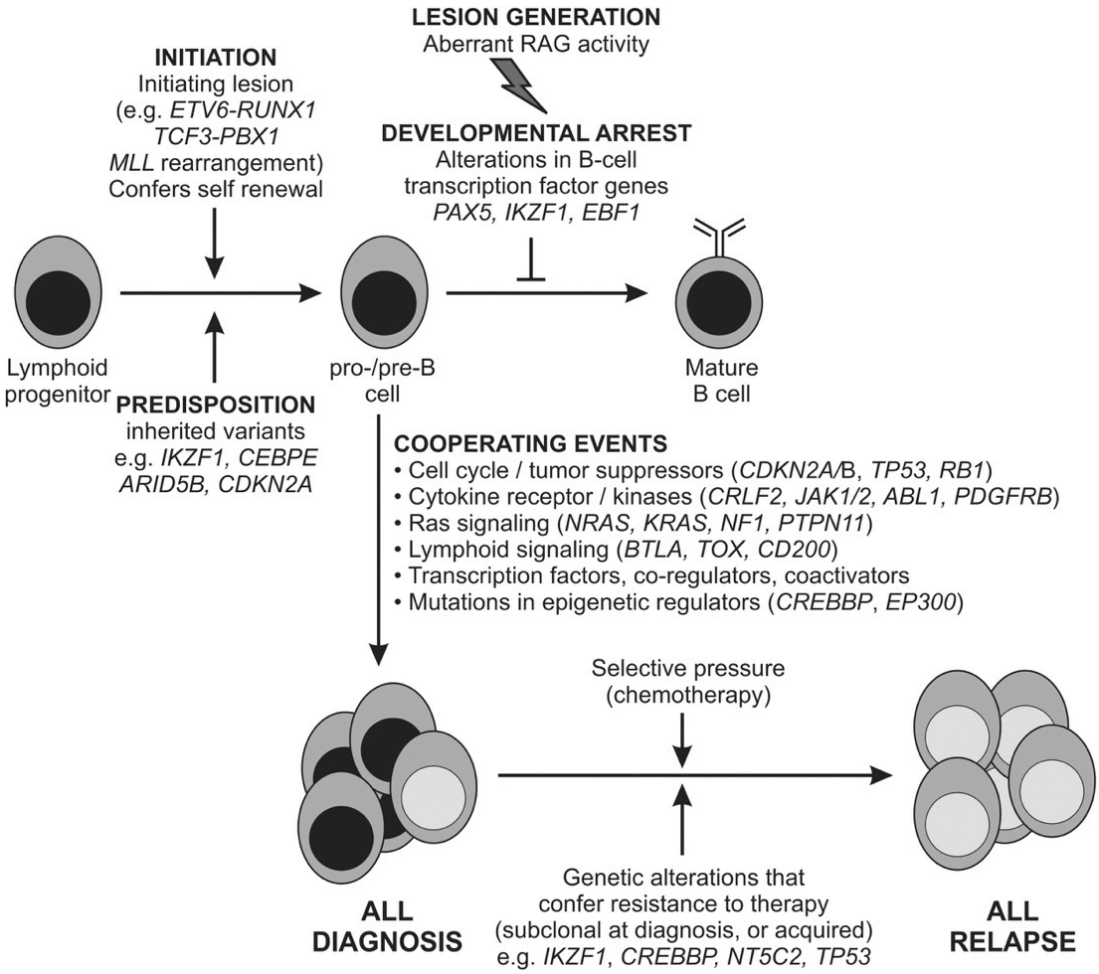

Figure 1 – Schematic illustration of B-cell development and the timing of genetic alterations implicated in BCP ALL, including predisposing and leukemia-initiating lesions, followed by relapse-causing lesions that are either acquired during treatment or present in subclones at diagnosis; subclonality is represented by white and black cores in the ALL cells. Reprinted from **Seminars in Hematology, Volume 50, Mullighan, Genomic characterization of childhood acute lymphoblastic leukemia, 314-324, 2013**, with permission from Elsevier.

1.3.2 Diagnosis and treatment

The first reported case of childhood leukemia was diagnosed and described by Dr. Henry Fuller in 1846.[48] By that time, no treatment was available and children with leukemia did not survive the disease. In 1947, the pediatric oncologist Dr. Sidney Farber discovered that the drug aminopterin, a folic acid antagonist, induced temporary remission in children.[49] The real treatment break-through came with the introduction of combinatory treatment, with different cytotoxic drugs used in different stages of the disease process.[50]

1.3.2.1 Symptoms and diagnosis

A child with leukemia presents with symptoms of a suppressed bone marrow, these include bone pain, fatigue, paleness, bruises, bleedings and infections. The diagnosis of ALL is verified by detection of a pathological number of immature B- or T-lymphoblast in the bone marrow. Analysis of the morphology and immunophenotype of the leukemic cells allows for subclassification according to the maturation stage of the leukemic blasts; about 80% of childhood cases have a precursor-B cell immunophenotype (BCP ALL).[51]

1.3.2.2 Risk-assessment and treatment strategy

The structure of the contemporary chemotherapy treatment strategy, including induction of remission, consolidation/intensification, and maintenance therapy, was established in the 1980s.[31] Several common treatment protocols have since been established around the world with the aim to unify diagnostics and treatment of childhood ALL. In Sweden, all children with ALL are diagnosed, risk-assessed and treated according to the common Nordic treatment protocol, Nordic Society of Pediatric Hematology and Oncology (NOPHO).[32,52]

Induction treatment is started as soon as the bone marrow examination has verified the diagnosis, and the preliminary risk-assessment is based on clinical risk factors, including white blood cell (WBC) count, age and the presence of extra-medullary disease.[32] Response to treatment is monitored by minimal residual disease (MRD), i.e. amount of remaining leukemic blast in bone marrow samples taken at specific time points.[32,53] MRD is currently considered one of the most powerful indicators for outcome in childhood ALL.[54-56] Also, cytogenetic analyses provide crucial information about prognosis and risk for relapse.[29] At day 29 of treatment, the risk profile is re-assessed based on results from cytogenetic analysis and MRD, and the treatment intensity is adjusted accordingly.[53] As a result of this risk-adapted treatment strategy, as well as improved supportive care, overall survival rates exceed 80% in modern protocols.[31,32] However, 10–20% of patients relapse, and the survival rate for relapsed ALL is poor.[57]

1.3.3 Risk-associated cytogenetic markers in childhood ALL

The presence of acquired genetic aberrations in the leukemic blasts constitutes the basis for risk stratification and guides treatment intensity.[29,32,58-60] Targeted genetic analyses of the diagnostic bone marrow sample are performed based on the pathology findings. The current standard genetic work-up for ALL includes G-banding for detection of aneuploidies, and FISH or RT-PCR analysis for detection of translocations. The frequencies of cytogenetic markers detected in childhood ALL at diagnosis are depicted Figure 2.

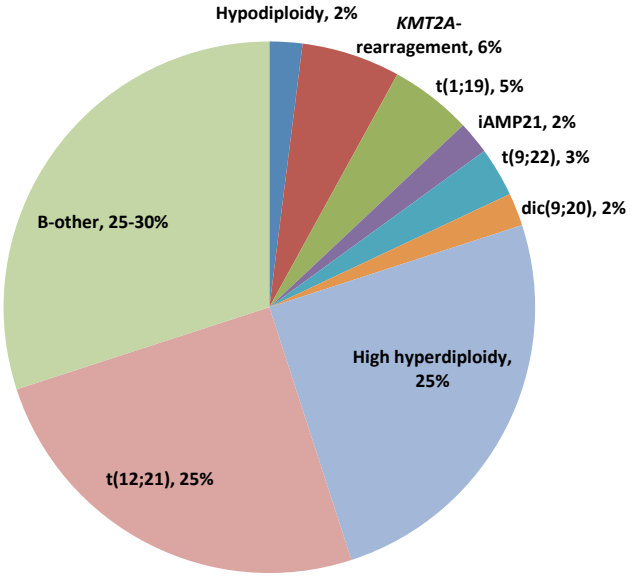

Figure 2 – Pie chart showing distribution of risk-stratifying genetic markers in childhood BCP ALL at diagnosis; B-other represents cases without any of the currently used markers

1.3.3.1 Standard-risk genetic markers

High hyperdiploidy (HeH) and t(12;21) translocation are the most common cytogenetic subtypes in childhood ALL, present in 25% of cases respectively. The HeH subtype is defined as 50-67 chromosomes; the gained chromosomes are typically chromosome X, 4, 6, 10, 14, 17, 18 and 21, but why these chromosomal gains give rise to leukemia is not known.[61] The t(12;21) translocation causes the *ETV6/RUNX1* fusion; both *ETV6* and *RUNX1* are hematopoietic transcription factors, but the disease causing mechanism is not known. As mentioned above, this fusion can be present in bone marrow cells already *in utero*,[37,38] and additional genetic lesions are needed to cause overt leukemia. Additional leukemia-associated genetic alterations are present in a majority of HeH and t(12;21) cases.[62,63] Both these subtypes are associated with favorable clinical parameters and outcome, however if other clinical or molecular signs of higher risk are present, the patient will be stratified according to these.

1.3.3.2 Intermediate-risk genetic markers

The t(1;19) translocation, the dicentric (9;20), and intrachromosomal amplification of chromosome 21 (iAMP21) are the three markers associated with intermediate risk in NOPHO. The t(1;19) translocation, present in 5-6% of childhood ALL, causes the *TCF3/PBX1* fusion and gives rise to an aberrant protein with oncogenic properties.[64] The dicentric (9;20)(p13;q11) subtype is present in 5% of childhood cases in the Nordic countries and associated with poor prognosis if treated according to standard risk protocols.[65] The aberration is characterized by additional genetic alterations, i.e. homozygous loss of *CDKN2A* and heterozygous loss of *PAX5*,[66] but the mechanism behind the higher risk profile is not fully understood. IAMP21 is present in 2% of pediatric BCP ALL cases and is associated with high age at diagnosis (median age 9) and low WBC.[67,68] IAMP21 is considered a primary event, with the amplification present in all subclones and always retained at relapse.[69] Although studies have described the genetic composition and the likely mechanisms of formation in iAMP21,[35,69-72] the mechanism underlying the higher relapse risk of this subtype is not understood. Recurrent alterations, such as deletion of *ETV6*, *IKZF1* and *RB1*, as well as *P2RY8/CRLF2* translocations have been reported.[69,73] However, the only genetic alteration consistently present in all iAMP21 cases is additional copies of the region of amplification.[69] The minimal region of amplification encompasses several protein coding genes, including *RUNX1*, but to this date, no causative gene has been identified in the region.

Retrospective studies have demonstrated that both the iAMP21 and t(1;19) subtype are associated with a high relapse rate if treated according to standard-risk protocols,[60,74,75] but the prognosis improves when they are treated according to intermediate or high-risk protocols.

1.3.3.3 High-risk genetic markers

Three cytogenetic markers are currently associated with high risk of relapse or treatment failure, these are hypodiploidy, rearrangements of the *KMT2A* gene at 11q23 (previously known as *MLL*) and the t(9;22) translocation (Philadelphia chromosome). The hypodiploid subtype is further subclassified into hypodiploid (44 chromosomes), high hypodiploid (40 to 43 chromosomes), low hypodiploid (33 to 39 chromosomes) and near-haploid (23 to 29 chromosomes). The two latter are associated with a very poor outcome while the high hypodiploid group is more heterogenous with varying outcome.[76,77] Studies have shown that both the near-haploid and low hypodiploid subtype harbor recurrent activating mutations in kinase genes, some sensitive to PI3K/mTOR inhibitors,[42] suggesting a role for kinase inhibitors in the treatments of these patients. Moreover, studies have shown that near-haploid cases often have a fraction of cells with duplication of the haploid clone, masking as hyperdiploid clones.[78] Methods that detect uniparental disomies (UPDs) are needed to identify these, in order to prevent misclassification.

Rearrangements of *KMT2A* are present in approximately 80% of infant ALL cases and affect a total of 5-6% of childhood ALL cases.[79] *KMT2A* encodes a transcriptional regulator that

affects the expression of several important hematopoietic development genes. Over 80 different genes have been reported to form fusions with, and affect the activity of, the *KMT2A* gene in this rearrangement, and the most common partners are *AFF1* (4q21) and *MLLT1* (19p13.3).[80]

The t(9;22) translocation, which causes the *BCR/ABL1* fusion and subsequent over-activity of the ABL1 tyrosine kinase, is one of the best-understood genetic subtypes in leukemia. While the translocation is associated with favorable prognosis and long-term survival when treated with kinase inhibitors in CML,[26] the prognosis for *BCR/ABL1*-positive ALL and CML in blast crisis is poor.[81] The *BCR/ABL1* fusion is present in 3% of childhood ALL, and these patients receive additional treatment with a tyrosine kinase inhibitor according to a special protocol.[32]

1.3.3.4 Challenges in genetic diagnosis of ALL

Several issues remain to be solved in the clinical genetic diagnostics of childhood ALL. First, approximately 25% of cases harbor none of the currently used genetic risk markers; these are referred to as B-other and constitute a substantial proportion of cases that subsequently relapse.[82] B-other is a heterogeneous group with either seemingly normal or failed karyotypes or aberrations of unknown or undetermined significance in their leukemic cells. For these cases, routine genetic work-up fails to contribute to the risk assessment. Second, genetic markers associated with high risk are infrequent and do not correspond to all cases that relapse or fail treatment. Third, relapses occur across all genetic subtypes, including those associated with a favorable outcome, such as HeH and t(12;21). Fourth, the molecular mechanisms responsible for relapse risk or treatment failure in the intermediate/high-risk subtypes remain largely unknown. Taken together, these issues indicate that additional or more adequate prognostic markers are needed.

