Thesis for doctoral degree (Ph.D.)
2007

Genetic Studies of Acute Lymphoblastic Leukemia

Ekaterina Kuchinskaya
GENETIC STUDIES OF
ACUTE LYMPHOBLASTIC LEUKEMIA

Ekaterina Kuchinskaya

Stockholm 2007
All previously published papers were reproduced with permission from the publisher
©Ekaterina Kuchinskaya, 2007

Published and printed by
REPROPRINT AB
Stockholm 2007
www.reproprint.se
Gårdsvägen 4, 169 70 Solna
A bend in the road is not the end of the road

To my family
ABSTRACT

Acute lymphoblastic leukemia (ALL) is the most common pediatric malignancy, comprising approximately 25% of all childhood cancers. A huge improvement in outcome of children with ALL during the last 20 years has resulted in remission rates exceeding 95% and a 5 year event-free-survival approaching 80%. Tailored and individualized treatment with prolonged regimens of multi-agent chemotherapy based on clinical and cytogenetic prognostic factors may explain this development. However, although ALL may be considered as the same disorder in adults and children, the improvement in outcome has been less prominent in adults, and relapse remains the most important cause of treatment failure in all age groups. Therefore, there is a need for more detailed and functional characterization of leukemic clones with already established recurrent cytogenetic abnormalities, as well as for detection and investigation of previously unidentified recurrent changes.

In the first study we used a whole genome expression array to study and compare gene expression profiles in children and adults in order to find an explanation for the difference in outcome. However, no specific expression signatures, characteristic for the different age groups, were observed. Surprisingly, we detected some similarities in the gene expression pattern between infants with MLL gene translocation and older adults. We also observed that prognostically important cytogenetic aberrations determined the pattern of the expression profile and defined genes, in which the expression pattern was characteristic of cytogenetic subgroups independent of age. Gene expression profiling turned out to be a sensitive tool and helped us to refine the karyotype in a few patients.

Some recurrent chromosomal aberrations in ALL remain hard to discover and their prognostic importance is still unclear. Intrachromosomal amplification of 21q is often detected as multiple RUNX1 gene signals on interphase FISH analysis, but the size of the amplicon and the mechanism of the amplification have not been thoroughly studied. We investigated eight patients with 21q amplification using array-CGH. A step-wise pattern of amplification was found in all patients, and the amplification was followed by a terminal deletion of 21q in some of them. All imbalances at the genome level were confirmed by FISH. Breakage-fusion-bridge cycles were suspected as the mechanism of amplification, and this hypothesis was supported by the observation of anaphase bridge structures including DNA from chromosome 21 in two cases.

The same approach was used to characterize seven cases with dic(9;20). Although no identical breakpoint was found, clustering of the breakpoints, covering a 1.5 Mb segment of 9p13.2 and a 350 kb segment of 20q11.2, was observed. All cases were unbalanced with loss of 9p13.2-pter and five had loss of 20q11.2-qter, whereas two displayed gain of 20cen-pter. This means that the cytogenetic description of this abnormality should be dic(9;20)(p13.2;q11.2). It remains to be elucidated whether the
rearrangement leads to a chimeric gene or if the functionally important outcome is loss of 9p and/or 20q.

In the forth study, 28 patients with ALL and normal or failed karyotype on G-banding were studied with a 33K, tiling BAC array. This approach allowed us to revise the karyotypes in 75%, and genetic changes were found in 86% of the patients. Most of these patients showed copy number alterations that were below the resolution of G-banding. The frequency of ETV6 gene deletion was found to be underestimated. In three patients no aberrations were found, which may be explained by the limitation of the method, absence of copy number alterations in these samples, e.g. balanced chromosome rearrangements, or other genetic mechanisms responsible for the development of leukemia.
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications, which are referred to in the text by their Roman numerals:


TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS ........................................................................ 7
LIST OF ABBREVIATIONS .................................................................................. 10
INTRODUCTION ................................................................................................. 11
  Etiology ............................................................................................................. 11
  Hematopoiesis ................................................................................................. 12
  Age related differences in ALL ....................................................................... 13
CLINICAL DIAGNOSIS OF ALL ....................................................................... 15
CHROMOSOME ABNORMALITIES IN ALL ......................................................... 16
  Numerical abnormalities ................................................................................. 17
  High hyperdiploidy ........................................................................................ 17
  Hypodiploidy .................................................................................................. 18
  Moderate hyperdiploidy ................................................................................ 18
  Prognostically important structural aberrations ......................................... 18
  in BCP-ALL .................................................................................................. 18
  ETV6/RUNX1 rearrangement ........................................................................ 18
  BCR/ABL1 rearrangement .............................................................................. 20
  PBX1/TCF3 rearrangement .............................................................................. 21
  MLL (mixed lineage leukemia gene) rearrangements .................................. 22
  Intrachromosomal amplification of 21q ......................................................... 22
  Other recurrent structural abnormalities ....................................................... 23
T-ALL .................................................................................................................. 24
  Normal karyotype in ALL .............................................................................. 25
  Breakage-Fusion-Bridge Cycles .................................................................... 25
MODERN DIAGNOSTIC APPROACHES IN ALL ........................................... 27
  Gene expression studies of ALL ................................................................... 28
  Array-CGH .................................................................................................... 29
AIMS OF THE STUDY ...................................................................................... 31
MATERIALS AND METHODS .......................................................................... 32
  Patients ........................................................................................................... 32
  Material .......................................................................................................... 32
  Methods ......................................................................................................... 33
    FISH ........................................................................................................... 33
    SKY ............................................................................................................. 33
    RT-PCR ....................................................................................................... 34
    Array-CGH ................................................................................................. 34
  Large scale gene expression analysis ......................................................... 35
RESULTS ........................................................................................................... 38
  Gene expression analysis in children and adults with ALL (Paper I) .......... 38
  Delineation of intrachromosomal amplification of 21q in patients with ALL (Papers II, IV) ................................................................. 38
  Characterization of dic(9;20)(p11-13;q11)(Paper III) ................................... 40
  Characterization of patients with normal karyotype by G-banding (Paper IV) ........ 40
DISCUSSION........................................................................................................................................42
  Application of whole genome expression arrays............................................................................42
  Array-CGH to study genomic imbalances in ALL .............................................................................43
    Intrachromosomal amplification of 21q .......................................................................................43
    Characterization of dic(9;20). .......................................................................................................44
    Characterization of patients with a normal karyotype .................................................................44
FUTURE DEVELOPMENT .....................................................................................................................46
ACKNOWLEDGEMENTS ......................................................................................................................47
REFERENCES..........................................................................................................................................50
## List of 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>AML</td>
<td>acute myeloid leukemia</td>
</tr>
<tr>
<td>Array-CGH</td>
<td>array comparative genomic hybridization</td>
</tr>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosome</td>
</tr>
<tr>
<td>BCP-ALL</td>
<td>B-precursor ALL</td>
</tr>
<tr>
<td>CGH</td>
<td>comparative genomic hybridization</td>
</tr>
<tr>
<td>CNA</td>
<td>copy number alteration</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>CRA</td>
<td>common region of amplification</td>
</tr>
<tr>
<td>EFS</td>
<td>event-free survival</td>
</tr>
<tr>
<td>EST</td>
<td>expressed sequence tag</td>
</tr>
<tr>
<td>FISH</td>
<td>fluorescent <em>in situ</em> hybridization</td>
</tr>
<tr>
<td>HSR</td>
<td>homogeneously staining regions</td>
</tr>
<tr>
<td>Kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>Mb</td>
<td>megabase</td>
</tr>
<tr>
<td>MRD</td>
<td>minimal residual disease</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>SKY</td>
<td>spectral karyotyping</td>
</tr>
<tr>
<td>WBC</td>
<td>white blood cells</td>
</tr>
<tr>
<td>AYA</td>
<td>adolescents and young adults</td>
</tr>
<tr>
<td>BFB cycles</td>
<td>breakage-fusion-bridge cycles</td>
</tr>
<tr>
<td>CNV</td>
<td>copy number variation</td>
</tr>
</tbody>
</table>
INTRODUCTION

