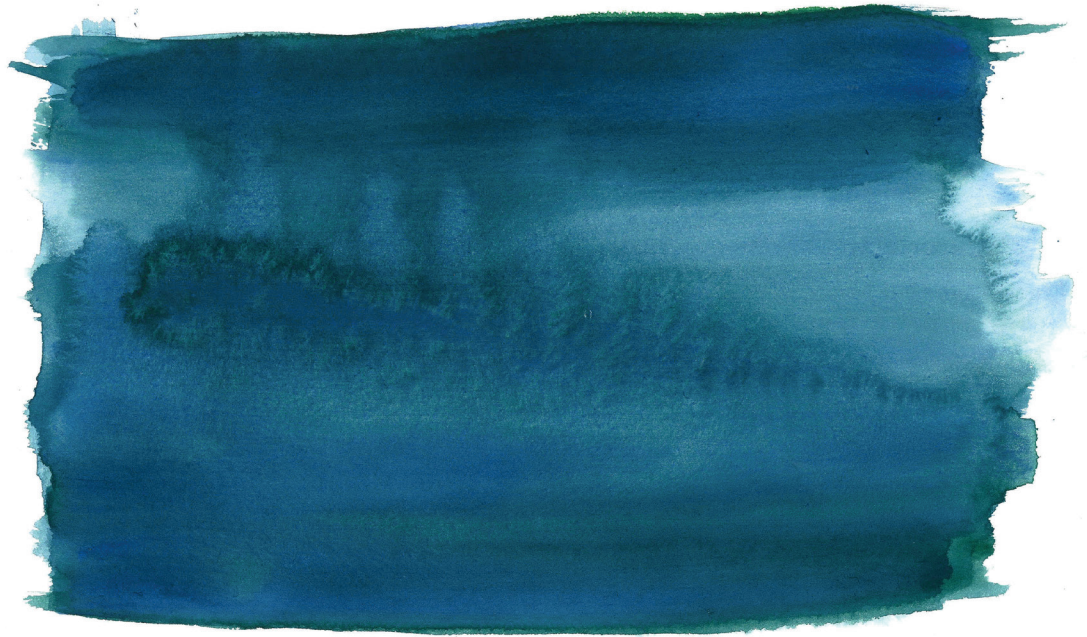


Molecular studies of prognostic and etiological factors in childhood leukemia



Vasilios Zachariadis



**Karolinska
Institutet**

From the DEPARTMENT OF MOLECULAR MEDICINE AND
SURGERY

Karolinska Institutet, Stockholm, Sweden

MOLECULAR STUDIES OF PROGNOSTIC AND ETIOLOGICAL FACTORS IN CHILDHOOD LEUKEMIA

Vasilios Zachariadis



**Karolinska
Institutet**

Stockholm 2015

Cover art by Ulrika Runius

Other artwork produced in collaboration with Alexis Zachariadis

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by AJ E-Print AB

© Vasilios Zachariadis, 2015

ISBN 978-91-7549-918-5

MOLECULAR STUDIES OF PROGNOSTIC AND ETIOLOGICAL FACTORS IN CHILDHOOD LEUKEMIA

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Vasilios Zachariadis

Principal Supervisor:

Associate professor Ann Nordgren
Karolinska Institutet
Department of Molecular Medicine and Surgery

Opponent:

Associate professor Roland Kuiper
Radboud University Nijmegen Medical Centre
Department of Human Genetics

Co-supervisor(s):

Professor Magnus Nordenskjöld
Karolinska Institutet
Department of Molecular Medicine and Surgery

Examination Board:

Professor Björn Andersson
Karolinska Institutet
Department of Cell and Molecular Biology

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

Dr Per-Erik Sandström
Umeå University
Department of Clinical Sciences

Associate professor Erik Forestier
Umeå University
Department of Medical Biosciences

Associate professor Dawei Xu
Karolinska Institutet
Department of Medicine

ABSTRACT

Acute lymphoblastic leukemia (ALL) is the most common cancer in childhood. It is caused by the somatic acquisition of genetic abnormalities and malignant transformation of immature lymphocytes in the bone marrow, most commonly of B-cell lineage. Chromosomal translocations are a hallmark of childhood ALL, constituting different subtypes of disease in terms of clinical characteristics and treatment outcomes. More than that, these gross chromosomal changes are often directly linked to specific disruptions at the molecular level through resulting fusion genes, aberrant transcriptional activation or associated structural and single nucleotide variants.

The aim of this thesis was to establish the frequency and prognostic impact associated with the chromosomal abnormality $\text{dic}(9;20)(\text{p}13.2;\text{q}11.2)$ in childhood B-cell precursor (BCP) ALL, and to better our understanding of the genetic basis underlying this disease. Thereby, we aimed to improve the diagnostics and risk-stratification in the context of existing anti-leukemic treatments, while potentially highlighting new rational strategies of therapy in $\text{dic}(9;20)$ and childhood ALL in general.

In **paper I** we found that the $\text{dic}(9;20)$ was present in almost five percent of BCP ALL cases, making it the third most common subgroup in the cohort. Furthermore, we showed that $\text{dic}(9;20)$ -positive cases treated on the NOPHO ALL-2000 protocol had a lower event-free survival than the most common subtypes of ALL. In **paper III**, we designed and validated a method for the detection of $\text{dic}(9;20)$ in a clinical setting using FISH. In **papers II and IV**, we characterized the genetic basis of disease in cases carrying the $\text{dic}(9;20)$, discovering first that homozygous deletions of tumor suppressor *CDKN2A* are present in almost all cases, but that the heterogeneity of the translocation breakpoints did not support the consistent formation of a fusion gene. Further, in **paper IV**, through the application of multiple genome-wide techniques, we presented a full spectrum of acquired structural and sequence level variation in $\text{dic}(9;20)$ -positive ALL, as well as the integrated analysis of DNA methylation, gene expression and anti-leukemic drug sensitivity. Together, these data revealed a genetic profile distinct from that of other ALL subtypes, not accounted for by individual fusion genes or single gene abnormalities alone. Importantly, we found evidence of altered expression of several key genes governing cell survival and programmed cell death, attributable to changes in promoter DNA methylation; some affecting the response to existing anti-leukemic agents, and others highlighting specific pathways that may be of value in developing new therapies.

Together, these studies add to our understanding of the clinical relevance and underlying biology of $\text{dic}(9;20)$ -positive BCP ALL and provide a basis for the rational exploration of new treatment options for children with this disease.

PUBLICATIONS

This thesis is based on the following papers, referred to by roman numerals in text.

- I. **Zachariadis V**, Gauffin F, Kuchinskaya E, Heyman M, Schoumans J, Blennow E, Gustafsson B, Barbany G, Golovleva I, Ehrencrona H, Cavelier L, Palmqvist L, Lönnerholm G, Nordenskjöld M, Johansson B, Forestier E, and Nordgren A. (2011). The frequency and prognostic impact of dic(9;20)(p13.2;q11.2) in childhood B-cell precursor acute lymphoblastic leukemia: results from the NOPHO ALL-2000 trial. *Leukemia* 25, 622–628
- II. **Zachariadis V**, Schoumans J, Barbany G, Heyman M, Forestier E, Johansson B, Nordenskjöld M, and Nordgren A. (2012). 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* 159, 488–491
- III. **Zachariadis V**, Schoumans J, Öfverholm I, Barbany G, Halvardsson E, Forestier E, Johansson B, Nordenskjöld M, and Nordgren A. (2014). 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* 28, 196–198
- IV. **Zachariadis V**, Nordlund J, Taylan F, Tran A-N, Öfverholm I, Tesi B, Dahlberg J, Saft L, Heyman M, Pokrovskaja K, Grandér D, Nilsson D, Vezzi F, Nordenskjöld M, Lönnerholm G, Forestier E, Barbany G, Syvänen A-C, and Nordgren A. Genetic characterization of dic(9;20)-positive B-cell precursor acute lymphoblastic leukemia. *Manuscript*

RELATED PUBLICATIONS

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

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

- VII. Olsson L, Ivanov Öfverholm I, Norén-Nyström U, **Zachariadis V**, Nordlund J, Sjögren H, Golovleva I, Nordgren A, Paulsson K, Heyman M, Barbany G, and Johansson B. The clinical impact of *IKZF1* deletions in paediatric B-cell precursor acute lymphoblastic leukaemia is independent of minimal residual disease stratification in NOPHO treatment protocols used between 1992 and 2013. *British Journal of Haematology*. In press

CONTENTS

1	Introduction.....	3
1.1	Cancer is a disease of the genome.....	3
1.2	Acute lymphoblastic leukemia.....	3
1.2.1	Treatment and risk stratification of childhood ALL.....	4
1.2.2	Etiology of childhood ALL.....	5
1.2.3	Cytogenetic subtypes in BCP ALL are important predictors of outcome.....	7
1.2.4	Childhood BCP ALL in the post-genomic era.....	9
1.2.5	The dic(9;20)(p13.2;q11.2) in childhood BCP ALL.....	10
2	Aims of the thesis.....	12
3	Overview of studies.....	13
4	Materials and methods.....	14
4.1	Patients and clinical data.....	14
4.1.1	Patient cohort, paper I.....	14
4.1.2	Patient cohort, paper IV.....	15
4.2	Cytogenetic analyses.....	15
4.2.1	G-banding and metaphase FISH analysis.....	15
4.2.2	Design of FISH probe kit.....	16
4.3	Statistical analyses.....	17
4.4	Genomic analyses.....	18
4.4.1	Detection of copy number alterations.....	18
4.4.2	Whole genome sequencing.....	19
4.4.3	RNA sequencing.....	19
4.4.4	DNA methylation analysis.....	21
4.4.5	<i>In vitro</i> drug sensitivity assay.....	21
5	Results and discussion.....	22
5.1	Frequency and clinical characteristics of dic(9;20) BCP ALL.....	22
5.2	Detecting the dic(9;20) rearrangement.....	23
5.3	Treatment outcome.....	26
5.4	Breakpoint mapping and fusion gene detection.....	27
5.5	Secondary genetic aberrations associated with dic(9;20).....	29
5.6	Integrated genomic profiling of dic(9;20)-positive cells.....	31
6	Conclusions.....	34
7	Future perspectives.....	35
8	Acknowledgements.....	36
9	References.....	37

LIST OF ABBREVIATIONS

ALL	Acute Lymphoblastic Leukemia
AMKL	Acute Megakaryoblastic Leukemia
AT	Ataxia Telangiectasia
BAF	B-allele frequency
BCP ALL	B-cell precursor acute lymphoblastic leukemia
BM	Bone marrow
CEP9	Centromere of chromosome 9
CEPH	Utah residents with ancestry from Northern and Western Europe
CpG	CG dinucleotide
chr	Chromosome
CNA	Copy number alteration
dic(9;20)	Dicentric rearrangement between chromosomes 9 and 20
DS	Down Syndrome
EFS	Event-free survival
FISH	Fluorescence <i>in situ</i> hybridization
FMCA	Fluorometric microculture cytotoxicity assay
G-banding	Giemsa banding
GWAS	Genome-wide association study
HeH	High hyperdiploidy
HR	High risk
iAMP21	Intrachromosomal amplification of chromosome 21
IR	Intermediate risk
LRR	Log R ratio
MRD	Minimal residual disease
NOPHO	Nordic Society of Pediatric Hematology and Oncology
OMIM	Online Mendelian Inheritance in Man
OS	Overall survival
PCR	Polymerase chain reaction

RAG	Recombination-activating gene
RNA-seq	RNA sequencing
RT-PCR	Reverse transcriptase polymerase chain reaction
SR	Standard risk
WHO	World Health Organization

PREFACE

Cancer is a disease of the genome. It is the somatic acquisition of mutations, sometimes acting in concert with inherited genetic variation that initiates and drives the development of malignant disease. The sum of these aberrations often comes to define the clinical course of disease and response to therapy. Resolving their constitution and origin are therefore key to understanding the basis of cancer and ensure adequate treatment for all patients.

Today, more than 90 percent of children diagnosed with acute lymphoblastic leukemia (ALL) are cured with contemporary treatment. This is the result of careful, iterative adjustments to multi-agent chemotherapy protocols over the course of decades. While the treatment of childhood ALL is one of the foremost success stories in cancer therapy, a cure for all children remains elusive. Also considering today's survivors there is little reason to rest. While the treatment of childhood ALL in many ways is a true endeavor of personalized medicine, guided in large part by both patient and disease characteristics, it remains one of desperate imprecision. To cure all children diagnosed with ALL, avoid treatment-related toxicities, and identify factors predisposing to leukemia development, we need to understand the genetic aberrations that initiate and maintain leukemic disease.

This thesis summarizes a series of studies on patients with childhood ALL carrying a specific chromosomal rearrangement, namely $\text{dic}(9;20)$. The first part provides a general introduction to the field of cancer genetics and childhood ALL. The next part summarizes and discusses the results of our studies on $\text{dic}(9;20)$ -positive ALL. Lastly, included are the four original research studies on which this thesis is based.

Stockholm, April 18th 2015

A handwritten signature in black ink, appearing to read 'V Zachariadis', with a long horizontal flourish extending to the right.

Vasilios Zachariadis

1 INTRODUCTION

Over a decade has now passed since the initial sequencing and successful draft assembly of the human genome.^{1,2} Since then, primarily through feats of engineering and technical developments outside the field of biology, we have quickly gained the ability to read genetic sequence at a large scale, in different forms and in greater detail.³ While these advances have arguably only begun to impact the clinical practice of medicine as a whole, some areas of medical research have rapidly transformed as a result. In short, we have gained the ability to explore more of our genome, its expression,⁴ and its regulatory components⁵ in a larger number of patients. In particular, the areas of rare genetic disease and cancer have seen rapid gains. We now have the ability to more accurately diagnose patients and catalog a large amount of germline and somatic variation directly linked to the pathogenesis of disease.⁶⁻⁸

1.1 CANCER IS A DISEASE OF THE GENOME

While only a fraction of cases can be attributed to inherited genetic variation, all cancer results from the acquisition of somatic genetic variation, conferring proliferative and other selective advantages over neighboring cells.⁹ These events can occur in sequence over a longer period of time,¹⁰ or in a handful of catastrophic cellular events.¹¹ This does not imply that, by being able to read the genetic changes in a cancer cell, will we gain a complete understanding of its biology. It does, however, present a new opportunity to study individual cancer types at scale, looking for recurrent mutations driving disease, some of them already the target of approved drugs and others possible future molecular targets.¹² Sequencing tumor material across histologically different cancers can also reveal previously unrecognized principles of cancer biology.¹³⁻¹⁵

Just as different cancers, historically defined by their tissue of origin, can be very similar at the genomic level, there is mounting evidence for the presence of significant genetic heterogeneity also within an individual tumor or population of cancer cells.¹⁶⁻¹⁸ This adds another layer of complexity, both in solid tumors when selecting therapy to prevent metastatic disease,¹⁸ and in hematological malignancies to avoid the emergence of treatment-resistant clones.¹⁷ In particular, this phenomenon stipulates that rational treatments, guided by specific targets rather than intensity alone, are a necessity as the added selective pressure that is chemotherapy may in some cases only serve to worsen the course of the disease.¹⁶

1.2 ACUTE LYMPHOBLASTIC LEUKEMIA

ALL is a malignant disease of the bone marrow, characterized by a block in lymphoid differentiation and the rapid clonal expansion of immature, non-functioning immune cells of T- or B-cell lineage – termed lymphoblasts.^{19,20} This population of cells quickly overwhelms the normal bone marrow compartment, suppressing other lineages of the hematopoietic system. This gives rise to the common symptoms of acute leukemia including pallor, easy bruising, bleeding, fevers and infections.

ALL is the most common cancer affecting children worldwide.²¹ In Sweden this corresponds to between 60-80 children per year, and to about 200 in the Nordic countries as a whole.^{21,22} In the USA the equivalent number is 3000 new cases per year.²³ In most countries around the world, the incidence is remarkably consistent of 3-4 cases per 100,000 individuals, and one that has remained without significant change over the last decades.^{21,24}

1.2.1 Treatment and risk stratification of childhood ALL

Ever since Sidney Farber and colleagues first reported transient remissions in children with acute leukemia following treatment with aminopterin,²⁵ an antifolate, the development of new cytotoxic drugs and their efficient application has mirrored incremental gains in the treatment of childhood ALL (figure 1).²⁶

While Farber's approach was an empirical one, and raised controversy at first, it was soon clear that the rational evaluation of a few chemical compounds, like aminopterin, could bring great treatment gains to cancers of different types, and to childhood leukemias in particular.²⁷ What was also clear to researchers and clinicians already at the time, was that to rationally evaluate and compare individual chemotherapy drugs, one had to take into account the particular type of cancer and its developmental stage.²⁷

Less than a decade later, methotrexate was introduced (known as amethopterin at the time) and by early 1960, its combination with 6-mercaptopurine induced sustained remissions in 20 percent of cases.²⁸ With the addition of cranial irradiation and intrathecal methotrexate to previous combination treatments, the concept of being able to cure many children with leukemia was solidified.²⁹ The following decades saw rapid improvements in outcome across treatment protocols in many countries, mainly through the intensification of therapy, but also the addition of new drugs such as asparaginase.²⁶ By the 1980s, the structure of contemporary chemotherapy treatment was firmly established; induction of remission, intensification/consolidation, and continuation or maintenance therapy. At the same time, pediatric oncologists in the Nordic countries, like elsewhere in the world, organized to establish common protocols, unifying diagnostics and treatment across the five Nordic countries, starting with the NOPHO ALL-92 protocol.^{22,30}

Beginning with the NOPHO ALL-2008 protocol, cranial irradiation was omitted from the treatment of childhood ALL. It also established the standardized use of early minimal residual disease (MRD) monitoring and stratification by cytogenetic markers.