1.3.4 Emerging subtypes in ALL

During the last decade, the development of large-scale methods for analysis of the genome and the global gene expression pattern in malignant cells have provided insights to the molecular landscape of childhood ALL. Novel prognostic markers have been proposed, however the prognostic impact of these markers could not always be confirmed in other patient cohorts. This discrepancy is likely explained by differences in risk profiles of the different cohorts, but also by differences in treatment intensity and possibly even variation in the germline genetic background of patient cohorts. This implies that potential markers have to be tested and verified within the respective treatment protocols before they can be implemented in routine diagnostics.

1.3.4.1 The copy number landscape of ALL

In 2007, the first genome-wide copy number studies of ALL, using array-based comparative genomic hybridization (array-CGH) or single nucleotide polymorphism (SNP) arrays, could demonstrate that copy number alterations (CNAs) are frequent in ALL.[83-85] Since then, numerous studies have shown that CNAs play an important role in leukemia development and several potential prognostic markers have been proposed, mainly deletions in genes involved in lymphoid development, cell cycle regulation and apoptosis.[17,71,86-89] Deletions of the lymphoid transcription factor genes *IKZF1* and *VPREB1* have been associated to poor prognosis,[88-90] while deletions of the transcriptional regulator gene *ERG* have been suggested to correlate with favorable outcome in cases with *IKZF1* deletions.[91,92] Deletions of *CREBBP*, *TBL1XR1* and *SPRED1* have been detected in a high proportion of paired diagnostic and relapse samples;[93-96] this may suggest a relapse-prone effect of these aberrations. Interestingly, *CREBBP* deletions are present in relapsed HeH cases, indicating that this deletion might counteract the otherwise favorable prognosis for this subtype.[97] Similarly, deletions of the lymphoid development gene *VPREB1* have been associated to relapse and poor prognosis in cases with the standard-risk marker t(12;21).[90] Deletions of these genes have also been reported in a newly described type of leukemia, i.e. "early-T-cell", a subtype characterized by poor prognosis and deletions of tumor suppressors, activating mutations in oncogenes and inactivating mutations in epigenetic regulators.[42]

1.3.4.2 The Philadelphia-like subtype

In 2009, two independent research groups reported on a novel high-risk subtype of B-ALL, with a gene expression profile similar to *BCR/ABL1*-positive ALL but lacking the canonical fusion gene.[88,98] This subtype, termed "Philadelphia-like ALL", harbored aberrations in tyrosine kinase genes, as well as recurrent deletions of the transcription factor *IKZF1*. Patients with this profile in the leukemic cells had a poor outcome, with high MRD levels at the end of induction and resistance to the cytotoxic drugs asparaginase and daunorubicin. The profile was present in ~10% of childhood ALL cases and a majority of these harbored no known risk-stratifying genetic markers, pointing to a prognostic relevance of this subtype. However, the clinical relevance of this subtype has been debated; while several studies have observed a poor prognosis for this subtype,[99,100] others have indicated that MRD-based risk-stratification is sufficient to counteract this negative prognostic effect.[101] Further studies have shown that activating mutations in kinase genes, including *ABL1*, *CRLF2*, *PDGFRB* and *JAK2*, are present in over 90% of the cases,[102,103] and many of these are targetable for kinase inhibitors. These novel insights in the aberrant kinase signaling in ALL suggest that a substantial proportion of ALL cases could potentially be treated with tyrosine kinase inhibitors.

1.3.4.3 Overexpression of CRLF2

Overexpression of the JAK-STAT kinase gene *CRLF2* is present in approximately 6% of childhood BCP ALL cases, and 50% of these harbor additional high-risk aberrations such as *JAK*-mutations or *IKZF1* deletions.[104] Overexpression of *CRLF2* can be caused either by

deletion of the pseudoautosomal region 1 (PAR1) at Xp22/Yp22, which juxtaposes the *P2RY8* gene to *CRLF2*, or by illegitimate recombination between the immunoglobulin gene *IGH* and the *CRLF2* gene; in both instances, the new partner gene causes an overexpression of *CRLF2*. In rare cases, overexpression is caused by gain-of–function mutations in *CRLF2*. However, while this subtype was associated with inferior outcome in the first studies that described the alteration,[105] this correlation has not always been reproducible in other cohorts.[100]

1.3.4.4 DUX4-rearranged and ETV6/RUNX1-like ALL

A recent study of the fusion gene landscape of BCP ALL showed that 4% (8/195) of the total cohort, and 16% (8/50) of B-other cases, harbored fusion rearrangements involving the *DUX4* gene, i.e. *DUX4/IGH* or *DUX4/ERG*.[106] All rearrangements caused overexpression of *DUX4* and aberrant expression of *ERG*. *DUX4* encodes a transcription factor that regulates the expression of stem-cell development genes, which suggests that its overexpression might cause activation of stem cell programs.[106] Other studies have supported the finding of a *DUX4*-rearranged subtype,[107,108] and the *DUX4* rearrangement has been proposed as an early, leukemia-initiating event, where binding of DUX4 to the *ERG* gene causes deregulated expression of coding and noncoding *ERG* transcripts. *DUX4*/*ERG* ALL has been associated with *ERG* deletions and favorable outcome, irrespective of co-occurring negative prognostic genetic alterations such as *IKZF1* deletion.[108]

Another finding in the first study[106] was a novel subtype with an *ETV6/RUNX1*-like expression profile but lacking the canonical fusion, present in 12% of B-other cases; all cases harbored aberrations in *ETV6* and *IKZF1*, and the subtype was associated with favorable outcome despite the presence of *IKZF1* deletions.[106]

1.3.5 *IKZF1* deletions in childhood ALL

Deletion of the *IKZF1* gene (Δ*IKZF1*), encoding the transcription factor IKAROS, occurs in approximately 15% of pediatric BCP ALL.[88,89,95,109,110] Δ*IKZF1* was originally described in association with the high-risk subtype *BCR/ABL1*-positive ALL, and was thought to contribute to the poor prognosis of this subtype.[111] In 2008, a large study on copy number alterations in ALL could demonstrate that Δ*IKZF1* was significantly associated with relapse in *BCR/ABL1*-negative BCP ALL,[17] suggesting a negative prognostic impact in all types of BCP ALL.

1.3.5.1 Functional effect of IKZF1 deletions

The *IKZF1* gene is located on chromosome 7q11, and the majority of reported aberrations in this locus are deletions; sequencing of *IKZF1* has shown a low frequency of point mutations.[89,100,111,112] Heterozygous deletions of the entire *IKZF1* locus or individual exons are predicted to cause haploinsufficiency, whereas deletion of a specific subset of exons (4-7) result in expression of IK6, a shorter, dominant-negative isoform of IKAROS.[81,111,113]

In normal B-cell development, different isoforms of *IKZF1*, generated through alternative splicing, are expressed during different development stages.[114] However, the IK6 isoform is considered a pathogenic isoform caused by deletion only.[115] The most frequently occurring intragenic Δ*IKZF1* in childhood BCP ALL are shown in Figure 3.

While some studies have suggested that haploinsufficiency of IKAROS contributes to leukemia development through disruption of maturation signaling, with subsequent uncontrolled growth of immature B-cells,[116] others have indicated that Δ*IKZF1* might occur only as a symptom of a general genomic instability, and that the genetic instability is the true cause of relapse and treatment failure.[117] The latter theory is contradicted by the results from a recent SNP-array study, in which no significant difference in the frequencies of CNAs between Δ*IKZF1*-positive and Δ*IKZF1*-negative cases was detected.[95]

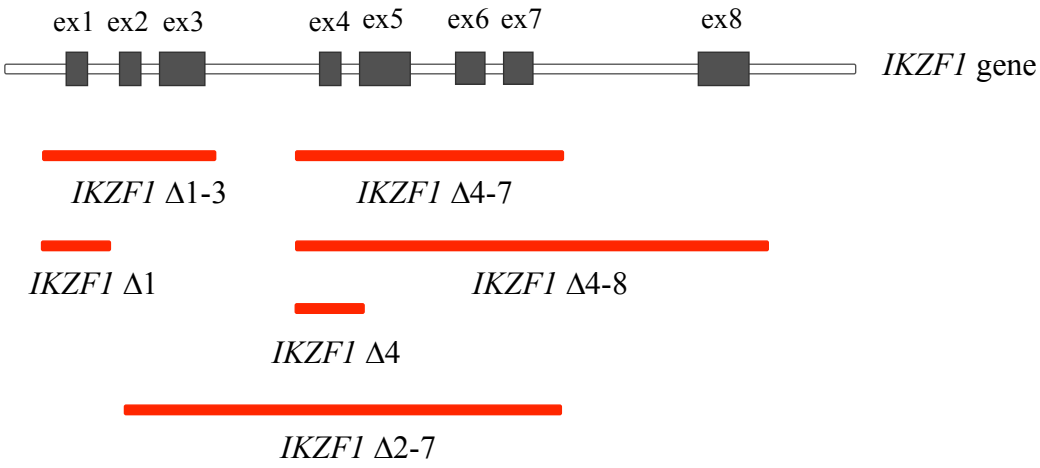

Figure 3 – Schematic illustration of the *IKZF1* gene; red bars represent the exonic regions most commonly affected by intragenic deletions in childhood BCP ALL

1.3.5.2 IKZF1 deletion as a prognostic marker

Since the study in 2008, several research groups have verified a significant association between Δ*IKZF1* and reduced event-free survival (EFS) and overall survival (OS) in ALL.[88,89,95,100,109,110,112] Of particular interest is the observation that Δ*IKZF1* confers a poor prognosis in cases without established risk-stratifying aberrations.[88,89,100] Moreover, Δ*IKZF1* is frequently occurring in the Philadelphia-like subtype mentioned in the previous section.[88,98] However, studies have also suggested that the prognostic impact of Δ*IKZF1* might be modified by other co-occurring genetic aberrations, such as deletions of the *ERG* gene and rearrangements of the *CRLF2* gene, or by treatment intensity, as well as by MRD stratification.[91,92,118-120] Thus, different studies have been conflicting as to whether Δ*IKZF1* represents an independent risk factor or not.[100,109,112,118,120] Despite numerous studies, there is no consensus regarding the relevance in using Δ*IKZF1* as a prognostic marker in the routine diagnostic setting.

2 AIM OF THESIS

The overall purpose of this thesis was to characterize genetic alterations underlying leukemia in children, with the primary aim to identify novel recurrent genetic lesions and evaluate their relevance for disease progression and outcome in pediatric ALL. These novel genetic markers might complement or replace current prognostic markers and facilitate clinical decision-making. A secondary aim was to evaluate alternative genetic analysis methods to detect novel as well as established markers in the routine laboratory diagnostics.

Specifically, the aims of thesis were to:

- Determine the frequency and prognostic impact of *IKZF1* deletions in childhood ALL treated according to NOPHO protocols

- Characterize and compare copy number profiles across cytogenetic subtypes of childhood ALL and identify focal alterations associated with leukemia development

- Investigate the genomic structure and transcriptional effects of the iAMP21 alteration in childhood ALL using massively parallel sequencing and genome-wide arrays

- Improve the sensitivity and efficiency of the genetic diagnostics of ALL

3 MATERIAL AND METHODS

3.1 PATIENTS AND CLINICAL DATA

All samples used in the studies were retrospectively collected from biobanks; residual bone marrow samples from children diagnosed with ALL from 1995 and onwards have been routinely saved at the Karolinska University Hospital as well as in a common NOPHO biobank. Clinical data for all cases, including clinical parameters, molecular findings and survival data, were extracted from the NOPHO database.

3.1.1 Patient cohort paper I and II

In **paper I**, DNA from a consecutive cohort of children (age 1–18) diagnosed with BCP ALL between 2001 and 2011 at the Karolinska University Hospital in Stockholm was used. During these years, a total of 148 children were diagnosed with BCP ALL in Stockholm; DNA from diagnostic bone marrow aspirates was available from 120 of those cases. Infant ALL cases were excluded. Among the 120 cases included in the study, 86 were treated according to the NOPHO-2000 protocol and the remaining 34 were treated according to the current NOPHO-2008 protocol. Median age was 5 years, with a male to female ratio of 1:1 and a median diagnostic WBC count of 9.5 x 10^9/l. Risk group assignment within the NOPHO protocol had classified 40% of the cohort as standard-risk (SR; n=48), 44% as intermediate-risk (IR; n=53) and 16% as high-risk (HR; n=19) ALL. A majority of cases (72/120) harbored low-risk cytogenetic markers, i.e. HeH and t(12;21), a few (13/120) had intermediate risk markers, i.e. iAMP21, t(1;19) or dic(9;20), and one patient was diagnosed with a high-risk, hypodiploid leukemia. In total, 34 out of 120 patients (28%) showed no risk-stratifying genetic markers at diagnosis. Two of these had Down syndrome ALL (DS-ALL). Only two *BCR/ABL1*-positive ALL cases were diagnosed in Stockholm during this time period and no DNA was available from these patients.

Paper II included all BCP ALL cases diagnosed in Sweden between 1992 and 2013 that had available *IKZF1* gene dose data (n=354). Thirteen of the 354 cases were infant ALL and seven were *BCR/ABL1*-positive; these 20 cases were excluded from further analysis. Among the remaining 334 cases, 23% (76/334) were treated according to NOPHO-1992, 43% (143/334) according to NOPHO-2000 and 34% (115/334) according to NOPHO-2008. Risk stratification at diagnosis had divided the patient into standard-risk (SR; n=136), intermediate-risk (IR, n=140) and (HR; n=58). The median age was 4 years (range 1–17 years), the male to female ratio 1:1 and the WBC count 9.7 x 10^9/l (range 0.9–1161 x 10^9/l). The cohort comprised 95 (28%) B-other cases, the remaining 239 (72%) cases harbored stratifying cytogenetic markers.

In the years between 1992 and 2013, a total of 1 133 Swedish children (1–17 years) were diagnosed with *BCR/ABL1*-negative BCP ALL and treated according to NOPHO protocols.