Acute lymphoblastic leukemia (ALL) is a malignant disorder that originates in a single B- or T-lymphocyte progenitor. The term “leukemia”, meaning “white blood” was suggested by R. Virchow in 1847. It reflects the increase of white cells in the blood of the patient. By the end of the nineteenth century the bone marrow was described as the source of the blood and its role in leukemia development was established. Leukemia was classified into myeloid and lymphatic groups and acute leukemia was separated from chronic leukemia based on a rapid development of the disease and a short time between presentation and death. In spite of these scientific advances, leukemia was a fatal disease, inevitably leading to death. Treatment with arsenic, blood transfusions and radiation led only to short remissions (Piller 2001).

The new era in the treatment of leukemia started during the Second World War with the introduction of the experimental medicine and chemically synthesized compounds: alcylation agents and folic acid inhibitors. In addition, at the same time the therapeutic value of adrenal corticosteroids was discovered (Beutler 2001, Thomas 2006). However, the main breakthrough in the treatment of leukemia took place during the 1970s, when continuous regimens utilizing combinations of a few drugs were introduced into clinical practice. Nowadays, the treatment of leukemia and especially childhood acute lymphoblastic leukemia is the greatest success in cancer treatment with remission rates in children exceeding 95% and an event-free survival (EFS) of around 80 (Schrappe, et al 2000).

Etiology

ALL develops as the result of a combination of genetic and environmental factors. Some of these factors are well established, e.g. genetic predisposition (including certain genetic conditions, such as Down’s syndrome and Fanconi’s anemia), viral infection (HTLV-1), ionizing radiation, chemical agents (benzene), drugs (alkylating agents, topoisomerase II inhibitors). However, the link between a certain known predisposing agent and development of ALL can only be shown in a small proportion
of cases and in the majority the question of the cause of ALL remains unresolved. Genetic studies of identical twins, detection of leukemia-specific fusion genes or immunoglobulin genes rearrangements in neonatal blood have established the prenatal origin of leukemia in many childhood cases (Greaves, et al 2003, Maia, et al 2004). It is believed that the first genetic event necessary for the development of childhood ALL frequently originates \textit{in utero}, and is followed by additional postnatal mutations that lead to clinically overt leukemia. (Mori, et al 2002, Stam, et al 2006).

**Hematopoiesis**

After birth, hematopoiesis in the human is restricted to the bone marrow. Infants have hematopoietic bone marrow in all bones and in adults it is in the central skeleton and in proximal ends of long bones. Hematopoiesis results in the production of various blood cell types. Lineage specification and cell commitment is achieved by the precise activation and/or repression of specific genes (Bonnet 2005). Hematopoiesis is a continuous process, but can be separated into distinct stages. Depending on the need, the bone macro- and microenvironment and growth factors influence pluripotent stem cells to differentiate into committed stem cells of either the myeloid or lymphoid type. They have a limited capacity for self-renewal, but have the potential to differentiate and develop mature progeny. Next, committed stem cells, influenced by various growth factors, differentiate into lineage-specific progenitor cells. Myeloid progenitor cells give rise to erythrocytes, granulocytes and megakaryocytes. Lymphoid progenitor cells differentiate into T-cells and B-cells that have different protective functions in the immune system (Travlos 2006) (Figure 1). Within the hematopoietic microenvironment, early progenitors are maintained in specific compartmentalized niches where they interact with other cell types and components of extracellular matrix (Bonnet 2005). Transcription factors represent nodal points of control of hematopoiesis. Different families of transcription factors regulate the developmental program of stem cells and when their behavior is disrupted, leukemic proliferation may be initiated.

In acute leukemias the pre-leukemic blast cell undergoes genetic changes,
leading to malignant transformation, and successive mitotic divisions give rise to subclones. Transformed cells are frozen at a particular stage of differentiation. The fact that the most frequent targets of chromosomal translocations in acute leukemias are transcription factors emphasizes their critical role in hematopoiesis (Lecuyer and Hoang 2004). The degree of differentiation and clinical behavior of the leukemia are dependent on the stage of maturation arrest of the cells in the bone marrow (Sell 2004).

Age related differences in ALL

A large number of different genetic changes are found in ALL, and there are also differences in outcome of the disease. More than 95% of children with ALL enter remission on induction protocols and up to 75-80% never relapse and can be considered
as cured (Schultz, et al 2007). The situation in the adult setting is less encouraging. Most of the adults (up to 85%) enter a first remission, but more that half of them relapses early and the 5-year event-free survival in most clinics seldom exceeds 40% (Moorman, et al 2007a). A third age group may be referred to as adolescent and young adult (AYA) group (Bleyer 2007). The age until which patients are treated in pediatric clinics, and according to pediatric protocols, differs between 15 and 21 years. The analysis of survival in this age group has suggested that adolescents and young adults do better when treated with pediatric protocols (Boissel, et al 2003, Hallbook, et al 2006, Ramanujachar, et al 2006). In general, the cure rate of AYAs is worse than that of small children, but better than in adults (Jeha 2003, Nachman 2005).

Moreover, the distribution of prognostic molecular genetic abnormalities differs between age groups (Table 1), which led to the suggestion that ALL in children and adults represents two different nosologic entities with a similar clinical picture (Forestier and Schmiegelow 2006). The proportion of cases bearing unfavorable prognostic abnormalities is higher in adults, and those with favorable ones are more

