Molecular Genetic Characterization of Acute Lymphoblastic Leukemia with a Poor Prognosis

Setareh Safavi

LUND UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended in Belfragesalen, BMC, Lund.
On Friday October 16, 2015, at 13:00 PM.

Faculty opponent
Dr Julie Irving

Northern Institute for Cancer Research, Newcastle University
Newcastle upon Tyne, United Kingdom
Abstract
Acute lymphoblastic leukemia (ALL) affects individuals at all ages, with peak incidences in children <4 years and adults >50 years. ALL is broadly categorized into B-cell precursor (BCP) and T-cell ALL with specific clinical features associated with outcome. In contrast to pediatric ALL, which has a favorable prognosis, adult ALL is associated with a much poorer outcome with less than 40% overall survival rates, decreasing with higher age. The presence of specific acquired genetic abnormalities is important for diagnosis, prognostication, and treatment stratification. ALL can be further categorized into subgroups defined by structural or ploidy abnormalities. One such subgroup, hypodiploid ALL (<46 chromosomes) is seen in 5-8% of all cases, and associated with a very dismal prognosis. It can be further subdivided into two distinct genetic and clinical subgroups, namely near-haploidy (24-31 chromosomes) and low hypodiploidy (32-39 chromosomes), and, comprising cases with a more heterogenous background, high hypodiploidy (40-43 chromosomes) and cases with 44 and 45 chromosomes. Near-haploid and low hypodiploid ALL are very rare, comprising less than 1% of BCP ALL, with overall survival rates of <30%.

The general aim of my PhD project has been to characterize ALL patients with a poor prognosis, including adult ALL (article I) and hypodiploid ALL (article II-IV). To investigate the genetic landscape of adult ALL, we performed SNP array analysis on 126 ALL cases. Characteristic deletions seen in pediatric ALL were detected, furthermore, comparison of diagnostic and relapse clonal relationship showed evolution from an ancestral clone in the majority of cases, highlighting similarities in childhood and adult disease. In addition, the analysis revealed several recurrent cryptic genetic events not previously implicated with adult ALL, including the BCAT1, BTLA, NR3C1, PIK3AP1 and SERP2 genes. In articles II-IV the genetic and epigenetic background of hypodiploid ALL was further investigated using SNP array analysis, exome and RNA sequencing, methylation array analysis and FISH analysis. Characteristic chromosomal patterns were confirmed and subtype specific alterations targeting IKZF3, NF1, FLT3 and TP53 were identified near-haploid and low hypodiploid respectively. Furthermore, due to the specific pattern of CDKN2A deletions in one case, we could conclude that chromosomal loss was the primary event with further microdeletions occurring after the near-haploidization. Combining SNP array and FISH analysis, a subclonal pattern was detected in three cases harboring >79 chromosomes, showing a possible hypodiploid origin due to the extensive loss of heterozygosity identified in such cases. That all three cases harbored TP53 mutations emphasized similarities to low hypodiploid ALL. In conclusion, screening for specific genetic abnormalities routinely in the clinic may improve prognostication and treatment stratification in cases with a poor prognosis.

Key words: Acute lymphoblastic leukemia, adult ALL, genetic characterization, hypodiploid ALL

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.
Molecular Genetic Characterization of Acute Lymphoblastic Leukemia with a Poor Prognosis

Setareh Safavi

Division of Clinical Genetics
Department of Laboratory Medicine
Faculty of Medicine
Lund University
2015
Contents

Original Articles 7
Abbreviations 8
Preface 9
Acute lymphoblastic leukemia
A brief introduction 11
Acquired chromosomal abnormalities 11
Cytogenetic and molecular genetic techniques 12
Adult acute lymphoblastic leukemia
The genetic landscape of ALL 15
Prognosis and therapy 17
Hypodiploid acute lymphoblastic leukemia 19
A glance into the intricate realm of hypodiploidy
A tale of disomies 20
Delineation of hypodiploid ALL 21
Clinical features of hypodiploid ALL 22
Masked hypodiploidy and its clinical significance 23
Origins and consequences of hypodiploidy 24
Acquired genetic aberrations in hypodiploid ALL 26
The present study 29
Focus and aims 29
Material and methods
Patient material 29
SNP arrays 30
Exome and RNA sequencing 32
Methylation array 32
FISH analyses 33
Results, discussion and concluding remarks
Article I 33
Articles II & III 34
Article IV 37
Original Articles

This thesis is based on the following articles, which will be referred to in the text by their Roman numerals.


## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>Acute lymphoblastic leukemia</td>
</tr>
<tr>
<td>BCP</td>
<td>B-cell precursor</td>
</tr>
<tr>
<td>BM</td>
<td>Bone marrow</td>
</tr>
<tr>
<td>EFS</td>
<td>Event free survival</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridization</td>
</tr>
<tr>
<td>LOH</td>
<td>Loss of heterozygosity</td>
</tr>
<tr>
<td>MRD</td>
<td>Minimal residual disease</td>
</tr>
<tr>
<td>NGS</td>
<td>Next-generation sequencing</td>
</tr>
<tr>
<td>NOPHO</td>
<td>Nordic society of Pediatric Hematology and Oncology</td>
</tr>
<tr>
<td>OS</td>
<td>Overall survival</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>RNA sequencing</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>UPID</td>
<td>Uniparental isodiosomy</td>
</tr>
<tr>
<td>WBC</td>
<td>White blood cell</td>
</tr>
</tbody>
</table>
Preface

The German pathologist Rudolph Virchows theory that cells ascend from other cells – *omnis cellula e cellula* - has provided the foundation of our understanding of neoplastic transformation.

Cells generating more cells reflect a normally very tightly regulated process, i.e., cell division, stimulated by certain signals and halted by other signals. Cancer cells have acquired specific changes, allowing them to ignore these signals leading to uncontrolled self-renewal, proliferation, differentiation and survival. Uncontrollable growth is one feature all cancers have in common, producing generations of clones, with one common genetic ancestor. We would be underestimating the cancer by just referring to it as a clonal disease. This disease entity is truly a clonally evolving disease. With every new generation of descendants, spawning clones genetically diverse, with novel acquired aberrations, allowing them to survive, grow and become therapy resistant. Much like a *Trojan horse*, the ancestral clone, with its acquired abilities, relentlessly grows and expands, taking over normal tissue and spreading to distant parts - a threat unforeseen. For years, we have been mimicking the cancer cells environment, following its every move, learning and analyzing its changes, and finding ways to install drugs to fool the cancer with our very own “*Trojan Horse*”, to overcome the battle. In many cases, we are victorious, but the battle continues in those where we are not successful.

The main goals of my PhD study have been to genetically characterize cases with a poor prognosis, comprising of (i) adult ALL, with current cure rates of <40%, investigated in article I and (ii) hypodiploid ALL in children and adults, seen in up to 5-8% of cases, associated with a very poor prognosis and dismal overall survival rates, investigated in articles II, III and IV, in order to detect the underlying genetic mechanisms.

*Lund, September 2015*
Acute lymphoblastic leukemia

A brief introduction

Leukemia has been coined as *neoplasia of blood* defined by the abnormal growth of white blood cells. By the early 1900s leukemia had been characterized depending on its proliferative rate, to be acute with a rapid growth, or, it could develop slowly and was later termed as chronic. Cell of origin can further subdivide cases into myeloid or lymphoid leukemia.

Acute lymphoblastic leukemia (ALL) is a heterogeneous genetic disease caused by acquired genetic aberrations in hematopoietic stem cells or precursor cells, resulting in the rapid expansion of immature white blood cells, i.e., blasts, in the bone marrow (BM) and in the peripheral blood (Pui, 2008). The disease is classified broadly as B- or T-lineage ALL (T-ALL). The former can be subcategorized into B-cell precursor ALL (BCP ALL), and cases with mature B immunophenotype. BCP ALL is the most common childhood malignancy, comprising a majority of cases and associated with a good prognosis (Schmiegelow *et al*; 2010, Pui, 2012; Pui & Evans, 2013). A distinct peak of incidence between 2 to 6 years of age and another peak in older patients over 50 years of age is seen (Hjalgrim *et al*, 2003; Forestier & Schmiegelow, 2006; Pui *et al*, 2008; Inaba *et al*, 2013). In contrast to childhood cases, adults have a very poor prognosis with long-term survival rates of 40% decreasing with higher age (Chessells *et al*, 1998; Rowe, 2010; Faderl *et al*, 2010).