A comparison of the 334 (29%) cases used in this study with the 799 (71%) cases with unknown *IKZF1* status did not show significant differences with regard to clinical/molecular parameters or survival frequencies; the only exception was the median WBC count, which was significantly higher in the investigated cohort.

3.1.2 Patient cohort paper III

In **paper III**, 50 childhood BCP ALL cases diagnosed between 2001 and 2013 at the Karolinska University Hospital were included. Cases diagnosed between 2001 and 2008 (n=23) were selected to include common cytogenetic subtypes but also to maximize for cases without known risk-associated genetic aberrations, i.e. "B-other". Cases diagnosed 2009 and onwards (n=27) were consecutive. Thirty of the 50 cases had known cytogenetic markers, i.e. HeH (n=12), t(12;21) (n=9), t(1;19) (n=3), iAMP21 (n=3), t(9;22) (n=2), and *KMT2A* rearrangement (n=1). The remaining 20 cases were classified as B-other. All patients were treated according to NOPHO protocols except for one infant ALL and two t(9;22)-positive ALL cases. Median age at diagnosis was 5.5 years (range 0–18 years) and the median WBC count was 20.1 x 10^9/l (range 1.2–775 x 10^9/l). Two patients had DS-ALL. MRD values from end of induction (day 29) were available for 37 of the 50 cases; a majority of these cases had a positive MRD value (23/37).

The study also included a validation cohort, comprising 191 consecutive pediatric cases diagnosed with BCP ALL at the Departments of Pediatric Oncology and Hematology, Lund and Linköping University Hospitals, Sweden. A majority of these harbored known cytogenetic risk-markers, i.e. HeH (n=56), t(12;21) (n=40), t(1;19) (n=8), iAMP21 (n=2), t(9;22) (n=5), *KMT2A* rearrangements (n=9), hypodiploidy (n=4), while 35% (n=67) corresponded to B-other or failed karyotypes.

3.1.3 Patient cohort paper IV

For **paper IV**, 16 cases diagnosed with iAMP21-positive BCP ALL between 1992 and 2014 were selected; one case did not have sufficient material available and the remaining 15 were included together with relapse samples from two of the patients. The cohort represented all known iAMP21-positive ALL cases available (n=12), with the addition of three cases for which the iAMP21 was identified retrospectively through a BCP ALL methylation classifier described in a study by Nordlund et al.[121] Patients were treated according to the NOPHO ALL-1992 (n=2), ALL-2000 (n=11) and ALL-2008 (n=2) protocols. The median age at diagnosis was 9 years (range 5-17 years) and the median WBC count was 5.5 x 10^9/l (range 1.7 - 61.5 x 10^9/l). An additional cohort of 34 BCP ALL cases without iAMP21, i.e. dic(9;20) (n=14), HeH (n=6), 11q23 translocation (n=4), t(12;21) (n=4), t(9;20) (n=4) and B-other (n=2), was included to study differential gene expression.

3.2 MOLECULAR ANALYSES

3.2.1 Diagnostic cytogenetic analyses

At the time of diagnosis, the bone marrow samples of patients had been analyzed using chromosome banding and FISH to detect aneuploidies/gross rearrangements and risk-associated structural abnormalities respectively. FISH analysis with commercial probes had been used to screen for the t(1;19)(q23;p13) *TCF3/PBX1*, t(9;22)(q34;q11) *BCR/ABL1*, t(12;21)(p13;q22) *ETV6/RUNX1* and *KMT2A* (*MLL*) rearrangements, as well as iAMP21; iAMP21 was defined as >5 signals from the *RUNX1* probe, given that polysomy 21 had been excluded. The FISH probes used for detection of dic(9;20) had been designed and evaluated in **paper VI** before implemented in the routine diagnostics of ALL. All cytogenetic findings had been centrally reviewed by the NOPHO cytogenetic group.

3.2.1.1 DNA and RNA extraction

The DNA used for **papers I** and **III** had been extracted from bone marrow aspirates at the time of diagnosis using the Gentra Puregene Tissue kit (QIAgen, Hilden, Germany) for cases diagnosed prior to 2009, and the QIAgen AllPrep (QIAgen) for cases from 2009 and onwards. For **paper IV**, DNA and RNA from diagnostic bone marrow samples were extracted for the purpose of the study, using the QIAgen All-Prep DNA/RNAMini kit. RNA sample quality was measured using the RNA Nano Assay 6000 on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA); only RNA samples with a RNA integrity number (RIN) above 7 were used in study IV.

3.2.2 Targeted copy number analysis: MLPA

In **paper I**, multiplex ligation-dependent probe amplification (MLPA) was used to detect *IKZF1* deletions in 120 consecutive childhood ALL cases. MLPA is a multiplex variant of polymerase chain reaction (PCR), which allows for multiple genetic sites to be tested in the same run. Two oligonucleotide probes are designed for each target site; when both probes bind correctly, they are ligated to a complete probe, which is subsequently amplified by PCR. Each ligated probe has a unique length discernable by electrophoresis, and the forward primer is labeled with fluorescent dye, making it detectable on a capillary sequencer. The relative quantity of each amplified probe, i.e. the peak height, is normalized against control samples to generate relative copy number (CN) states of the target site in the sample. In our study, we used the commercial probe kits P335 and P202 (MRC-Holland, Amsterdam, The Netherlands). The P335 probe mix contains one probe per coding exon of the *IKZF1* gene, as well as probes for exons in the *PAX5*, *ETV6*, *RB1*, *BTG1*, *EBF1*, *CDKN2A/B* and PAR1 (*SHOX*, *CRLF2*, *CSF2RA*, *IL3RA*, and *P2RY8*) genes. Relative peak heights between 0.75 and 1.35 were considered normal (CN=2); values below 0.75 and above 1.35 were considered as deletions (CN ≤1) or duplications/amplifications (CN ≥3) respectively. Values below 0.25

indicated biallelic loss (CN=0). Four samples failed MLPA due to poor sample quality. Samples with deletions in *IKZF1* were further analyzed with the probe kit P202, which contains two probes per exon of *IKZF1*.

3.2.3 Genome wide copy number analyses: array-CGH and SNP-array

In **paper II-IV**, genome-wide array methods were used to detect copy number alterations (CNAs). In **paper III**, we used array comparative genomic hybridization (array-CGH) to investigate the study cohort (n=50). With this method, patient reference DNA are labeled with two different fluorescent dies respectively and hybridized to probes attached to a glass slide or a filter; the signal intensities of patient and reference DNA are compared and translated into a log2 ratio in order to detect copy number imbalances in the patient DNA. The resolution of array-CGH is determined by the size of the probes and the genomic distance between the probes. In our study, samples diagnosed between 2001 and 2008 (n=23) and samples from 2009 and onwards (n=27) were run on slightly different 4 x 180K array-CGH platforms from OGT (Oxford Gene Technology, Oxfordshire, UK), however average probe spacing for both platforms was 20 kb and the resolutions were comparable. To allow detection of subclonal CNAs, a cutoff log2 ratio corresponding to heterozygous loss or gain in >50 % of input DNA was set. Focal CNAs were defined as CNAs spanning less than 0.3 Mb and including only one or few genes. All data analyses were performed using the CytoSure Interpret software (OGT).

In papers II, **III** and **IV**, single nucleotide polymorphism (SNP) arrays were used. Similar to array-CGH, sample probe intensities are normalized against reference samples to produce log2 ratios centered at zero for a diploid sample, and log2 ratios are grouped using circular binary segmentation. However, in addition to CNAs, the SNP-array can detect SNPs and loss of heterozygosity (LOH). Of the 334 cases in **paper II**, 218 (65%) had been analyzed by SNP arrays at the respective center, using Human-OmniExpress BeadChip (Illumina, San Diego, CA, USA), Human 610-Quad BeadChip (Illumina), HumanOmni1-Quad BeadChip (Illumina), Human1M-Duo BeadChip (Illumina) or Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA). The CN state was determined using the GenomeStudio 2011.1 (Illumina), Chromosome Analysis Suite (Affymetrix), or the Nexus Copy Number (BioDiscovery, Hawthorne, CA, USA) software. Deletions affecting segments of ≥3 consecutive probes were included in the study. The remaining 116 cases (35%) included in **paper II** represented the cohort analyzed with MLPA in **paper I**; a comparison between cases identified by SNP arrays (n=218) and MLPA (n=116) showed no significant differences regarding clinical features or deletion types.

The validation cohort in **paper III** had been analyzed using the HumanOmni1-Quid and Human1MDuo array systems (Illumina) and CNAs were calculated using The Genome Studio Software 2011.1 (Illumina). In **paper IV**, the cohort was analyzed with SNP array platforms Omni 2.5M (n=9) or Omni 2.5M+Exome (n=4) (Illumina Inc, San Diego, CA,

USA), containing 2.5 million probes each. Segments of ≥10 consecutive probes and spanning ≥20 kb were included in the analysis. Annotation and filtering was performed using BEDOPS v2.4.2[122] R v3.1.0 and processed data was visualized using the Integrative Genomics Viewer (IGV) v2.3.42.[123]

In all these papers, CNAs previously reported as benign copy number variants in the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/home) and/or in local in-house hospital databases were generally excluded from further analysis.

3.2.4 Whole genome sequencing

In **paper IV**, DNA from diagnostic iAMP21-positive ALL cases (n=3) was analyzed with mate-pair whole genome sequencing (MP-WGS). The mate-pair technique generates long reads, suitable for structural variant detection. In mate pair library preparation, an initial fragmentation generates relatively long DNA fragments; these are end-labeled and circularized before subsequent fragmentation and paired-end sequencing. The sequenced short reads correspond to the ends of a longer DNA fragment, and thus the structure of the longer DNA sequence can be deduced from the genomic positions of the end sequences. This method was also used in **paper VIII**, with the purpose to evaluate the method for the routine diagnostics of ALL. In both **paper IV** and **VIII**, mate-pair libraries were prepared from diagnostic DNA using Illumina's Nextera Mate Pair Sample Preparation Kit, according to manufacturers instruction for a gel-free preparation of 2 kb effective insert size library. The libraries were sequenced on an Illumina HiSeq 2500 sequencer, 2x100 bp to an average raw coverage depth of ~5x. Adapter sequences were removed using Trimmomatic v0.32 and remaining pairs were aligned to the hg19 GRCh37 human reference genome sequence using bwa-mem version 0.7.4[124] resulting in a 3X mapped coverage. Mapped reads were analyzed for intra- and interchromosomal rearrangements using FindTranslocations (https://github.com/vezzi/Find Translocations), a locally developed software implementing a sliding windows analogue of a previously published procedure.[125] All rearrangements and translocations were manually inspected in IGV version 2.42.18.[123]

3.2.5 DNA methylation analysis

In **paper IV**, previously published data on DNA methylation levels from eight diagnostic iAMP21 samples were reanalyzed. In the previous study, the iAMP21 cases had been included as part of a training set to build subtype-specific DNA methylation classifiers, which were further tested in **paper VIII**. In **paper IV**, the iAMP21 samples (n=8) were compared to diagnostic BCP ALL samples without iAMP21 (n=665) and to 25 CD19+ B-cells from healthy blood donors using the non-parametric Wilcoxon rank-sum test to investigate differential methylation in iAMP21. The mean, standard deviation (SD), and mean methylation difference between the two groups was measured. Minimal cut-off value

for the mean absolute differences in DNA methylation (Δβ) was set to ≥20% in order to identify CpG sites with large difference between the groups. The p-values were corrected for False Discovery Rate (FDR) due to multiple testing.

3.2.6 RNA sequencing

In **paper IV**, RNA from diagnostic (n=12) and relapse (n=3) iAMP21-positive ALL samples was analyzed using whole transcriptome sequencing (RNA sequencing). Samples were first treated with RiboZero (Epicentre, Madison, WI, USA) in order to eliminate ribosomal RNA. Next, strand-specific RNA sequencing libraries were prepared using ScriptSeq v2 (Epibio.com) and sequenced 50 bp paired-end on Illumina HiSeq 2000/2500, producing ~100 million read pairs per sample. Sequence reads were aligned to the human genome reference build GRCh38 (hg38) using Spliced Transcripts Alignment to Reference (STAR) version 2.4.0j[126] with exon junction support from Ensembl gene annotation version 77. The trimmed mean of M-values normalization method[127] was used for normalization of raw read counts, and Voom[128] was used for variance normalization. Genes with a count of ≥ 1 per million mapped reads (CPM), in ≥ 2 samples, were included for further analysis. Gene expression levels were also normalized to reads per kilobase per million mapped reads (RPKM). Raw sequence reads in fastq format were further analyzed for the presence of fusion genes using FusionCatcher.

3.2.6.1 Differential gene expression

Analysis of differential gene expression was performed using the R/Bioconductor package limma.[129] Twelve iAMP21-positive cases were contrasted against the combined average expression of a total of 46 BCP ALL cases. Unsupervised hierarchical clustering was performed using Eucledian distances and complete linkage and significance tested by F test statistic in Limma.

3.2.6.2 Sequence variant analysis

The cDNA sequencing reads were aligned to the human genome reference GRCh37 (hg19) and nucleotide variants were detected by using STAR[126] with exon junction support from Ensembl gene annotation version 75. Sequence reads that overlapped exon junctions were split and trimmed to exon boundaries using GATK version 3.2-2.[130] Duplicate reads, clustered SNPs (≥3 within 35bp), variants with high Fisher strand values (FS>30) or low depth-adjusted quality scores (QD<2), were excluded from further analysis. No constitutional DNA was available for filtering. Variants were annotated and assessed using Ensembl Variant Effect Predictor[131] and GEMINI[132], with the following criteria for putative somatic variants in the minimal region of amplification on chromosome region 21q: 1) non-synonymous variant; 2) in a coding region; 3) supported by ≥ 10 variant reads, with a minimum of 20 total reads; 4) not in a repeat masked region. Variants outside of the MRA were additionally filtered on potential pathogenicity using Scaled Combined Annotation

Dependent Depletion (CADD),[133] with a cut-off score of ≥15. Filtered variants were further tested against previously registered cancer relevant variants in the COSMIC (http://cancer.sanger.ac.uk/cosmic/) and the Pediatric Cancer Genome Project (https://pecan.stjude.org/) databases.