<table>
<thead>
<tr>
<th>Karyotype</th>
<th>Genes involved*</th>
<th>Adults, %§</th>
<th>Children, %§</th>
<th>Adults*</th>
<th>Children*</th>
<th>Imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperdiploid</td>
<td>Unknown</td>
<td>2-11</td>
<td>25</td>
<td>F</td>
<td>F</td>
<td>B</td>
</tr>
<tr>
<td>Near tetraploid</td>
<td>Unknown</td>
<td>2</td>
<td>&lt;1</td>
<td>F</td>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>t(12;21)(p13;q22)</td>
<td>ETV6/RUNX1</td>
<td>0-3</td>
<td>20-25</td>
<td>NS</td>
<td>F</td>
<td>B</td>
</tr>
<tr>
<td>12p aberrations</td>
<td>Unknown</td>
<td>4-6</td>
<td>7-9</td>
<td>F</td>
<td>NS</td>
<td>B/T</td>
</tr>
<tr>
<td>t(10;14)(q24;q11)</td>
<td>Ectopic HOX11</td>
<td>2-3</td>
<td>NA</td>
<td>F</td>
<td>NS</td>
<td>T</td>
</tr>
<tr>
<td>14q11 deletion</td>
<td>T-cell receptor</td>
<td>5-7</td>
<td>3-4</td>
<td>F</td>
<td>NS</td>
<td>T</td>
</tr>
<tr>
<td>Normal</td>
<td>Unknown</td>
<td>15-35</td>
<td>31-40</td>
<td>I</td>
<td>I</td>
<td>B/T</td>
</tr>
<tr>
<td>Pseudodiploid</td>
<td>Unknown</td>
<td>30-50</td>
<td>18-26</td>
<td>A</td>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>Near triploid</td>
<td>Unknown</td>
<td>3</td>
<td>&lt;1</td>
<td>A</td>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>t(1;19)(q23;p13.3)</td>
<td>PBX1/TCF3</td>
<td>2-3</td>
<td>4-5</td>
<td>A</td>
<td>F</td>
<td>B</td>
</tr>
<tr>
<td>9p deletion</td>
<td>CDKN2A locus</td>
<td>6-30</td>
<td>7-11</td>
<td>I</td>
<td>A</td>
<td>B/T</td>
</tr>
<tr>
<td>t(9;22)(q34;q11)</td>
<td>BCR/ABL1</td>
<td>37</td>
<td>2-6</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>t(4;11)(q21;q23)</td>
<td>MLL/AF4</td>
<td>3-7</td>
<td>2</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>-7, del(7q), del(7p)</td>
<td>Unknown</td>
<td>6-11</td>
<td>4</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>del(5q)</td>
<td>Unknown</td>
<td>&lt;2</td>
<td>1</td>
<td>NS</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>+8</td>
<td>Unknown</td>
<td>10-12</td>
<td>2</td>
<td>A</td>
<td>NS</td>
<td>B</td>
</tr>
<tr>
<td>del(6q)</td>
<td>Unknown</td>
<td>3-16</td>
<td>6-9</td>
<td>NS</td>
<td>NS</td>
<td>T</td>
</tr>
</tbody>
</table>

§ Frequency with which aberrations are found in adult and pediatric populations; * Prognostic impact of aberrations in adults and children: favorable prognosis (F), intermediate risk (I), adverse prognosis (A), and prognosis is not specified (NS); Imm indicates whether the aberration is associated with B-cell (B) or T-cell (T) immunophenotype.
common in children (Nachman 2005). Some aberrations, such as the ETV6/RUNX1 translocation, intrachromosomal amplification of 21q and dic(9;20) are almost never diagnosed in adults. Acute leukemias share common, but not uniform clinical features, such as resistance to chemotherapy or the potential for extramedullary or central nervous system (CNS) infiltration. It has been suggested that the gene expression of blast cells reflects differences in the underlying mechanisms involved in leukemogenesis and thus the expression pattern of leukemic cells could be used to classify the cells into functionally distinct groups (Scandura 2005).

**CLINICAL DIAGNOSIS OF ALL**

At the time of the initial diagnosis, the discrimination of acute lymphoblastic from acute myeloid leukemia (AML) is vital because the treatment strategies for these diseases are distinct (Scandura 2005). Prognostically important clinical parameters in ALL are age, sex, white blood cells (WBC) count, bulky extramedullary disease and CNS involvement (Smith, et al 1996). Immunophenotype is another important factor affecting both the choice of treatment, the methods of follow-up and the outcome. Therefore, detailed immunophenotyping of the leukemic clone, both for the subclassification of cases according to the maturation-stage of normal B- and T-lineage lymphocytes, and as baseline investigation for monitoring of minimal residual disease (MRD), is essential (Pui and Jeha 2007). Treatment response is another important prognostic parameter, which also strongly modulates the effect of all other risk-factors, since the disease is uniformly fatal without treatment. A rapid response to treatment is an indicator of a favorable prognosis while slow responders with >5% blasts in the bone marrow at day 29 have a higher risk of relapse. Response to treatment measured as MRD has also proven to be highly predictive of outcome (Izraeli and Waldman 2004).

Analysis of the patients karyotype provides crucial information of clinical and prognostic importance, and allows a more individualized approach to the choice of treatment strategy. More than half of the newly diagnosed ALL cases display detectable and usually single cytogenetic abnormalities on G-banding (Mrozek, et al 2004), and this method still remains the golden standard in the genetic diagnosis of ALL. Other
molecular cytogenetic methods, such as FISH and SKY, provide additional information, which in many cases may lead to the revision of the initial karyotype (Nordgren 2003). Balanced translocations are the most common genetic lesions in both ALL and AML, and represent critical early events in the genesis of the leukemic clone. Genes encoding transcription factors are nearly always found at one of the breakpoints, and the fusion proteins that are formed by the translocations generally interfere with the normal function of one or both of the rearranged genes (Greaves and Wiemels 2003, Scandura 2005). The translocations often give rise to unique leukemia-associated fusion transcripts that can be identified using reverse transcriptase polymerase chain reaction (RT-PCR). These transcripts can be detected with high sensitivity and specificity, thus making them very useful to follow minimal residual disease (MRD).

**CHROMOSOME ABNORMALITIES IN ALL**

Chromosome abnormalities in ALL are found in 65-70% of the patients by standard karyotyping (Nordgren 2003). They are important indicators of prognosis and may be divided into two groups: numerical and structural abnormalities. Common numerical abnormalities include high hyperdiploidy, moderate hyperdiploidy and hypodiploidy. Many recurrent structural aberrations can be found in ALL, most of them are translocations that give rise to fusion genes with different prognostic impact. Common pronostically important structural aberrations include t(9;22)(q34;q11), t(12;21)(p13;q22), t(1;19)(q23;p13), MLL rearrangements (with t(4;11)(q21;q23) and t(11;19)(q23;p13) as the most common) and intrachromosomal amplification of 21q in the BCP-ALL group (Moorman, et al 2007b, Rubnitz and Pui 2003). There are many recurrent structural aberrations in T-ALL, but their prognostic impact has not been determined in the majority of cases (Ballerini, et al 2002, Gottardo, et al 2007, Graux, et al 2004, Uckun, et al 1998). The frequency of aberrations associated with adverse prognosis is higher in adults, and this may partly explain the worse prognosis compared to children (Plasschaert, et al 2004). In a number of patients with ALL, chromosome
analysis fails or results in a normal karyotype findings. This prevents correct risk stratification and may lead to under- or overtreatment of patients.