T-lineage ALL represents a heterogeneous high-risk disease entity targeting the thymocytes (Pui, 2004). In contrast to BCP ALL, a male predominance is seen and the median age is 9 years.

Acquired chromosomal abnormalities

That malignant cells harbor chromosomal abnormalities promoting malignant transition was already stated in 1914 by Theodor Boveri (Boveri, 1914). With time and progress his idea could be tested, and cytogenetic studies on dividing cells, i.e., metaphases, allowed the subsequent characterization of many acquired genetic aberrations. Cytogenetic and
molecular genetic studies have revolutionized our understanding of acquired genetic aberrations and are pivotal for diagnostic and prognostic purposes (Forestier et al., 2000; Johansson et al., 2004; Schultz et al., 2007; Harrison, 2009; Moorman et al., 2010). Characteristic hallmarks of acute leukemia are structural and numerical abnormalities (Mitelman, 2015). The former includes balanced and unbalanced rearrangements; balanced rearrangements are in many cases the primary aberration, an exclusive finding in the leukemic cells and thus vital for leukemogenesis (Johansson et al., 1996; Greaves & Wiemels, 2003). These aberrations can be interchromosomal i.e., translocations, or intrachromosomal exemplified by inversions. Consequently, the effects of such rearrangements result in the activation of oncogenes, and may produce novel fusion genes with oncogenic functions (Rabbitts, 1994; Mitelman et al., 2007). Unbalanced rearrangements are caused by unbalanced translocations, deletions, or loss of whole chromosomes, i.e., monosomies.

Numerical abnormalities comprise cytogenetic subgroups characterized by changes in ploidy, i.e., modal chromosome numbers with loss or gain of whole chromosomes. The reasons for gain or loss of chromosomal sets are unknown; however, it is believed that gene dosage effects play a key role (Paulsson & Johansson, 2009).

Finally, amplifications/duplications, cryptic microdeletions, epigenetic alterations and sequence mutations are common in leukemia and have been shown to promote leukemic transformation (Pui et al., 2004; Kuiper et al., 2007; Mullighan et al., 2007; Harrison et al., 2009; Zhang et al., 2011; Downing et al., 2012; Inaba et al., 2013).

**Cytogenetic and molecular genetic techniques**

Cytogenetic studies thrived in the 1970s with the introduction of chromosome banding techniques. The first balanced translocation identified in cancer was the t(9;22)(q34;q11) with the BCR/ABL1 fusion, also known as the Philadelphia chromosome. Although a hallmark of chronic myeloid leukemia, this translocation is also present in ALL (Propp & Lizzi 1970; Rowley, 1973; Pui et al., 2004).

Fluorescence in situ hybridization (FISH) and reverse transcription-polymerase chain reaction (RT-PCR) allowed the subsequent identification of many cytogenetically cryptic abnormalities, further characterizing subgroups. The former technique applies fluorescently labeled probes, hybridizing to target sequences in interphase or metaphase cells, the latter technique is advantageous in detecting gene fusions (Pinkel et al., 1986; Cremer et al., 1988; Langabeer et al., 1997). Significantly, both methods are
used routinely for diagnostic purposes in the clinic to detect common fusions, such as the t(12;21)(p13;q22).

Ploidy abnormalities are a common finding in ALL, and DNA indexing has been an immensely useful technique for detecting such changes.

With the development of array-based high resolution screening techniques, such as single nucleotide polymorphism arrays (SNP arrays), characterization of alterations on gene level and on single base pairs was made possible (Schaaf et al, 2011). Another milestone was reached in the understanding we have today of cancer development, when next generation sequencing (NGS) was introduced; since then numerous variations implicated in ALL have been identified (Mullighan et al, 2007; Downing et al, 2012; Holmfeldt et al, 2013; Paulsson et al, 2015). Such techniques are visualized in Table 1, including sensitivity and targets of each test.
Table 1. Overview of cytogenetic and molecular genetic techniques

<table>
<thead>
<tr>
<th>TECHNIQUE</th>
<th>TARGET TEST</th>
<th>SENSITIVITY</th>
<th>ABNORMALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-BANDING</td>
<td>Chromosome number, morphology</td>
<td>Low/medium</td>
<td>Translocations, deletions, amplifications, aneuploidy</td>
</tr>
<tr>
<td>LOCUS SPECIFIC FISH</td>
<td>Localization of specific DNA target sequences</td>
<td>Medium/high</td>
<td>Gene fusions, deletions, amplifications, aneuploidy</td>
</tr>
<tr>
<td>DNA INDEX</td>
<td>Measurement of DNA content</td>
<td>High</td>
<td>Aneuploidy</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Assessment of fusion transcripts</td>
<td>High</td>
<td>Gene fusions</td>
</tr>
<tr>
<td>SNP ARRAYS</td>
<td>Evaluation of millions of SNPs across the genome</td>
<td>Medium</td>
<td>Copy number alterations, deletions, amplifications, aneuploidy, copy number neutral loss of heterozygosity</td>
</tr>
<tr>
<td>NGS</td>
<td>Evaluation of the genome, epigenome and transcriptome</td>
<td>High</td>
<td>Detection of SNPs (homozygous and heterozygous), indels, rare variant detection, novel RNA variants, splice sites, gene expression analysis, genome-wide methylation.</td>
</tr>
</tbody>
</table>

*Adapted from Moormal et al, 2012

Indels, small deletions and insertions; NGS, next generation sequencing; RT-PCR, reverse transcription-polymerase chain reaction; SNP arrays, single nucleotide polymorphism array.

**Adult acute lymphoblastic leukemia**

Age is a prognostic factor in ALL, with poorer outcome being associated with increasing age. There is an obvious correlation between age of onset and subtype-specific leukemia; for example, high hyperdiploidy is more common in younger cases and low hypodiploidy in older. Likewise, such age restriction is also seen in solid tumors; for example Wilms tumor and retinoblastoma are restricted to pediatric cases.

Pediatric ALL has been thoroughly investigated with cytogenetic and molecular genetic techniques; SNP array analyses has been a well utilized tool in pediatric ALL, however, only a few larger studies have focused on adult cases. Cryptic hidden aberrations may be the key force promoting
leukemogenesis in adult ALL. Thus, more studies are needed to decipher this riddle.

In this chapter, clinical and genetic features of adult ALL are summarized.

**The genetic landscape of ALL**

**BCP ALL**

A majority of adult ALL cases are of BCP origin, accounting for 75% of cases. BCP ALL comprises specific cytogenetic subgroups, which are in turn associated with outcome.

**Structural abnormalities**

Translocations are a hallmark of leukemia, below are some examples of typical rearrangements in ALL.

The t(1;19)(q23;p13.3) involves TCF3 located on chromosome 19p13, which is fused with PBX1, located on 1q23, generating the TCF3/PBX1 fusion gene (Carroll et al., 1984; Michael et al., 1984). Both balanced and unbalanced forms are found, with the latter being most commonly detected. One possible route of origin is believed to be initiated by trisomy 1, with subsequent rearrangement resulting in the fusion, and followed by loss of the derived chromosome 1 (Paulsson et al., 2005). This leads to the exclusive presence of the derivative of chromosome 19 (Secker-Walker et al., 1992). It has been reported that better outcome is seen in patients harboring the unbalanced type (Secker-Walker et al., 1992).

A rearrangement associated with a good prognosis is the t(12;21)(p13;q22) translocation, producing the ETV6/RUNX1 chimeric gene fusion, albeit, only found in 2% of adult cases.

On the other hand, the most common translocation in adult cases, t(9;22)(q34;q11) (Propp & Lizzi, 1970), results in the BCR/ABL1 fusion gene, where part of the ABL1 gene in 9q34 is translocated to the BCR gene in 22q11. This fusion results in upregulation of tyrosine kinase activity and is found in about 30% of cases, increasing with higher age, and is associated with a poor prognosis (Secker-walker et al., 1997; Charrin et al., 2004; Pui & Evans, 2006; Moorman et al, 2010). There are several differently sized transcripts but the most common one found in ALL is the P190 variant (Radich et al., 1994; Faderl et al, 2003). BCR/ABL1 positive ALL with IKZF1 microdeletions has been observed in both pediatric and adult cases (Mullighan et al., 2008; Iacobucci et al, 2009) and are associated with a more aggressive disease (Martinelli et al, 2009; Mullighan et al, 2009; van der Veer et al, 2014).
A promiscuous gene known to undergo fusion with many partners, namely \textit{KMT2A}, previously known as \textit{MLL}, is frequently rearranged and results in the t(4;11)(q21;q23) producing \textit{KMT2A}/\textit{AFF1} fusion (Meyer \textit{et al}, 2006). This is one of the most common fusions, resulting in about 11\% of adult cases (Group Francais de Cytogénétique Hématologique, 1996), with an incidence increasing with age.