3.3 VALIDATIONS

In **paper I**, all detected deletions affecting the *IKZF1* gene were validated with the P202 kit; deletions that affected single *IKZF1* exons with the P335 kit were only included if validated by the P202 kit. In **paper III**, all CNAs were manually inspected using CytoSure (OGT) and IGV[134], and a consecutive cohort of 191 cases analyzed by SNP-array was used for validation of novel CNAs. Validated CNAs were less frequent in the validation cohort, likely due to a higher proportion of B-other cases as well as to the inclusion of subclonal CNAs in the study cohort. In **paper IV**, the rearrangement breakpoints in chromosome 21 detected by MP-WGS were manually inspected at the nucleotide level using IGV, and validated using PCR for sample KSALL11. The PCR primers were designed using Primer3Plus.[135] Recurrent fusion transcripts were validated using RT-PCR and Sanger sequencing of the cDNA fusion breakpoint.

3.4 STATISTICAL ANALYSES

All clinical data were retrieved from the NOPHO database. In **papers I** and **III**, statistical calculations were performed using the SPSS software (SPSS Inc., Chicago, IL, USA). The significance limit for two-sided p-values was set to <0.05. In **papers I** and **II**, the Kaplan-Meier method was used to calculate the probabilities of event-free survival (pEFS) and overall survival (pOS) at 10 years after diagnosis; groups were compared using the log rank test. The events in the pEFS analysis were defined as induction failure, relapse, death in first remission and second malignant neoplasm. In the analysis of pOS, death was the only end point. To assess the independent impact of *IKZF1* deletion on survival, multivariate analyses adjusting for common risk factors were performed using the Cox proportional hazard regression analysis.

In **paper II**, clinical and genetic features between cases with known *IKZF1* status (n=334) and those with unknown *IKZF1* status (n=799) as well as between *IKZF1* deleted (n=50) and non-deleted cases (n=284) were compared using the Mann-Whitney U-test for continuous variables and the two-tailed Fisher's exact probability test for discrete variables. Also, in **paper II**, the 10-year probability of relapse (pRel) was calculated using the Kaplan-Meier method; only relapses were included in the estimations of cumulative incidence of relapse.

The frequencies of specific CNAs in different cytogenetic subtypes described in **paper III** were compared using the two-tailed Fisher's exact probability test for discrete variables. The

David Functional Annotation Bioinformatics tool (http://david.abcc.ncifcrf.gov) was used for gene cluster and pathway analyses in **papers III** and **IV**.

In **paper IV**, statistic analyses were performed using R, and the statistical methods used for the genome and transcriptome analyses were implemented as described in the previous method sections.

4 RESULTS AND DISCUSSION

4.1 FREQUENCY OF Δ*IKZF1* IN NOPHO COHORTS

In **paper I**, we investigated the frequency and prognostic impact of *IKZF1* deletions (Δ*IKZF1*) in patients diagnosed with ALL in the Stockholm region between 2001 and 2011; this was the first published study assessing the impact of Δ*IKZF1* within the NOPHO protocols. Deletions were detected in 16% (19/116) of *BCR/ABL1*-negative BCP ALL cases; a minority (5/19) were intragenic deletions of exon 4-7, predicted to result in the dominant-negative isoform IK6. Median WBC count in the group with Δ*IKZF1* (n=19) and the group without Δ*IKZF1* (n=97) was 20.3 and 8.1 respectively, which correlated with a slightly larger proportion of high-risk (HR) patients in the Δ*IKZF1* group (21% vs. 15%).

In **paper II**, the cohort was extended to include all *BCR/ABL1*-negative BCP ALL cases with available Δ*IKZF1* data in Sweden (n=334). In this cohort, 15% (50/334) harbored deletions of the *IKZF1* locus, a majority of these (58%) were focal (only including the *IKZF1* gene) and there were no significant differences in clinical or molecular parameters between deletion types. Also in this study, the Δ*IKZF1* -positive cases (n=50) had a significantly higher median WBC count than cases without Δ*IKZF1* (n=284) and the Δ*IKZF1* cases had more frequently been stratified as HR (36% vs. 14%; p<0.001).

4.2 PROGNOSTIC IMPACT OF Δ*IKZF1* IN NOPHO PROTOCOLS

In **paper I**, we could show that the probability for event free survival (pEFS) and overall survival (pOS) were significantly reduced in the Δ*IKZF1*-positive group compared to the group with intact *IKZF1* (p=0.02 and p=0.001, respectively) (Figure 4). These associations remained significant in a multivariate analysis adjusting for known clinical and molecular risk factors (pEFS; P=0.028, pOS; p=0.005), however MRD data was not available for the analysis. In addition, we observed a high prevalence of Δ*IKZF1* (26%) in B-other cases, and Δ*IKZF1* was present in all cases that died in this group (4/34). Our results indicated that Δ*IKZF1* was an independent risk factor for poor outcome and that Δ*IKZF1* might represent a negative prognostic marker for childhood leukemia treated accorded to NOPHO.

The rationale for **paper II** was to further investigate the prognostic impact of Δ*IKZF1* but also to determine if Δ*IKZF1* is an independent risk factor also in the context of MRD and other prognostic markers. This study could confirm a significant association between Δ*IKZF1* and decreased pEFS (p<0.001) and pOS (p=0.001) as well as increased probability of relapse (pRel) (35% vs. 12%; p<0.001) (Figure 5) in the total cohort. However, the impact varied among the different NOPHO protocols and risk groups. Δ*IKZF1*-positive cases had a significantly increased pRel in the 1992 protocol but not in the 2000 and 2008 protocols (Figure 5). Also, *ΔIKZF1* was associated with higher relapse rate in the SR and HR groups, but not in the IR group. Co-occurrence of Δ*IKZF1* and PAR1 deletions (n=4) was

significantly associated with a higher risk of relapse compared to Δ*IKZF1* without PAR1 deletion (pRel 75% vs. 30%, p=0.045), while the presence of *ERG* deletions did not affect relapse risk for Δ*IKZF1*-positive cases (n=3). Multivariate analysis including the whole cohort showed that Δ*IKZF1* was the strongest independent risk factor for relapse and poor outcome when age, WBC count, treatment protocol, MRD, PAR1 deletions and cytogenetic groups were included as variables in the model (p=0.001). This was true also when only patients with available MRD data (n=219) were included in the analyses.

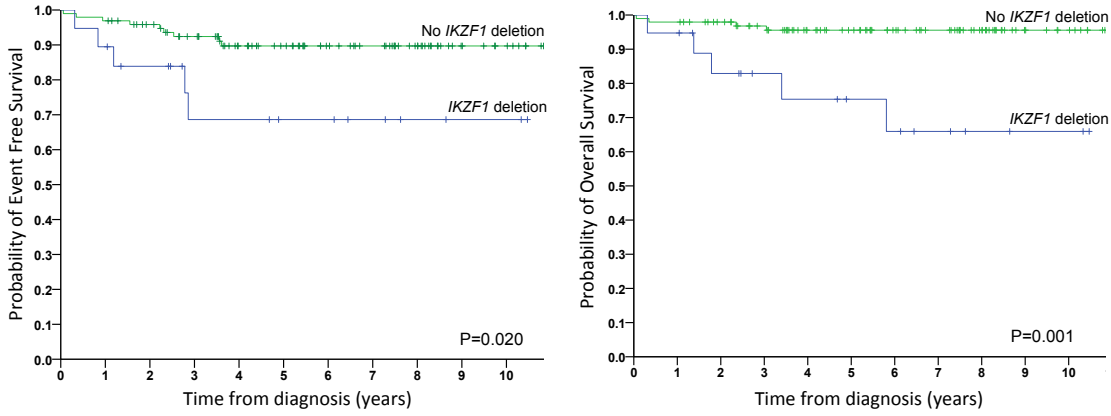

Figure 4 – Kaplan-Meier estimates of probability of A) event free and B) overall survival of B-cell precursor acute lymphoblastic leukemia cases (n=116) with and without *IKZF1* deletions (blue and green lines respectively) treated according to NOPHO protocols in **paper I**

Paper II could confirm that the frequency and prognostic effect of Δ*IKZF1* was most pronounced in the B-other group; *ΔIKZF1* was detected in 27% in this group and was strongly associated with increased risk of relapse (pRel 52% vs. 4%; p<0.001) (Figure 5) and decreased survival (pOS 58% vs. 90%; p<0.001). Notably, 83% of the relapsed B-other cases were *ΔIKZF1*-positive. The significant association with relapse risk and poor outcome in the B-other group was consistent through all three of the treatment protocols and remained significant in a multivariate analysis including only B-other.

In an ongoing study, we are investigating the prognostic impact of Δ*IKZF1* in approximately 400 patients treated according to the current NOPHO-2008 protocol; preliminary data on 272 cases show a potential significant association between Δ*IKZF1* and relapse risk, as well as a high frequency (28%) of Δ*IKZF1* in the B-other group (n=80), and a significant risk for relapse and poor outcome in this subtype.

Taken together, these findings support that Δ*IKZF1* represents a negative prognostic marker for childhood ALL treated according to NOPHO protocols, and highlight the importance of including Δ*IKZF1* in the diagnostic evaluation of patients presently grouped as B-other.

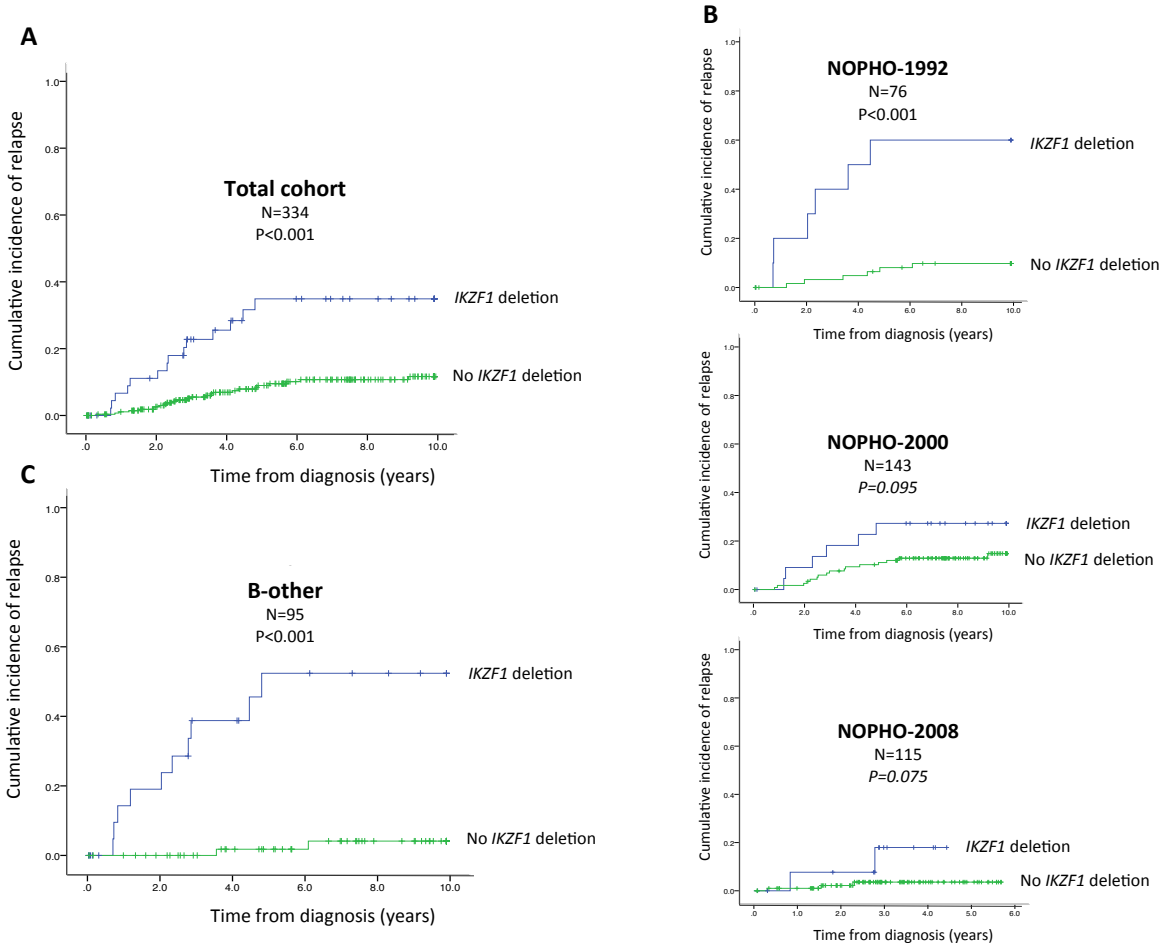

Figure 5 – Kaplan-Meier estimates of probability of relapse of B-cell precursor acute lymphoblastic leukemia cases with and without *IKZF1* deletions (blue and green lines respectively) A) in the total cohort (n=334), and B) in relation to treatment protocols, and C) in the B-other group (n=95) in **paper II**

4.3 *PAX5*^{AMP} – A NOVEL SUBTYPE OF ALL?

An unexpected finding in **paper I** was a recurrent intragenic amplification of the *PAX5* gene *(PAX5*^{AMP}) in 4% (5/116) of the cohort. *PAX5* encodes the transcription factor PAX5, an important regulator of B-cell development. The gene is frequently deleted or mutated in BCP ALL,[84,98,136] however intragenic amplifications had previously only been reported in isolated cases,[137,138] indicating that this aberration might escape detection using other methods than MLPA. The amplifications affected exons 2 and/or 5 (Figure 6), encoding the DNA-binding domains of PAX5. As all *PAX5*^{AMP} cases were B-other and 40% (2/5) had relapsed, we proposed this as a potential novel genetic subtype of ALL.