**Numerical abnormalities**

**High hyperdiploidy**

High hyperdiploidy in ALL is usually defined as 52-60 chromosomes and is characterized by multiple nonrandom trisomies and tetrasomies. It occurs in approximately 30% of pediatric and 2-11% of adult B-precursor ALL (BCP-ALL), and constitutes a distinct subgroup associated with low-risk features and is characterized by a favorable prognosis (Raimondi, et al 1996). The most common gains are X, 4, 6, 10, 14, 17, 18, and 21 (Heerema, et al 2007, Moorman, et al 2007a, Nordgren, et al 2001). A particularly favorable prognosis has been described in association with gain of chromosomes 4, 10, 6 and 17, whereas a gain of chromosome 5 and i(17q) have been associated with poor prognosis (Heerema, et al 2000). No specific structural rearrangements have been found in ALL with high hyperdiploidy (Mitelman, et al 2007) and no specific pattern of subtle gains or losses was detected in this cytogenetic subtype using array-CGH (Mullighan, et al 2007, Paulsson, et al 2006). The small unbalanced changes never involved chromosomes that are frequently gained in hyperdiploid ALL (Paulsson, et al 2006, Strefford, et al 2006). The mechanism behind high hyperdiploidy and the pathogenic consequences remains unknown. Recent findings of clonotypic immunoglobulin genes rearrangements in neonatal blood spots, cord blood and twin studies indicated that ALL with high hyperdiploidy has a prenatal origin, at least in some cases, but that one or more independent additional postnatal genetic events would be required for development of overt leukemia (Maia, et al 2004, Panzer-Grumayer, et al 2002, Taub, et al 2002, Yagi, et al 2000). Missense mutations of FLT3, detected in around 20% of hyperdiploid leukemia has been suggested as one of the secondary genetic events in this subgroup (Taketani, et al 2004). A recent report demonstrated evidence that hyperdiploidy arises through simultaneous gain of chromosomes in a single abnormal mitosis in a majority of the cases (Paulsson, et al 2005), and expression analyses have shown that hyperdiploid childhood leukemias
have a unique gene expression pattern, associated with the specific trisomies (Yeoh, et al 2002).

**Hypodiploidy**

Hypodiploidy with less than 45 chromosomes, and particularly near-haploidy (24-28 chromosomes), is associated with an extremely poor prognosis in both children and adults (Moorman, et al 2007a, Moorman, et al 2007b). The results of a multicenter study including children with ALL from 10 national studies showed that hypodiploidy occurred in about 5% of children (excluding patients with established chromosomal structural abnormalities). Of these only 1% had less than 45 chromosomes. The event-free survival in this group of children is about 38%, and the overall survival only 50% (Nachman, et al 2007). Adults with low hypodiploidy and near triploidy have recently been shown to represent a single distinct subtype of adult ALL (Charrin, et al 2004). The proportion of such patients is comparable with children (4%), and the 5-year overall survival is very low (22%) (Moorman, et al 2007a).

**Moderate hyperdiploidy**

The group with 47-51 of chromosomes is often characterized by presence of the concurrent structural abnormalities, which are detected in 75% of all cases and is associated with intermediate prognosis (Forestier, et al 2000). Most cases have 47 chromosomes, fewer have 48 chromosomes and a modal number of 49-51 chromosomes is rather rare. The most common chromosome gains are +21, +X, +8, and +10.

**Prognostically important structural aberrations in BCP-ALL**

**ETV6/RUNX1 rearrangement**

The t(12;21)(p13;q22), resulting in ETV6/RUNX1 gene rearrangement, fuses ETV6, which encodes an ETS family transcription factor, and RUNX1, which encodes a transcription factor with a DNA binding domain. The t(12;21) is the most frequent
structural aberration in childhood ALL, and is found in around 25% of the cases (Forestier, et al 2007), but it is rarely (3-5%) diagnosed in adults (Lee, et al 2005, Moorman, et al 2007a). Favorable prognostic factors associated with this aberration are age between 2 and 5 years, and low leukocyte counts. This abnormality was initially associated with a favorable prognosis (Maloney, et al 1999, Rubnitz, et al 1999, Uckun, et al 2001), although later studies observed no significant survival differences (Hann, et al 2001, Hubeek, et al 2001, Takahashi, et al 1998) and late relapses are not infrequent. Twin studies and analyses of neonatal blood spots have shown that ETV6/RUNX1 often arises prenatally as an early or initiating mutation, but additional genetic events are required for the development of overt leukemia (Greaves, et al 2003, Greaves and Wiemels 2003, Zelent, et al 2004). The ETV6/RUNX1 fusion is considered to arise in a B-cell committed progenitor (CD34⁺CD33⁻CD19⁺ cells), and is never found in myeloid progenitor cells (Castor, et al 2005).

In addition to the cryptic t(12;21), more than 50-70% of patients display different chromosome aberrations, mainly unbalanced abnormalities (Forestier, et al 2007), with loss of 12p and gain of chromosome 21 as the most common, identified in almost 30% of the patients. The role of 12p13 deletion, including the ETV6 gene, is not clear. Based on the high frequency of 12p13 deletion in patients with ETV6/RUNX1 translocation (about 50%), it has been suggested to represent an important secondary event, required for the development of overt leukemia (Zuna, et al 2004). Recent studies employing array-CGH showed that the frequency of the deletion is higher than reported and often, only a few exons of the gene are deleted interfering with its function (Mullighan, et al 2007, Tsuzuki, et al 2007). 12p13 deletions are found not only in ALL patients with ETV6/RUNX1 translocation, but also in several solid neoplasias and in ALL patients with a normal karyotype and high hyperdiploidy, which has led to the hypothesis that a tumor suppressor gene is located in this region, with ETV6 as the most obvious candidate (Wood, et al 2003).

RUNX1 is an important player in different pathways, and its rearrangements are detected in different hematopoietic malignancies. RUNX1 proteins bind to a nucleotide sequence TGT/cGGT, found in the promoter of many genes that are important in
hematopoiesis (Mikhail, et al 2002). Its target genes include growth factors (GM-CSF, IL-3), growth factor receptors, signaling molecules (bcl-2, p21), transcription factors (c-myb) and effector molecules (granzyme B, myeloperoxidase) (Michaud, et al 2003). Fusion proteins resulting from a RUNX1 translocation actively repress expression of a variety of target genes due to the high affinity of the fusion protein to CFBβ, which is greater than that of the wild-type RUNX1 protein. In addition, germline mutations of RUNX1 have been associated with rare cases of a familial platelet disorder with predisposition to acute myelogeneous leukemia (FPD/AML) (Song, et al 1999). RUNX1 is a common translocation partner in myeloid malignancies, e.g. t(8;21), t(3;21), and more rarely t(16;21) and t(19;21) (Michaud, et al 2003), but the only translocation involving RUNX1 in ALL is t(12;21(p13;q22). A number of studies have identified point mutations of RUNX1 in various sporadic myeloid malignancies, but additional chromosomal abnormalities have always been present (Imai, et al 2000, Osato, et al 1999) and just an altered transcription of RUNX1 was not sufficient for leukemogenesis.

**BCR/ABL1 rearrangement**

The t(9;22)(q34;q11) occurs in 3-5% of children, the frequency increases with age and it is found in about 25% of adults with ALL. It is associated with a very poor prognosis (Pui and Jeha 2007) in all age groups. The BCR/ABL1 fusion protein acts as an oncoprotein by activating several signalling pathways, which contribute to the malignant transformation (Advani and Pendergast 2002). The size of the BCR fragment included in the fusion protein differs, because of splicing at different breakpoints in the BCR locus. Three distinct protein products have been described: p190 (e1a2 junction), p210 (b2a2 or b3a2 junction), and p230 (e19a2 junction). This difference in structure influences the biological and clinical phenotypes associated with the BCR/ABL1 variants (Melo 1996) (Figure 2). In children with ALL, P190 is more frequent and represents 80-90% of all childhood Philadelphia positive ALL. In adults, the distribution of P190 and P210 is more equal (Arico, et al 2000). The difference in BCR breakpoints also reveals the stem cell origin of some leukemias. The cases with a P210 BCR/ABL1 fusion originate in a CD34+CD38−CD19 candidate stem cell population corresponding to a lymohomyeloid
hematopoietic stem cell. In contrast, ALL with a P190 BCR/ABL1 fusion arises in a progenitor cell committed to B-cell lineage (Castor, et al 2005). Nowadays, patients with both breakpoints are regarded as a single disease entity and treated with the same protocols. However, a recent study indicated a better prognosis in patients with P190 BCR/ABL1 ALL (Cimino, et al 2006).