\textbf{Numerical abnormalities}

Distinct ploidy groups defined by nonrandom chromosomal patterns and mutations, can further subcategorize BCP ALL. One prominent example is high hyperdiploidy (51-67 chromosomes), associated with a good prognosis; however, a rare finding in adult ALL. This subgroup is characterized by a specific pattern of gains resulting in trisomies for chromosome X, 4, 6, 10, 14, 17, and 18 and tetrasomy 21. Dosage effects of additional gains on the chromosomal set are believed to play a role in leukemogenesis (Paulsson & Johansson, 2009; Paulsson \textit{et al}, 2010). Recently it was shown that gains were an early event occurring before other mutational events (Paulsson \textit{et al}, 2015).

Hypodiploidy (<46 chromosomes) is defined by chromosomal losses. It is a rare phenomenon in ALL, associated with a very dismal outcome (Raimondi \textit{et al}, 2003; Harrison \textit{et al}, 2004; Nachman \textit{et al}, 2007), further discussed below (see section Hypodiploid ALL). The low hypodiploid category (32-39 chromosomes) is restricted mainly to adolescents and adults (Callen \textit{et al}, 1989; Group Francais de Cytogénétique Hématologique, 1996). Near-triploidy is defined by having a modal chromosomal number ranging from 66 to 79, and is detected in about 4\% cases (Charrin \textit{et al}, 2004). Near-tetraploidy (84-100 chromosomes) is uncommon finding with a 2\% incidence (Group Francais de Cytogénétique Hématologique, 1996). Both latter ploidy subgroups are associated with \textit{ETV6-RUNX1} positive ALL in childhood cases (Attarbaschi \textit{et al}, 2006; Raimondi \textit{et al}, 2006).

\textit{Copy number alterations and copy neutral loss of heterozygosity}

Studies have shown that adult ALL harbors a similar pattern of deletions to that seen in pediatric ALL (Paulsson \textit{et al}, 2008). Microdeletions have been found to be more common than amplifications/duplications, and comprise genes involved in cell-cycle regulation such as \textit{CDKN2A} located on 9p21.3. This gene is reported deleted or silenced in many different tumor types, encoding a tumor suppressor, it is a feasible target (Olsson \textit{et al}, 2011; Mullighan \textit{et al}, 2007). Deletions of \textit{CDKN2A} are found in about 30\% of BCP ALL cases (Faderl \textit{et al}, 2003; Paulsson \textit{et al}, 2008; Ribera \textit{et al}, 2015). Other recurrently deleted genes include \textit{PAX5, EBF1, LEF1, ETV6,}
RB1 and IKZF1 (Paulsson et al, 2008; Moorman et al, 2012). The latter gene, located at 7p12.2 encodes a zinc finger protein. Approximately 30-50% of adult cases harbor focal or larger deletions of this gene, and studies have reported higher incidence in BCR-ABL1 positive cases (Ribera et al, 2015).

T-ALL, in brief
The remaining 25% of adult cases are of T-lineage origin. In contrast to BCP ALL that can be divided into genetic subgroups, T-ALL is a very heterogeneous disease. Below are examples of the most prominent pathway and gene targets in T-ALL.

Rearrangements in the T-cell receptor (TCR) loci leading to the aberrant expression of TCR genes are a common finding in T-ALL cases (Onciu et al, 2002; Cauwelier et al, 2006). Exemplified by the t(1;14)(p32;q11), in which TAL1, located on 1p32 is juxtaposed to TRA/D at 14q11. The aberrant expression of transcription factor oncogenes in T-ALL, due to such rearrangements, frequently involve homebox and basic helix-loop-helix genes, such as the TAL1, TAL2 and LYL1 genes (Mellentin et al, 1989; Xia et al, 1991; Xia et al, 1994; Keersmacker et al, 2005).

Two defining alterations in T-ALL, which are seen in a majority of cases comprise activating mutations in NOTCH1 in approximately 60% of cases (Weng et al, 2004). The role NOTCH1 plays in cancer is well established, and has been shown to function as an oncogene in T-ALL, promoting tumorigenesis (Roy et al, 2007). In addition, mutations in the FBXW7 gene are also common, resulting in overexpression of NOTCH1 (Neumann et al, 2015). Secondly, deletions of the CDKN2A locus is another characteristic of T-ALL found in up to 70% of cases (Keersmacker et al, 2005).

Prognosis and therapy
The prognosis of ALL is dependent on prognostic factors such as age, white blood cell counts (WBC) and acquired genetic abnormalities (Rowe, 2010).

In recent years, superior outcomes in young adults treated with chemotherapy regimen has been reported, and thus, decreasing a transplant-related mortality rate which has been high in adult ALL (Litzow, 2015).

The survival rates at five years for adult cases ranges between 30 to 40% in patients aged 60 years and less. In patients between 60 and 70 years it is less than 15% and less than 5 % in patients over 70 years of age (Hoelzer et
al, 1988; Thomas et al, 2004; Rowe et al, 2005; Pui & Evans, 2006; Rowe, 2010).

Studies have reported that younger adults treated on pediatric protocols with a more intensive regimen show improved response to treatment, with increased overall survival compared with those on adult protocols. (Boissel et al, 2003; de bont et al, 2004; Nachman, 2005; Stock et al, 2008). In the Nordic countries adults aged 45 years and less are now treated on the Nordic Society for Paediatric Haematology and Oncology ALL 2008 protocol, which is the common protocol initially used to treat childhood cases. However, even though adults are treated on pediatric protocols, it should be emphasized that adult patients differ significantly from childhood patients as regards complete remission rates, minimal residual disease (MRD) response and risk group assignments even when treated on the pediatric protocols (Toft et al, 2012). The reasons for the poorer outcome seen in adult cases may be reflected by the differences in frequencies in subgroups between childhood and adult ALL, such as higher incidence of BCR-ABL1 positive cases and fewer high hyperdiploid cases.
Hypodiploid acute lymphoblastic leukemia

A glance into the intricate realm of hypodiploidy

The word aneuploid originates from ancient Greek, translating to “not good fold”, reflecting an erroneous chromosomal count (Pfau & Amon, 2012). Chromosomal count, or ploidy, represents an important prognostic factor in BCP ALL; hypodiploidy is rare, seen in 5-8% of childhood and adult BCP cases and associated with a very poor outcome.

Hypodiploidy is defined as having fewer than the diploid number of chromosomes (<46 chromosomes). The majority of cases harbor 45 chromosomes, thus, cases with <45 chromosomes are even rarer, constituting about 1% of BCP ALL. This subgroup can be further divided into categories with distinct clinical and genetic features; near-haploidy (24-31 chromosomes) and low hypodiploidy (32-39 chromosomes) specifically, and, comprising a heterogeneous category, high hypodiploidy (40-43 chromosomes).

Two typical peaks clustering around 26-27 and 35-36 chromosomes represent the majority of near-haploid and low hypodiploid ALL cases; the reason for this specific clustering is unknown. It can be due to the fact that a certain combination of chromosomes is advantageous for the leukemic cells, i.e., specific karyotypic patterns are delineated by clonal selection. Alternatively, the reason for the specific modal numbers may result from the underlying genetic mechanism of chromosomal loss in hypodiploidy. That chromosomal loss initiates an aneuploid state resulting in chromosomal modal numbers in such narrow ranges, with subsequent clonal evolution promoting for doubled populations, reflect the selective pressures pushing for further divergence and ultimately allowing for more heterogeneous subclones to predominate. This scenario can be visualized by the fact that the frequency of hyperdiploid cells increases progressively and over time dominate in cell culture (Kohno et al, 1980; Aburawi et al, 2011). Thus, the ”perfect’ near-haploid or low hypodiploid karyotype with
such modal numbers, clustering around the specific peaks observed may in fact promote for the "ultimate" doubled clones with favorable fitness.