The frequency and prognostic impact of *PAX5*^{AMP} were further explored in **paper IX** (under review in Blood), including MLPA data from 15 international study groups in Europe. Amplifications were validated using quantitative genomic PCR, arrays and FISH, and population-based screening showed that *PAX5*^{AMP} occurred in ~1% of BCP ALL (52/5 535), with an incidence of 3% in the B-other group (33/1 271). *PAX5*^{AMP} was associated with a

high risk for relapse, with 5-year event-free and overall survival rates of 49% (95% CI 36-61%) and 67% (95% CI 54-77%), respectively. The amplifications remained stable between diagnosis and relapse, indicating that this abnormality might represent a primary event in ALL. The high incidence in B-other cases supports that *PAX5*^{AMP} may define a distinct subtype that should be assessed as a potential risk-stratifying marker in future ALL protocols.

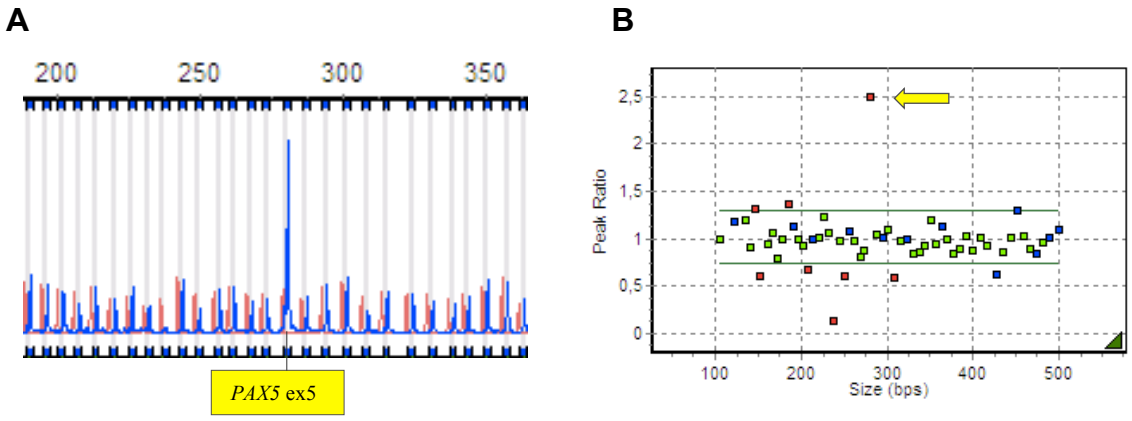

Figure 6 – Images from MLPA analysis of one of the cases with *PAX5*^{AMP} in paper I, the large peak in A and the high red probe marked with an arrow in B both correspond to amplification of *PAX5* exon 5

4.4 THE COPY NUMBER LANDSCAPE OF B-OTHER

Cases without cytogenetic risk markers, i.e. B-other, constitute a diagnostic problem in all current treatment protocols, and the search for novel genetic markers is particularly relevant for this group. As described in previous sections, **papers I** and **II** demonstrated that Δ*IKZF1* is frequent in B-other cases and that the negative prognostic effect is particularly significant for this group. Also, **papers I** and **IX** have proposed *PAX5*^{AMP} as a rare but recurrent negative prognostic marker for the B-other group.

In **paper III**, where we investigated the CN landscape of ALL, Δ*IKZF1* was detected in 35% of the B-other group; other frequent deletions affected *VPREB1* (35%), *CDKN2A/B* (35%), *PAX5* (20%) and *ERG* (15%). Deletions of the B-cell development gene *VPREB1* (Δ*VPREB1*) have recently been reported in association with poor prognosis[90] and the Philadelphia-like subtype.[98] While Δ*IKZF1* consistently occurred together with other CNAs in our study, Δ*VPREB1* was the only somatic imbalance detected in two B-other cases, suggesting that this deletion could be important for malignant transformation.

Taken together, these four studies have shown that B-other cases harbor recurrent focal deletions with potential prognostic significance that are not detected and/or reported in the current genetic work-up of ALL. Detection of these deletions would provide information for risk stratification in NOPHO treatment protocols, particularly for the B-other group.

4.5 DELETIONS IN THE CYTOGENETIC SUBTYPES OF ALL

According to the current view, most of the risk-stratifying genetic markers need secondary genetic alterations in order to cause leukemia. In **paper II**, we found no evidence for a prognostic impact of Δ*IKZF1* in any of the cytogenetic subtypes except for B-other, and a negative correlation between t(12;21) and Δ*IKZF1* was detected; only 6% of the cases with t(12;21) had Δ*IKZF1*. In **paper III**, we further characterized the genome-wide copy number landscape in different cytogenetic subtypes of BCP ALL. Overall, we identified over 400 CNAs, and the vast majority of these lesions were subchromosomal deletions (Figure 7). The frequencies of recurrent CNAs in the different cytogenetic subtypes are shown in Figure 8; in agreement with previous studies, deletions of *CDKN2A/B* (n=15), *IKZF1* (n=15), *VPREB1* (n=15) and *PAX5* (n=10) were common, but we also detected a few rare and recurrent deletions in genes not previously implicated in BCP ALL, including *IRF1*, *PDE4B* and *INIP*.

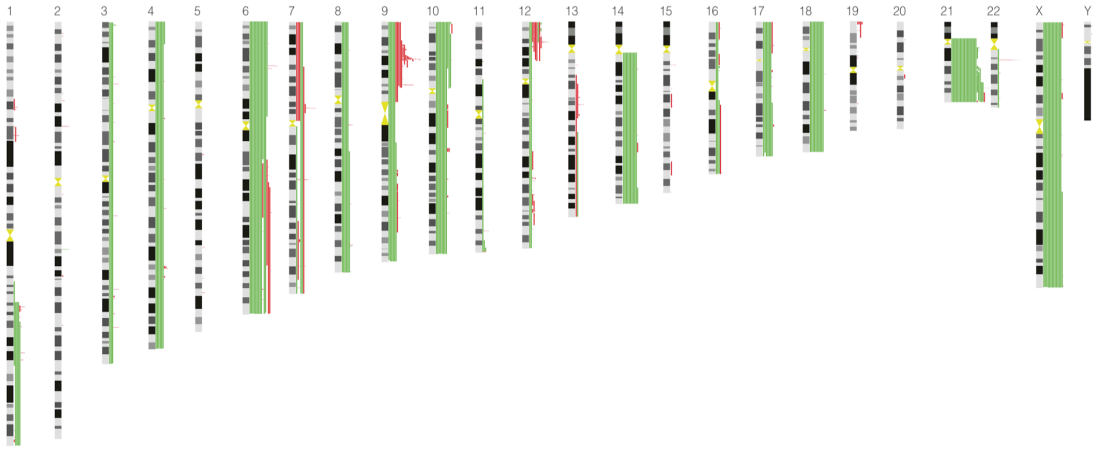

Figure 7 – Ideogram summarizing somatic copy number alterations detected in **paper III**, with gains in green and losses in red

IRF1 encodes a tumor suppressor protein involved in the regulation of normal hematopoiesis and leukemogenesis; deletions of *IRF1* are frequent in myelodysplastic disorders[139] and may thus constitute a possible mechanism of tumor suppression inactivation in ALL. Pharmacologic inhibition of PDE4B increases glucocorticoid sensitivity[140] and constitutional variants have been associated with relapse in BCP ALL,[141] which indicates that somatic loss of *PDE4B* could be relevant for therapy response and relapse in ALL. Deletion of the DNA repair gene *INIP* was unique for the t(12;21) subtype. The encoded protein is involved in the maintenance of genome stability and DNA damage response,[142,143] and *INIP* deleted cases showed few additional CNAs, suggesting that this alteration could be important for leukemia development.

Moreover, we found recurrent deletions recently described in association with high-risk and relapsed ALL that affected *VPREB1*[90] as well as the chromatin remodeling genes

CREBBP[93,94,97] and *TBL1XR1*;[95,96] the two latter are predicted to result in glucocorticoid resistance. *CREBBP* deletions (n=5) were exclusive to cases with polysomy of chromosome 21, i.e. HeH and DS-ALL, and the HeH cases with focal *CREBBP* deletions had a positive MRD value at the end of induction. The presence of this high-risk alteration in the otherwise low-risk HeH subtype indicates a prognostic value in detecting and reporting this alteration in the clinical setting.[97] Similarly, Δ*VPREB1* was common in the t(12;21) group (55%) (Figure 8) and associated with a higher WBC count in this otherwise low-risk subtype. Deletions of *CREBBP* and *BTLA* clustered in subtypes with extra chromosome 21 material, i.e. iAMP21, HeH, unbalanced t(12;21) and DS-ALL, suggesting that these deletions might be associated with both germline and somatic gain of chromosome 21 in BCP ALL.

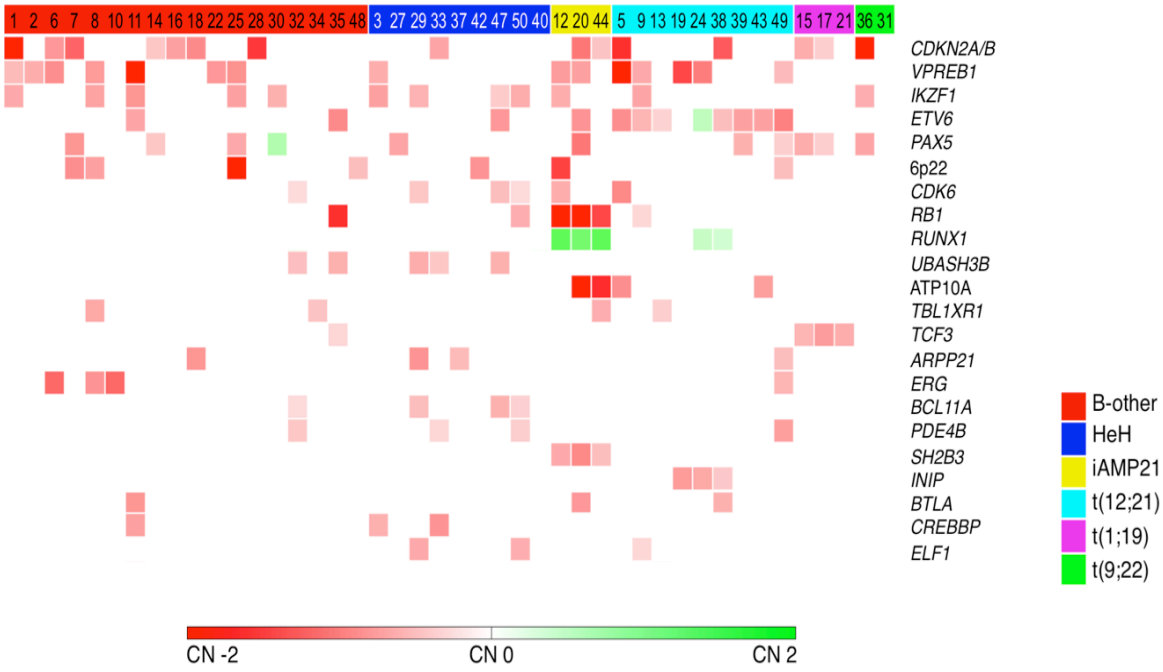

Figure 8 – Heat map showing relation between genes/regions affected by recurrent (>2 samples) CNAs (rows; red=deletion, green=duplication) and cytogenetic subtypes (columns) in **paper III**

4.6 INTEGRATED ANALYSIS OF IAMP21

Although the iAMP21 aberration only constitutes 2% of childhood BCP ALL, other alterations affecting chromosome 21, such as polysomies and translocations, are often seen in the disease. Moreover, iAMP21-positive BCP ALL is associated with a high risk of relapse when treated according to standard-risk protocols, and the biological mechanism behind the high relapse risk is still unknown.

Papers I and **II**, which included 5 and 7 iAMP21 cases respectively, could show that Δ*IKZF1* were infrequent in iAMP21 and did not add prognostic information for this subtype. In **paper III**, including three iAMP21 samples, a significant association between iAMP21

and deletions of the tumor suppressor gene *RB1* and the JAK2 regulating gene *SH2B3* ($p=0.001$ and $p<0.001$ respectively) was detected. Also, the largest number of subchromosomal CNAs was found in the iAMP21 subtype, even after the 21q amplifications were excluded from the analysis (median 17, range 10–19). In **paper IV**, we further investigated the genomic structure and transcriptional effects of the iAMP21 rearrangement. SNP-array and WGS demonstrated a large complexity and heterogeneity of the iAMP21 rearrangement between analyzed cases (Figure 9), however no obvious explanation to the pathogenicity could be found. The study could verify a high frequency of deletions in *RB1* (42%) and *SH2B3* (42%), however none was present in all iAMP21 samples and thus the role of these CNAs in the pathogenicity of iAMP21 is not clear.