In 2001, Druker *et al* described the complete hematological response in 53 of 54 chronic myeloid leukemia (CML) patients and significant activity in Ph+ blast crisis, all treated with a specific BCR-ABL1 inhibitor.^{31,32} It proved that targeting specific molecular aberrations in cancer was not only feasible but potentially transformative for cancer care. While the effect of Imatinib and subsequent iterations was not as dramatic in Ph+ ALLs, the addition of Imatinib to existing protocols in children with Ph+ ALL has proved beneficial.³³ Importantly, the successful targeting of the BCR-ABL1 fusion protein has spurred the pursuit to

understand the molecular outcome also of other abnormalities in ALL and other cancers, with the aim of developing targeted treatments.

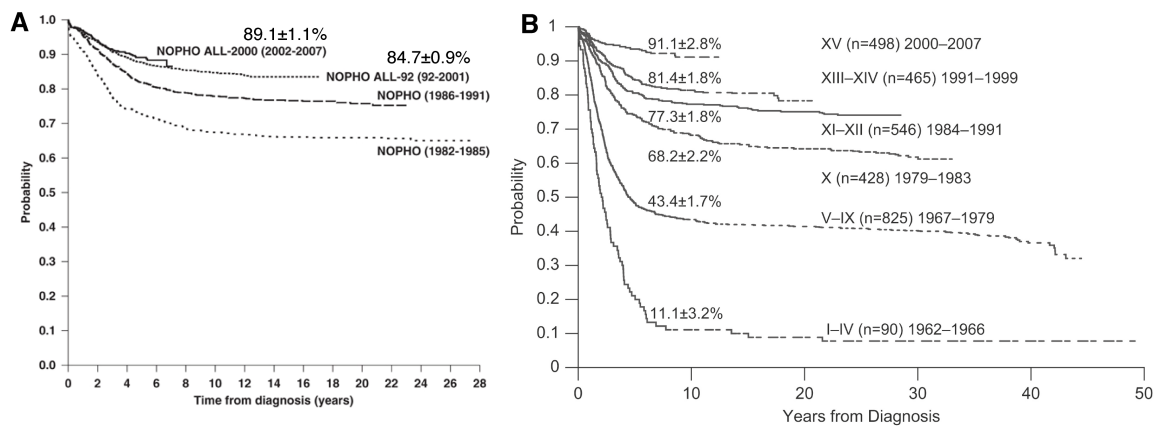


Figure 1: Overall survival over consecutive protocols in the Nordic countries and at the StJude Children's Research Hospital. A) Consecutive NOPHO ALL (1982-2007) protocols, with 2668 children enrolled in the ALL-92 and ALL-2000 protocols. B) Consecutive StJude treatment protocols (1962-2007), including a total of 2852 children. Adapted and reprinted from Schmiegelow et al³⁰ and Pui and Evans²⁶, with permission from the respective publishers.

1.2.2 Etiology of childhood ALL

With few exceptions, the etiology of childhood ALL is not known. Ionizing radiation, prior chemotherapy or Down Syndrome and other genetic disorders are known to confer an increased risk, but together only explain a fraction of cases.^{19,24}

Assuming a model of cancer development with the sequential acquisition of somatic variants, the cumulative risk of developing cancer increases with age and may in large part be attributed to the rate of stem cell turnover in the tissue of origin.^{34,35} It follows that manifestations of cancer early in life should be different from those presenting with increased frequency towards the end of life.

Indeed, from a genetic perspective, childhood cancer is by all accounts a special case of cancer biology, as evidenced by the consistently low number of protein altering mutations detected compared to adult cancers.³⁶ As more studies leverage whole exome or whole genome sequencing in the evaluation of somatic mutations in childhood cancer, it is also clear that many of them will be carriers of known cancer-predisposing gene variants in their germline.³⁷ Furthermore, even in well-characterized subtypes of BCP ALL it is revealed that many patients may develop leukemia as a manifestation of a common underlying predisposition. This is the case with low hypodiploid ALL (32-39 chromosomes) and germline *TP53* mutations, the cause of Li Fraumeni syndrome (OMIM#151623).³⁸

1.2.2.1 Genetic syndromes and childhood ALL

Some genetic disorders such as Down syndrome (DS), are not only predisposing to leukemia in childhood, but to specific types of leukemia such as acute megakaryoblastic leukemia.^{39,40} Indeed, even within ALL patients with DS, they present with high frequencies of very specific genetic lesions, particularly JAK-STAT activating *JAK2* mutations and *CRLF2* fusions.^{41,42}

There are a host of other genetic disorders, like Noonan syndrome (OMIM#163950), Fanconi anemia (OMIM #227650), Bloom syndrome (OMIM #210900) and ataxia-telangiectasia (AT; OMIM #208900) that further highlight the connection between leukemia development and predisposing mutations.⁴³ These syndromes are all associated with an increased risk of acquiring somatic lesions, either through an overall propensity for chromosomal breakage (AT, Fanconi) or more subtly acting gain-of-function sequence mutations (Noonan). This highlights common genes and signaling pathways being causative of the germline disorder, as well as being the frequent targets of acquired aberrations in ALL and other childhood cancers.⁴⁴

1.2.2.2 Rare variants and familial predisposition to childhood ALL

It would not be surprising then, that an appreciable number of individuals that develop cancer in childhood, including ALL, are carriers of genetic variants predisposing to cancer development.³⁷ Previous epidemiological data in non-syndromic ALL sibships appear to support this, revealing subtype concordance between siblings more often than expected which suggests shared factors underlying leukemia development.⁴⁵

It is only recently, as the scale and scope of genetic analyses has increased, that direct evidence has begun to emerge of the presence of individuals and families, not only in the context of recognized genetic disorders, that are carriers of variants strongly predisposing to childhood ALL.⁴⁶⁻⁴⁹ Interestingly, these studies also indicate that while some associated hematological manifestations are evident throughout life, the risk of developing leukemia for mutations in for example *PAX5* and *ETV6*, may be limited to childhood and adolescence.^{46,50}

1.2.2.3 Common variants associated with the development of childhood ALL and response to therapy

For higher frequency alleles in human populations, genome-wide association studies (GWAS) have been successful in associating germline genetic variation with the predisposition to a whole range of complex and malignant disease.⁵¹ Also in childhood ALL, GWAS studies have succeeded in highlighting genetic variants in a handful of genes that are consistently associated with the risk of developing childhood ALL.⁵²⁻⁵⁸ More often than not, these variants are found within or in close proximity to known leukemia-associated genes, such as *IKZF1*, *CDKN2A* and *CEBPE*. Others like *GATA3*, a key transcription factor in T-cell development, and *TP63*, a p53-family transcription factor, have been linked to particular subtypes of childhood ALL.^{55,57,59} These variants, even in aggregate, can only account for small increases in the risk of developing ALL, and may not be enough to warrant general

screening efforts or genetic counseling. However, they may prove useful in highlighting biological pathways whose role would not otherwise be recognized as part of leukemia development.⁵¹

For the individual patients, common variation in genes involved in the metabolism of anti-leukemic drugs is already of clinical importance. For example, the metabolism of thiopurines, some of the most commonly used drugs in the treatment of ALL, varies significantly between patients as a result of common genetic variation in the thiopurine *S*-methyltransferase (*TPMT*) gene. Carriers of the null gene variant exhibit increased chemosensitivity and a risk of excessive treatment related complications at standard doses.^{60–62} Genotyping or direct measurement of *TPMT* activity is therefore implemented in many protocols today as a means of guiding thiopurine concentrations.⁶² This, and other examples of germline variation influencing chemosensitivity, serve to further highlight the importance of detecting them, in addition to somatically acquired variation, in advancing the care of children with ALL.⁶³

Finally, it seems clear that additional genetic variants of low to medium allele frequencies, both in genes predisposing to disease as well as those influencing treatment, are likely to be discovered as the number of patients analyzed increases the statistical power to do so.⁵¹ It is not evident whether these variants will ever be of value as direct predictors of childhood ALL, or whether their functional impact can be elucidated in detail. Still, the combination of rare and common germline variation, including pharmacogenomic information, can be successfully be added to established predictors of treatment outcome, such as MRD, to introduce toxicity-aware and other rational therapy adjustments.⁶⁴

1.2.3 Cytogenetic subtypes in BCP ALL are important predictors of outcome

In childhood BCP ALL, it is clear that largely non-overlapping subtypes of disease may be identified based on the genetic abnormalities they acquire.^{19,20,65} In the current WHO classification, released in 2008, seven subtypes of BCP ALL are recognized: t(9;22)(q34;q11.2)/*BCR-ABL1*, *MLL*/11q23 translocations, t(12;21)(p13;q22)/*ETV6-RUNX1*, t(1;19)(q23;p13.3)/*TCF3-PBX1*, t(5;14)(q31;q32)/*IGH@-IL3*, high hyperdiploidy and hypodiploidy.⁶⁶

More broadly, subtypes of childhood BCP ALL can be divided into three cytogenetically different groups: a) cases with chromosomal translocations, often resulting in activating gene fusions, such as cases with t(12;21)/*ETV6-RUNX1* or *MLL*-rearrangements; b) cases with non-random loss or gain of whole chromosomes, such as cases of high hyperdiploidy (HeH) or hypodiploid ALL; c) other cases with or without detectable cytogenetic aberrations.⁶⁵

These subtypes constitute clinically and biologically distinct entities of disease, many of them resulting in fusion genes or other associated abnormalities that can directly initiate and/or propagate leukemic disease.²⁰ Most common are cases with HeH and the t(12;21)/*ETV6-RUNX1* rearrangement, together accounting for just over half of all cases of BCP ALL (figure 2). The distribution of is not equal across age groups, however. Within the common age peak of ALL, at 2-5 years of age, the proportion of cases with t(12;21)/*ETV6-RUNX1* or HeH is

very high, while rapidly decreasing with age of diagnosis.⁶⁵ The t(12;21)/*ETV6-RUNX1* is, for example, almost completely absent in adult ALL cases, while aberrations such as t(9;22)/*BCR-ABL1* increase in frequency with age and others, like *MLL* rearrangements, are predominant in very young children.⁶⁵

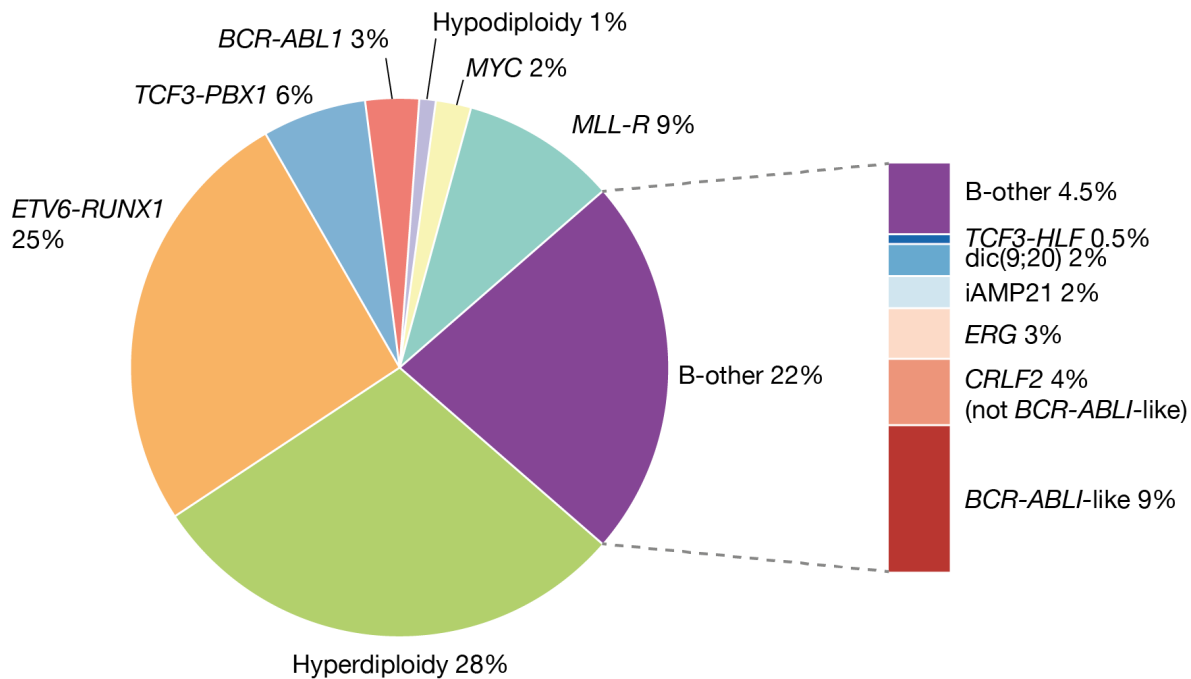


Figure 2: Recognized subtypes of BCP ALL. The pie chart corresponds to the WHO classification of 2008,⁶⁶ not accounting for abnormalities that were discovered or gained recognition after 2004 and the application of genome-wide techniques in childhood ALL. Right bar shows the development since, including subtypes defined by copy number profiling or gene expression alone. Adapted from Bhojwani et al²⁰ with permission from the publisher.

Importantly, specific genetic subtypes are often associated to initial treatment response, risk of relapse while on therapy and overall survival (Figure 3).^{63,65} For this reason, many of them are routinely screened for at diagnosis using targeted cellular or molecular assays, such as fluorescence *in situ* hybridization (FISH) and reverse-transcriptase PCR (RT-PCR). While the precise approach to screening for genetic aberrations differs between laboratories and between treatment protocols, almost all genetic laboratories perform some targeted analysis in addition to traditional karyotyping by G-banding.⁶⁵ The NOPHO ALL-2000 protocol stipulated the targeted analysis of 11q23/*MLL* rearrangement, t(1;19)/*TCF3-PBX1*, t(9;22)/*BCR-ABL1* and t(12;21)/*ETV6-RUNX1*.³⁰ The NOPHO ALL-2008 protocol saw the addition of several new cytogenetic markers to treatment stratification, requiring, for example, the exclusion of dic(9;20) or intrachromosomal amplification of chromosome 21 (iAMP21) by some method, before allocation to standard-risk treatment.³⁰

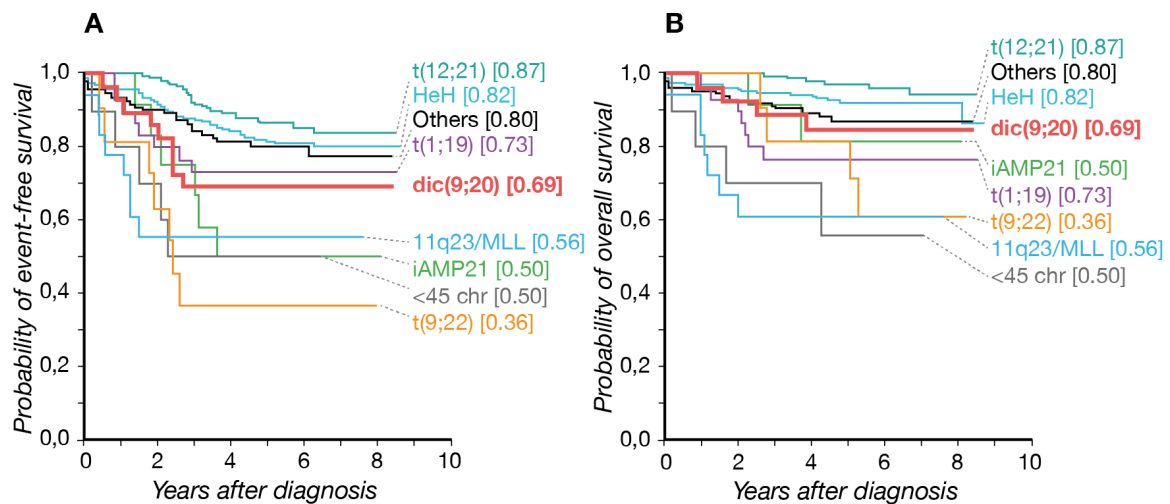


Figure 3: Kaplan-Meier survival curves, showing the probability of (A) event-free and (B) overall survival in 882 BCP ALL cases treated according to the NOPHO ALL-2000 protocol by cytogenetic subgroup. Adapted from figure 2 in paper I.