Regardless of the reason, chromosomal modal numbers ranging from 30-32 and 40-42 are seldom identified. Furthermore, cases with 40-45 chromosomes are very heterogeneous, harboring genetic alterations and clinical features that cannot be categorized as the former two subtypes. It is feasible to debate the significance of categorizing such ALLs by modal chromosome number due to the extensive case variability. Hence, ALLs with chromosome numbers in the upper low hypodiploid range, or cases within the lower high hypodiploid range may in fact not belong in their respective modal number group. Significantly, genomic and epigenetic studies on hypodiploid ALL can help stratify cases according to their true mutational backgrounds and not just by chromosomal count.

A tale of disomies

Tumors harboring chromosomal model numbers ranging between 23 to 34 are very rare, comprising <0.5% of cytogenetically investigated malignancies and found in 1% of ALL cases (Mitelman et al, 2015). Moreover, regardless of tumor type certain chromosomes are frequently found in a heterodisomic state (Mandahl et al, 2012). For example, heterodisomies 5, 7, 19, 20, and 21 are common in hyperhaploid chondrosarcomas (Bove et al, 2000; Olsson et al, 2011), and chromosomes 5, 19 and 21 are also commonly retained in two copies in low hypodiploid ALL (Harrison et al, 2004). The non-random pattern of heterodisomies highlights the pathogenetic significance of retained heterozygosity in hypodiploidy. However, it is important to stress that hypodiploid malignancies harbor tumor specific profiles; hence, chromosomes that are often heterodisomic in one tumor type may be monosomic in another. An exception to this rule is chromosome 21, found in a disomic state in one-fourth of cases, and in almost 100% of cases in ALL (Pui et al, 1990; Heerema et al, 1999; Raimondi et al, 2003; Harrison et al, 2004; Nachman et al, 2007; Mandahl et al, 2012). Significantly, genes on chromosome 21 must be vital for survival. However, in a study by Kohno et al (1980) clonal evolution of a near-haploid cell line, namely NALM-16, yielded eventual loss of chromosome 21 in culture, showing that in vitro maintenance of disomy 21 is not vital for the growth of cells, and survival. Thus, tracking the evolution of retained chromosomes in near-haploid cells may allow for better understanding of the pathogenetic impact of retained heterodisomic chromosomes.
Delineation of hypodiploid ALL

The hypodiploid subgroups have been defined differently in the literature; in this thesis the definitions used are based on recent mutational findings further distinguishing such cases from another.

Near-haploid ALL with 24-31 chromosomes
The near-haploid subgroup was first identified in 1975, when Kessous et al described a unique karyotypic pattern in a five year old girl with leukemic cells harboring 27 chromosomes. Two years later, Oshimura et al (1977) published a case report of a 12 year old girl with a similar pattern, with a karyotype harboring 27 chromosomes. Since then, several studies have reported cases with near-haploidy and found that this disease entity is defined by a modal chromosome number ranging from 25-29, with a majority of cases clustering around 26 chromosomes (Harrison et al, 2004; Nachman et al, 2007; Holmfeldt et al, 2013). Near-haploid ALL is characterized by massive chromosomal loss and a non-random pattern of retained heterodisomic chromosomes. The most frequently retained chromosomes are 21, X/Y, 14, and 18 (Heerema et al, 1999; Raimondi et al, 2003; Harrison et al, 2004; Nachman et al, 2007).

Low hypodiploid ALL with 32-39 chromosomes
In 1989, Callen et al reported the first low hypodiploid cases thus dividing cases with 30-39 chromosomes apart from cases with less than 30 chromosomes. Clinically, an apparent age difference was eminent in the two groups; younger age predominated in cases with <30 chromosomes, in stark contrast to older age in the second group. Similar to cases with near-haploidy, a non-random pattern of retention of heterodisomies X/Y, 14, 18 and 21 is seen in low hypodiploid ALL. However, additional gains specific for these cases are also seen, such as retention of chromosomes 1, 5, 6, 8, 10, 11, 19 and 22, in two copies (Harrison et al, 2004; Nachman et al, 2007). As seen in near-haploid ALL, a majority of low hypodiploid cases harbor a doubled line resulting in a population in the near-triploid range (Charrin et al, 2004; Mulbacher et al, 2015). Charrin et al (2004) could show that the latter ploidy category derived from a low hypodiploid clone, based on the identical profiles of numerical abnormalities seen in both populations. They further suggested that near-triploidy and low hypodiploid should be characterized as one entity and treated as such.
High hypodiploid ALL with 40-43 chromosomes and cases with 44 and 45 chromosomes

High hypodiploid ALL and cases with 44 and 45 chromosomes harbor a much more heterogeneous group and include cases with T-ALL (Pui et al, 1987; Raimondi et al, 2003; Harrison et al, 2004). In the few cases with 40-43 chromosomes investigated, disomies 2, 6, 10, 11, 20 and 22 are common (Nachman et al, 2007). Cases within the high hypodiploid range have been classified differently in the literature; some studies have grouped patients with 42-44 chromosomes apart from those with 45 chromosomes (Harrison et al, 2004). Duplication of the hypodiploid stemline is usually not observed, in contrast to the other hypodiploid subgroups.

Clinical features of hypodiploid ALL

Age, sex and immunophenotype

Near-haploid and low hypodiploid ALL have a common/pre-B immunophenotype, and usually low white blood cell (WBC) count of <50x10^9/L (Callen et al, 1989; Harrison et al, 2004; Nachman et al, 2007). Near-haploid ALL predominates in childhood cases ranging up to 15 years of age, with a median age of 7 years. Low hypodiploid ALL, on the other hand, is seen in older childhood cases and adult patients (Callen et al, 1989; Pui et al, 1990; Harrison et al, 2004; Nachman et al, 2007). Near-haploid ALL displays a sex ratio close to one, whereas some studies have reported low hypodiploidy to be more common in males (Callen et al, 1989; Nachman et al, 2007).

High hypodiploid ALL and cases with 44 and 45 chromosomes have a pre-B immunophenotype, and in contrast to near-haploid and low hypodiploid ALL also include T-lineage immunophenotypes. ALL with 40-45 chromosomes do not show any distinct association with age or gender (Harrison et al, 2004).

Outcome

That near-haploidy is associated with a poor outcome was first described by Brodeur et al (1981), and confirmed in subsequent studies (Gibbons et al, 1991; Chessels et al, 1997; Harrison et al, 2004). Decreasing chromosomal count has been associated with worsening outcome, and both near-haploid and low hypodiploid ALL frequently relapse and have low survival rates (Pui et al, 1990; Gibbons et al, 1991; Heerema et al, 1999; Forestier et al, 2000; Raimondi et al, 2003; Harrison et al, 2004; Schultz et al, 2007; Nachman et al, 2007; Moorman et al, 2010). Therapy is successful in eradicating the doubled mirror clones, however, the treatment is ineffective.
in targeting the stemline, which simmers in the background in a quiescent state, proving difficult to target.

Clinical studies on childhood near-haploid and low hypodiploid ALL have reported event free survival (EFS) rates of 25-40% (ranging from three to eight years) (Heerema et al, 1999; Raimondi et al; 2003; Harrison et al, 2004; Nachman et al, 2007). Adult cases with low hypodiploidy also have an extremely poor overall survival, with three year EFS rates of <30% (Group Francais de Cytogenetique Hématologique, 1996; Harrison et al, 2004; Moorman et al, 2007). Some studies have reported better outcome in ALLs with 42-44 chromosomes, and pediatric cases with 45 chromosomes. Such studies have shown cases to have EFS rates of 65-66% (Heerema et al, 1999; Harrison et al, 2004). For adult cases with 42-45 chromosomes the five year EFS rates have been reported to be 45% (Harrison et al, 2004).

In the current US protocols (Schultz et al, 2007) the breakpoint for stratifying cases to a very high risk group is set to <44 chromosomes. In the Nordic countries, however, the protocol used for a more intensive treatment includes cases with <45 chromosomes.