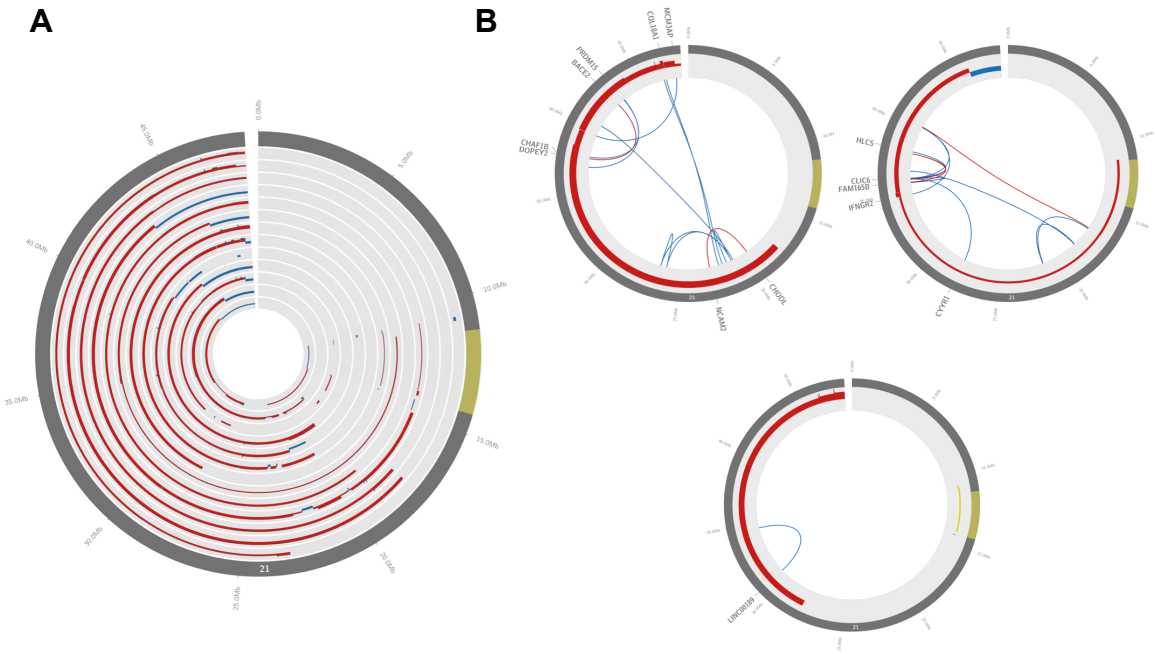

Figure 9 – Circos plots of chromosome 21, showing A) CN data for all analyzed samples (n=12), and B) intrachromosomal rearrangements detected by mate-pair WGS (n=3); blue links represent rearrangements retaining the original strand orientation, red links represent inverted rearrangements, CN changes are shown with red (amplifications) and blue (deletions) lines in the circos edges

To study the transcriptional effects underlying the pathogenicity of the 21q amplification, 12 iAMP21-positive cases were compared with 34 iAMP21-negative BCP ALL cases using RNA sequencing. The analysis focused on the minimal region of amplification (MRA), determined by a previous study to a 5,1 Mb region on 21q22.3.[69] Our results showed unique and significant overexpression of 13 MRA genes in iAMP21; the most significant of these were the tyrosine kinase gene *DYRK1A* and the chromatin remodeling genes *CHAF1B* and *SON* (Figure 10).

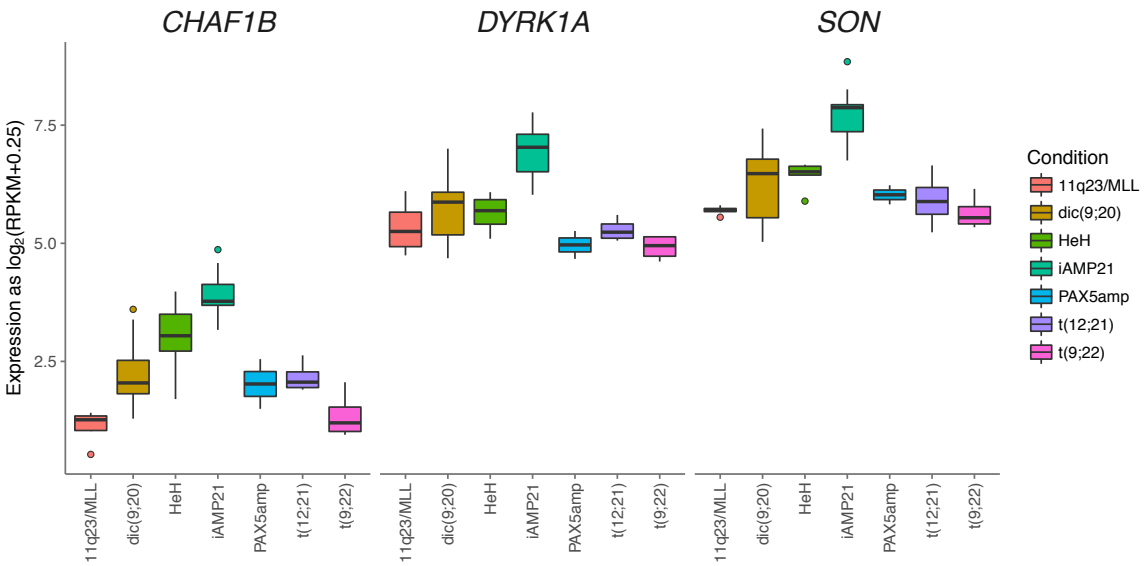

Figure 10 – Box plot showing mean expression of genes *CHAF1B, DYRK1A* and *SON* across the different subtypes of BCP-ALL in **paper IV**; these three genes were the top significantly overexpressed genes in the minimal region of amplification in the iAMP21 subtype

Fusions including *DYRK1A* are present in BCP ALL,[102] further strengthening the role of DYRK1A in this disease, and *DYRK1A* overexpression has been suggested as a potential tumor-promoting factor in Down syndrome associated ALL and acute megakaryocytic leukemia (AKML).[144-146] DYRK1A promotes cell quiescence in normal lymphocyte development,[147] and, interestingly, the cell quiescence mechanism of DYRK1A has been shown to cause relapses in gastrointestinal stromal tumors treated with Imatinib.[148] Thus, the quiescence promoting function could be relevant for the relapse risk in patients with iAMP21, and the tyrosine kinase properties of DYRK1A makes it a potential candidate for targeted treatment.

Both *CHAF1B* and *SON* are involved in chromatin remodeling. Overexpression of *CHAF1B* has been associated with unfavorable prognosis in several malignancies[149-153] and short SON isoforms are upregulated in leukemic blasts and acute myeloid leukemia.[154,155]

In **paper IV**, we also investigated the presence of sequence variants in the RNA sequencing data; many of the variants affected genes with cell cycle and/or kinase properties, and functional analysis demonstrated that all top significant clusters represented different aspects of cell division, including chromosome segregation, chromatin assembly and cell cycle regulation. However, no constitutional DNA was available for filtering of somatic variants, and thus it remains to be elucidated if the detected variants represent acquired mutations or constitutional variants with leukemia predisposing properties.

4.7 METHODOLOGICAL ASPECTS OF BCP ALL DIAGNOSTICS

In **paper I**, we investigated the potential of MLPA as a method for targeted CN analysis. Comparison between FISH and MLPA results for alterations in *CDKN2A* and *ETV6* showed a high concordance, with a slightly higher detection rate by MLPA. The likely explanation for this discrepancy is that the MLPA probes target smaller stretches of the genome and will thus detect CNAs below the resolution of FISH. The robustness of MLPA is affected by sample quality factors, in particular salt concentration, but with current DNA extraction methods and sufficient experience, the method had a high accuracy and low rate of false negative findings for CNAs present in >30% of cells. MLPA is inadequate to detect CNAs present in smaller subclones.

In **papers III** and **IV**, we used array-CGH and SNP-array, respectively, to detect genome-wide CNAs. Both methods have been validated in the clinical setting and array-CGH is now routinely used in the clinical diagnostics of childhood ALL at the Karolinska University Hospital. Both methods have a high sensitivity for detection of CNAs present in a majority of cells, however the array-CGH platform contains fewer probes, and thus CNAs affecting very small parts of the genome escape detection. For example, the *PAX5*^{AMP} detected by MLPA in **paper I** was also detected by the SNP-array but not by array-CGH. Compared with MLPA, the SNP-array is more robust and can detect subclonal CNAs present in >20 % of cells, while the array-CGH has a low sensitivity for subclonal CNAs.

To detect the novel prognostic CNAs in B-other cases, as well as the relapse-associated CNAs in standard-risk subtypes, high-resolution CN analysis at diagnosis is recommended. Importantly, relapse-prone alterations can be present in small subclones at diagnosis, and thus genomic array platforms used in the clinic should have a high probe density for the relevant genes, and ideally allow robust detection of small subclones.

In **paper IV**, we used mate-pair WGS (MP-WGS) to investigate the iAMP21 subtype, and in **paper VIII** (submitted to Genes, chromosomes and cancer), this method was applied on BCP ALL cases with other cytogenetic alterations, i.e. t(1;19), t(9;22), t(12;21) and *KMT2A* rearrangement. Both studies showed that the method could detect structural rearrangements with high resolution. In **paper VIII**, MP-WGS accurately detected all known rearrangements, including determination of the exons involved and sometimes the exact fusion sequence, as well as previously undetected genetic events that could be confirmed by Sanger sequencing. The results indicate that MP-WGS combines the screening character of karyotyping together with the specificity and resolution offered by FISH and/or RT-PCR analysis. The WGS method offers large potential for leukemia diagnostics, however, the sequencing depth must be sufficient to detect subclonal alterations, and, ideally, information on constitutional variants should be used for filtering out somatic variants.

5 CONCLUSIONS

In this thesis project, we have used a variety of methods to search for genetic alterations relevant for disease progression and outcome in childhood ALL. The main conclusions presented in this thesis are as follows:

- *IKZF1* deletion is a negative prognostic marker for childhood BCP ALL treated according to NOPHO protocols, with a pronounced effect in the B-other group
- Intragenic amplifications of *PAX5* might represent a novel genetic subtype associated with a high relapse risk in the B-other group
- A majority of BCP ALL presently classified as B-other harbor recurrent focal CNAs with prognostic significance; screening for Δ*IKZF1*, $PAX5^{AMP}$ and PAR1 deletions at diagnosis would contribute to the diagnostic evaluation of these patients
- CNAs associated with high relapse risk, i.e. deletions of *CREBBP*, *VPREB1* and *TBXLR1*, are rare but recurrent in the low-risk cytogenetic subtypes, suggesting that detection of these would contribute to the clinical evaluation at diagnosis
- Customized SNP-array platforms detect the novel prognostic CNAs with high accuracy, and MLPA represents a cost-effective alternative to this analysis. MP-WGS has the ability to detect both the current risk-stratifying alterations as well as the described novel CNAs in one single analysis
- The iAMP21 subtype has a unique transcriptional profile, with significant overexpression of three potential candidate genes, i.e. the tyrosine kinase *DYRK1A* and the chromatin remodeling genes *SON* and *CHAF1B*; each of these genes is involved in leukemia development and could represent a potential target for therapy

6 FUTURE PERSPECTIVES

While the novel markers proposed in this thesis are potential indicators of poor prognosis, no standardized treatment strategy is yet available for these cases. For example, relapses of Δ*IKZF1*-positive patients occur in all risk groups, including the HR group, which indicates that the currently used HR protocols are insufficient to prevent relapses for this subtype. Further insights into causative mechanisms behind the novel and currently used risk markers might open for targeted treatment approaches for these groups.

The recent use of massively parallel sequencing methods in the research field has increased our knowledge of the molecular background in ALL. Rather than providing simple answers, studies have unveiled previously unknown complexities of the disease and revealed intricate correlations between constitutional and somatic genetic lesions, aberrant epigenetic patterns, transcriptional effects and treatment response. A better understanding of this interplay is likely needed in order to improve diagnostic and treatment procedures. Importantly, constitutional variants have been shown to play a role in therapy response and risk for therapy resistance and relapse. While the main focus of cancer diagnostics has been to find somatic drivers by filtering out constitutional variants, this new knowledge indicates that information on these variants might actually facilitate treatment decisions and help predict treatment response.

Survival rates in modern protocols are approaching 90% and thus any change to the treatment approach must be carried out with precaution. Nevertheless, a reduction of treatment intensity would reduce the risk of side effects and improve quality of life for the patients. Further improvement of survival and quality of life in childhood ALL will likely be obtained through the development of targeted compounds combined with a more personalized diagnostic approach, where treatment decisions are based on the presence of targetable somatic mutations and therapy-relevant constitutional genetic variants in each child rather than on risk-group stratification.

7 ACKNOWLEDGEMENTS

I am deeply grateful to all who have supported the work behind this thesis, and to the patients, their parents and physicians, who have made this research possible.

My main supervisor, associate professor Gisela Barbany, for your positivity, intelligence and humor in guiding me through this research education with the perfect balance of support and freedom. Your passion and excitement for our research have inspired me and I could not have wished for a better supervisor.

My co-supervisors: Magnus Nordenskjöld, for welcoming me into the fascinating world of genetics, for finding me just the right research project, and for all your support through these years. Ann Nordgren, for all your positivity, smart ideas and encouraging pep talks. Mats Heyman, for all your help with the clinical data and for your enthusiasm for our research.

The leukemia-research group, for all the help with planning and executing the projects: Anh Nhi Tran, for your excellent guidance in the lab, and all the nice talks in between. Vasilios Zachariadis, for all your contributions and for answering all my bioinformatics questions. Fulya Taylan, for help with primers and figures, and for your positive attitude in general.

All researchers, clinicians and co-authors who have contributed to the projects: Linda Olsson and Bertil Johansson for successful collaborations. Jessica Nordlund, for your essential contributions and for guiding me in the interpretation of methylation data. Ann-Christine Syvänen, Gudmar Lönnerholm, Yanara-Marincevic-Zuniga, Dan Grandér, Katja Pokrovskaja and Leonie Saft for productive collaborations. All pediatric oncologists and researchers associated with the NOPHO Biology group, in particular Johan Malmros, Stefan Söderhäll, Erik Forestier, Ulrika Norén Nyström and Kield Schmiegelow, who have shown encouraging interest in our research and contributed with a clinician's perspective on the projects.

All past and present co-workers and co-authors at the Clinical Genetics division: the laboratory staff, for all your help and support with sample and lab issues. The physicians, for welcoming me into the clinic, and for patiently answering all my questions regarding genetics. A special thanks to Erik Iwarsson and Erik Björck for supervising me in the clinic and for encouragement in the research, to Agne Liedén for help with interpretation of array data, and to Eva Rudd for being a great friend and for supporting my research studies with endless amounts of great coffee and inspirational discussions.