1.2.4 Childhood BCP ALL in the post-genomic era

The development of genome-wide microarray based techniques, and later, high-throughput sequencing methods, has significantly expanded the range of genetic alterations we know to be contributing to the development of ALL.^{20,67} It was clear from gene expression profiling that the recognized cytogenetic subgroups of disease were also associated to molecular profiles, and importantly, that entirely new groups of patients could be characterized.^{68,69} Genome-wide analyses of copy number alterations in ALL revealed that a number of genes crucial to normal B-cell development and differentiation, like *PAX5* and *IKZF1*, were recurrently lost or otherwise disrupted in high frequencies in BCP ALL.^{70–72} These were most often secondary events to defined cytogenetic abnormalities, contributing to the development and progression of disease. Importantly, some of them, including *IKZF1* and *CREBBP*, have been shown associate with high risk disease and poor treatment outcomes.^{73,74}

From this work, new subtypes of distinct biology and associated clinical outcome have begun to emerge (figure 2, sidebar). In particular, the heterozygous deletion of *IKZF1*, a gene encoding the lymphoid transcription factor Ikaros, has consistently been associated with poor outcome in BCP ALL.^{73,75–78} While deletions of *IKZF1* are enriched in high-risk and cytogenetically non-stratifying ALLs, they can be found across established subtypes of disease, and increases the risk of relapse independent of other established risk-factors, such as MRD.^{75,76,78} This highlights the added value of detecting secondary genetic aberrations, also within established cytogenetic groups.

Genome-wide techniques that assay the somatic state epigenetics of childhood ALL have added an additional layer of information to our understanding of leukemia development. In particular, the genome-wide interrogation of methylation levels at CpG islands, enriched at the site of gene promoters, has revealed that different cytogenetic subtypes are highly concordant also in their patterns of DNA methylation.^{79,80} In fact, analysis of diagnostic

samples by DNA methylation alone will accurately classify them by subtype, and often describe ALLs in which cytogenetics could not be formed or was inconclusive.⁸¹ DNA methylation can thereby be of value also as a prospective diagnostic tool in ALL.

1.2.5 The dic(9;20)(p13.2;q11.2) in childhood BCP ALL

The dic(9;20) was first recognized as a non-random chromosomal abnormality in 1995, with case series reporting its recurrence in both adults and children with BCP ALL.^{82,83} Over the next decade, a number of reports brought the total number of cases to 60, and it was now evident that the dic(9;20) abnormality occurred preferentially in childhood BCP ALL, accounting for over 90 percent of cases.⁸⁴⁻⁹⁰ It occurs most often within the peak age range of childhood ALL, between 2-5 years, and has consistently been found to occur with a female predominance.^{85,90}

1.2.5.1 Detection and molecular consequences of the dic(9;20) rearrangement

The dic(9;20) results in loss of the short arm of chromosome 9 and the long arm of chromosome 20, joining 9q and 20p with the respective centromeres in between (figure 4). In most cases with dic(9;20), this can render as monosomy 20, and was previously considered a good indicator for when to suspect an otherwise subtle dic(9;20)-rearrangement.⁸⁵ However, it would be easy to mistake it for monosomy 20 alone, or in combination with deletion 9p, a common event in BCP ALL. Therefore, many cases go undetected using G-banding techniques alone.⁹⁰

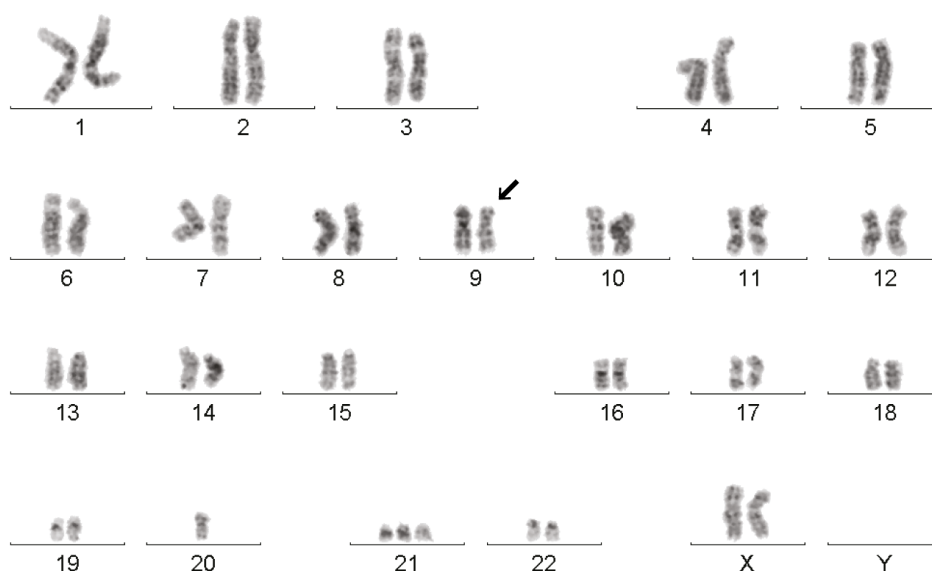


Figure 4: Karyotype visualized by Giemsa staining (G-banding) in a female with 46,XX,dic(9;20),+21. A black arrow indicates the subtle abnormality that is dic(9;20).

Using array comparative genomic hybridization (aCGH) in seven cases with dic(9;20), the breakpoints on chr 9 and 20 were shown to cluster within 9p13.2 and 20q11.2, respectively.⁸⁹ However, both aCGH and higher resolution SNP array assays revealed that the breakpoints,

while clustered, were not identical and likely did not form a consistent fusion gene as a result of the dic(9;20) rearrangement.^{89,91} Still, some dic(9;20)-positive cases do form fusion genes involving *PAX5* on 9p13.2 and a handful of partners on 20q, as a result of the rearrangement.⁹²⁻⁹⁴ Other cases are found to harbor *PAX5* sequence mutations, or additional structural variants.^{38,71,93} Thus, the suggestion is that disruption of *PAX5*, through different mechanisms, may be the primary functional outcome of the dic(9;20) rearrangement.⁹²

1.2.5.2 Subtype identity and outcome in dic(9;20)-positive BCP ALL

To date, over 160 cases of childhood BCP ALL cases with dic(9;20) have been reported, including many presented in this thesis. In about half of the cases, dic(9;20) is the only chromosomal abnormality.⁹⁵ While dic(9;20) is detected at comparable frequencies across major ALL study groups, with comparable treatment outcomes,^{20,85,90,96,97} it has only served as a risk-stratifying marker in the NOPHO ALL-2008 protocol.³⁰

2 AIMS OF THE THESIS

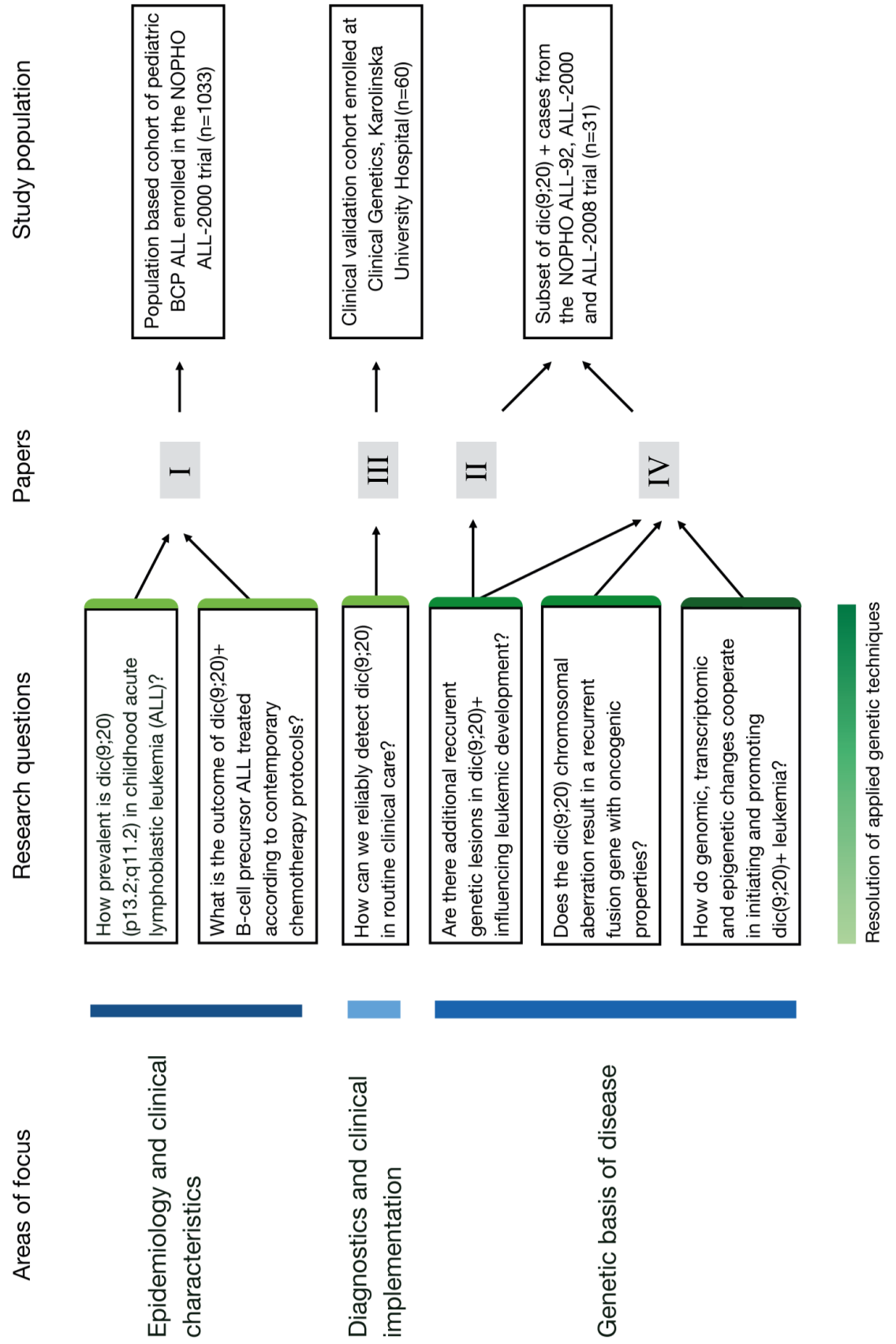
The overall aim of this thesis was to characterize genetic abnormalities of prognostic significance in childhood ALL and develop genetic tests for use in the routine laboratory, thereby advancing the care of children with leukemia.

Specifically, our aim was to:

- Determine the frequency and prognostic impact of the dic(9;20) rearrangement
- Develop better methods of detecting dic(9;20) BCP ALL in a clinical setting
- Understand the molecular consequences of the dic(9;20) rearrangement

3 OVERVIEW OF STUDIES

Molecular studies of prognostic and etiological factors in childhood leukemia

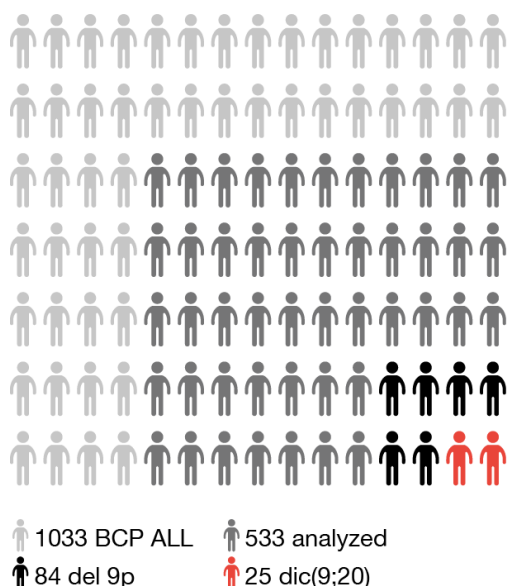


4 MATERIALS AND METHODS

4.1 PATIENTS AND CLINICAL DATA

4.1.1 Patient cohort, paper I

FISH analysis



Survival analysis



Figure 5: Schematic overview of cohorts for FISH and survival analysis in **paper I**, treated according to the NOPHO ALL-2000 protocol

Between January 1, 2001 and December 31, 2006, 1174 infants, children, and adolescents <18 years were diagnosed with ALL in the Nordic countries (Denmark, Finland, Iceland, Norway and Sweden). Among these, 1033 (88%) were BCP ALL, of which 882 (85%) were treated according to the NOPHO ALL 2000 protocol. Risk stratification divided patients into standard intensity (SR: WBC $\geq 10 \times 10^9/l$, 1-9.9 years, no high risk (HR) features), intermediate intensity (IR: WBC $> 10 \times 10^9/l$, age > 10 years, no HR features) and three HR groups (intensive, very intensive, extra intensive). HR features were WBC $\geq 100 \times 10^9/l$, 11q23/MLL rearrangement, t(9;22)(q34;q11), t(1;19)(q23;p13), and hypodiploidy (<45 chromosomes). The dic(9;20) was not a risk-stratifying aberration in the NOPHO ALL-2000 protocol.³⁰ Treatment intensity for patients with dic(9;20)-positive ALL was made solely based on age, WBC count, the presence of extra-medullary leukemia and morphologic response during induction therapy. Karyotyping and targeted analyses was done within the NOPHO ALL-2000 protocol for: 11q23/MLL rearrangement, t(1;19)(q23;p13)/TCF3-PBX1, t(9;22)(q34;q11)/BCR-ABL1 and t(12;21)(p13;q22)/ETV6-RUNX1. Diagnostic BM smears from 533 (52%) of the 1033 BCP ALL patients were successfully analyzed using interphase FISH. In the 491 BCP ALLs, which were not analyzed, dic(9;20)-positive cases were ascertained using G-banding and metaphase FISH analyses only.

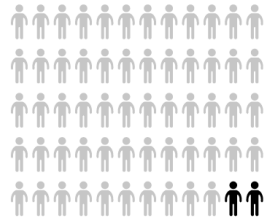
4.1.2 Patient cohort, paper IV

Additional cohorts and analyses

FMCA analysis (n=546)

532 BCP ALL

14 dic(9;20)

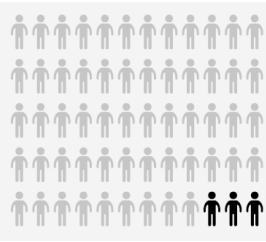
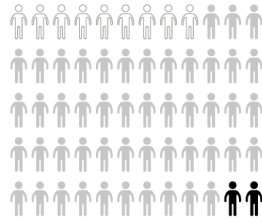


DNA methylation (n=900)

136 controls

24 dic(9;20)

740 BCP/T-ALL



External cohort (n=118)

112 BCP ALL

6 dic(9;20)

RNA-seq analysis (n=46)

42 BCP ALL

14 dic(9;20)



dic(9;20)-positive cases



67 dic(9;20)

31 genetically analyzed

Figure 6: Schematic overview of dic(9;20)-positive and other BCP ALLs included in paper IV and analyzed by multiple genomic techniques

In **paper IV**, we included all 67 cases of dic(9;20) BCP ALL diagnosed between 1996 and 2013 in the Nordic countries, and that were treated according to the NOPHO ALL-92, ALL-2000 and ALL-2008 protocols. We analyzed 31 of these cases, from which material was available, by one or multiple genomic analyses. The analyzed cases were from all treatment protocols; ALL-92 (n=3), ALL-2000 (n=25) and ALL-2008 (n=3). The dic(9;20) rearrangement had previously been confirmed in all cases through centrally reviewed karyotype data or FISH analysis in **paper I**, or a combination of DNA-methylation and copy number profiling (four cases in **paper IV**).⁸¹

4.2 CYTOGENETIC ANALYSES

4.2.1 G-banding and metaphase FISH analysis

G-banding analyses were performed using standard methods in 15 cytogenetic laboratories in the Nordic countries and all abnormal karyotypes were centrally reviewed. In **paper I**, suspected dic(9;20)-positive cases with cells in fixative were analyzed further by metaphase FISH analyses to confirm the presence of dic(9;20), using CEP 9 and CEP 20 (Abbott Molecular Inc., Des Plaines, IL) and/or WCP 9 and WCP 20 probes (Cytocell Ltd, Cambridge, UK),

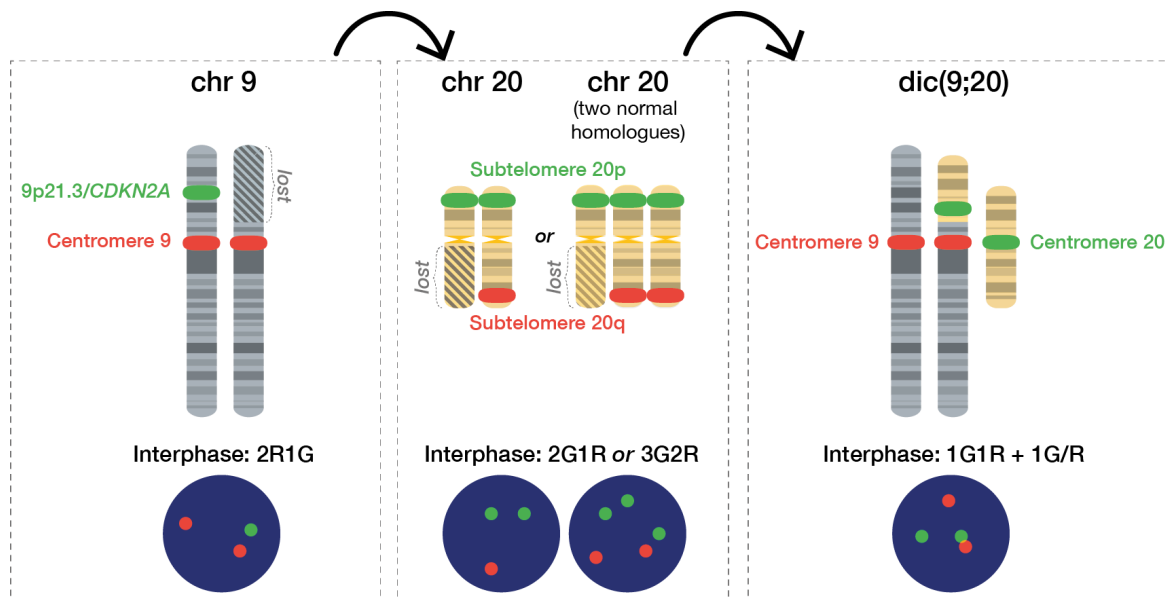


Figure 7: Three-step FISH analysis for the detection of *dic(9;20)* in paper I

Bone marrow smears from 533 BCP ALL patients were analyzed in a three-step manner (figure 7). The LSI p16 (9p21) FISH probe (Abbott Molecular Inc.), representing a mixture of the p16 (official gene symbol *CDKN2A*) probe labeled with Spectrum Orange and a CEP 9 probe labeled with Spectrum Green, was used for identifying 9p deletions. Cases with loss of *CDKN2A*, which all *dic(9;20)*-positive ALL cases have, were subsequently screened, according to the manufacturer's instructions (Abbott Molecular Inc.), with the Vysis ToTelvysion probes that are specific for the subtelomeres of 20p and 20q. Cases displaying imbalances between the number of signals for 20p and 20q (figure 7, second panel) were further analyzed using CEP 9 and CEP 20 probes to confirm the presence of a dicentric rearrangement (figure 7, panel 3).