**Masked hypodiploidy and its clinical significance**

That near-haploid and low hypodiploid ALL frequently harbor duplicated mirror clones in the hyperdiploid and near-tripliod ranges have been reported in many studies (Oshimura et al, 1977; Brodeur et al, 1981; Callen et al, 1989; Gibbons et al, 1991; Holmfeldt et al, 2013). Fluorescence in situ hybridization analyses and measurements of DNA index are helpful in detecting the masked near-haploid and low hypodiploid (Brodeur et al, 1981; Ma et al, 1998; Stark et al, 2001) as the doubled near-haploid clone may been mistaken for high hyperdiploid ALL associated with a favorable prognosis and treated on standard risk protocols (Paulsson & Johansson, 2009). Adding to the complexity of these cases, the doubled clones may dominate at diagnosis, hence masking the underlying true hypodiploid clones, which in turn, dominates at relapse (Brodeur et al, 1981; Stark et al, 2001; Raimondi et al, 2003). Clonal evolution may hence not only be frequent but also an early event. The presence of a dual population of small and large blasts in the BM further distinguishes these cases from classic high hyperdiploid ALL (Oshimura et al, 1977). The coinciding, as well as exclusive, “hyperdiploid” line harbours two and four copies of chromosomes corresponding to one and two copies in the original near-haploid clone, thus, all chromosomes in the doubled population are uniparental isodisomies, as a result of the extensive loss of heterozygosity (LOH) in the near-haploid stemline (figure 1) (Stamberg et al, 1986; Onodera et al, 1992; Aburawi et al, 2011). This is in contrast to true
hyperdiploid ALL, which harbors trisomies for all gained chromosomes, and is not associated with LOH (Paulsson & Johansson, 2009) however, it is worth mentioning common gains in hyperdiploid ALL results in trisomy 10, 14, 18, 21 and X/Y, corresponding to the specific retention of disomies for just these chromosomes in near-haploid and low hypodiploid ALL. Careful and correct analysis of near-haploid and low hypodiploid ALL is of high clinical significance considering that such ALLs are stratified to high risk groups in most current treatment protocols (Schmieglow et al, 2010).

![Figure 1. Near-haploid karyotype and its masked mirror clone.](image)

**Figure 1. Near-haploid karyotype and its masked mirror clone.** Near-haploid karyotype with 26 chromosomes, with disomies 18, 21 and X. The remaining chromosomes are monosomies. In the duplicated clone, with 52 chromosomes, all monosomies are found in two copies (UPIDs) and retained disomies (18, 21 and X) are found in four copies (tetradasomies).

**Origins and consequences of hypodiploidy**

In 1977, Oshimura et al were the first to postulate the origins of a hypodiploid cell; it may be due to misdivision of a diploid cell as a result of multipolar mitosis. Although many have attempted to understand the intricate enigma of the mechanism resulting in massive chromosomal loss, or why some chromosomes display retained heterozygosity, its effects still remains unclear. Since the leukemic genome always contains heterodisomies, an intermediate haploid stage with 23 chromosomes can be excluded. They might originate through successive loss of chromosomes, or by a single erroneous meiosis-like somatic event, such as tripolar division.

How hypodiploidy is advantageous for growth of malignant cells is unclear; massive chromosome loss will likely affect the global gene expression.
Genes that are expressed in a monoallelic manner, due to random inactivation or imprinting, may be affected (Goldmit et al, 2004). Furthermore, haploinsufficiency of genes on the monosomic chromosomes may also contribute to leukemic transformation; moreover, chromosomes rarely retained in two copies may harbor genes that counteract tumor development.

The non-random retention of disomies may be due to key regulatory genes, essential for cell survival, located on these chromosomes. It is feasible to postulate that the expression of recessive genes, due to the extensive loss of heterozygosity, and the non-random pattern of retention of specific chromosomes promotes and strengthens the pathogenetic impact.

It is believed that the duplication of the near-haploid or low hypodiploid stemline occurs through endoreduplication (Pui et al, 1990; Gibbons et al, 1991). Cancer is an evolutionary disease, in which founder clones gives rise to generations of new descendants adding to tumor heterogeneity, reflecting the advantage of the doubling event in near-haploid and low hypodiploid ALL. It has been established that clonal evolution is developed through acquiring selective advantageous driver mutations, neutral and deleterious lesions (Greaves et al, 2012). The doubled populations in hypodiploid malignancies allow for clones with a more normal copy number, more prone to give rise to subclonal heterogeneity.

In chondrosarcoma, clonal evolution through loss of chromosomes and subsequent polyploidization, possibly occurring more than once, allows the malignant cells to gain further imbalances such as gains, losses, translocations, supporting the latter theory (Olsson et al, 2011). Although cell-to-cell variability has been encountered in solid tumors, this phenomenon is generally not seen in ALL. Significantly, few additional aberrations are seen in the duplicated populations in near-haploid and low hypodiploid ALL. Hence, the complex pattern of aberrations identifiable in chondrosarcoma arising subsequently to the hypodiploidy has not been identified in hypodiploid ALL, indicating, that other hidden genetic mechanisms may be in play. Little is known about the pathogenetic effect of the specific pattern of retained heterodisomies in hypodiploidy. In inflammatory leiomyosarcoma disomy 5, 20, 21 and 22 are common, and exhibit higher gene expression levels than monosomic chromosomes (Nord et al, 2012). Thus, a direct consequence of hypodiploidy is gene dosage effects on retained chromosomes.
Acquired genetic aberrations in hypodiploid ALL

The division between near-haploid and low hypodiploid ALL was initially based on chromosome count, gender and age differences (Callen et al., 1989; Harrison et al., 2004), but has recently been confirmed in a large next generation sequencing study (Holmfeldt et al., 2013) showing distinct mutational profiles.

Structural aberrations

Near-haploid ALL rarely harbor any structural aberrations and no fusion gene that is specific for this subgroup has been detected to date. Structural aberrations are infrequent also in low hypodiploid ALL, however, more common when compared to near-haploid cases; deletions of the short arm of chromosome 6, 9 and 12 has been identified in low hypodiploid cases (Harrison et al., 2004).

High hypodiploid ALL and cases with 44-45 chromosomes, in stark contrast to the other hypodiploid groups, frequently harbor primary translocations and complex chromosomal karyotypes (Pui et al., 1990; Heerema et al., 1999). For example, ETV6-RUNX1 and BCR-ABL1 positive ALL often have karyotypes with 44 and 45 chromosomes. In the latter group, dicentric chromosomes, resulting from unbalanced rearrangements, such as dic(9;20)(p13;q11), and other frequently dicentric chromosomes involving chromosomes 7 and 12 are common.

Gene targets

The quest for specific gene targets in hypodiploid ALL has until recently been a perplexing and challenging mission. Cases harbor a relatively low frequency of microdeletions compared with other subtypes of BCP ALL (Mullighan et al., 2007; Paulsson et al., 2008; Holmfeldt et al., 2013). This could be due to the fact that the leukemic cell has already lost massive amounts of genetic material; additional deletions may therefore not be necessary for leukemogenesis or be incompatible with cell viability.

However, in a recent next generation sequencing study by Holmfeldt et al. (2013) the mutational landscapes of near-haploid and low hypodiploid ALL were unraveled; near-haploid and low hypodiploid ALL harbor distinct mutational and expression profiles. For example, alterations in receptor tyrosine kinase and RAS pathways targeting NRAS, KRAS, MAPK1, FLT3, and PTPN11, and mutations in NF1, coding for a tumor suppressor and negative regulator of RAS signaling were characteristic targets in near-haploid ALL. In addition, a finding not previously implicated in leukemia but found specifically in near-haploid cases was deletions of a putative RAS-signaling inhibitor, namely PAG1. Holmfeldt et al. (2013) further
reported focal deletions of a histone gene cluster at 6p22 to be frequent in near-haploidy, with the most common target being the transcriptional coactivator and histone acetyltransferase encoding gene CREBBP.

Low hypodiploid ALL, on the other hand, harbor alterations in TP53, IKZF2 and RB1. The latter gene encodes for a tumor suppressor and is widely associated in ALL and other malignancies (Holmfeldt et al, 2013).

IKAROS gene family and hypodiploidy
Interestingly, near-haploid and low hypodiploid ALL harbor alterations in IKZF3 and IKZF2, respectively. This is in contrast to ALL in general, where IKZF1 alterations constitutes characteristic finding in high-risk BCP ALL (Mullighan et al, 2009). The members of this family are involved in lymphopoeisis, B-cell differentiation, T-cell development and encode zinc finger transcription factors (Molnár et al, 1994; Winandy et al, 1995; Karlsson et al, 2002). It is believed that near-haploid ALL harboring IZKF3 deletions ascend from more mature lymphoid precursors and low hypodiploid ALL from less mature progenitor cells, in relation to IZKF2 being expressed in common lymphoid progenitor cells and pre B cells (Holmfeldt et al, 2013).