Friends and fellow researchers at CMM: Alice, Anders, Benedicte, Bianca, Cecilia, Jesper, Josephine, Karin, Malin, Miriam, Nina, Samina, Sofia, Tobias, and Wolfgang. A special thanks to Alexandra Löfstedt, for our many inspiring discussions, Maria Pettersson, for all fun talks during lunch, Ellika Sahlin, for being a great friend and travel partner and for excellent proof reading of the thesis, Johanna Winberg, for being a great friend and support

through these years, and for sharing the angst of the clinical and research work, Katarina Haapaniemi Kouru for all the fun times at the clinic and nice dinners.

The organizers of the cancer research school NatiON: Claes Karlsson, Dan Grandér and Ingemar Ernberg, for providing interesting and important research courses, and for all the inspiring discussions.

The administrative staff at MMK: Ann-Britt Wikström and Britt-Marie Witasp, for your invaluable help with applications, course registrations and grants. The administrative staff at Clinical Genetics: Karin Kindberg and Eva Ekblom, for all your help and quick problem solving. The CMM IT department, in particular Christer Uvehag, for helping out with license registrations and computer problems.

I would also like to express my gratitude to Karolinska Institutet, Stockholms Läns Landsting (ALF and forskar-AT), Barncancerfonden and Mary Bevé foundation for financial support of this work.

A warm thanks to all my friends, with a special thanks to Louise Brattström Stolt for painting the fantastic arts for this thesis and for being a great friend and partner in crime since med school, and to my dear friend Josefin van der Meer for sharing so many thoughts, ideas and crazy times since the “naturvetenskapliga basår”.

To my family: George, Marc, Ulrika, Stefan, Harald, Amina and Isabel. Mariana, Volen, Elin and Edvin. A special thanks to Stefan, for support and for being a warm and generous dinner-host, and to Ulrika, for babysitting Marc during this thesis-writing period, and for being such a strong and tough female role model.

To George, for all your love and support, for encouraging me to start with research in the first place and for being who you are. To Marc, for being the most fun and wonderful baby ever.

8 REFERENCES

1. Wirchow R. Zur pathologischen Physiologie des Blutes. II. Weisses Blut. . *Archives of Pathology, Anatomy and Physiology* 1847; **I**: 563-72.

2. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. *Journal of the National Cancer Institute* 1960; **25**: 85-109.

3. Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012; **481**(7381): 306-13.

4. Jones S, Chen WD, Parmigiani G, et al. Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences of the United States of America* 2008; **105**(11): 4283-8.

5. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009; **458**(7239): 719-24.

6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; **144**(5): 646-74.

7. Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 2011; **144**(1): 27-40.

8. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature* 2013; **502**(7471): 333-9.

9. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nature reviews Genetics* 2010; **11**(10): 685-96.

10. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number alteration. *Nature genetics* 2013; **45**(10): 1134-40.

11. Lander ES. Initial impact of the sequencing of the human genome. *Nature* 2011; **470**(7333): 187-97.

12. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008; **456**(7218): 53-9.

13. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 2009; **106**(45): 19096-101.

14. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* 2008; **5**(7): 621-8.

15. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013; **499**(7457): 214-8.

16. Mullighan CG, Phillips LA, Su X, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* 2008; **322**(5906): 1377-80.

17. Yang JJ, Bhojwani D, Yang W, et al. Genome-wide copy number profiling reveals molecular evolution from diagnosis to relapse in childhood acute lymphoblastic leukemia. *Blood* 2008; **112**(10): 4178-83.

18. Ding L, Ley TJ, Larson DE, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012; **481**(7382): 506-10.

19. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England journal of medicine* 2012; **366**(10): 883-92.

20. Travlos GS. Normal structure, function, and histology of the bone marrow. *Toxicologic pathology* 2006; **34**(5): 548-65.

21. Bonnet D. Normal and leukaemic stem cells. *British journal of haematology* 2005; **130**(4): 469-79.

22. Oettinger MA, Schatz DG, Gorka C, Baltimore D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 1990; **248**(4962): 1517-23.

23. Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983; **302**(5909): 575-81.

24. Mullighan CG. Genomic characterization of childhood acute lymphoblastic leukemia. *Seminars in hematology* 2013; **50**(4): 314-24.

25. Papaemmanuil E, Rapado I, Li Y, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nature genetics* 2014; **46**(2): 116-25.

26. Hehlmann R, Hochhaus A, Baccarani M, European L. Chronic myeloid leukaemia. *Lancet* 2007; **370**(9584): 342-50.

27. Bejar R, Levine R, Ebert BL. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011; **29**(5): 504-15.

28. Dohner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *The New England journal of medicine* 2000; **343**(26): 1910-6.

29. Harrison CJ, Haas O, Harbott J, et al. Detection of prognostically relevant genetic abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: recommendations from the Biology and Diagnosis Committee of the International Berlin-Frankfurt-Munster study group. *British journal of haematology* 2010; **151**(2): 132-42.

30. Hjalgrim LL, Rostgaard K, Schmiegelow K, et al. Age- and sex-specific incidence of childhood leukemia by immunophenotype in the Nordic countries. *Journal of the National Cancer Institute* 2003; **95**(20): 1539-44.

31. Pui CH, Evans WE. A 50-year journey to cure childhood acute lymphoblastic leukemia. *Seminars in hematology* 2013; **50**(3): 185-96.

32. Schmiegelow K, Forestier E, Hellebostad M, et al. Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2010; **24**(2): 345-54.

33. Kratz CP, Stanulla M, Cave 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* 2016; **59**(3): 111-5.

34. Ripperger T, Bielack SS, Borkhardt A, et al. Childhood cancer predisposition syndromes-A concise review and recommendations by the Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology. *American journal of medical genetics Part A* 2017.

35. Li Y, Schwab C, Ryan SL, et al. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* 2014; **508**(7494): 98-102.

36. Little MP, Tawn EJ, Tzoulaki I, et al. A systematic review of epidemiological associations between low and moderate doses of ionizing radiation and late cardiovascular effects, and their possible mechanisms. *Radiation research* 2008; **169**(1): 99-109.

37. Greaves MF, Wiemels J. Origins of chromosome translocations in childhood leukaemia. *Nature reviews Cancer* 2003; **3**(9): 639-49.

38. Mori H, Colman SM, Xiao Z, et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proceedings of the National Academy of Sciences of the United States of America* 2002; **99**(12): 8242-7.

39. Kinlen L, Doll R. Population mixing and childhood leukaemia: Fallon and other US clusters. *British journal of cancer* 2004; **91**(1): 1-3.

40. Gilham C, Peto J, Simpson J, et al. Day care in infancy and risk of childhood acute lymphoblastic leukaemia: findings from UK case-control study. *Bmj* 2005; **330**(7503): 1294.

41. Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. *Nature reviews Cancer* 2006; **6**(3): 193-203.

42. Holmfeldt L, Wei L, Diaz-Flores E, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nature genetics* 2013; **45**(3): 242-52.

43. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nature genetics* 2009; **41**(9): 1006-10.

44. Trevino LR, Yang W, French D, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nature genetics* 2009; **41**(9): 1001-5.

45. Perez-Garcia A, Ambesi-Impiombato A, Hadler M, et al. Genetic loss of SH2B3 in acute lymphoblastic leukemia. *Blood* 2013; **122**(14): 2425-32.

46. Shah S, Schrader KA, Waanders E, et al. A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nature genetics* 2013; **45**(10): 1226-31.

47. Pui CH, Mullighan CG, Evans WE, Relling MV. Pediatric acute lymphoblastic leukemia: where are we going and how do we get there? *Blood* 2012; **120**(6): 1165-74.

48. Fuller H. Particulars of a case in which enormous enlargement of the spleen and liver together with dilation of all the blood vessels of the body were found co-incident with a peculiarly altered condition of the blood. *Lancet* 1846; **2**: 43-4.

49. Farber S, Diamond LK. Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid. *The New England journal of medicine* 1948; **238**(23): 787-93.

50. Frei E, Freireich EJ, Gehan E, et al. Studies of Sequential and Combination Antimetabolite Therapy in Acute Leukemia: 6-Mercaptopurine and Methotrexate. *Blood* 1961; **18**(4): 431-54.

51. Bhojwani D, Yang JJ, Pui CH. Biology of childhood acute lymphoblastic leukemia. *Pediatric clinics of North America* 2015; **62**(1): 47-60.

52. Frandsen TL, Heyman M, Abrahamsson J, et al. Complying with the European Clinical Trials directive while surviving the administrative pressure - an alternative approach to toxicity registration in a cancer trial. *European journal of cancer* 2014; **50**(2): 251-9.

53. Toft N, Birgens H, Abrahamsson J, et al. Risk group assignment differs for children and adults 1-45 yr with acute lymphoblastic leukemia treated by the NOPHO ALL-2008 protocol. *European journal of haematology* 2013; **90**(5): 404-12.

54. Borowitz MJ, Devidas M, Hunger SP, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood* 2008; **111**(12): 5477-85.

55. Bjorklund E, Matinlauri I, Tierens A, et al. Quality control of flow cytometry data analysis for evaluation of minimal residual disease in bone marrow from acute leukemia patients during treatment. *Journal of pediatric hematology/oncology* 2009; **31**(6): 406-15.

56. Coustan-Smith E, Sancho J, Hancock ML, et al. Use of peripheral blood instead of bone marrow to monitor residual disease in children with acute lymphoblastic leukemia. *Blood* 2002; **100**(7): 2399-402.

57. Bhojwani D, Pui CH. Relapsed childhood acute lymphoblastic leukaemia. *The lancet oncology* 2013; **14**(6): e205-17.

58. Schultz KR, Pullen DJ, Sather HN, et al. Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG). *Blood* 2007; **109**(3): 926-35.

59. Forestier E, Johansson B, Borgstrom G, Kerndrup G, Johansson J, Heim S. Cytogenetic findings in a population-based series of 787 childhood acute lymphoblastic leukemias from the Nordic countries. The NOPHO Leukemia Cytogenetic Study Group. *European journal of haematology* 2000; **64**(3): 194-200.

60. Moorman AV, Ensor HM, Richards SM, et al. Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial. *The lancet oncology* 2010; **11**(5): 429-38.

61. Paulsson K, Johansson B. High hyperdiploid childhood acute lymphoblastic leukemia. *Genes, chromosomes & cancer* 2009; **48**(8): 637-60.

62. Paulsson K, Forestier E, Lilljebjorn H, et al. Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 2010; **107**(50): 21719-24.

63. Forestier E, Andersen MK, Autio K, et al. Cytogenetic patterns in ETV6/RUNX1-positive pediatric B-cell precursor acute lymphoblastic leukemia: A Nordic series of 245 cases and review of the literature. *Genes, chromosomes & cancer* 2007; **46**(5): 440-50.

64. Smith KS, Jacobs Y, Chang CP, Cleary ML. Chimeric oncoprotein E2a-Pbx1 induces apoptosis of hematopoietic cells by a p53-independent mechanism that is suppressed by Bcl-2. *Oncogene* 1997; **14**(24): 2917-26.

65. Zachariadis V, Gauffin F, Kuchinskaya E, 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 : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2011; **25**(4): 622-8.

66. Zachariadis V, Schoumans J, Barbany G, et al. Homozygous deletions of CDKN2A are present in all dic(9;20)(p13.2;q11.2)-positive B-cell precursor acute lymphoblastic leukaemias and may be important for leukaemic transformation. *British journal of haematology* 2012; **159**(4): 488-91.

67. Harewood L, Robinson H, Harris R, et al. Amplification of AML1 on a duplicated chromosome 21 in acute lymphoblastic leukemia: a study of 20 cases. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2003; **17**(3): 547-53.

68. Robinson HM, Broadfield ZJ, Cheung KL, et al. Amplification of AML1 in acute lymphoblastic leukemia is associated with a poor outcome. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2003; **17**(11): 2249-50.

69. Rand V, Parker H, Russell LJ, et al. Genomic characterization implicates iAMP21 as a likely primary genetic event in childhood B-cell precursor acute lymphoblastic leukemia. *Blood* 2011; **117**(25): 6848-55.

70. Robinson HM, Harrison CJ, Moorman AV, Chudoba I, Strefford JC. Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage-fusion-bridge cycle. *Genes, chromosomes & cancer* 2007; **46**(4): 318-26.

71. Kuchinskaya E, Nordgren A, Heyman M, et al. Tiling-resolution array-CGH reveals the pattern of DNA copy number alterations in acute lymphoblastic leukemia with 21q amplification: the result of telomere dysfunction and breakage/fusion/breakage cycles? *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2007; **21**(6): 1327-30.

72. Sinclair PB, Parker H, An Q, et al. Analysis of a breakpoint cluster reveals insight into the mechanism of intrachromosomal amplification in a lymphoid malignancy. *Human molecular genetics* 2011; **20**(13): 2591-602.

73. Harrison CJ, Moorman AV, Schwab C, et al. An international study of intrachromosomal amplification of chromosome 21 (iAMP21): cytogenetic characterization and outcome. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2014; **28**(5): 1015-21.

74. Moorman AV, Robinson H, Schwab C, et al. Risk-directed treatment intensification significantly reduces the risk of relapse among children and adolescents with acute lymphoblastic leukemia and intrachromosomal amplification of chromosome 21: a comparison of the MRC ALL97/99 and UKALL2003 trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2013; **31**(27): 3389-96.

75. Heerema NA, Carroll AJ, Devidas M, et al. Intrachromosomal amplification of chromosome 21 is associated with inferior outcomes in children with acute lymphoblastic leukemia treated in contemporary standard-risk children's oncology group studies: a report from the children's oncology group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2013; **31**(27): 3397-402.

76. Harrison CJ, Moorman AV, Broadfield ZJ, et al. Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia. *British journal of haematology* 2004; **125**(5): 552-9.