Recently, a tyrosine kinase inhibitor (imatinib), has been effectively utilized in the treatment of BCR/ABL1 positive childhood and adult ALL (Jabbour, et al 2007, Thomas, et al 2004, Towatari, et al 2004), offering a unique targeted treatment to this cytogenetic subgroup, which may help to overcome the adverse prognosis in patients with this abnormality.

**PBX1/TCF3 rearrangement**

t(1;19)(q23;p13) fuses the TCF3 (E2A) gene on 19p13 to the PBX1 gene on 1q23, and occurs in a balanced and an unbalanced form, which arises from non-disjunction leading to loss of the der(1) chromosome (Boomer, et al 2001, Shikano, et al 1986). This aberration is found in 5% of children and <5% of adults with ALL (Foa, et al 2003). The translocation in children has been associated with inferior prognosis when treated with less intensive therapy (Boomer, et al 2001, Crist, et al 1990, Raimondi, et al 1990). It has been shown that blast cells carrying the aberration accumulate less methotrexate polyglutamates, compared to those with other subtypes of BCP-ALL. The expression of the SLC19A1 gene, which encodes the methotrexate uptake transporter, is reduced in patients with t(1;19) compared to patients with high hyperdiploidy (Kager, et al 2005),
and this has suggested that higher doses of methotrexate would lead to a better outcome in these patients. At present the 5-year event-free survival is 92% in children (Gadner, et al 2006), but the translocation is associated with an adverse prognosis in adults (Piccaluga, et al 2006).

**MLL (mixed lineage leukemia gene) rearrangements**

Rearrangements of the *MLL* gene at 11q23 are found in 75% of infants with ALL, with t(4;11)(q21;q23) as the most common one. *MLL* rearrangements involve about 50 different partner genes (Pui and Campana 2007). They are also found in older children and adults, but at a lower frequency (2% and 7%, respectively) (Johansson, et al 1998, Moorman, et al 2007a, Pui and Evans 2006, Pui, et al 1995). The prognosis is poor in infants and adults, but better in older children (Mann, et al 2007, Pui, et al 2003). In most cases the translocations are present at birth, establishing the prenatal origin of leukemia in infants (Greaves and Wiemels 2003). *MLL* rearrangements are, as the name suggests, found in both ALL and AML and in a proportion of mixed lineage leukemia (Dimartino and Cleary 1999). Gene expression studies have shown that patients with ALL and *MLL* rearrangements have unique expression signatures, which are different from patients with AML and other patients with ALL (Armstrong, et al 2002). Further gene expression studies have been able to separate infants with *MLL* rearrangements into two subgroups, one of which had an extremely poor prognosis (Tsutsumi, et al 2003).

**Intrachromosomal amplification of 21q**

Intrachromosomal amplification of 21q has been recently described as a recurrent abnormality in childhood ALL (Harewood, et al 2003, Soulier, et al 2003), with a frequency of 2% in a large series of patients (Moorman, et al 2006). No adult ALL cases have been reported so far. Children with this aberration are older, have a lower WBC count and a lower platelet count compared to the general ALL population. The amplification has been associated with an adverse prognosis, and the patients with this aberration are currently treated as high risk in some protocols (Moorman, et al 2007b, Robinson, et al 2003). The 5 year EFS has been described to be as low as 29%, which is
significantly worse than would be expected based on other prognostic parameters (Moorman, et al 2006). The abnormality was initially described as RUNX1 amplification, as it is often diagnosed as a cluster of signals from RUNX1, using FISH to detect the ETV6/RUNX1 rearrangement (Niini, et al 2000). Array-CGH has allowed detailed characterization of the amplicon, which varies significantly in size between patients from 15 to 30 Mb, with a common region of amplification (CRA) of 6-6.5 Mb, including RUNX1 and a multitude of genes (Kuchinskaya, et al 2007, Strefford, et al 2007). The amplification has a characteristic step-wise increase towards the telomere, followed by a deletion of the very telomeric part in half of the patients, and has not been associated with any other specific aberration. Gene expression analysis showed that RUNX1 was not differentially expressed in patients with intrachromosomal amplification of 21q compared to other ALL cases, and no other candidate genes were identified within the CRA that might be driving the leukemia in these patients (Strefford, et al 2006). Thus, whether RUNX1 or any other gene within the CRA is the target of amplification in this cytogenetic subgroup remains to be established.

Breakage-fusion-bridge (BFB) cycles have been suggested as the mechanism of formation of this abnormality (Robinson, et al 2007). The distribution of signals from locus-specific probes on 21q indicates that amplification is formed as a result of multiple breaks and reunions, leading to complex chromosomal rearrangements, including inversions, deletions, amplifications and duplications. Deletions of the subtelomeric region in at least half of the patients are indicative of telomere disruption. Structures similar to anaphase bridges have been demonstrated in bone marrow smears from patients bearing the aberration (Kuchinskaya, et al 2007, Robinson, et al 2007).

**Other recurrent structural abnormalities**

Other recurrent structural abnormalities in BCP-ALL include for example dic(9;20)(p13;p11), dic(9;12)(p13;p13), dic(7;9)(p11;p11), t(12;17)(p13;q12), and t(8;14)(q11.2q32) (Pan, et al 2006, Reid, et al 2006, Strehl, et al 2003). These aberrations occur at a frequency of <2% in ALL, and their prognostic importance is not yet established.
T-ALL

Approximately 12% of childhood ALL can be classified as T-lineage (Gadner, et al 2006). T-cell ALL is more common in males than in females and affects older children and young adults. The disease is characterized by a high white blood cell count, lymphoadenopathy, splenomagaly, mediastinal mass and CNS involvement, and is often characterized by an aggressive course (Uckun, et al 1998). Despite intensive high-risk adapted therapy, treatment failure occurs in about 25% of the patients, and outcome remains dismal after relapse (Goldberg, et al 2003). The cytogenetics and molecular genetics of T-ALL is poorly understood, and clinical features are unreliable predictors of outcome compared to BCP-ALL. Genetic analysis of T-ALL has led to the identification of a large amount of genetic abnormalities, including chromosomal translocations, deletions, amplifications and mutations. Recurrent chromosomal abnormalities are detectable in about 50% of T-ALL using conventional cytogenetics, and can be divided into two main subgroups: deletions and translocations. These abnormalities result in aberrant expression of general transcription factors, like C-MYB, MYC, TAL1, TAL2, LMO1, LMO2, BCL11B, HOX11, HOX11L2, HOXA, and ABL1 (Ballerini, et al 2002, Clappier, et al 2007, Graux, et al 2004). Other translocations lead to formation of specific fusion proteins, for example CALM/AF10, and MLL rearrangements (Asnafi, et al 2003). Activating mutations of NOTCH1 genes have been identified in about 50% of T-ALLs (Weng, et al 2004). A favorable prognostic impact has been described for this aberration (Breit, et al 2006). Activating NOTCH1 mutations predict favorable early treatment response and long-term outcome in childhood precursor T-cell lymphoblastic leukemia. However, the prognostic value of the remaining aberrations has not yet been well defined and is not used for treatment stratifications in any clinical protocols.