4.2.2 Design of FISH probe kit

After mining breakpoint data of all previously published cases of *dic(9;20)*, we designed a set of probes that would collectively detect: (i) the unbalanced loss of 20q material, regardless of whether one or two normal chromosome 20 homologs are present, and (ii) the co-localization of centromeres 9 and 20 (figure 8). Specifically, the probe on 20q is located just distal of the most telomeric breakpoints reported in *dic(9;20)* resulting, if positive, in loss of one signal from 20q while still being close enough to ensure the probe's specificity. Several iterations of the co-localization signal was tested on a set of positive and negative controls, altering positions and probe size to consistently produce strong and homogeneous signals. BACs covering the specified area on 20q11 were sent to Kreatech Diagnostics (Amsterdam, the Netherlands) where they were labeled with green and mixed with a centromere-specific probe for chromosome 9 (SE9) labeled with blue and a 20p11 probe labeled with red (figure 8). All bone marrow slides were prepared according to the manufacturer's instructions (Kreatech Diagnostics) and analyzed using a Zeiss Axioskop 2 (Carl Zeiss, Göttingen, Germany) and Smart Capture 3 software (Digital Scientific Ltd, Cambridge, UK).

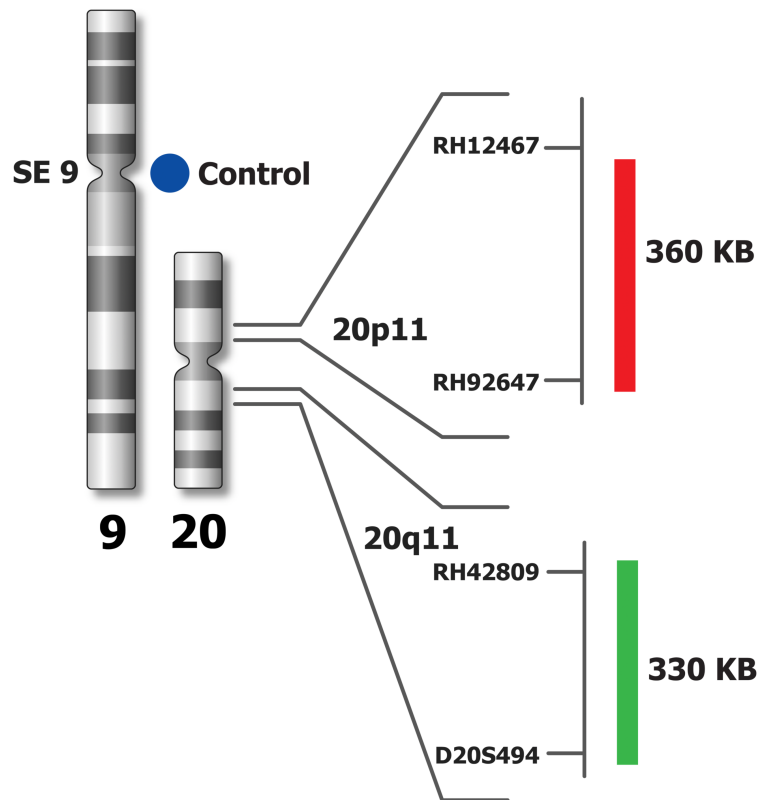


Figure 8: Ideogram of chromosomes 9 and 20, illustrating the *dic(9;20)* FISH probe design. Adapted from figure 1 in paper III.

4.3 STATISTICAL ANALYSES

In **paper I**, we calculated the probabilities of EFS (pEFS) and OS (pOS) at five years after diagnosis using the Kaplan-Meier method. The different cytogenetic subgroups were compared using the log rank test (figure 3). A multivariate analysis was also performed, including all cytogenetically defined subgroups and clinical parameters (WBC and age at diagnosis). The significance limit for two-sided P-values was set to <0.05 in all tests. Time in first remission was defined as time from diagnosis until first event, comprising induction failure, relapse, death of disease, death in remission, or second malignant neoplasm. In the OS analysis, death of any cause was the endpoint. The median observation time for patients in continuous complete first remission (CR1) was 67 months (range 28-108 months). The NOPHO leukemia registry is updated annually and follow-up data were extracted from the registry on April 1, 2010. The Statistical Package for the Social Sciences (SPSS Inc., Chicago, IL) software for Macintosh was used for the statistical analyses.

In **paper IV**, we plotted the pEFS, pOS and cumulative incidences of relapse and death, of all 67 *dic(9;20)*-positive BCP ALLs using the Kaplan-Meier method in R.⁹⁸ We compared the outcome in NOPHO ALL-92 and ALL-2000 protocols, to the outcome of the ALL-2008 protocol using the log rank test. Additional statistical methods in genomic analyses were implemented as described below.

4.4 GENOMIC ANALYSES

In **paper IV**, we used multiple genome-wide techniques to profile *dic(9;20)*-positive BCP ALL. We compared them to other subtypes of ALL, both within NOPHO ALL treatment protocols and an independent cohort. Figure 6 summarizes the cohorts studied, and figure 9 the methods used and their corresponding read-out.

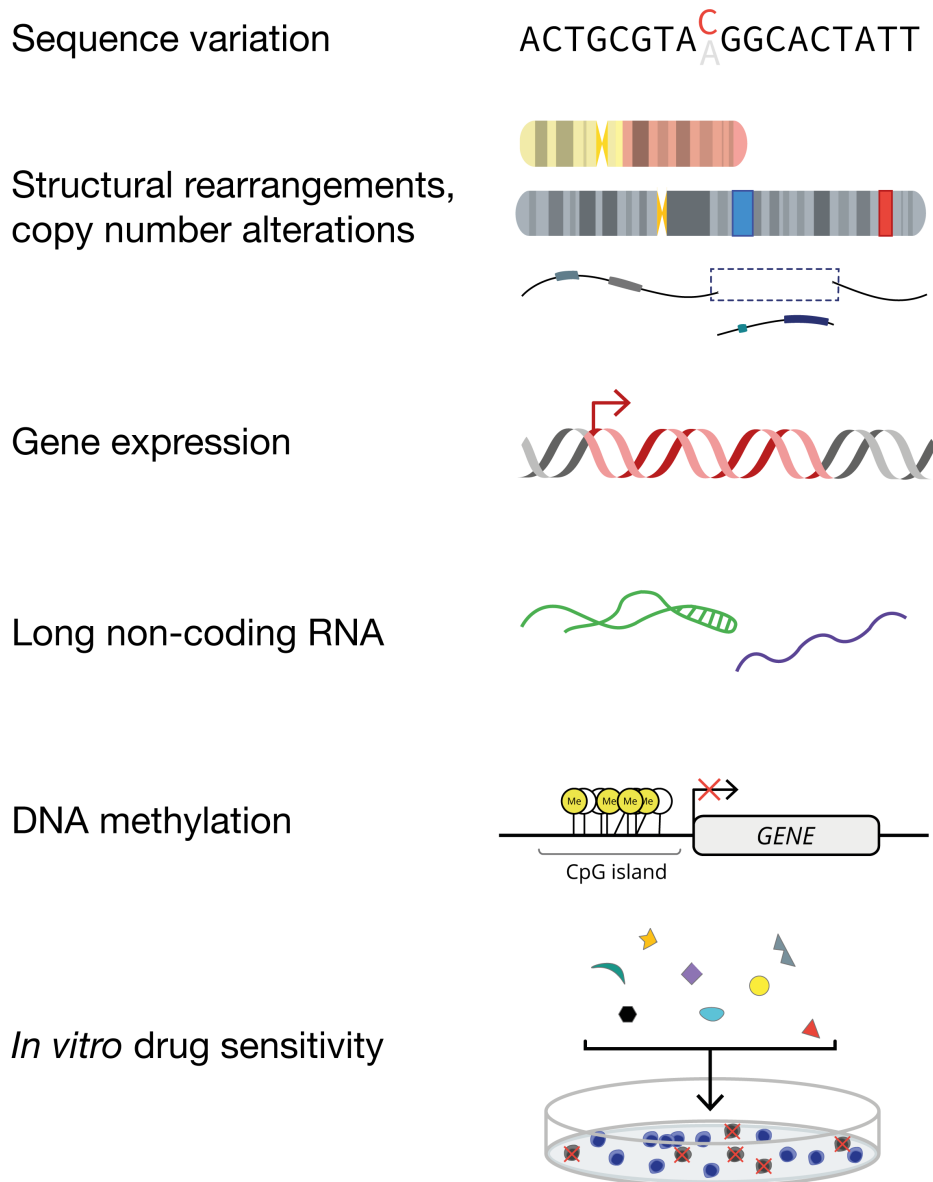


Figure 9: Overview of genomic and functional analyses used in paper IV to characterize the genetic basis of *dic(9;20)*-positive BCP ALL

4.4.1 Detection of copy number alterations

Copy number alterations (CNA) were detected using Omni 610k (n=1), Omni 2.5M (n=17) or 2.5M+Exome (n=7) genotyping arrays (Illumina). Probe intensities were normalized against a panel of internal controls to produce log transformed ratios of experimental versus

normal reference intensities centered at zero for a diploid sample (log2 ratios, LRR). Quality controls included >99% genotype call rate and an LRR standard deviation of <0.15.

Chromosome X in males was normalized by centering on mean LRR for each chr X position ($x[\text{sample}] - \text{mean}[\text{ref}]$), created from a set of 5 normal germline samples of males (3 CEPH males and 2 internal controls of unrelated males). LRR values were segmented using circular binary segmentation with default parameters.⁹⁹ Segmented copy number data was combined with probe-level log2 ratios and B-allele frequencies (BAF) to detect allele-specific CNAs using the Tumor Aberration Prediction Suite.¹⁰⁰ Segments encompassing less than 10 probes and smaller than 20kbp were filtered out. Regions of gain, loss and copy neutral loss of heterozygosity (CNLOH) were annotated against Ensembl v75, common CNVs from the Database of Genomic Variants¹⁰¹ and an in-house database of normal variation from ~4000 cases evaluated at the Clinical Genetics Service, Karolinska University Hospital. CNAs overlapped to more than 90% by a normal CNV were excluded. All annotation, filtering and visualization was performed using BEDOPS v2.4.2,¹⁰² R v3.1.0⁹⁸ and IGV v2.3.42.¹⁰³

In cases lacking SNP array data but with DNA methylation arrays available, intensities from the Infinium 450k array (Illumina) were used to map the 9p and 20q breakpoints in IGV, as previously described.⁸¹

4.4.2 Whole genome sequencing

Sequencing libraries were prepared from high molecular weight DNA using Nextera Mate-Pair Sample Preparation Kit (Illumina Inc, San Diego, CA, USA). DNA sequencing libraries with an effective insert size of ~2kb were created for each sample using the Nextera Mate-Pair sample preparation kit (Illumina). These were then sequenced on an Illumina HiSeq 2500, 100bp paired-end, to an average raw coverage depth of ~5x. After adapter trimming using Trimmomatic and kmer analysis for quality, cleaned reads were aligned to human genome reference build GRCh37 using bwa-mem version 0.7.4.¹⁰⁴ Mapped reads were analyzed for intra- and interchromosomal translocations using a sliding window method implemented in FindTranslocations (<https://github.com/vezzi/FindTranslocations>). In addition, putative breakpoints informed by previous patient-specific cytogenetic analysis or copy number alterations were directly analyzed for the presence of linking mates using IGV.¹⁰³

4.4.3 RNA sequencing

RNA was extracted from diagnostic bone marrow cells after Ficoll gradient separation using the AllPrep DNA/RNA or AllPrep DNA/RNA/Protein extraction kit (Qiagen). Sample quality and RNA quality was ensured by only including samples with RIN values above 7, measured using the RNA Nano Assay 6000 on a Bioanalyzer 2100 (Agilent). Samples were depleted for ribosomal RNA using RiboZero (Epicentre) and sequencing libraries were then prepared using ScriptSeq v2 (Epicentre), a strand specific protocol. The libraries were sequenced on Illumina HiSeq 2000 or 2500 machines with two samples per flowcell lane, 50bp paired-end, producing an average of 100 million read pairs per sample. Two additional sequencing

libraries from RNA samples of lower quality were prepared using TruSeq RNA Access (Illumina). These samples were not included in the gene expression analysis.

4.4.3.1 Gene expression profiling

RNA sequence reads were aligned to the human genome reference build GRCh38 (hg38) using STAR version 2.4.0j,¹⁰⁵ with exon junction support from Ensembl gene annotation version 77. Reads overlapping Ensembl features were summarized in each sample using featureCounts, as implemented in the subread package version 1.4.6.^{106,107} Raw read counts were normalized by the trimmed mean of M-values normalization method,¹⁰⁸ and variance normalized using voom.¹⁰⁹ Genes with a count of at least 1 per million mapped reads (CPM), in at least 2 samples were included for further analysis. Differential gene expression was performed by fitting a linear model informed by subgroups, using the R/bioconductor package limma.¹¹⁰ Fourteen dic(9;20)-positive cases were contrasted against the combined average expression of 32 non-dic(9;20) cases, including HeH, t(12;21)/*ETV6-RUNX1*, *MLL*-rearranged, t(9;22)/*BCR-ABL1* and iAMP21. Unsupervised hierarchical clustering, using Euclidean distances and complete linkage, was based on the top 1000 (~5%) most variably expressed genes across the whole cohort by *F* test statistic in limma.

An independent cohort of 118 BCP ALLs, including six dic(9;20), was used to validate the dic(9;20) specific expression pattern.³⁸ Gene expression signatures trained to detect *BCR-ABL1*-like BCP ALL in previous studies of the DCOG¹¹¹ and StJude/COG⁷³ groups were evaluated in this cohort by unsupervised hierarchical clustering. Affymetrix gene set IDs from the original classifiers were converted to EnsemblIDs and used to subset the complete set of expressed genes in the cohort.

Gene set enrichment analysis was performed against Hallmark, C2 and C5 gene sets in MSigDB.¹¹² Gene sets including less than 10 genes expressed in this cohort were excluded. Where applicable, gene ids or Affymetrix expression array probe sets were converted to Ensembl IDs using Biomart.¹¹³ In all instances, self-contained gene set testing using rotation with was applied, with 10000 permutations, implemented as the mroast function in limma.¹¹⁴

4.4.3.2 Sequence variant detection

RNA sequence reads were aligned to the human genome reference GRCh37 (hg19) using STAR version 2.4.0j,¹⁰⁵ with exon junction support from Ensembl gene annotation version 75. Duplicate sequence reads were marked using Picard version 1.129 (<http://picard.sourceforge.net>) and excluded from further analysis. Using GATK version 3.2-2,¹¹⁵ sequence reads overlapping exon junctions were split and trimmed to exon boundaries and base quality score recalibration applied. Finally, SNV and indel discovery and genotyping was performed in individual samples. To reduce likely false positive variant calls, hard quality filters were applied; excluding clustered SNPs (≥ 3 within 35bp), variants with high Fisher strand values ($FS > 30$) or low depth-adjusted quality scores ($QD < 2$). To validate the overall quality and reliability of the variant calls, we estimated the concordance of genotype calls comparing, in each sample, the genotypes from SNP array and RNA-seq

genotypes for all positions covered by at least 10 RNA sequence reads. All variants were annotated using Ensembl Variant Effect Predictor¹¹⁶ and explored using GEMINI.¹¹⁷ Filtering for putative somatic variants were applied as follows: 1) Non-synonymous, stop-gain or frameshift variants in coding regions; 2) Supported by at least 10 variant reads, and a minimum of 20 total reads; 3) Not in repeat masked region; 4) Previously confirmed as somatic in at least two cases reported in COSMIC (<http://cancer.sanger.ac.uk/cosmic/>), in the Pediatric Cancer Genome Project (<http://explore.pediatriccancergenomeproject.org>) or in previous large scale BCP ALL sequencing studies;^{118,119} 5) Visually inspected supporting reads in IGV.