77. Nachman JB, Heerema NA, Sather H, et al. Outcome of treatment in children with hypodiploid acute lymphoblastic leukemia. *Blood* 2007; **110**(4): 1112-5.

78. Safavi S, Forestier E, Golovleva I, et al. Loss of chromosomes is the primary event in near-haploid and low-hypodiploid acute lymphoblastic leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2013; **27**(1): 248-50.

79. Pieters R, Schrappe M, De Lorenzo P, et al. A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial. *Lancet* 2007; **370**(9583): 240-50.

80. Meyer C, Hofmann J, Burmeister T, et al. The MLL recombinome of acute leukemias in 2013. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2013; **27**(11): 2165-76.

81. Iacobucci I, Storlazzi CT, Cilloni D, et al. Identification and molecular characterization of recurrent genomic deletions on 7p12 in the IKZF1 gene in a large cohort of BCR-ABL1-positive acute lymphoblastic leukemia patients: on behalf of Gruppo Italiano Malattie Ematologiche dell'Adulto Acute Leukemia Working Party (GIMEMA AL WP). *Blood* 2009; **114**(10): 2159-67.

82. Pui CH, Carroll WL, Meshinchi S, Arceci RJ. Biology, risk stratification, and therapy of pediatric acute leukemias: an update. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011; **29**(5): 551-65.

83. Kuiper RP, Schoenmakers EF, van Reijmersdal SV, et al. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2007; **21**(6): 1258-66.

84. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007; **446**(7137): 758-64.

85. Strefford JC, Worley H, Barber K, et al. Genome complexity in acute lymphoblastic leukemia is revealed by array-based comparative genomic hybridization. *Oncogene* 2007; **26**(29): 4306-18.

86. Bungaro S, Dell'Orto MC, Zangrando A, et al. Integration of genomic and gene expression data of childhood ALL without known aberrations identifies subgroups with specific genetic hallmarks. *Genes, chromosomes & cancer* 2009; **48**(1): 22-38.

87. Kawamata N, Ogawa S, Zimmermann M, et al. Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* 2008; **111**(2): 776-84.

88. Mullighan CG, Su X, Zhang J, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *The New England journal of medicine* 2009; **360**(5): 470-80.

89. Kuiper RP, Waanders E, van der Velden VH, et al. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2010; **24**(7): 1258-64.

90. Mangum DS, Downie J, Mason CC, et al. VPREB1 deletions occur independent of lambda light chain rearrangement in childhood acute lymphoblastic leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2014; **28**(1): 216-20.

91. Clappier E, Auclerc MF, Rapion J, et al. An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2014; **28**(1): 70-7.

92. Zaliova M, Zimmermannova O, Dorge P, et al. ERG deletion is associated with CD2 and attenuates the negative impact of IKZF1 deletion in childhood acute lymphoblastic leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2014; **28**(1): 182-5.

93. Ma X, Edmonson M, Yergeau D, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nature communications* 2015; **6**: 6604.

94. Mullighan CG, Zhang J, Kasper LH, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature* 2011; **471**(7337): 235-9.

95. Olsson L, Castor A, Behrendtz M, et al. Deletions of IKZF1 and SPRED1 are associated with poor prognosis in a population-based series of pediatric B-cell precursor acute lymphoblastic leukemia diagnosed between 1992 and 2011. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2014; **28**(2): 302-10.

96. Zhang J, Mullighan CG, Harvey RC, et al. Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood* 2011; **118**(11): 3080-7.

97. Malinowska-Ozdowy K, Frech C, Schonegger A, et al. KRAS and CREBBP mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2015.

98. Den Boer ML, van Slegtenhorst M, De Menezes RX, et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *The lancet oncology* 2009; **10**(2): 125-34.

99. Loh ML, Zhang J, Harvey RC, et al. Tyrosine kinome sequencing of pediatric acute lymphoblastic leukemia: a report from the Children's Oncology Group TARGET Project. *Blood* 2013; **121**(3): 485-8.

100. van der Veer A, Waanders E, Pieters R, et al. Independent prognostic value of BCR-ABL1-like signature and IKZF1 deletion, but not high CRLF2 expression, in children with B-cell precursor ALL. *Blood* 2013; **122**(15): 2622-9.

101. Roberts KG, Pei D, Campana D, et al. Outcomes of children with BCR-ABL1-like acute lymphoblastic leukemia treated with risk-directed therapy based on the levels of minimal residual disease. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2014; **32**(27): 3012-20.

102. Roberts KG, Li Y, Payne-Turner D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *The New England journal of medicine* 2014; **371**(11): 1005-15.

103. Roberts KG, Morin RD, Zhang J, et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer cell* 2012; **22**(2): 153-66.

104. Russell LJ, Capasso M, Vater I, et al. Deregulated expression of cytokine receptor gene, CRLF2, is involved in lymphoid transformation in B-cell precursor acute lymphoblastic leukemia. *Blood* 2009; **114**(13): 2688-98.

105. Harvey RC, Mullighan CG, Chen IM, 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* 2010; **115**(26): 5312-21.

106. Lilljebjorn H, Henningsson R, Hyrenius-Wittsten A, et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. *Nature communications* 2016; **7**: 11790.

107. Liu YF, Wang BY, Zhang WN, et al. Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia. *EBioMedicine* 2016; **8**: 173-83.

108. Zhang J, McCastlain K, Yoshihara H, et al. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nature genetics* 2016; **48**(12): 1481-9.

109. Dorge P, Meissner B, Zimmermann M, et al. IKZF1 deletion is an independent predictor of outcome in pediatric acute lymphoblastic leukemia treated according to the ALL-BFM 2000 protocol. *Haematologica* 2013; **98**(3): 428-32.

110. Ofverholm I, Tran AN, Heyman M, et al. Impact of IKZF1 deletions and PAX5 amplifications in pediatric B-cell precursor ALL treated according to NOPHO protocols. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2013; **27**(9): 1936-9.

111. Mullighan CG, Miller CB, Radtke I, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* 2008; **453**(7191): 110-4.

112. Waanders E, van der Velden VH, van der Schoot CE, et al. Integrated use of minimal residual disease classification and IKZF1 alteration status accurately predicts 79% of relapses in pediatric acute lymphoblastic leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2011; **25**(2): 254-8.

113. Virely C, Moulin S, Cobaleda C, et al. Haploinsufficiency of the IKZF1 (IKAROS) tumor suppressor gene cooperates with BCR-ABL in a transgenic model of acute lymphoblastic leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2010; **24**(6): 1200-4.

114. Molnar A, Georgopoulos K. The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. *Molecular and cellular biology* 1994; **14**(12): 8292-303.

115. Collins-Underwood JR, Mullighan CG. Genomic profiling of high-risk acute lymphoblastic leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2010; **24**(10): 1676-85.

116. Tonnelle C, Dijon M, Moreau T, Garulli C, Bardin F, Chabannon C. Stage specific over-expression of the dominant negative Ikaros 6 reveals distinct role of Ikaros throughout human B-cell differentiation. *Molecular immunology* 2009; **46**(8-9): 1736-43.

117. Qazi S, Uckun FM. Incidence and biological significance of IKZF1/Ikaros gene deletions in pediatric Philadelphia chromosome negative and Philadelphia chromosome positive B-cell precursor acute lymphoblastic leukemia. *Haematologica* 2013; **98**(12): e151-2.

118. Chen IM, Harvey RC, Mullighan CG, et al. Outcome modeling with CRLF2, IKZF1, JAK, and minimal residual disease in pediatric acute lymphoblastic leukemia: a Children's Oncology Group study. *Blood* 2012; **119**(15): 3512-22.

119. Palmi C, Lana T, Silvestri D, et al. Impact of IKZF1 deletions on IKZF1 expression and outcome in Philadelphia chromosome negative childhood BCP-ALL. Reply to "incidence and biological significance of IKZF1/Ikaros gene deletions in pediatric Philadelphia chromosome negative and Philadelphia chromosome positive B-cell precursor acute lymphoblastic leukemia". *Haematologica* 2013; **98**(12): e164-5.

120. Volejnikova J, Mejstrikova E, Dorge P, et al. Ikaros (IKZF1) alterations and minimal residual disease at day 15 assessed by flow cytometry predict prognosis of childhood BCR/ABL-negative acute lymphoblastic leukemia. *Pediatric blood & cancer* 2013; **60**(3): 420-7.

121. Nordlund J, Backlin CL, Wahlberg P, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome biology* 2013; **14**(9): r105.

122. Neph S, Kuehn MS, Reynolds AP, et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* 2012; **28**(14): 1919-20.

123. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 2013; **14**(2): 178-92.

124. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**(14): 1754-60.

125. Talkowski ME, Ordulu Z, Pillalamarri V, et al. Clinical diagnosis by whole-genome sequencing of a prenatal sample. *The New England journal of medicine* 2012; **367**(23): 2226-32.

126. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; **29**(1): 15-21.

127. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology* 2010; **11**(3): R25.

128. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* 2014; **15**(2): R29.

129. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 2015; **43**(7): e47.

130. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* 2011; **43**(5): 491-8.

131. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010; **26**(16): 2069-70.

132. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS computational biology* 2013; **9**(7): e1003153.

133. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* 2014; **46**(3): 310-5.

134. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nature biotechnology* 2011; **29**(1): 24-6.

135. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research* 2007; **35**(Web Server issue): W71-4.

136. Schwab CJ, Chilton L, Morrison H, et al. Genes commonly deleted in childhood B-cell precursor acute lymphoblastic leukemia: association with cytogenetics and clinical features. *Haematologica* 2013; **98**(7): 1081-8.

137. Familiades J, Bousquet M, Lafage-Pochitaloff M, et al. PAX5 mutations occur frequently in adult B-cell progenitor acute lymphoblastic leukemia and PAX5 haploinsufficiency is associated with BCR-ABL1 and TCF3-PBX1 fusion genes: a GRAALL study. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2009; **23**(11): 1989-98.

138. Schwab CJ, Jones LR, Morrison H, et al. Evaluation of multiplex ligation-dependent probe amplification as a method for the detection of copy number abnormalities in B-cell precursor acute lymphoblastic leukemia. *Genes, chromosomes & cancer* 2010; **49**(12): 1104-13.

139. Willman CL, Sever CE, Pallavicini MG, et al. Deletion of IRF-1, mapping to chromosome 5q31.1, in human leukemia and preleukemic myelodysplasia. *Science* 1993; **259**(5097): 968-71.

140. Ogawa R, Streiff MB, Bugayenko A, Kato GJ. Inhibition of PDE4 phosphodiesterase activity induces growth suppression, apoptosis, glucocorticoid sensitivity, p53, and p21(WAF1/CIP1) proteins in human acute lymphoblastic leukemia cells. *Blood* 2002; **99**(9): 3390-7.

141. Yang JJ, Cheng C, Devidas M, et al. Genome-wide association study identifies germline polymorphisms associated with relapse of childhood acute lymphoblastic leukemia. *Blood* 2012; **120**(20): 4197-204.

142. Huang J, Gong Z, Ghosal G, Chen J. SOSS complexes participate in the maintenance of genomic stability. *Molecular cell* 2009; **35**(3): 384-93.

143. Li Y, Bolderson E, Kumar R, et al. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *The Journal of biological chemistry* 2009; **284**(35): 23525-31.

144. Lee P, Bhansali R, Izraeli S, Hijiya N, Crispino JD. The biology, pathogenesis and clinical aspects of acute lymphoblastic leukemia in children with Down syndrome. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* 2016; **30**(9): 1816-23.

145. Letourneau A, Santoni FA, Bonilla X, et al. Domains of genome-wide gene expression dysregulation in Down's syndrome. *Nature* 2014; **508**(7496): 345-50.

146. Malinge S, Bliss-Moreau M, Kirsammer G, et al. Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome. *The Journal of clinical investigation* 2012; **122**(3): 948-62.

147. Thompson BJ, Bhansali R, Diebold L, et al. DYRK1A controls the transition from proliferation to quiescence during lymphoid development by destabilizing Cyclin D3. *The Journal of experimental medicine* 2015; **212**(6): 953-70.

148. Boichuk S, Parry JA, Makielski KR, et al. The DREAM complex mediates GIST cell quiescence and is a novel therapeutic target to enhance imatinib-induced apoptosis. *Cancer research* 2013; **73**(16): 5120-9.

149. de Tayrac M, Saikali S, Aubry M, et al. Prognostic significance of EDN/RB, HJURP, p60/CAF-1 and PDLI4, four new markers in high-grade gliomas. *PloS one* 2013; **8**(9): e73332.

150. Staibano S, Mascolo M, Mancini FP, et al. Overexpression of chromatin assembly factor-1 (CAF-1) p60 is predictive of adverse behaviour of prostatic cancer. *Histopathology* 2009; **54**(5): 580-9.

151. Staibano S, Mascolo M, Rocco A, et al. The proliferation marker Chromatin Assembly Factor-1 is of clinical value in predicting the biological behaviour of salivary gland tumours. *Oncology reports* 2011; **25**(1): 13-22.

152. Polo SE, Theocharis SE, Grandin L, et al. Clinical significance and prognostic value of chromatin assembly factor-1 overexpression in human solid tumours. *Histopathology* 2010; **57**(5): 716-24.

153. Polo SE, Theocharis SE, Klijanienko J, et al. Chromatin assembly factor-1, a marker of clinical value to distinguish quiescent from proliferating cells. *Cancer research* 2004; **64**(7): 2371-81.

154. Kim JH, Baddoo MC, Park EY, et al. SON and Its Alternatively Spliced Isoforms Control MLL Complex-Mediated H3K4me3 and Transcription of Leukemia-Associated Genes. *Molecular cell* 2016; **61**(6): 859-73.

155. Ahn EE, Higashi T, Yan M, et al. SON protein regulates GATA-2 through transcriptional control of the microRNA 23a~27a~24-2 cluster. *The Journal of biological chemistry* 2013; **288**(8): 5381-8.