TP53 and low hypodiploidy
TP53 is a well-studied tumor suppressor gene, frequently altered in many different malignancies (Leroy et al, 2014). Holmfeldt et al (2013), reported that TP53 mutations are a hallmark of low hypodiploid ALL, identifiable in 90% of pediatric and adult cases, comprising missense, nonsense and insertion-deletion mutations resulting in loss-of-function mutations (Holmfeldt et al, 2013). In addition, they demonstrated that nearly half of the pediatric cases with such mutations harbored heterozygous TP53 mutations in remission bone marrow or peripheral blood. Significantly, the presence of a common Li Fraumeni syndrome (LFS) substitution, seen in three cases further stressed that pediatric low hypodiploid ALL may be a manifestation of LFS, due to the high prevalence of germline TP53 mutations. Li-Fraumeni syndrome is an autosomal dominant disorder, predominating in childhood and young adult cases; it increases the risk of developing several different types of malignancies (Li et al, 1988; Nichols et al, 2001; Olivier et al, 2010; Kim, 2015). Moreover, two cases from this cohort had known familial history of cancer (Holmfeldt et al, 2013). In contrast to pediatric low hypodiploid ALL, none of the adult cases analyzed for TP53 mutations with available remission material to date have been associated with LFS, i.e., adult low hypodiploid ALL harbours acquired TP53 mutations (Holmfeldt et al, 2013; Mulbacher et al, 2014).
**Expression profile**

Holmfeldt *et al* (2013) investigated the gene expression profiles of near-haploid and low hypodiploid cases and could confirm their distinct genomic profiles with yet another technique. Furthermore, they reported that near-haploid and low hypodiploid ALL harbor global expression differences distinguishing them from another, and concluded that aneuploidy is not accountable for the major differences observed in the expression profile between these two subtypes.
The present study

Focus and aims

The aim of this thesis was to gain a better understanding of the basic genetic mechanisms in ALL with a poor prognosis. Furthermore, characterizing genetic abnormalities that could be linked to outcome may provide us with potential markers that could be used in a clinical setting.

The aims were more specifically:

- To identify novel gene targets in a population based study on adult ALL cases (article I)
- To characterize the genetic and epigenetic landscape of hypodiploid ALL (articles II, III and IV)

Material and methods

This section contains a brief overview of the material and methods used in the study. For a more detailed description, see the Materials and methods section and the corresponding supplement of each article. All studies were approved by the regional ethical committee of Lund University.

Patient material

Article 1

This study comprised 126 cases of adult ALL cytogenetically investigated as part of clinical routine between 1985 and 2012 at the Department of Clinical Genetics, University and Regional Laboratories, Region Skåne, Lund, Sweden.
Article II

Here, we investigated 12 pediatric and adult hypodiploid ALL cases obtained from collaborators in the United Kingdom and Sweden and from Skåne University Hospital.

Article III

The study comprised patient material from cases cytogenetically investigated as part of clinical routine at the Department of Clinical Genetics, University and Regional Laboratories, Region Skåne, Lund, Sweden. (Partly overlapping with cases in articles I and II).

Article IV

All hypodiploid (<46 chromosomes) ALL cases diagnosed between 1992 and 2013, aged 1-18 years in the Nordic countries comprising 188 cases were included.

SNP arrays

A single nucleotide polymorphism (SNP) is a variation occurring with an incidence of >1% in a population, where one of the nucleotides differ in the homologous chromosomes. SNP array analyses takes advantage of these naturally occurring variations and enable detection of base pair variation in leukemic samples. In this thesis, arrays from Illumina have been utilized, containing SNPs ranging from 700 000 to 5 million. Probes are immobilized on silica beads and distributed in microwells covering the array surface. The probe-covered bead targets the sequence adjacent to the SNP of interest and hybridizes the parental homologues in a nonspecific manner. The hybridization signals are scanned and measured. The analysis allows for detection of deletions (hetero and homozygous deletions), gains or LOH (UPIDs). However, balanced translocations cannot be detected using this technique. SNP arrays are specifically advantageous for detection of LOH; an UPID is detected when a region contains no heterozygous SNPs with any change in copy number. Such changes may involve whole chromosomes or partial segments and have been identified in ALL (Paulsson et al, 2010).

The analysis provides information on genotypes and copy numbers corresponding to B allele frequencies and log₂ ratios. In a diploid segment, the B-allele frequency detects homozygous SNPs (value of 0 or 1) and heterozygous SNPs (value of 0.5). The log₂ ratio detects the average copy number correlating to 0. Near-haploid/low hypodiploid samples are normalized to have a log₂ ratio of 0; their respective hyperdiploid/near-
triploid mirror clones will also be normalized to have a log2 ratio of 0. In these particular cases the analysis cannot distinguish between the stemline and duplicated mirror clones (figure 2).

Figure 2. SNP array analysis results visualizing near-haploidy followed by clonal evolution. A normal diploid cell contains homzygous and heterozygous SNPs throughout the whole chromosome representing both parental copies. Near-haploidy results in massive chromosomal loss; a majority of chromosomes will therefore exist as monosomies, i.e., no heterozygous signals. The few chromosomes retained in two copies, harbor a “normal” diploid profile. Following clonal evolution the duplicated mirror clone produces cells harboring chromosomes with total UPDity. All monosomies exist in two copies, i.e., two copies of one parental chromosome, and all disomies are tetrasomies in the doubled clone with identical patterns in both lines.
Exome and RNA sequencing

Next generation sequencing (NGS) comprises analyses on the genome (whole genome and exome sequencing), the epigenome (methylation sequencing) and the transcriptome (RNA sequencing), resulting in high-throughput sequencing with billions of DNA strands sequenced in parallel. Significantly, this allows for detection of homozygous and heterozygous SNPs, insertion-deletions, and missense and nonsense mutations in cancer (Meyerson et al., 2010; Stransky et al., 2011; Stratton, 2011; Andersson et al., 2015). This technique produces millions of output reads that are mapped to the genome, and allows for high sensitivity.

In brief, the NGS library is constructed by fragmentation of genomic DNA or RNA followed by adaptor ligation. The library is loaded on a flow cell allowing complementary binding of fragments, and subsequent bridge amplification of the bound fragments yields clonal clusters. When this step is complete, sequencing is initiated by adding regents and fluorescently tagged nucleotides. The flow cell can be imaged and emission from each cluster recorded. Reads are aligned to a reference genome (Illumina).

In this thesis, the exome and RNA sequencing analyses were performed by BGI Tech Solutions (see material and methods section in Article III for detailed procedure). Exome sequencing identifies protein coding variants throughout the genome, which constitute about 2% of the genome. Many disease-associated alterations are linked to just these regions; hence rare mutations can be unraveled. RNA sequencing allows identification of fusion genes and splicing events. In addition, gene expression analysis allows further characterization and may reveal novel aberrant oncogenes promoting tumorigenesis.

Methylation array

Epigenetic changes such as DNA methylation represent another example of cooperating alterations in cancer. Abnormal methylation patterns in cancer are well-studied (Boultonwood & Wainscoat, 2007, Garcia-Manero et al., 2009). Hypermethylation of CpG islands are detected at gene promoters, resulting in the silencing of gene expression (Esteller & Herman, 2002; Jones & Baylin, 2002; Herman & Baylin, 2003; Galm et al., 2006). No studies so far have investigated the methylation status of hypodiploid cases. In this thesis, the Human 450K array (Illumina), was utilized (Article III), converting DNA with sodium bisulfite to find C-to-T alterations at defined genomic positions (performed by Sciblue Genomics, Lund, Sweden).
FISH analyses

FISH is a well-utilized method where fluorescently tagged probes complementary bind sequences of interest (either whole chromosomes, or parts of chromosomes). Rearrangements, gain or loss of material can be detected with this method. In articles II and III, interphase and metaphase FISH analyses were performed in order to investigate chromosomal copy numbers. For near-haploid and low hypodiploid cases combining SNP array and FISH analyses is highly advantageous; the former identifies LOH and thus the underlying near-haploidy, whereas, FISH detects copy number changes, which is difficult to determine in these particular cases with SNP array analyses.

Results, discussion and concluding remarks

Article I

Novel gene targets detected by genomic profiling in a consecutive series of 126 adults with acute lymphoblastic leukemia

This is the largest series to date, reporting SNP array analysis results on adult ALL. Different ploidy subgroups could be identified immediately using this technique; such as the detection of three additional high hyperdiploid and five additional low hypodiploid cases, stressing that SNP array analysis is a sensitive tool for detection of aneuploidy.