4.4.3.3 Detection of fusion transcripts

Raw sequence reads in fastq format were analyzed for evidence of fusion transcripts using FusionCatcher v0.99.3e.¹²⁰ Briefly, the algorithm functions by automatically adjusting parameters for adapter sequence removal and read trimming according to read length, and by mapping sequences to all possible exon-exon junctions leveraging multiple aligners. Putative fusions with common mapping reads, indicating a high probability of being false positive because of sequence homology, or fusions with both ends mapping to repeat regions were excluded.

4.4.4 DNA methylation analysis

Previously published DNA methylation levels across 450k CpG sites from twenty cytogenetically defined dic(9;20)-positive cases and four characterized by methylation profile and CNAs were included in this study.^{80,81} The twenty cases were previously used as part of a training set totaling over 600 childhood ALL samples to build subtype-specific DNA methylation classifiers, including a 37-CpG classifier for dic(9;20). Four additional dic(9;20) cases were detected by using this classifier to describe a set of cytogenetically uninformative cases, and confirmed by the detection of copy number loss of chromosome arms 9p and 20q.⁸¹

4.4.5 In vitro drug sensitivity assay

In vitro drug sensitivity data was obtained for all available dic(9;20) and other BCP ALL cases for which RNA-seq data was also available. These data were previously published as part of a larger NOPHO ALL cohort including 546 BCP ALL samples.¹²¹ In short, a fluorometric microculture cytotoxicity assay (FMCA) was used to measure the treatment response of primary leukemic blasts, obtained at leukemia diagnosis and put in short term culture, against a panel of the most commonly used cytotoxic drugs in current leukemia treatment protocols. All drugs were tested in triplicate and compared against six non-treated wells (controls) and six with medium only (blanks) to calculate a survival index. The FMCA thus provides a reliable measure of the overall capability of specific drugs to kill off malignant cells, presented here as percent surviving cells.¹²¹

5 RESULTS AND DISCUSSION

5.1 FREQUENCY AND CLINICAL CHARACTERISTICS OF dic(9;20) BCP ALL

In **paper I**, we screened 533 cases of childhood BCP ALL diagnosed between 2001-2006 for the presence of dic(9;20)(p13.2;q11.2) using interphase FISH (figure 7). During this time period, a total of 1174 children were diagnosed with ALL in the Nordic countries, of which 1033 were BCP ALL. The 491 cases where BM smears were unavailable for FISH analysis did not differ significantly in terms of subtype distribution, age, risk grouping and other clinical characteristics (not shown). Using a three-step screening approach, deletions of 9p21.3/CDKN2A were detected in 84/533 (15.7%) of cases, of which 25/84 (29.7%) had dic(9;20). In total, the frequency of dic(9;20) was 4.7% (25/533), the third most common cytogenetic abnormality after HeH (34%) and t(12;21) (23%) in the cohort.

The clinical characteristics of dic(9;20)-positive and other BCP ALLs screened by interphase FISH were analyzed in paper I and are summarized here in table 1. The median age of patients with dic(9;20)-positive BCP ALL was 3.3 years, and the rearrangement was associated with slightly higher white blood cell (WBC) counts at diagnosis compared to most other subtypes (median 19.2, range 1.6-336). In addition, there was a strong female predominance (68%), confirming previous reports of this unusual characteristic for childhood ALL.^{90,97} Importantly, while dic(9;20) was not a risk-stratifying aberration in this protocol, the majority of cases (18/25, including one infant) were assigned to non-standard risk treatments based on their clinical characteristics alone.

The dic(9;20) rearrangement was the principal abnormality detected in all but two cases; one with t(9;22)/*BCR-ABL1* and one with t(12;21)/*ETV6-RUNX1*. Of the 25 dic(9;20)-positive cases detected by interphase FISH, only 11 were also detected at diagnosis by G-banding or metaphase FISH. In total, cytogenetic information was available at diagnosis for 20 of the dic(9;20)-positive cases, with 15 of them (75%) shown to harbor additional numerical or structural changes. Recurrent abnormalities included gains of chr 21 and chr 20, that is two normal chr 20 homologues in addition to the dic(9;20). We and others have reported the presence of established, primary aberrations in addition to the dic(9;20), also detected in **paper I** in two cases.^{90,96,97} In **paper IV**, an additional case harboring both t(12;21)/*ETV6-RUNX1* and dic(9;20) was detected by targeted FISH at diagnosis. Thus, the association with both t(9;22)/*BCR-ABL1*, reported previously in two pediatric cases,^{88,97} and t(12;21)/*ETV6-RUNX1*, reported in two cases in these papers, can be considered recurrent.⁹⁰ This is important to recognize, as the presence of for example *BCR-ABL1* dictates a different treatment in the current protocols.³³ Whether there is an actual difference in outcome for cases harboring dic(9;20) alone, compared to its presence in combination with other rearrangements, will require the further study of more than the handful of such cases currently known.

In summary, these data show that the frequency of dic(9;20) in BCP ALL is close to 5 percent, at least twice that detected by G-banding/metaphase analysis alone. This is also

consistent with the lower frequency of 1-2.5% reported in all previous studies, all relying on chromosome banding analyses alone.^{85,96,97} While the frequency of dic(9;20)-positive BCP ALL was not systematically studied in **paper IV**, the number of cases with dic(9;20) in the NOPHO ALL-2008 protocol were fewer than those reported using the retrospective FISH analysis applied in **paper I**, considering an equivalent number of years and cases (2009-2013). The total number of dic(9;20)-positive cases during this time period were 21. This is comparable to the frequency detected by chromosome banding techniques alone in the ALL-2000 study (20/1033 in **paper I**). Assuming the underlying incidence of dic(9;20) did not vary between protocols, this suggests that many BCP ALLs with dic(9;20) go undetected also in the current protocol iteration.

5.2 DETECTING THE DIC(9;20) REARRANGEMENT

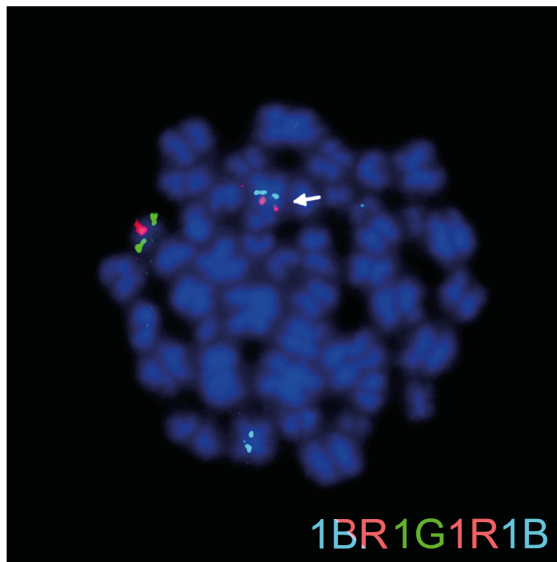
From the study in **paper 1**, it is clear that a significant proportion of cases with dic(9;20) go undetected using traditional cytogenetic techniques alone. Of the 25 cases with dic(9;20) detected by FISH in **paper I**, 19 (76%) were scored as having two signals from chromosome arm 20p and one from 20q. This pattern is consistent with the presence of the dic(9;20) in addition to one normal homologue of chr 20. The other six cases displayed a pattern of three signals from 20p and two from 20q, indicating the presence of two normal chr 20 homologues, in addition to the dic(9;20) rearrangement. In total, we show that almost half of dic(9;20) cases harbor an additional chromosome 21 and, in a quarter of cases, two normal homologues of chromosome 20 in addition to the dic(9;20). Most dic(9;20)-positive cases will therefore go undetected if preferentially screened for in a hypodiploid (<46 chr) cohort or focusing solely on cases with monosomy 20.^{38,85}

For these reasons, we designed a FISH probe kit that could be used to directly screen for the dic(9;20) in a clinical setting. In **paper III**, we combined a set of probes informed by all previously reported breakpoints of the dic(9;20) rearrangement, to show the reliable detection of dic(9;20) in a single step, suited for clinical diagnostic use (figure 10). We validated the probe kit by screening metaphase chromosomes and interphase nuclei from five previously identified dic(9;20), including cases both with monosomy 20 and +20. Further, we screened a cohort of 50 negative controls, comprising both normal blood and BM smears from diagnosis in other childhood ALL and different adult hematological diseases, all of which had supporting karyotype data available. In most cases, additional aCGH and/or spectral karyotyping had been performed previously and used to support the analysis. No false positive cases were detected at a 3% threshold (+3 s.d. from the mean) and metaphases showed good specificity and signal (figure 10). Implementing the FISH probe kit for routine diagnostics at the Clinical Genetics hospital service, at Karolinska Hospital, two dic(9;20) cases were identified prospectively, including one in which the dic(9;20) was present in a minor clone at diagnosis and not detected by chromosome banding.

While the unbalanced nature of the dic(9;20) rearrangement allows its inference from genomic profiling, using for example SNP arrays, the dicentric nature of the rearrangement cannot be confirmed this way. Further, as we showed using SNP array and WGS analysis in

paper IV, the dic(9;20) breakpoints are often located in low-mappability regions on chr 9 and chr 20, precluding their detection by WGS or RNA-seq in such cases. In addition, the single-cell nature of FISH analysis allows the reliable detection of minor clones harboring the rearrangement, useful both at diagnosis and during follow-up of treatment. For these reasons, we find that the described probe kit is useful in routine clinical diagnostics, either alone or complemented with other high-throughput methods. Combining FISH and other methods, we expect this will render an accurate diagnosis in more cases with dic(9;20). This is important to ensure adequate treatment in protocols using the dic(9;20) as a risk-stratifying marker, like the NOPHO ALL-2008 protocol.³⁰ Further, better methods of detection will also improve our ability to comprehensively study the clinical implications of dic(9;20) in larger and more representative cohorts.

dic(9;20), metaphase FISH



dic(9;20), interphase FISH

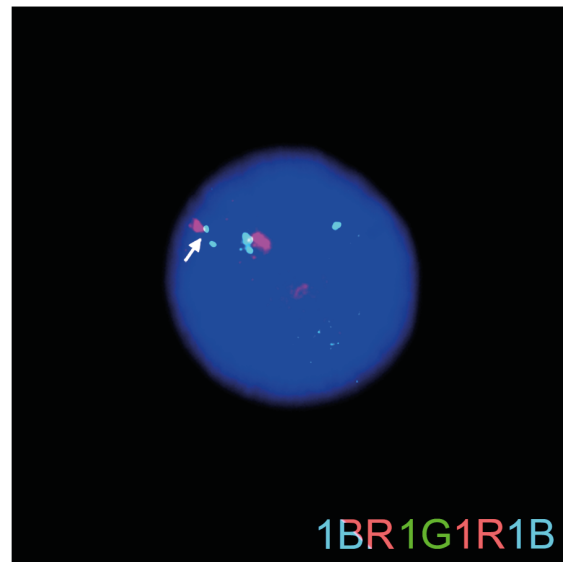


Figure 10: Metaphase chromosomes and interphase nuclei of dic(9;20)-positive cells hybridized with a dic(9;20)-targeted probe kit including: centromere 9 (B; blue), 20p (R; red) and near centromere 20 (G; green). Positive cells are scored as 1BR, representing the fusion, and 1G1R1B, representing normal homologues of chromosome 9 and 20. White arrows indicate the dic(9;20) rearrangement in the respective experiments.

Table 1: Patient characteristics, clinical risk stratification and outcome of 533 BCP ALLs analyzed by interphase FISH, subdivided by cytogenetic aberrations

Features	HeH	<i>t</i> (12;21)*	<i>dic</i> (9;20)*#	<i>t</i> (1;19)	<i>11q23/MLL</i>	<i>t</i> (9;22)#	<i>iAMP21</i>	<45 chr	Normal	Others	Failure
No. of cases (%)	181 (34)	120 (23)	25 (4.7)	18 (3.4)	19 (3.6)	12 (2.3)	10 (1.9)	7 (1.3)	56 (11)	66 (12)	19 (3.6)
<i>Gender</i>											
Female (%)	82 (45)	56 (47)	17 (68)	11 (61)	9 (47)	4 (33)	2 (20)	5 (71)	25 (45)	28 (42)	7 (37)
Male (%)	99 (55)	64 (53)	8 (32)	7 (39)	10 (53)	8 (67)	8 (80)	2 (29)	31 (55)	38 (58)	12 (63)
<i>Age (years)</i>											
Max	17.9	15.6	15.1	13.1	13.1	17.9	17.3	12.8	14.8	17.7	14.4
Median	3.7	4.5	3.3	4.3	1.0	9.1	9.7	4.1	4.9	6.4	6.7
Min	1.0	1.4	1.3	1.9	0.1	4.0	3.1	3.0	1.0	1.4	1.6
<i>WBC count (x 10⁹/l)</i>											
Max	171	260	336	159	528	101	30.7	100	170	172	231
Median	7.0	7.9	19.2	34.7	76.0	15.1	3.2	15.7	12	13.4	12.3
Min	0.6	0.9	1.6	4.1	3.1	1.3	1.3	3.9	0.3	0.8	1.7
EML (%)	6 (3.3)	3 (2.5)	1 (4.0)	0	2 (11)	1 (8.3)	0	0	1 (1.8)	0	0
<i>Classification</i>											
SR (%)	94 (52)	56 (47)	7 (28)	1 (5.6)	1 (5.3)	0	3 (30)	0	13 (23)	15 (23)	6 (32)
IR (%)	61 (34)	41 (34)	5 (20)	1 (5.6)	1 (5.3)	0	6 (60)	0	27 (48)	34 (52)	5 (26)
HR (%)	24 (13)	23 (19)	12 (48)	16 (89)	7 (37)	4 (33)	1 (10)	7 (100)	14 (25)	16 (24)	8 (42)
Other (%)	2 (1.1)	0	1 (4.0)	0	10 (53)	8 (67)	0	0	2 (3.6)	1 (1.5)	0
<i>Primary events</i>											
CR1 (%)	150 (83)	105 (88)	17 (68)	15 (83)	10 (53)	8 (67)	5 (50)	4 (57)	49 (88)	57 (86)	14 (74)
Induction failure (%)	2 (1.1)	0	0	0	1 (5.3)	0	0	0	0	1 (1.5)	1 (5.3)
Resistant disease (%)	2 (1.1)	0	0	0	0	0	0	0	1 (1.8)	1 (1.5)	0
Relapse (%)	19 (10)	13 (11)	6 (24)	3 (17)	8 (42)	4 (33)	5 (50)	1 (14)	4 (7.1)	7 (11)	4 (21)
Dead in CR1 (%)	5 (2.8)	0	1 (4.0)	0	0	0	0	2 (29)	0	0	0
SMN	3 (1.7)	2 (1.7)	1 (4.0)	0	0	0	0	0	2 (3.6)	0	0

Abbreviations: chr, chromosomes; CR1, complete first remission; EML, extra-medullary leukemia at diagnosis (spread to central nervous system and/or mediastinum and/or testis); HR, high risk; IR, intermediate risk; SMN, second malignant neoplasm; SR, standard risk; WBC, white blood cell. In this table, *one case with *dic*(9;20) and *t*(12;21) and #one with *dic*(9;20) and *t*(9;22) are included in the *dic*(9;20) group. The classification 'Other' includes patients treated according to the Interfant-99, EsPhALL, or modified BFM.

5.3 TREATMENT OUTCOME

To study the treatment outcome of dic(9;20)-positive childhood BCP ALL, in **paper I** we analyzed the survival of 882 children uniformly treated according to the NOPHO ALL-2000 protocol. Combining the dic(9;20) cases detected by FISH with additional cases detected at diagnosis by G-banding or metaphase FISH alone, there were 29 dic(9;20)-positive BCP ALLs treated according to the ALL-2000 protocol. Of these, seven cases suffered relapses (24%): four in the BM (7-33 months after diagnosis) and three with isolated CNS relapses (22-29 months after diagnosis). All cases with BM relapses were classified as HR at initial risk stratification, while the CNS relapses were SR or IR. Further, the proportion of CNS relapses in dic(9;20) was higher than in the cohort as a whole, 3/7 (43%) in dic(9;20) versus 19/128 in other BCP ALLs (15%), but did not reach statistical significance (chi square 3.81; $p=0.0506$). Two additional events occurred in cases with dic(9;20); one infection-related death in CR1 (HR treatment) and one patient that developed a non-Hodgkin lymphoma 30 months after diagnosis (SR treatment).