One common finding in T-ALL is deletion of the 9p21.3 region, harboring the tumor suppressor genes CDKN2A, ARF and CDKN2B. The deletion leads to loss of G1 control of the cell cycle (CNKN2A and CDKN2B) and possible inactivation of the p53 pathway (ARF). Cryptic deletions are detected in up to 65% of children and adults with
T-ALL, and a homozygous deletion has been established as an adverse prognostic marker with high risk of early relapse (Bertin, et al 2003, Calero Moreno, et al 2002, Ramakers-van Woerden, et al 2001). However, such deletions are not exclusively found in T-ALL. They are also found at a lower frequency in BCP-ALL cases. Although 9p21 deletions in BCP-ALL have been associated with an adverse prognosis, it was recently demonstrated that 9p21.3 deletion is not an independent adverse prognosis risk factor in BCP-ALLs (Mirebeau, et al 2006, van Zutven, et al 2005).

**Normal karyotype in ALL**

Conventional cytogenetic analysis of G-banded metaphase chromosomes is generally more demanding to perform in pediatric ALL than in other types of leukemia. Metaphases are often of poor quality and preferential growth of non-malignant cells may lead to false-negative normal karyotypes (Nordgren 2003). Genetic changes may also remain undetected due to the fact that they are below the resolution of chromosome banding. Leukemic patients with a normal karyotype may therefore harbor various genetic changes and consequently, the prognostic and/or diagnostic significance of many of chromosomal aberrations have not been reliably evaluated, leading to an under- or overestimation of the relapse risk and thus wrong treatment strategies (Larramendy, et al 1998a). The introduction of FISH and PCR into routine clinical practice for the most important prognostic aberrations significantly decreased the proportion of patients with normal karyotype. Recent analyses of ALL samples with high-resolution array-CGH further diminished this figure. The accumulating knowledge about different unbalanced aberrations diagnosed with high-resolution array-CGH leads to re-evaluation of mechanisms underlying leukemogenesis and suggestion of potential targets of treatment and follow-up.

**Breakage-Fusion-Bridge Cycles**

Amplification of oncogenes or drug resistance genes plays an important role during the malignant transformation of human cells. One form of cancer instability is intrachromosomal amplification of large regions, containing oncogenes (Brison 1993).
There are two cytogenetic manifestations of gene amplification: extrachromosomal double minutes and homogeneously staining regions (HSRs) (Schwab 1998). Many hypotheses have been proposed to explain how double minutes and HSRs are generated. The “episome model” suggests that circular molecules excised from chromosomes replicate autonomously because they contain the origin of replication. Another hypothesis is the breakage-bridge-cycles (BFB) model, proposed more than 50 years ago by McClintok (Figure 3).

The sequence of the events in BFB cycles was suggested as follows. The initiating event is loss of a telomere. The initial break sets the telomeric boundary of the amplified unit and occurs spontaneously or as a part of general chromosomal instability that characterizes the early stages of many cancers. During replication the uncapped sister chromatids fuse. At the same time cells undergo additional genetic changes, which induce fragile site expression. The chromatin organization of fragile sites predisposes them to break during anaphase. The resulting dicentric chromosome forms a bridge during anaphase and then breaks when the two centromeres are pulled in opposite directions. Such anaphase bridges serve as a hallmark of BFB cycles and can be detected by Giemsa staining and FISH. The BFB cycle is continued during the next cell cycle.

Figure 3. The mechanism of gene amplification involving BFB cycles (Adopted from (Lo, et al 2002)).
when sister chromatids fuse following DNA replication. Breakage at locations other than the site of fusion results in amplification of sequences in one daughter cell and deletions in the other daughter cell. Clustering of the recurrent breaks within chromosome fragile sites was demonstrated when studying the model of BFB cycles (Hellman, et al 2002). The distance between the amplified arrays is dependent upon the distance of the break from the site of fusion. The addition of telomeres to the end of broken chromosomes can promote chromosome stability and terminate the BFB cycles. Broken chromosomes can acquire telomeres through a number of mechanisms: e. g. added directly to the end of chromosomes by telomerase or acquired through translocation of the ends of other chromosomes (Diede and Gottschling 1999, Fouladi, et al 2000). Evidence that BFB cycles play a role in low-copy gene amplification is provided by the observation of structures consistent with BFB cycles in some tumors (Rudolph, et al 2001). The presence of inverted repeats, large duplications, and prolonged periods of chromosome instability is consistent with BFB.

**MODERN DIAGNOSTIC APPROACHES IN ALL**

Techniques applied for the diagnosis of ALL have quickly developed since 1970s, and nowadays represent a set of tests, including cytomorphological examination, immunophenotyping, conventional and molecular cytogenetics, as well as molecular genetics methods. These techniques has changed our understanding of ALL as a single nosologic entity to a heterogeneous complex of cytogenetic and molecular genetic subgroups leading to ALL as a result of different mechanisms of leukemogenesis (Gilliland 2001, Haferlach, et al 2007). During the last years, additional tests to study genetic changes of importance for leukemia have been introduced, including array platforms to study small gene dose changes and large scale expression variation.
Gene expression studies of ALL

Platforms to study whole genome expression patterns were introduced in 1995 and raised expectations in providing insight into the pathophysiology of acute leukemias (Schena, et al 1995). It was hypothesized that differences in the prognosis and different pathways involved in leukemogenesis would find their reflection in certain changes of the gene expression on the whole genome level. Disruption of the pathways would lead to an increase or decrease in the expression of sets of genes, and diseases with different underlying mechanisms would differ in their gene expression pattern. It was suggested that whole genome expression profiling is capable of two major functions: class discovery, meaning classifying of previously unrecognized tumor subtypes; and class prediction, which refers to the assignment of tumor samples to previously recognized classes.

The first gene expression study of leukemic samples was conducted on samples from patients with acute leukemia, both ALL and AML (Golub, et al 1999). It showed that expression analysis was indeed a reliable tool to assign samples into known classes as well as predict the class of unknown samples. Around 1100 genes were found to differ in their expression pattern in ALL and AML, and 50 selected genes were correctly predicting the assignment of the test case (AML or ALL). Later it was demonstrated that acute lymphoblastic leukemia with MLL translocation has a unique gene expression profile, distinct from that of AML, and ALL without MLL rearrangement (Armstrong, et al 2002, Rozovskaia, et al 2003) and information was provided about genes in which the expression pattern was characteristic to this subgroup (e.g. FLT3, HOX9, MEIS1). Yeoh and colleagues were the first to demonstrate that children with newly diagnosed ALL could be classified into six subgroups with a predictive accuracy of 96%: T-ALL, ETV6/RUNX1 positive, MLL rearrangement, PBX1/TCF3 positive, high hyperdiploidy and BCR/ABL1 positive (Yeoh, et al 2002). Further studies, conducted on childhood and adult ALLs, always showed a main role of lineage or primary genetic lesion in establishing the expression pattern, and demonstrated the ability of expression arrays to identify new potential subclasses of ALL samples, e.g. T-ALLs with high MRD level on

The value of expression arrays is not only the assignment of samples, but it also allows an insight into gene regulation and identification of groups of genes with aberrant expression, novel clusters segregating with specific mutations in acute leukemias, such as activating mutations of RAS and FLT3, implying that these mutations underlie fundamental aspects of the biology of leukemogenesis (Scandura 2005). Such an approach may lead to the identification of new targets to gene therapy, e.g. the successful use of FLT3 inhibitors.