Microdeletions are more common than gains

Characteristic deletions, comprising CDKN2A, PAX5, IKZF1, ETV6, RB1, and EBF1 genes were detected by the SNP array analyses. Moreover, several recurrent cryptic genetic events not previously implicated in adult ALL, including the BCAT1, BTLA, NR3C1, PIK3AP1 and SERP2 genes, were identified in 2-6% of cases.

Statistical analyses

Cases with IKZF1 deletions had higher WBC counts, indirectly suggesting a more aggressive disease. Moreover, a higher incidence of RB1 deletions in women was detected.
Clonal relationship

We suggest that mechanisms of relapse may be similar in adult and pediatric ALL. Different evolutionary genetic patterns could be ascertained from the analyses of paired diagnostic and relapse samples, showing identical genetic changes in 27%, clonal evolution in 22%, and ancestral clones in 50%. This is the first investigation of the clonal relationship between diagnostic and relapse cases that has been reported in adult ALL using SNP array analysis. Our findings agree well with most, albeit not all, studies of childhood ALL and indicate that the mechanisms of relapse may be similar in adult and pediatric ALL.

The genetic backbone of adult ALL has been elucidated, showing a similar genetic mechanism to pediatric cases. The novel target genes identified in this study may be explored as therapeutic targets in order to investigate if they have an effect on the pathogenetic impact.

Articles II & III

Genetic and epigenetic investigations on hypodiploid ALL

In these studies we investigate the mechanism of formation in hypodiploidy, characterize the chromosomal patterns and mutational profiles by applying various techniques.

Copy number variations detected by SNP array analyses

The SNP array analysis identified the extensive LOH due to chromosomal loss, and non-random pattern of retained heterodisomies specific for near-haploid and low hypodiploid ALL. In addition, the analysis proved to be a sensitive tool in detecting the underlying near-haploidy/low hypodiploidy immediately, showing the potential of this technique for correct risk stratification.

In both subgroups deletions in CDKN2A/B were identified. The pattern of deletions and LOH indicated that microdeletions of CDKN2A/B occurred in a subsequent step in one case, strongly suggesting that the chromosome loss is the primary event in near-haploid and low hypodiploid ALL.

As regards to near-haploid ALL, deletion of ETV6 was detected in a single case. Two of three adult low hypodiploid cases analyzed, harbored CREBBP deletions, something previously believed to be restricted to pediatric near-haploid ALL (Holmfeldt et al, 2013).
Finally, amplifications of the following genes were detected in two cases; MIR17HG, GPC5 at 13q31.3 and SOX9 at 17q24.3. Of which none have been implicated in BCP ALL.

**SNP and FISH analyses confirm chromosomal instability in low hypodiploid/high hypodiploid ALL**

Of the four high hypodiploid cases included in this study, three cases with modal chromosome numbers ranking from 79-84 were classified as low hypodiploid/high hypodiploid due to the extensive LOH detected in the SNP array analyses. Unambiguously, these cases seemed to comprise a majority of UPIDs, a few retained chromosomes, with few further deletions, typical for near-haploid and low hypodiploid ALL. In addition, all three harbored apparent subclonality for some of the typically retained heterodisomic chromosomes; a pattern not previously reported in hypodiploid ALL. The fourth high hypodiploid case, on the other hand, harbored only three whole chromosomal losses, no UPIDs, and many more deletions, clearly distinguishing it from the former cases. These novel findings prompted us to further investigate such cases with FISH analyses consequently detecting high cell to cell variability in all three cases, a finding indicative of chromosomal instability (CIN) (Figure 3). Significantly, our results show that high hypodiploid karyotypes within the upper near-triploid and near-tetraploid range may in fact originate from a low hypodiploid stemline, however, specific for these cases is further clonal evolution adding more dimensions of genetic heterogeneity not seen in typical low hypodiploid ALL. The three chromosomal patterns identified by the SNP array analyses and chromosomal counts confirmed by FISH are depicted in figure 3 below.
Figure 3. Different chromosomal patterns identified by SNP array analysis in low hypodiploid/high hypodiploid ALL. Typical low hypodiploid chromosomal patterns with extensive loss of heterozygosity was detected in three cases with >79 chromosomes, indicating a hypodiploid origin. In addition, the analysis detected subclonality as a split in the B-allele frequency indicating more than two copies of chromosomes. FISH analysis detected a high cell-to-cell variability for all such cases, visualized by the chromosomes in the figure next to the SNP array results.

Exome sequencing analyses confirms previous findings

Our results confirm previous findings in near-haploid and low hypodiploid ALL; alterations in NF1, IKZF3, FLT3 and CREBBP was detected in the former subgroup and, mutations in TP53 were specific in the latter subtype. The possibility that the low hypodiploid/high hypodiploid cases may have originated from a low hypodiploid stemline was further highlighted by the presence of TP53 mutations in all three cases. Our results suggest that
ALLs with modal numbers in the upper low hypodiploid and high hypodiploid range may be characterized by chromosomal doubling followed by CIN, consequently, discussing modal numbers for these cases might not be of such importance, rather than characterizing and treating these cases dependent on their chromosomal and mutational profiles.

Expression and methylation profiles

We performed unsupervised principal component analysis on methylation array data and could show that the specific subgroups clustering remained even when restricting the analyses to genes on commonly lost or retained chromosomes. No differences in methylation between chromosomes present in one or two copies could be detected; hence, methylation levels were not correlated to chromosomal copy number. Furthermore, we performed gene set enrichment analysis (GSEA) to calculate whether a defined set of genes showed statistically significant differences between the groups (Mootha et al, 2003; Subramanian et al, 2005), and could show possible enrichment of genes on chromosome X, Y, 14 and 21 in near-haploid cases and of genes on chromosomes X, Y, 1, 5, 6, 8, 10, 11, 18, 19, 21 and 22 in low hypodiploid cases. Significantly, these results emphasize the effects of hypodiploidy, namely, dosage effects on the retained chromosomes. That clonal evolution yields extra copies of chromosomes harboring oncogenes and favorably eliminates those harboring tumor suppressor genes promotes cancer fitness, and is thus a consequence of aneuploidy.

Article IV

Hypodiploid childhood acute lymphoblastic leukemia in the Nordic countries

We present a population-based study of 188 hypodiploid childhood ALL, and confirm the specific cytogenetic and clinical features previously reported. However, we stress that ALLs with 40-45 chromosomes are heterogeneous and should probably not be considered to constitute a separate genetic subtype of childhood ALL. Our data suggest that such cases lack distinct genetic and clinical profiles and should be classified according to primary translocations and/or mutations.
Future perspectives

Genetic characterization of adult and hypodiploid ALL has unraveled the presence of specific targeted genes and pathways. We have a better understanding of the genetic profiles and moving forward by focusing on screening for such target genes routinely in a clinical setting will hopefully aid in improved treatment stratification and ultimately in a better outcome in cases with a poor prognosis in ALL.

For hypodiploid cases specifically, such ALLs should be defined depending on mutational profiles rather than chromosomal count. In this scenario, recurrently targeted genes such as *IKZF3*, *IKZF2*, *RB1*, *CREBPP*, *NF1*, *TP53* and genes involved in RAS and RTK pathways should be routinely analyzed and thus included in the initial risk assessment of hypodiploid ALL. Furthermore, by focusing on the mutational profiles, therapeutic drugs can be developed targeting specific pathways.

Significantly, it is of biological interest to elucidate the order of events in cancer evolution in hypodiploid ALL. Hyperdiploidy with a gain in chromosome sets is associated with a good prognosis. Recently, it was shown that these gains are an early event, arising before other mutations and thus the hyperdiploid pattern may not be a secondary event – contributing to the evolution – but a primary aberration initiating tumorigenesis, with secondary events needed for overt leukemia. Supporting a similar scenario, our data indicate that chromosomal loss is the primary event in near-haploid and low hypodiploid ALL, with copy number alterations occurring subsequently. This illuminates the fact that leukemogenesis may be initiated by chromosomal loss in an ancestral lymphoid progenitor stem cell, with secondary events targeting specific genes needed for overt leukemia. Following clonal evolution, the hypodiploid stemline spawns subclones, with possible additional aberrations. By isolating single blast cells from bone marrow samples at the time of diagnosis, before treatment, when treatment begins, at remission and at relapse, followed by NGS will allow for mapping of clonal evolution, detecting which aberrations dominate when and eventual evolution of subclones including their specific genetic aberrations.