In total, the nine primary events in dic(9;20) cases rendered a pEFS of 0.69 at 5 years after diagnosis. Analyzed by paired univariate analysis, this was significantly worse than for cases with t(12;21) (pEFS=0.87; $p=0.002$) and with HeH (pEFS=0.82; $p=0.04$) (figure 3A). By Cox multivariate analysis, accounting for cytogenetic subgroups, WBC, age and with dic(9;20)-positive cases as reference, only t(12;21)-positivity associated in a significantly superior outcome. In the whole cohort of 882 BCP ALLs, WBC count at diagnosis was the most powerful predictor of EFS.

Of the nine dic(9;20) cases with primary events, there were four deaths (13.7%) - three related to disease, and one in complete remission. The overall survival at 5 years after diagnosis was 0.85 (figure 3B) - worse than for t(12;21) cases in univariate analysis (pOS=0.97; $p=0.002$). Again in multivariate analysis, as described, WBC and age at diagnosis were found to be the strongest predictors of OS in the cohort as a whole ($p<0.01$). Of the cytogenetic abnormalities, only t(12;21) was significantly associated with OS ($p=0.035$).

Together, these data show that a significant proportion of dic(9;20)-positive BCP ALL cases relapsed when treated on the NOHO ALL-2000 protocol, resulting in a lower EFS compared to the most common, standard-risk aberrations. However, many dic(9;20) appeared to respond well to second-line treatment with an OFS that was only worse compared to cases with t(12;21). Comparable survival data in dic(9;20) in other treatment protocols are only available from smaller patient cohorts, often treated across protocol iterations^{96,97} In an ALL-BFM study including 19 dic(9;20)-positive cases, there were fewer relapses (16%) and somewhat better outcome measures than those described in **paper I** (5 year EFS rate 75%; 5 year OS rate 94%).⁹⁷ However, also in this study there were several cases with CNS involvement at diagnosis and CNS relapses. In a smaller number of dic(9;20) BCP ALLs treated on the UKALL97/99 protocols, again 5 year pEFS and pOS were relatively favorable at 77% and 92%, respectively.⁹⁶

In **paper IV**, we extended this analysis to include an additional 19 dic(9;20)-cases from the NOPHO ALL-2008 protocol. Here, we compared the clinical outcome of dic(9;20)-positive cases specifically, treated according to the ALL-92 or ALL-2000 protocol (without dic(9;20) stratification) to those on the ALL-2008 protocol (with dic(9;20) stratified to the IR treatment arm). While the number of dic(9;20) cases enrolled in the NOPHO ALL-2008 protocol were too few to allow statistical certainty, a single primary event was reported among the patients on the ALL-2008 protocol (a death in complete remission). Thus, there was a trend towards improved survival in the new protocol that requires further study. Considering that most dic(9;20)-positive cases were already stratified to non-standard risk treatment in the ALL-2000 protocol (table 1), based on clinical characteristics alone, the reason for this apparent difference in outcome is not immediately clear.

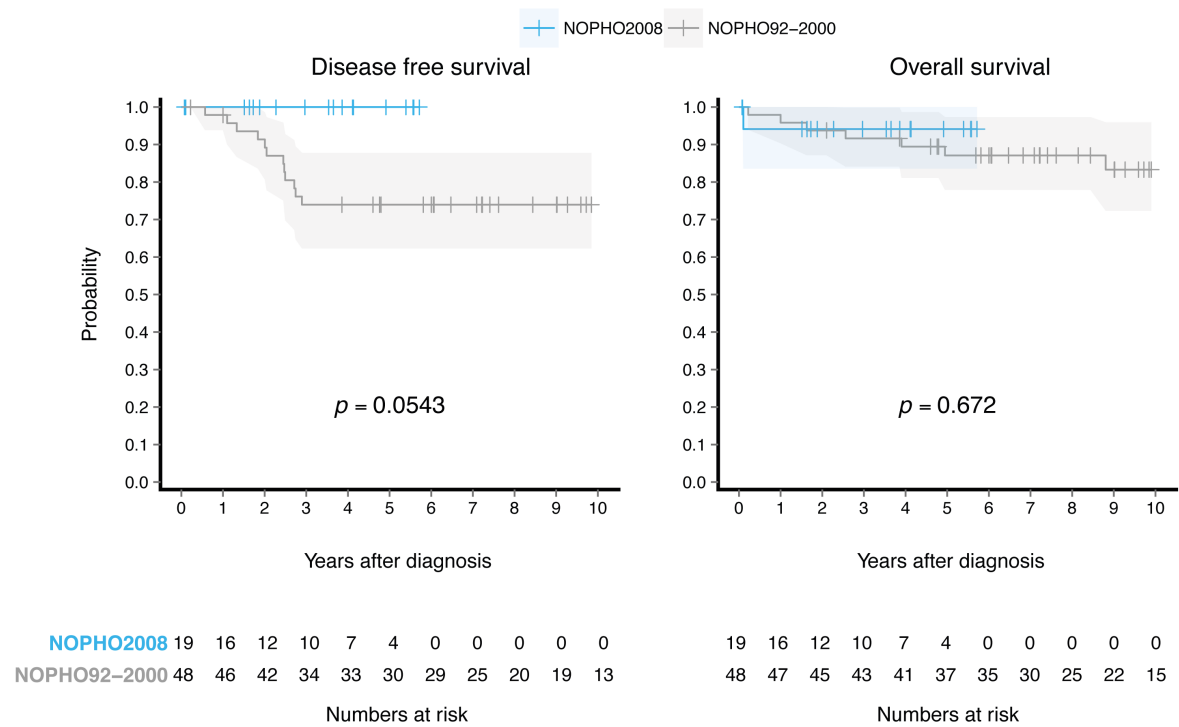


Figure 11: Probability of disease-free (left) and overall (right) survival of dic(9;20)-positive cases treated according to the NOPHO ALL-92 and 2000 (grey) compared to the ALL-2008 (blue) protocols, plotted as Kaplan-Meier curves and compared using the log-rank test. There was a single death among the ALL-2008 cases that occurred during first complete remission. Censored cases are marked with vertical bars and shaded areas for each protocol represent 95% confidence intervals.

5.4 BREAKPOINT MAPPING AND FUSION GENE DETECTION

To characterize the breakpoints in dic(9;20) in detail, we analyzed DNA from diagnostic BM samples using SNP arrays - first in **paper II**, then extended and with higher resolution in **paper IV**. As previously shown, dic(9;20) breakpoints are heterogeneous both on 9p and 20q.^{89,92,93} Here, most breakpoints on 9p were found to occur within or directly upstream of *PAX5*, with a second significant cluster in low-mappability and repeat regions near

centromere 9. The breakpoints on 20q were found to display more variability with no consistent correlation to the breakpoints on 9p, again in line with previous results.⁹³ In total, the breakpoints could be ascertained in 30 cases. In twelve of these, the break on 9p was located within *PAX5* - five of which also had intragenic *NOL4L* breaks at 20q11.2. By using RNA-sequencing and WGS in **paper IV**, we could detect the in-frame fusion gene *PAX5-NOL4L* (previously *C20orf112*) in 4 of 16 samples (figure 12). Two cases carried an identical fusion involving *PAX5* exon 5 and *NOL4L* exon 8, a configuration that has been reported previously.⁹² All four of the detected fusions were predicted to be in-frame and retain the DNA binding domain of *PAX5*. Together, these data show that *PAX5-NOL4L* is a recurrent fusion gene in dic(9;20), as detected by multiple techniques, but that *PAX5* fusions cannot account for all dic(9;20)-positive cases. Further, we detect a number of other cases with breakpoints within *PAX5* and where the corresponding breakpoints on chr 20 are in non-recurrent and intergenic region, questioning the formation a fusion gene even when *PAX5* is directly involved.

The *PAX5* gene, encoding a transcription factor central to B-cell development and lineage identity,¹²² is the most frequently altered gene in childhood BCP ALL.⁷¹ Fusion genes involving *PAX5* have been reported with a variety of 3' partners, both in other dic(9;20)-positive cases^{92,94} and childhood ALL in general.^{123–125} Some of the fusions are predicted to disrupt normal *PAX5* protein function, either through haploinsufficiency or by dominant-negative suppression of wt *PAX5* protein.¹²⁶ However, it is clear that the wide variety of 3' partners also results in considerable heterogeneity regarding the precise functional outcome of *PAX5* fusion genes.¹²³ For the *PAX5-NOL4L* fusion specifically, previous protein studies have shown that it is likely to result in suppression of wt *PAX5* function, binding with up to 10-fold higher affinity to *PAX5*-target sites.¹²⁶ Together with other structural variants and mutations commonly reported in *PAX5*, it is therefore thought that the dic(9;20) rearrangement ultimate serves to disrupt *PAX5*.⁹² To analyze this in our cohort, we investigated a well-characterized set of *PAX5* target genes for evidence of differential expression in dic(9;20)-positive cases compared to other BCP ALLs. Gene set enrichment analysis of known *PAX5* targets did not show any significant transcriptional deregulation of downstream genes, suggesting other aberrations may be important in cases with dic(9;20).

To characterize further cases with unclear breakpoints or suspected cryptic rearrangements, we applied low-coverage mate-pair sequencing in four cases. As expected, we could confirm the *PAX5-NOL4L* in a case with corresponding breakpoints on the SNP array and RNA-sequencing data supporting the same fusion. In another case, SNP array analysis had revealed a deletion of 9p extending beyond 9p13.2 and into 9q. Despite this, WGS revealed the presence of a *PAX5-NOL4L* fusion also in this case (figure 12, case KSALL4). It was part of a complex rearrangement, duplicated at the breakpoint sites at 9p13.2 and with links between each of 9p, 9q and 20q – all confirmed by Sanger sequencing.

In another two cases, one analyzed by WGS, there was evidence of a small, retained region within the extended deletion of 20q. In both cases, the retained region included only *PTPRT*,

a protein tyrosine phosphatase gene. Using WGS followed by Sanger sequencing, we found this ~1.4Mb fragment to be inserted between the 9p and 20q arms in the dic(9;20) rearrangement. The effect of this is unclear, however, as the *PTPRT* gene was not expressed in any of the cases subjected to RNA-sequencing in **paper IV**.

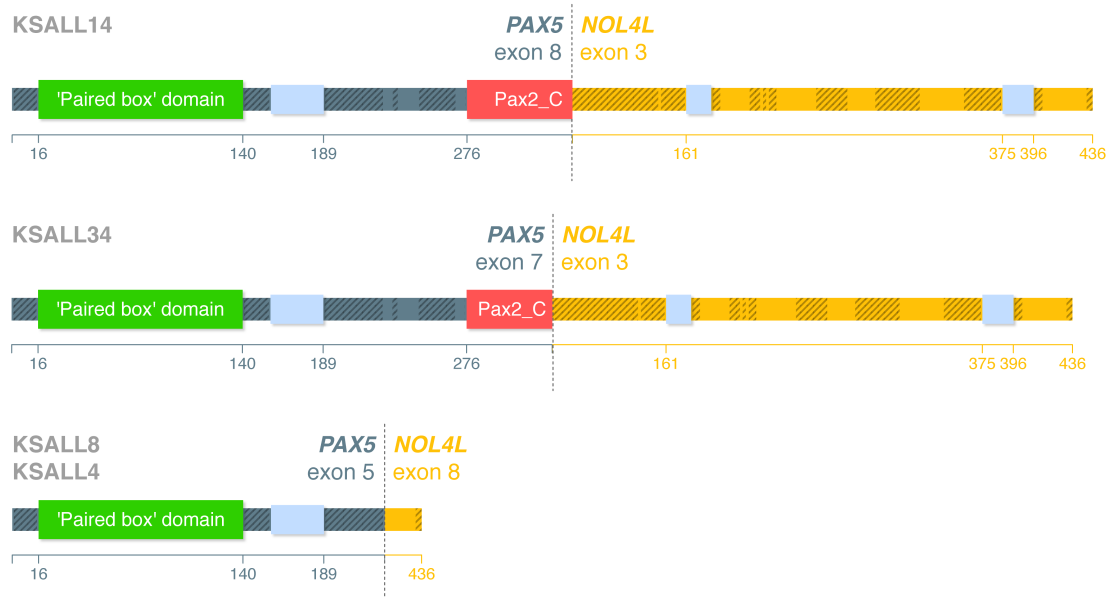


Figure 12: PAX5-NOL4L is a recurrent fusion gene in dic(9;20)-positive BCP ALL. Protein models of PAX5 (dark blue) and NOL4L (yellow) are shown in four cases, as detected by RNA-seq or WGS in paper IV and adapted from figure 2. The respective breakpoints are marked with vertical dotted lines, labeled with the corresponding exons involved in the fusion.

5.5 SECONDARY GENETIC ABERRATIONS ASSOCIATED WITH DIC(9;20)

In **papers II and IV**, we wanted to extend the spectrum of known genetic aberrations associated with the dic(9;20). Most strikingly, homozygous deletions of the tumor suppressor gene *CDKN2A* were present in all but one of 25 dic(9;20)-positive cases analyzed by SNP array (figure 13). Gene expression data also confirmed a significant suppression of *CDKN2A* expression in dic(9;20) as compared to other BCP ALLs ($-5.21 \log_2 \text{FC}$; adj $p < 0.001$). This locus encodes two independent tumor suppressors, P14ARF and P16INK4 from overlapping reading frames, and at least one copy of the gene is deleted in up to 35% of BCP ALL cases.^{71,127} Its association with treatment outcome has rendered conflicting results, possibly owing to the varying frequency of *CDKN2A* deletions between different subgroups.¹²⁷

In **paper II**, we showed that these *CDKN2A* deletions are present on the non-rearranged homologue of chromosome 9, one allele always lost as a result of the dic(9;20) rearrangement. The deletions were almost all below the resolution of FISH and their breakpoints appeared clustered. Extended to a larger number of cases in **paper IV**, the same clustering was evident. In 20 of 24 cases with homozygous deletion of *CDKN2A*, the lower end (centromeric) of the deletions all clustered tightly within *CDKN2B* (figure13). This is

suggestive of a common underlying mechanism, such as the ectopic activity of the RAG1/2 endonuclease complex.^{119,128} In one case, we could map the exact breakpoints of the deletion using WGS (figure 13, bottom panel). This revealed the presence of cryptic recombination signal sequences, required for RAG1 guiding, directly within the breakpoints of the deletion. The location was off by just a few base pairs from similar deletions in two previously published *ETV6-RUNX1* cases.¹¹⁹ In summary, homozygous deletions of *CDKN2A* deletions are present in almost all dic(9;20)-positive cases, more often than in other subgroups, and may be caused by aberrant RAG-mediated recombination.

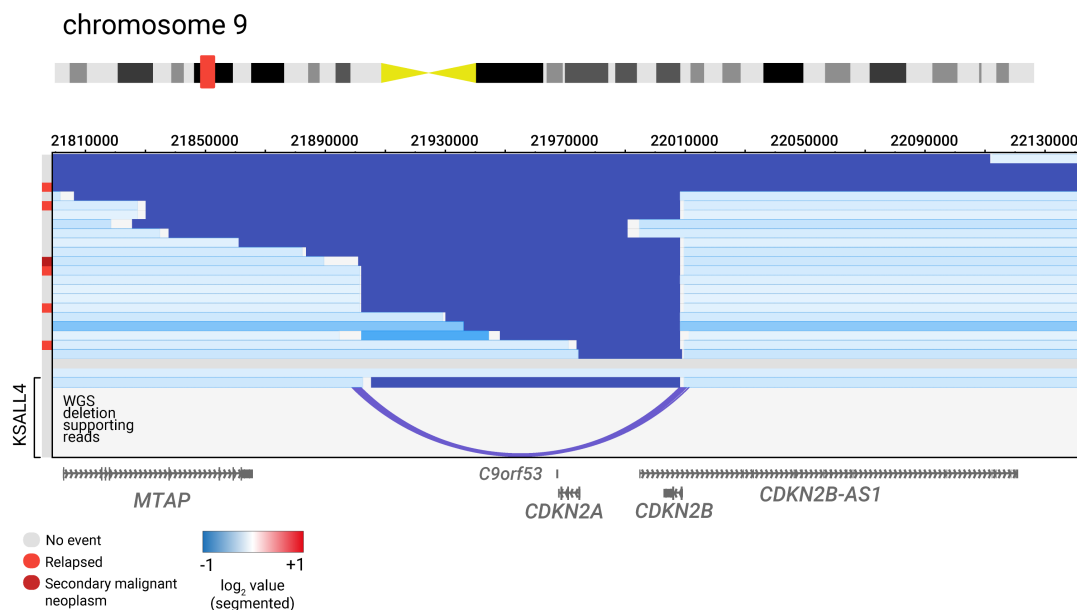


Figure 13: Homozygous deletions of *CDKN2A* occur in almost all dic(9;20)-positive BCP ALLs. Shown are 25 cases included in paper II and paper IV, with copy number states as segmented logR ratios (LRR) at 9p21.3. Light blue indicates heterozygous deletions, dark blue homozygous deletions. Clinical events are highlighted in the left column. Bottom panel shows case KSALL4, analyzed by WGS in paper IV, with read sequence pairs spanning the break in concordance with SNP array data.