Array-CGH

Several hundred chromosomal abnormalities have been identified in ALL, but most of them are very rare and of unclear prognostic significance (Mitelman, et al 2007, Scandura 2005). Numerical chromosomal aberrations affect gene dosage by which oncogenes may be activated. However, the resolution of G-banding in the majority of ALL samples seldom exceeds 20 Mb. In order to increase the resolution and overcome the problem with poor quality metaphases, comparative genomic hybridization (CGH) was introduced and revealed the presence of chromosome abnormalities in up to 90% of ALL patients (Kallioniemi, et al 1992, Karhu, et al 1997, Larramendy, et al 1998a). The technique requires genomic DNA, which is comparatively hybridized onto metaphase chromosomes with DNA isolated from a normal tissue. This approach detects gene dose changes, overcoming the limitation of poor-quality and normal metaphases. However, the resolution of this approach is 5-20 Mb, the quality of hybridization is often bad and the proportion of false-positive results is high (Larramendy, et al 1998b).

To overcome these limitations, array-CGH was introduced (Pinkel, et al 1998, Solinas-Toldo, et al 1997) and is now widely used in both research and in the clinical setting as an additional tool to characterize tumors and constitutional abnormalities (Bejjani, et al 2005, Pinkel and Albertson 2005). Instead of using metaphases, the test DNA is hybridized to arrays of BAC or oligonucleotide clones, and the resolution of the method depends on the density of the clones on the array. Array-CGH has allowed
more deep insight into the biology of ALL. The results have been in agreement with standard cytogenetic analyses, but have added important information concerning small gains and losses in almost every investigated sample (Kuiper, et al 2007, Mullighan, et al 2007, Paulsson, et al 2006, Strefford, et al 2006, Strefford, et al 2007, Tsuzuki, et al 2007). Findings of special interest are deletions of PAX5 in almost 50% of BCP-ALL, a higher incidence and various sizes of ETV6 deletion detected in samples with ETV6/RUNX1 rearrangement, and deletion of E2F, Ikaros and some other genes, leading to disruption of cell-cycle control (Kuiper, et al 2007, Mullighan, et al 2007). Many regions of interest are very small and the deletions may contain only a single clone.
AIMS OF THE STUDY

The principal goal of this study was to characterize children and adults with ALL with modern molecular genetic methods in order to provide better insight into genetic events underlying leukemogenesis. We aimed to:

1. Study whole genome expression profiles in children and adults with ALL in order to determine factors, guiding expression signatures and to explain differences in outcome of ALL in children and adults.

2. Find genes with expression pattern characteristic to different cytogenetic subgroups.

3. Characterize gene dose changes in children with ALL in order to establish their frequency, and to understand their significance and mechanism of formation.

4. Characterize the rearrangement dic(9;20) in children with ALL in order to increase the knowledge of its frequency, significance and genes of importance.
MATERIALS AND METHODS

Patients
Children and adults with ALL, treated at Karolinska University Hospital, Lund University Hospital, Umeå University Hospital, and St.-Petersburg Pavlov State Medical University were included into the study. The diagnosis of ALL was based on modern diagnostic criteria, including clinical examination, morphological examination of bone marrow, immunophenotyping, karyotyping and molecular cytogenetic investigation for a number of prognostically important aberrations (MLL rearrangement, BCR/ABL1, ETV6/RUNX1, PBX1/TCF3 rearrangements) where possible. All studies were approved by the ethics committee of Karolinska Institutet.

Material
Mononuclear cells from the bone marrow obtained at the time of diagnosis, relapse and in remission were purified by Ficoll-Paque TM (Amersham Biosciences, NJ, USA) density centrifugation prior to cryopreservation. Samples, obtained at the St.-Petersburg Pavlov State Medical University, were stored in preserving solution RNA later (Ambion, Austin, TX, USA), to ensure better RNA quality and to prevent possible RNA degradation during transport. Cryopreserved samples were later used for RNA and DNA isolation in expression and array-CGH experiments.

The most important requirement for gene expression studies is high quality of RNA and use of the same protocol for RNA extraction and purification of the samples, as these methods may have a significant influence on the gene expression profile (Staal, et al 2006). RNA was extracted from diagnostic mononuclear cell suspensions using the TRIzol reagent (Gibco-BRL Life Technologies, Gaithersburg, MD, USA) according to the manufacturer’s protocol and purified with the RNeasy purification kit (Qiagen GmbH, Germany) prior to preparation for hybridization. Differences in gene expression induced by different storage of ALL samples (liquid nitrogen and RNA later) were addressed in the statistical analysis of gene expression results.
Array-CGH requires high quality DNA. In our study DNA was isolated from fresh frozen bone marrow cells from diagnosis, relapse and remission. The DNA from the remission samples was used as a reference in order to avoid false positive results due to polymorphic changes in the genome (Iafrate, et al 2004). In cases where DNA from remission was not available, standard sex matched DNA was used (Promega, Madison, WI, USA). Bone marrow cells from diagnostic and relapse samples were used for cytogenetic analyses. Stored chromosome suspensions were used to confirm array-CGH findings by FISH.

**Methods**

Routine cytogenetic analysis by G-banding was performed as a part of routine investigation on all patients included into the study.

**FISH**

After the first application of FISH in 1986 (Pinkel, et al 1986), this technology has been widely applied in diagnostic and research settings. This rapid and simple technique allows detection of specific numerical aberrations, gene rearrangements, deletions and amplifications, not only on metaphase chromosomes, but also on interphase nuclei, both on smears and paraffin-embedded tissues. The basic principle of FISH is the hybridization of a fluorochrome-labelled DNA probe with a complementary target DNA sequence in the cells, tissue sections or metaphase spreads. A fluorescent counterstain (DAPI) is then added and the slide is viewed in fluorescent microscope with an appropriate set of filters. Application of interphase FISH to detect aberrations of prognostic importance in ALL has allowed identification of abnormalities also in cases with a normal karyotype by G-banding and with metaphases of bad quality, not feasible to analyse by G-banding.

**SKY**

Spectral imaging was first introduced by Schröck in 1996 and allows the simultaneous identification of all human chromosomes in different colours (Schrock, et al 1996). Spectral imaging utilizes a combination of epifluorescence microscopy, CCD
imaging, and Fourier spectroscopy to measure the entire fluorescence spectrum at all sample points of an image (Ried, et al 1997). Spectral karyotyping is useful to detect numerical and structural chromosomal abnormalities, identification of marker chromosomes and some cryptic rearrangements, as well as identification of the chromosomal origin of amplified sequences. Spectral karyotyping has the same limitations as G-banding: a low resolution of 10-20 Mb on tumour samples, inability to identify aberrations in the presence of only normal metaphases and low quality metaphase morphology (Schrock, et al 1996).