In addition, it would enable characterization of therapy- responsive and -resistant subclones. That therapy has shown to be successful in eradicating
the doubled “hyperdiploid” clones, actively dividing, in contrast to the quiescent hypodiploid clone, highlights the fact that more functional and translational research are needed, targeting the hypodiploidy at its roots. In conclusion, identifying which alterations that are present at a certain time may help improve therapy and will aid in understanding the clonal evolution in hypodiploid ALL.

We have characterized cases with chromosomal modal numbers in near-triploid/tetraploid ranges, harboring substantial LOH and similar profile to low hypodiploidy, including TP53 mutations. However, unique karyotypic findings for these cases include chromosomal instability – not identified in hypodiploid ALL previously. These genetic features may indeed represent a unique subcategory of patients and identifying such patterns in near-triploid and near-tetraploid cases can in turn reveal a true hypodiploid origin. Reported cases with high hypodiploid ALL also harbor TP53 mutations; however, they lack chromosomal instability patterns. Our analyses and published data on high hypodiploid cases comprise only a total of five cases. Hence, more studies are needed including more cases to truly genetically characterize such ALLs.
Leukemi kännetecknas av en okontrollerad tillväxt av omogna vita blodkroppar och orsakas av förvärvade genetiska förändringar. Genetiska förändringar eller avvikelse, som är drivande i cancerutvecklingen, uppstår när normala celler delar sig. Identifiering av sådana förändringar på gen- och kromosomnivå är av stor klinisk betydelse. Behandling styrs av vilka genetiska avvikelse som identifieras i leukemicellerna. Akut lymfatisk leukemi (ALL) är en av de vanligaste typerna av leukemi hos barn och uppvisar en god prognos. Vuxna patienter som drabbas har en betydligt sämre prognos.


Genetiskt bevarar den hypodiploida leukemicellen vissa kromosomer i ”normalt” tillstånd, alltså i två kopior. Resterande kromosomer finns i en kopia. Kromosom 21 är nästan alltid bevarad i två kopior. Förutom detta specifika kromosommönster detekteras ofta en dubblerad klon med två och fyra kopior av alla kromosomer som finns i en och två kopior i ursprungscellen. Det är av stor betydelse att identifiera den ursprungliga hypodiploida klonen för korrekt diagnostik och behandling.

förvärvade TP53-mutationer. Fall med 40-45 kromosomer skiljer sig åt genetiskt och upphandlar inte samma distinkta genetiska profil. Inga tidigare studier har således specifikt undersökt sådana fall med verktyg som NGS.

Syftet med mitt avhandlingsarbete har varit att kartlägga genetiska förändringar hos patienter med en dålig prognos. I avhandlingens första delarbete (artikel I) undersökte benmärgs- eller blodprover från 126 vuxna med ALL. I denna studie användes ”SNP array”, en metod som detekterar förändringar i leukemicellerna. Gener i flera kopior, eller få kopior, kan följaktligen identifieras. Jag fann återkommande genetiska förändringar i gener som visat sig vara cancerdrivande i ALL hos barn såsom förlust av CDKN2A, IKZF1, PAX5 och EBF1. Jag kunde även i 2-6% av fallen identifiera genförluster i tidigare fem okända gener, och utöver detta påvisa att återfallsmekanismer är genetiskt lika i barn och vuxen ALL.


kategoriserats i en annan subgrupp egentligen bör behandlas som hypodiploid ALL med högriskbehandling. Av denna anledning bör mutationsprofil vara viktigare än kromosommodaltal i dessa fall.

Sammanfattningsvis har jag i mitt avhandlingsarbete undersökt fall med en dålig prognos. Få stora studier har fokuserat på den genetiska bakgrunden i vuxen-ALL. Jag har påvisat att vuxna och barn bär på genetiska förändringar som liknar varandra. Vidare har resultaten från studierna i mitt avhandlingsarbete ökat förståelsen för hypodiploid ALL. Genetiska analyser på specifika gener som är kopplade till prognos kan förhoppningsvis leda till en bättre behandling av vuxen-ALL och hypodiploid ALL och möjliggöra utvecklingen av nya behandlingsmetoder.
Acknowledgements

I would like to express my deepest appreciation and gratitude to everyone who has been a part of this journey. In particular, I would like to thank:

My supervisor, Kajsa Paulsson. You have shown me how exciting the world of genetics can be, and you have truly inspired me with your vast knowledge and expertise. Thank you for all the support, motivation, excellent scientific meetings where we have not only discussed research but also every other possible topic! Thank you for being an outstanding supervisor, you have guided the way but allowed me to stand on my own, and always been there if I needed anything. Thank you for these amazing years, I could not have imagined this journey without you as my supervisor.

My co-supervisor, Bertil Johansson, thank you for your dedicated work and time put towards my research, and also thank you for sharing your vast and valuable knowledge, which has always helped my research to advance.

Eleanor Woodward, we have shared many fun memories and I want to thank you for your friendship and support. I will miss our stimulating scientific discussions and all our “fika” breaks and “chocolate-runs”. Thank you for being there at all hours, whenever I needed your help. Andrea Biloglav, thank you for everything you taught me in the lab, for answering any labrelated question (even on your holiday =)). Thank you for all the many laughs, conversations on just about everything and the support these past years. Linda Olsson, you welcomed me to Skåne when I first moved here, and made me feel right at home. You have made time for any research related issue I had, and encouraged me along the way. I will miss all our fun discussions about SNP, genetics and life.

Thank you Kristina K, for being an excellent conference companion, and for all the support in my project, when things seemed impossible a positive email from you made all the difference. To Kristina L, for taking your time
to listen to any problems I have and for always being positive, and for all the fun times! Thank you to Catarina and Rebeqa for your advice and encouragement in my projects.

Past and present PhD students; Anders, Charles, Daniel, Naveen, Pablo, Mattias and Ram, for all the good times and stimulating conversations. Specifically, thanks to Elsa, for your friendship these past years and for your support whenever I needed it. Our many scientific discussions (and on everything else from fashion to life) in the PhD room will be greatly missed! To Jakob, for being an excellent roommate and for promoting “SC”. To Axel, Mia, Niklas and Sofia (P.W), thank you for all the many laughs, for your support, and for listening to my hypochondriacal ramblings over these past years.

Thank you to all the principal investigators, specifically Felix Mitelman for creating such an excellent department, Fredrik Mertens, Thoas Fieretos, David Gisselsson Nord, Anna Andersson, Marcus Järås, and Karolin Hansen Nord, for contributing to the very creative research environment.

To All of the very skilled and helpful co-workers at the department of Clinical Genetics, specifically Calle, Linda HM, Jenny K, Jenny N, Nils, and Henrik, for helping me with any lab-related question and sharing your vast knowledge with me throughout the years. Marianne and Linda M, thank you for the supportive talks in the corridor about life and work.

To Anette Welin, for always being kind and helping me with any problems I have and always doing so with a smile.

Thank you to All of the very skilled co-workers at Clinical Genetics at the University Hospital. Specifically, to Mina, for always being so positive and encouraging my work.
To my family and friends

To All my friends for your support, specifically to Sara, Harri, Hannah, Jenny and Dilan, we have known each other since forever, and you have always have been there for me whenever I needed you! I couldn’t have imagined this journey without all of you by my side. Jenny, thank you also for being my amazing toastmaster, and Dilan, who always made time to listen to my research presentations, your encouragement has meant the world to me.

To Ninnie, we have been through everything together and you have been invested in both my research and in all other aspects of my life, I couldn’t be more grateful for our friendship.

To Marcus, thank you for making life outside work so much more fun. I’m grateful for all the times you kept me company (usually at midnight when I frantically needed to get something done) before presentations/thesis writing, waiting in the PhD room reassuring and encouraging me. Your support and love has made all the difference.

Thank you Thomas for inspiring me with your very profound knowledge about everything, and for all your help specifically these last months of intense writing.

To my uncle Magid, you have encouraged and motivated me in everything I set out to do since a young age, and I am so very appreciative for that. Thank you for always being there for me and I greatly value all the life lessons and beautiful travels we had throughout the years.

My sister, Sahar, you make me realize my true potential. Thank you for pushing me to excel and helping me reach my goals. You always stand by my side no matter what, and thank you for moving mountains for me. Whenever I feel lost, you help me find the way.

To my mother, Sara, you are my rock. Thank you for believing in me, when things seem impossible you always make me see the possibilities. During this amazing journey, you have shared all the ups and downs with me and your strength and positivity has guided the way when I have needed
it the most. I am so grateful for your endless support and love. Thank you for helping me aim for the stars.


Articles I-IV