In **paper IV**, we also detected an unexpectedly high frequency of heterozygous deletions of *IKZF1* (9/25; 36%) – a known predictor of poor outcomes across subgroups in childhood BCP ALL.^{73,75–78} We could not detect a difference in *IKZF1* expression as a result, consistent with a change in relative isoform distribution (increase in Ik6 levels) rather than overall expression.⁷³ In this dic(9;20) cohort, most primary events (4/6 total, including 5 relapses) occurred in cases with a deletion of *IKZF1* but was not statistically significant (chi-square 3.22; $p=0.072$). Considering the known prognostic effect of *IKZF1* deletions, it will be important to study this further in dic(9;20)-positive cases and understand to what extent they account for the risk of relapse in this subgroup.

Finally, we used RNA-sequence data to look for single-nucleotide variants of putative somatic origin in cases with dic(9;20). Five of 14 cases (35%) were shown to have mutations

in Ras/MAPK-pathway genes, including *KRAS*, *NRAS* and *PTPN11*. This is a comparable frequency to that of other BCP ALL subtypes,^{38,118,129} but it remains to be studied whether these represent subclonal events or important drivers in dic(9;20)-positive leukemia.

5.6 INTEGRATED GENOMIC PROFILING OF DIC(9;20)-POSITIVE CELLS

In **paper IV**, we also studied the overall gene expression and DNA methylation profiles of dic(9;20)-positive BCP ALL compared to other subgroups. Comparing RNA-seq data in 14 dic(9;20) cases against a set of 32 other BCP ALLs, including most major subtypes, hierarchical clustering revealed a distinct expression pattern in cases with dic(9;20). This was also true when limiting the analysis to non-coding genes alone, suggesting that the differentiating features of dic(9;20) are not limited to protein-coding genes.

Some of the most differentially expressed genes in dic(9;20)-positive cases were *CDKN2A*, *JAK2* and others affected by detected copy number alterations (homozygous deletion and loss of 9p, respectively). Other genes with significant changes in expression levels included lower levels of *DAPK1*, *SMADI* and increased expression of *KCNQ5*. Most of these findings were validated in an independent cohort of BCP ALLs, including six dic(9;20), that had gene expression data together with cytogenetic data available for analysis.³⁸ Further analysis of the gene expression data revealed four key gene sets (MSigDB, Hallmark set) that showed significant directional enrichment in dic(9;20)-positive cases, most significantly positive enrichment for genes involved in TNFalpha signaling via NFkB and apoptosis.

To understand the overall pattern of gene expression in dic(9;20)-positive BCP ALL, we integrated the gene expression data with DNA methylation data from the same cases.⁸⁰ Most of the differentially methylated CpGs in dic(9;20), as reported previously,⁸⁰ were annotated to genes not expressed in this cohort. However, comparing the genes included in a DNA methylation-based dic(9;20), 7 of 19 genes (37 CpG sites in total) also revealed significant differential expression in the RNA-seq data. Both *DAPK1* and *SMADI* showed increased levels of promoter hypermethylation in dic(9;20) – thus explaining the lower gene expression in the absence of structural or other sequence-level aberrations (figure 14A-B).

Death associated protein kinase 1 (DAPK1) is a key pro-apoptotic protein and regulator of autophagy, often hypermethylated in cancer and in particular in chronic lymphocytic leukemia.¹³⁰ *SMADI* is involved in the TGFbeta and BMP signaling cascade, and an important regulator of transcription. Interestingly, both *DAPK1* and *SMADI* were among the top downregulated genes in mice with short latency to leukemia development, when engrafted with primary cells from childhood BCP ALLs.¹³¹ Evaluating these sets of genes in dic(9;20)-positive cases showed a significant correlation for genes both up- and downregulated in the short-latency leukemia model. Taken together, these results reveal that somatic state epigenetic changes in dic(9;20)-positive BCP ALL confer significant changes to the expression of genes involved in cell survival signaling and regulation of programmed cell death. Further, the deregulation of these genes may be part of an expression profile,

constitutive to the leukemic cells, that is conducive to initiation and/or progression of disease.

Overexpression of *DAPK1* was previously reported to predict increased glucocorticoid resistance in childhood BCP ALL.¹³² Therefore, we investigated whether lower levels of *DAPK1* were correlated to differential sensitivity to a panel of anti-leukemic drugs. Comparing FMCA *in vitro* sensitivity data and gene expression in dic(9;20)-positive cases revealed an increased sensitivity to glucocorticoids with lower levels of *DAPK1* (figure 14C). Together with the data from the xenograft models, this suggests that the effect of *DAPK1* de-repression, using demethylating compounds for example, may render conflicting results in the context of existing chemotherapy regimens in childhood ALL – possibly removing strong growth advantages, but at the same time making cells less sensitive to common anti-leukemic agents.

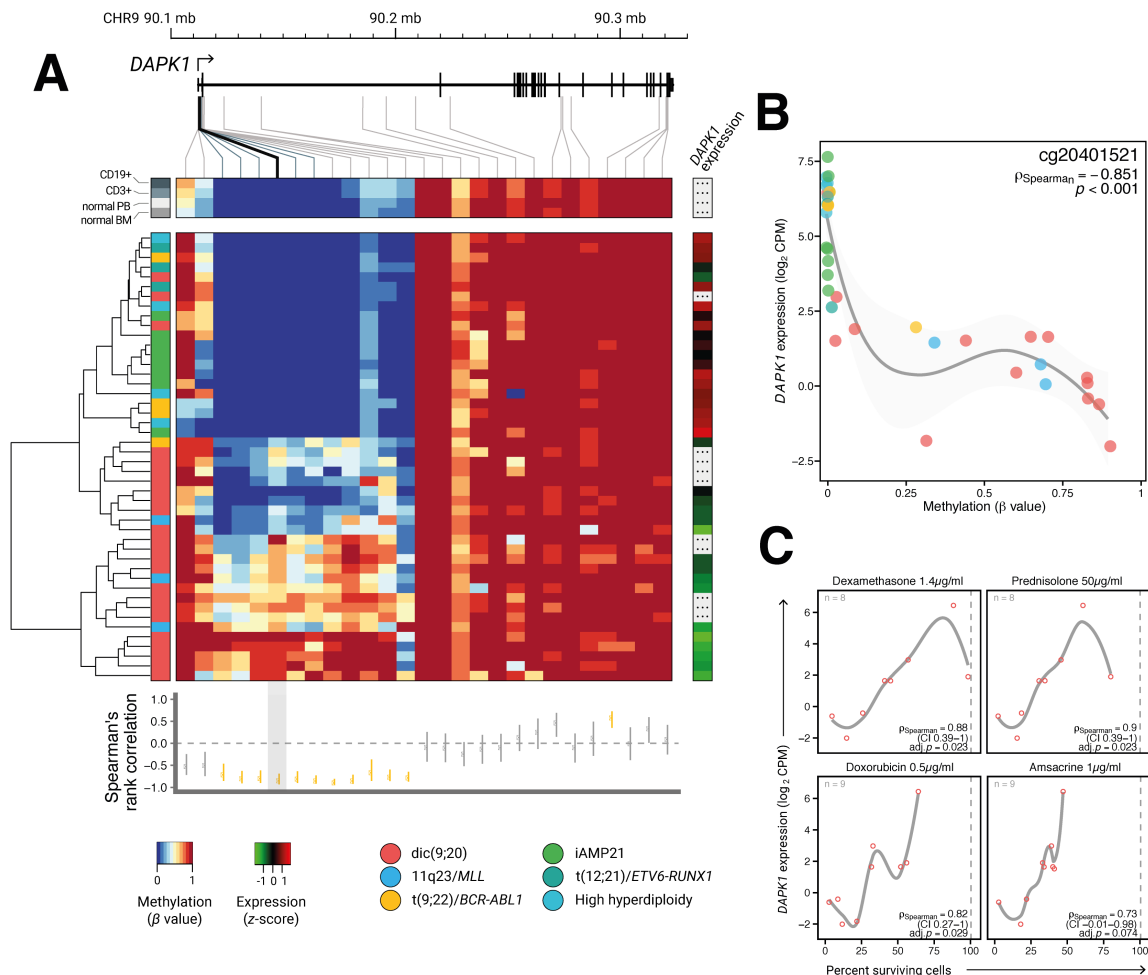


Figure 14: *DAPK1* is downregulated through promoter hypermethylation in dic(9;20)-positive BCP ALL and lower *DAPK1* expression correlates to increased glucocorticoid sensitivity. A) CpG methylation levels across the *DAPK1* gene model, with a heatmap of unsupervised hierarchical clustering on rows (blue – low methylation, red – high methylation). Below is the Spearman's rank correlation for each CpG to overall *DAPK1* expression, with confidence intervals as vertical bars and yellow marking significantly

correlated CpG sites. DAPK1 expression data in individual samples, were available, shown in a single column heatmap to the right. B) Correlation of DAPK1 expression and DNA methylation at one CpG site, highlighted with a thick black line in panel A. LOESS curves are fitted and individual samples colored by subgroup. Spearman rank correlation (-0.851, $p < 0.001$) is top right. C) DAPK1 expression compared to in vitro drug sensitivity, measured as percent surviving cells, in the top four drugs in the analysis.

In summary, in **paper IV**, we have presented an integrated analysis of structural and single nucleotide changes, combined with DNA methylation and gene expression profiling in dic(9;20)-positive BCP ALL. Together, these data extend the spectrum of known genetic aberrations in cases with dic(9;20) and highlight several key genes deregulated through epigenetic mechanisms – some of them conferring differential sensitivity to existing anti-leukemic agents.

6 CONCLUSIONS

The main conclusions of the studies presented in this thesis are:

- The dic(9;20)(p13.2;q11.2) chromosomal abnormality is more common than previously appreciated in childhood B-cell precursor ALL
- The dic(9;20) is often associated with non-standard risk clinical features at diagnosis, an increased risk of relapse and worse event-free survival compared to the most common BCP ALL subtypes
- The subtle nature of the rearrangement requires using a targeted FISH analysis for the reliable detection of all dic(9;20) cases in a clinical setting
- Fusion genes involving *PAX5* are recurrent in dic(9;20), but cannot account for the subtype or their overall expression profile
- The full spectrum of genetic abnormalities in dic(9;20)-positive cases include recurrent activating mutations in the Ras/MAPK pathway, near obligate loss of tumor suppressor *CDKN2A* by ectopic RAG endonuclease activity, and an enrichment of *IKZF1* deletions
- Genes controlling cell survival signaling and programmed cell death, including *DAPK1* and *UAP1*, are epigenetically downregulated in dic(9;20) – influencing the sensitivity to existing anti-leukemic agents and revealing rational targets of possible new therapies

7 FUTURE PERSPECTIVES

New molecularly targeted therapies and the emerging field of cancer immunotherapy promise to transform the way we treat and care for children with ALL, much in the way chemotherapeutic drugs first did in the previous century. Already today, an expanded portfolio of kinase inhibitors are proving useful in certain subtypes of ALL, and early data on chimeric antigen receptor T-cell engineering (CAR-T) in treating ALL are so promising that they look set to become the next landmark in childhood ALL therapy, possibly replacing stem-cell transplantations in those with the worst prognosis.

Resolving additional genetic aberrations, also in standard-risk ALLs, and how they relate to overall changes in somatic state epigenetic conformation, gene and protein expression, will continue in the years to come and render additional insights. With the techniques already available to us - and foreseeable further advancements in reading and editing the genome - these developments are squarely within a set of questions we already know to ask and will soon answer.

Together, an increased understanding of the genetic basis of disease, coupled with entirely new avenues of therapy, will provide never before seen opportunities to treat childhood ALL. The difference from previous advances is, of course, that childhood ALL today is a highly curable disease, far from the futile medical problem it once symbolized. Therefore, while emerging therapies have already started finding uses in high-risk cases, it is not immediately clear how to apply them for the benefit of the majority of children with ALL to afford a reduction in the intensity of chemotherapy treatment.

Our main challenge, as researchers and treating physicians, may in the end lie not only in finding all the different ways in which we can kill off a leukemic cell – but how to test and implement the best ones without jeopardizing previous gains; and to do so in an overall population of childhood ALLs of predominantly low-risk disease.

In parallel, we are tasked to use what is already known and available to us today, to increase access to modern diagnostics and adequate care for the 80 percent of children diagnosed with cancer that are treated outside contemporary protocols and care facilities. This highlights the fact that, globally, there is already today large outcome gains to be made in the care of children with cancer. In terms of diagnostics, new low-cost and widely applicable solutions for the analysis of genetic abnormalities are sure to be ahead of us, as the fields of molecular genetics, engineering and computer science intersect more often.

For all these reasons, it should be evident that a cure for all children with ALL is within reach, but that it will come only from a detailed understanding of the genetics of the disease, the determinants of therapy response, and the careful consideration of how to best leverage this clinically.

8 ACKNOWLEDGEMENTS

I am indebted to our patients, their families and treating physicians, without whom this research would not have been possible.

My main supervisor, **Ann Nordgren**, for her unwavering trust and support, leaving me to explore any and all ideas, and for bringing an unparalleled excitement to what we do, in every meeting, every single encounter, over the last four years. My co-supervisors, **Magnus Nordenskjöld**, **Gisela Barbany** and **Erik Forestier** for taking me on as a student. For always being generous with their time and patience through the course of this work.

The rest of the leukemia genetics working group: **Ingegerd Ivanov Öfverholm**, **Anh-Nhi Tran**, **Fulya Taylan**; for sharing all the gritty work in planning, executing and interpreting much of the work presented here.

To all the researchers and physicians that contributed to and co-authored the studies included in this thesis, proving that the most interesting and rewarding science is done in collaboration. In particular **Jessica Nordlund**, who has been instrumental to much of work presented in this thesis. **Ann-Christine Syvänen**, **Gudmar Lönnerholm**, **Yanara Marincevic-Zuniga** for successful collaborations. To **Dan Grandér** and **Katja Pokrovskaja**, for the collaboration and their time and many discussions.

To pediatric oncologists across the Nordic countries, and associated scientists within NOPHO and the NOPHO ALL Biology Working group. In particular **Mats Heyman**, **Arja Harila-Saari**, **Kjeld Schmiegelow**, **Olli Lohi** who have all been the very best representatives of physician scientists; for being passionate and unrelenting advocates of inclusive collaborations with the patients at heart, and for putting trust in me and other young scientists to help shape the pediatric cancer research of tomorrow.

The **Clinical Genetics research group** at CMM, past and present. In particular **Bianca Tesi** who, more than anyone, has had to listen to my incessant talking during the last few years; and for becoming data-literate together. **Tobias**, **Alexandra**, **Wolfgang**, **Daniel**, **Anna**, **Marie**, **Johanna**, **Anders**, **Malin**, **Ellika**, **Sigrid**, **Nina**, **Josefine**, **Maria**, **Miriam**, **Samina**, **Alice**, **Alisa**, **Paris**.

All clinicians and laboratory staff at the Clinical Genetics hospital service.

All the incredibly patient and helpful administrative personnel at the Department of Molecular Medicine and Surgery.

I am also grateful for the financial support of Karolinska Institutet, Stockholms läns landsting, Barncancerfonden, and Mery Bevé foundation. For funding this work and allowing me to travel and learn science around the world during these years.

My family.

My love.

9 REFERENCES

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304–1351 (2001).
3. Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* **470**, 187–197 (2011).
4. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
5. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome — biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734 (2011).
6. Yang, Y. *et al.* Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *N. Engl. J. Med.* **369**, 1502–1511 (2013).
7. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
8. Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat. Rev. Genet.* **advance online publication**, (2013).
9. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
10. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
11. Stephens, P. J. *et al.* Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* **144**, 27–40 (2011).
12. Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* **11**, 685–696 (2010).
13. Fredriksson, N. J., Ny, L., Nilsson, J. A. & Larsson, E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nat. Genet.* **46**, 1258–1263 (2014).
14. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
15. Kandoth, C. *et al.* Mutational landscape and significance across 12 major

cancer types. *Nature* **502**, 333–339 (2013).

16. Landau, D. A., Carter, S. L., Getz, G. & Wu, C. J. Clonal evolution in hematologic malignancies and therapeutic implications. *Leukemia* (2013). doi:10.1038/leu.2013.248
17. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature advance online publication*, (2012).
18. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
19. Pui, C.-H., Robison, L. L. & Look, A. T. Acute lymphoblastic leukaemia. *The Lancet* **371**, 1030–1043 (2008).
20. Bhojwani, D., Yang, J. J. & Pui, C.-H. Biology of Childhood Acute Lymphoblastic Leukemia. *Pediatr. Clin. North Am.* **62**, 47–60 (2015).
21. Howard, S. C. *et al.* Childhood cancer epidemiology in low-income countries. *Cancer* **112**, 461–472 (2008).
22. Gustafsson, G. *et al.* Improving outcome through two decades in childhood ALL in the Nordic countries: the impact of high-dose methotrexate in the reduction of CNS irradiation. Nordic Society of Pediatric Haematology and Oncology (NOPHO). *Leukemia* **14**, 2267–75 (2000).
23. Pui, C.-H. Genomic and pharmacogenetic studies of childhood acute lymphoblastic leukemia. *Front. Med.* 1–9 (2014). doi:10.1007/s11684-015-0381-3
24. Belson, M., Kingsley, B. & Holmes, A. Risk Factors for Acute Leukemia in Children: A Review. *Environ. Health Perspect.* **115**, 138–145 (2007).
25. Farber, S., Diamond, L. K., Mercer, R. D., Sylvester, R. F. & Wolff, J. A. Temporary Remissions in Acute Leukemia in Children Produced by Folic Acid Antagonist, 4-Aminopteroyl-Glutamic Acid (Aminopterin). *N. Engl. J. Med.* **238**, 787–793 (1948).
26. Pui, C.-H. & Evans, W. E. A 50-Year Journey to Cure Childhood Acute Lymphoblastic Leukemia. *Semin. Hematol.* **50**, 185–196 (2013).
27. Farber, S. *Advances in cancer research*. (Academic Press, 1956).
28. Frei, E. *et al.* Studies of Sequential and Combination Antimetabolite Therapy in Acute Leukemia: 6-Mercaptopurine and Methotrexate. *Blood* **18**, 431–454 (1961).
29. Aur, R. J. A. *et al.* Central Nervous System Therapy and Combination Chemotherapy of Childhood Lymphocytic Leukemia. *Blood* **37**, 272–281 (1971).
30. Schmiegelow, K. *et al.* Long-term results of NOPHO ALL-92 and ALL-2000

- studies of childhood acute lymphoblastic leukemia. *Leukemia* **24**, 345–54 (2010).
31. Druker, B. J. *et al.* Activity of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in the Blast Crisis of Chronic Myeloid Leukemia and Acute Lymphoblastic Leukemia with the Philadelphia Chromosome. *N. Engl. J. Med.* **344**, 1038–1042 (2001).
 32. Druker, B. J. *et al.* Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia. *N. Engl. J. Med.* **344**, 1031–1037 (2001).
 33. Biondi, A. *et al.* Imatinib after induction for treatment of children and adolescents with Philadelphia-chromosome-positive acute lymphoblastic leukaemia (EsPhALL): a randomised, open-label, intergroup study. *Lancet Oncol.* **13**, 936–945 (2012).
 34. Jacobs, K. B. *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* **44**, 651–658 (2012).
 35. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
 36. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
 37. Plon, S. E. *et al.* Abstract 11: Evaluating cancer susceptibility mutations and incidental findings from whole exome sequencing of sequentially diagnosed pediatric solid and brain tumor patients: Early results of the BASIC3 study. *Cancer Res.* **74**, 11–11 (2014).
 38. Holmfeldt, L. *et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat. Genet.* (2013). doi:10.1038/ng.2532
 39. Buitenkamp, T. D. *et al.* Acute lymphoblastic leukemia in children with Down syndrome: a retrospective analysis from the Ponte di Legno study group. *Blood* **123**, 70–77 (2014).
 40. Hasle, H., Clemmensen, I. H. & Mikkelsen, M. Risks of leukaemia and solid tumours in individuals with Down's syndrome. *The Lancet* **355**, 165–169 (2000).
 41. Hertzberg, L. *et al.* Down syndrome acute lymphoblastic leukemia, a highly heterogeneous disease in which aberrant expression of CRLF2 is associated with mutated JAK2: a report from the International BFM Study Group. *Blood* **115**, 1006–1017 (2010).
 42. Bercovich, D. *et al.* Mutations of JAK2 in acute lymphoblastic leukaemias associated with Down's syndrome. *The Lancet* **372**, 1484–1492 (2008).
 43. Seif, A. E. Pediatric leukemia predisposition syndromes: clues to understanding leukemogenesis. *Cancer Genet.* **204**, 227–244 (2011).
 44. Tartaglia, M. *et al.* Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat. Genet.* **34**, 148–150

(2003).

45. Schmiegelow, K. *et al.* High concordance of subtypes of childhood acute lymphoblastic leukemia within families: lessons from sibships with multiple cases of leukemia. *Leukemia* **26**, 675–681 (2012).
46. Zhang, M. Y. *et al.* Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat. Genet.* **47**, 180–185 (2015).
47. Noetzli, L. *et al.* Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat. Genet.* **advance online publication**, (2015).
48. Auer, F. *et al.* Inherited susceptibility to pre B-ALL caused by germline transmission of PAX5 c.547G>A. *Leukemia* (2013). doi:10.1038/leu.2013.363
49. Shah, S. *et al.* A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nat. Genet.* **45**, 1226–1231 (2013).
50. Shah, S. *et al.* A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nat. Genet.* **advance online publication**, (2013).
51. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five Years of GWAS Discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
52. Papaemmanuil, E. *et al.* Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat. Genet.* **41**, 1006–1010 (2009).
53. Treviño, L. R. *et al.* Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat. Genet.* **41**, 1001–1005 (2009).
54. Sherborne, A. L. *et al.* Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nat Genet* **42**, 492–494 (2010).
55. Perez-Andreu, V. *et al.* Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat. Genet.* (2013). doi:10.1038/ng.2803
56. Xu, H. *et al.* Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations. *J. Natl. Cancer Inst.* (2013). doi:10.1093/jnci/djt042
57. Migliorini, G. *et al.* Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype. *Blood* (2013). doi:10.1182/blood-2013-03-491316
58. Rudant, J. *et al.* ARID5B, IKZF1 and Non-Genetic Factors in the Etiology of Childhood Acute Lymphoblastic Leukemia: The ESCALE Study. *PLoS ONE* **10**, e0121348

(2015).

59. Ellinghaus, E. *et al.* Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic. *Leukemia* **26**, 902–909 (2012).
60. Krynetski, E. Y. *et al.* A single point mutation leading to loss of catalytic activity in human thiopurine S-methyltransferase. *Proc. Natl. Acad. Sci.* **92**, 949–953 (1995).
61. Relling, M. V. *et al.* Mercaptopurine Therapy Intolerance and Heterozygosity at the Thiopurine S-Methyltransferase Gene Locus. *J. Natl. Cancer Inst.* **91**, 2001–2008 (1999).
62. Schmiegelow, K. *et al.* Thiopurine methyltransferase activity is related to the risk of relapse of childhood acute lymphoblastic leukemia: results from the NOPHO ALL-92 study. *Leukemia* **23**, 557–564 (2008).
63. Pui, C.-H. & Evans, W. E. Treatment of Acute Lymphoblastic Leukemia. *N Engl J Med* **354**, 166–178 (2006).
64. Wesolowska-Andersen, A. *et al.* Genomic profiling of thousands of candidate polymorphisms predicts risk of relapse in 778 Danish and German childhood acute lymphoblastic leukemia patients. *Leukemia* **29**, 297–303 (2015).
65. Moorman, A. V. The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. *Blood Rev.* **26**, 123–135 (2012).
66. *WHO classification of tumours of haematopoietic and lymphoid tissues: [... reflects the views of a working group that convened for an Editorial and Consensus Conference at the International Agency for Research on Cancer (IARC), Lyon, October 25 - 27, 2007].* (Internat. Agency for Research on Cancer, 2008).
67. Harrison, C. J. Targeting signaling pathways in acute lymphoblastic leukemia: new insights. *ASH Educ. Program Book* **2013**, 118–125 (2013).
68. Yeoh, E.-J. *et al.* Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1**, 133–143 (2002).
69. Ross, M. E. *et al.* Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* **102**, 2951–2959 (2003).
70. Kuiper, R. P. *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* **21**, 1258–66 (2007).
71. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).

72. Kawamata, N. *et al.* Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* **111**, 776–84 (2008).
73. Mullighan, C. G. *et al.* Deletion of IKZF1 and Prognosis in Acute Lymphoblastic Leukemia. *N Engl J Med* **360**, 470–480 (2009).
74. Mullighan, C. G. *et al.* CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature* **471**, 235–239 (2011).
75. Dörge, P. *et al.* IKZF1 deletion is an independent predictor of outcome in pediatric acute lymphoblastic leukemia treated according to the ALL-BFM 2000 protocol. *Haematologica* **98**, 428–432 (2013).
76. Olsson, L. *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* **28**, 302–310 (2014).
77. Öfverholm, I. *et al.* Impact of IKZF1 deletions and PAX5 amplifications in pediatric B-cell precursor ALL treated according to NOPHO protocols. *Leukemia* (2013). doi:10.1038/leu.2013.92
78. Kuiper, R. P. *et al.* IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. *Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund UK* (2010). doi:10.1038/leu.2010.87
79. Chen, J., Odenike, O. & Rowley, J. D. Leukaemogenesis: more than mutant genes. *Nat. Rev. Cancer* **10**, 23–36 (2010).
80. Nordlund, J. *et al.* Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* **14**, r105 (2013).
81. Nordlund, J. *et al.* DNA methylation-based subtype prediction for pediatric acute lymphoblastic leukemia. *Clin. Epigenetics* **7**, (2015).
82. Rieder, H. *et al.* dic(9;20): a new recurrent chromosome abnormality in adult acute lymphoblastic leukemia. *Genes. Chromosomes Cancer* **13**, 54–61 (1995).
83. Slater, R. *et al.* A non-random chromosome abnormality found in precursor-B lineage acute lymphoblastic leukaemia: dic(9;20)(p1?3;q11). *Leukemia* **9**, 1613–1619 (1995).
84. Heerema, N. A. *et al.* Dicentric (9;20)(p11;q11) identified by fluorescence in situ hybridization in four pediatric acute lymphoblastic leukemia patients. *Cancer Genet. Cytogenet.* **92**, 111–115 (1996).
85. Clark, R. *et al.* Monosomy 20 as a pointer to dicentric (9;20) in acute lymphoblastic leukemia. *Leukemia* **14**, 241–6 (2000).

86. Raimondi, S. C. *et al.* Reassessment of the prognostic significance of hypodiploidy in pediatric patients with acute lymphoblastic leukemia. *Cancer* **98**, 2715–22 (2003).
87. Van Zutven, L. J. C. M. *et al.* CDKN2 deletions have no prognostic value in childhood precursor-B acute lymphoblastic leukaemia. *Leukemia* **19**, 1281–1284 (2005).
88. Jarosova, M. *et al.* Complex karyotypes in childhood acute lymphoblastic leukemia: cytogenetic and molecular cytogenetic study of 21 cases. *Cancer Genet Cytogenet* **145**, 161–8 (2003).
89. Schoumans, J. *et al.* Characterisation of dic(9;20)(p11-13;q11) in childhood B-cell precursor acute lymphoblastic leukaemia by tiling resolution array-based comparative genomic hybridisation reveals clustered breakpoints at 9p13.2 and 20q11.2. *Br. J. Haematol.* **135**, 492–499 (2006).
90. Forestier, E. *et al.* Clinical and cytogenetic features of pediatric dic(9;20)(p13.2;q11.2)-positive B-cell precursor acute lymphoblastic leukemias: a Nordic series of 24 cases and review of the literature. *Genes. Chromosomes Cancer* **47**, 149–158 (2008).
91. Strefford, J. C. *et al.* Genome complexity in acute lymphoblastic leukemia is revealed by array-based comparative genomic hybridization. *Oncogene* **26**, 4306–18 (2007).
92. An, Q. *et al.* Variable breakpoints target PAX5 in patients with dicentric chromosomes: A model for the basis of unbalanced translocations in cancer. *Proc. Natl. Acad. Sci.* **105**, 17050–17054 (2008).
93. An, Q. *et al.* Heterogeneous breakpoints in patients with acute lymphoblastic leukemia and the dic(9;20)(p11-13;q11) show recurrent involvement of genes at 20q11.21. *Haematologica* **94**, 1164–1169 (2009).
94. Kawamata, N. *et al.* Cloning of genes involved in chromosomal translocations by high-resolution single nucleotide polymorphism genomic microarray. *Proc. Natl. Acad. Sci.* **105**, 11921–11926 (2008).
95. Mitelman, F., Johansson, B. & Mertens, F. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. (2015). at <http://cgap.nci.nih.gov/Chromosomes/Mitelman>
96. Moorman, A. V. *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. *Lancet Oncol* **11**, 429–38 (2010).
97. Pichler, H. *et al.* Prognostic relevance of dic(9;20)(p11;q13) in childhood B-cell precursor acute lymphoblastic leukaemia treated with Berlin-Frankfurt-Münster (BFM) protocols containing an intensive induction and post-induction consolidation therapy. *Br. J.*

Haematol. **149**, 93–100 (2010).

98. Team, R. C. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. (ISBN 3-900051-07-0, 2012).
99. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657–663 (2007).
100. Rasmussen, M. *et al.* Allele-specific copy number analysis of tumor samples with aneuploidy and tumor heterogeneity. *Genome Biol.* **12**, R108 (2011).
101. MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* **42**, D986–992 (2014).
102. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
103. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotech* **29**, 24–26 (2011).
104. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
105. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
106. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
107. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* **30**, 923–930 (2014).
108. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
109. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
110. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* gkv007 (2015).
doi:10.1093/nar/gkv007
111. Den Boer, M. L. *et al.* A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol.* **10**, 125–134 (2009).

112. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
113. Kasprzyk, A. BioMart: driving a paradigm change in biological data management. *Database* **2011**, bar049 (2011).
114. Wu, D. *et al.* ROAST: rotation gene set tests for complex microarray experiments. *Bioinforma. Oxf. Engl.* **26**, 2176–2182 (2010).
115. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
116. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
117. Paila, U., Chapman, B. A., Kirchner, R. & Quinlan, A. R. GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations. *PLoS Comput Biol* **9**, e1003153 (2013).
118. Roberts, K. G. *et al.* Targetable Kinase-Activating Lesions in Ph-like Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* **371**, 1005–1015 (2014).
119. Papaemmanuil, E. *et al.* RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat. Genet.* (2014). doi:10.1038/ng.2874
120. Nicorici, D. *et al.* FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. *bioRxiv* 011650 (2014). doi:10.1101/011650
121. Lonnerholm, G. *et al.* Dic(9;20)(p13;q11) in childhood acute lymphoblastic leukaemia is related to low cellular resistance to asparaginase, cytarabine and corticosteroids. *Leukemia* **23**, 209–12 (2009).
122. Cobaleda, C., Schebesta, A., Delogu, A. & Busslinger, M. Pax5: the guardian of B cell identity and function. *Nat Immunol* **8**, 463–470 (2007).
123. Fortschegger, K., Anderl, S., Denk, D. & Strehl, S. Functional Heterogeneity of PAX5 Chimeras Reveals Insight for Leukemia Development. *Mol. Cancer Res.* molcanres.0337.2013 (2014). doi:10.1158/1541-7786.MCR-13-0337
124. Strehl, S., Konig, M., Dworzak, M. N., Kalwak, K. & Haas, O. A. PAX5/ETV6 fusion defines cytogenetic entity dic(9;12)(p13;p13). *Leukemia* **17**, 1121–3 (2003).
125. Schinnerl, D. *et al.* The role of the Janus-faced transcription factor PAX5-JAK2 in acute lymphoblastic leukemia. *Blood* blood–2014–04–570960 (2014). doi:10.1182/blood-2014-04-570960

126. Kawamata, N., Pennella, M. A., Woo, J. L., Berk, A. J. & Koeffler, H. P. Dominant-negative mechanism of leukemogenic PAX5 fusions. *Oncogene* **31**, 966–977 (2012).
127. Sulong, S. *et al.* A comprehensive analysis of the CDKN2A gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. *Blood* **113**, 100–107 (2009).
128. Waanders, E. *et al.* The Origin and Nature of Tightly Clustered BTG1 Deletions in Precursor B-Cell Acute Lymphoblastic Leukemia Support a Model of Multiclonal Evolution. *PLoS Genet* **8**, e1002533 (2012).
129. Paulsson, K. *et al.* Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci.* **107**, 21719–21724 (2010).
130. Raval, A. *et al.* Downregulation of Death-Associated Protein Kinase 1 (DAPK1) in Chronic Lymphocytic Leukemia. *Cell* **129**, 879–890 (2007).
131. Meyer, L. H. *et al.* Early Relapse in ALL Is Identified by Time to Leukemia in NOD/SCID Mice and Is Characterized by a Gene Signature Involving Survival Pathways. *Cancer Cell* **19**, 206–217 (2011).
132. Holleman, A. *et al.* The expression of 70 apoptosis genes in relation to lineage, genetic subtype, cellular drug resistance, and outcome in childhood acute lymphoblastic leukemia. *Blood* **107**, 769–776 (2006).