**RT-PCR**

RT-PCR is used in some laboratories to detect prognostically important aberrations in ALL. The method requires total or messenger RNA (mRNA) from leukemic blasts which is reverse transcribed in the presence of random primers to obtain cDNA which is then used in PCR with gene-specific primers to detect fusion genes.

**Array-CGH**

Array-CGH is a development of the metaphase-CGH technique (Pinkel, et al 1998, Solinas-Toldo, et al 1997) and overcomes limitations in resolution. The resolution of array platforms depends on the size of the probe sequences, the distance between them and their distribution throughout the genome. Today, two types of array platforms are available: BAC arrays and oligonucleotide arrays. DNA from a test tissue and from a reference (normal) tissue is differentially labelled with different fluorescent dyes in a random priming reaction, cleaned, denatured and co-hybridized on a glass slide, containing immobilized BAC clones or oligonucleotide sequences (Figure 4). After hybridization, the slides are washed, dried and scanned. The scanning procedure allows quantification of the signal intensity from different channels. The ratio of the signal intensities is log transformed and the result reflects the amount of DNA in the sample. Array CGH allows identification of unbalanced changes in the genome, i.e homozygous or hemizygous deletions. No balanced changes can be viewed with this technique.
Array-CGH has been applied in a wide range of studies. This has led to the identification of large-scale copy-number variations (CNV) in the genome that involve gains or losses of several kilobases to hundreds of kilobases of genomic DNA in phenotypically normal individuals (Iafrate, et al 2004), which have to be separated from disease causing mutations. Information, regarding CNVs is available in the Database of Genomic Variants (http://projects.tcag.ca/variation/) and presently includes 6482 CNVs in 3643 loci.

The microarrays with complete genome coverage, used in this thesis were produced from a 33K clone library at the Swegene DNA Microarray Resource Center, Department of Oncology, Lund University. This platform has the advantage that the BAC clones are uniformly distributed throughout the genome, achieving a resolution of up to 100 Kb and covers both genes and repetitive sequences. Analysis of the results has been performed in the Bio Array Software Environment (BASE) (Saal, et al 2002). Normalization of results was performed using the LOWESS algorithm (Yang, et al 2002), and smoothing was based on the signal intensity ratio of a few neighbour clones due to their partial overlap of the clones, and thus allowing exclusion of chance variations due to technical reasons. One of the first studies conducted on a similar platform detected more than 250 copy number variations (de Vries, et al 2005). In order to avoid false-positive findings due to CNVs, we used patient DNA isolated from bone marrow in remission as a reference in our array-CGH experiments when possible.

**Large scale gene expression analysis**

Two types of platforms to study gene expression are available: those containing arrayed cDNA clones of approximately 50-200 kb length (Schena, et al 1995) and oligonucleotide arrays where the arrayed probes consist of oligonucleotide fragments (approximately 20-80 mers) (Lockhart, et al 1996). Microarrays vary in the type of probes used, the manner in which the probes are arrayed onto a solid support, and the method of target preparation (Figure 4). These differences may have an impact on the interpretation of the results, and thus it is important to provide similar conditions for all experiments throughout one study.
In our study, oligonucleotide Hu95ver2a gene chips produced by Affymetrix were used to study gene expression. The platform includes around 12,500 genes and ESTs and represents a set of sixteen probe-pairs per gene, 25-mer long with representative sequences from the 3’ end of the gene. Each of the oligonucleotide probes has an almost identical duplicate, where only the central nucleotide in the duplicate (mismatch sequence) differs from the original sequence (perfect match sequence). After hybridization and scanning, the signal intensities from all the oligonucleotide sequences representing one probe are analysed and compared. Comparison of signal intensities from perfect match and mismatch probes estimates a sequence-dependent cross-hybridization and results in a better signal estimate. If the intensity of mismatch probe exceeds the perfect match, the expression level is estimated as absent.

Normalization of samples is required to remove array effects resulting from differences in overall intensities in different hybridizations. All normalisation methods are based on an assumption of the distribution of intensity values in different arrays. Normalization includes background subtraction and centering results across the experiments (Bolstad, et al 2003). The key objective in the gene expression experiments
is identification of genes that are differentially expressed between groups of biological samples (Claverie 1999). In expression array experiments, when thousands of genes are evaluated, the p-value threshold must be corrected for multiple comparisons, i.e. Bonferroni correction.

Clustering is widely used to explore the gene expression data and a number of algorithms exist. In general, cluster analyses may be divided into two groups: supervised and unsupervised analysis (Eisen, et al. 1998). In unsupervised hierarchical cluster analysis samples are assigned to groups based on similarities in gene expression pattern. In this analysis, samples with the same gene expression are placed closest to each other. This type of analysis is more useful to explore similarities between samples and to identify sub-classes. In supervised analysis, samples are assigned into predefined groups, based on known parameters (e.g. age, sex, cytogenetic subgroup, resistance to chemotherapy). Further analysis detects genes for which the expression pattern discriminates the groups under investigation. The reproducibility of the results may be assessed by different techniques, including bootstrapping and jack-knife tests.
RESULTS

Gene expression analysis in children and adults with ALL
(Paper I)

Whole genome gene expression analysis was performed on 44 patients with ALL, 29 children and 14 adults. The primary unsupervised hierarchical cluster analysis assigned samples with T-ALL and BCP-ALL into separate clusters. BCP-ALL samples with prognostically important cytogenetic abnormalities were assigned into separate subgroups (Figure 5). Probe sets discriminating patients with T-cell ALL from BCP-ALL and cytogenetic subgroups within BCP-ALL were identified. Using a predictor analysis we identified genes with expression patterns characteristic for known cytogenetic subclasses. We also clarified the karyotypes in three cases, based on cluster and gene expression results. Two patients were shown to have ETV6/RUNX1 rearrangement and one was diagnosed with a PBX1/TCF3 rearrangement. The molecular weight of the BCR/ABL1 fusion proteins was shown to influence the gene expression signatures and result in separate clustering of patients bearing BCR/ABL1 translocation with different breakpoints in BCR. We did not observe any difference between the expression profiles in children and adults, and the observed low sensitivity on post hoc power analysis of this comparison is mostly due to the small observed difference between the two groups.

Delineation of intrachromosomal amplification of 21q in patients with ALL (Papers II, IV)

We applied a tiling-resolution 33 K BAC array to investigate the size and pattern of amplification as well as additional genomic imbalances in eight children with ALL and intrachromosomal amplification of 21q. In one patient we also performed analysis on samples from two consecutive relapses. We observed that amplification has a characteristic step-wise rise of DNA gain toward the telomere, with the lowest level of amplification toward the centromere (Figure 6). We identified a common region of amplification (CRA) of 14.4 Mb, the whole size of the 21q amplicon was 15.5–30.2 Mb,
Figure 5. Unsupervised clustering of ALL samples correlated to leading cytogenetic aberration, immunophenotype and age.

Figure 6. A. (a) Metaphase and (b, c) interphase FISH of chromosome 21 with clones located in different regions showed unequal amplification of different parts of the chromosome and deletion of the telomeric region. Clones are labeled: RP11-539O13 (28.0-28.1 Mb) with Spectrum Orange (red), RP11-114H1 (41.1-41.3 Mb) with FITC (green), RP11-53E17 (44.9-45.0 Mb) with Cy5.5 (pink).

Figure 7. A typical array-CGH plot of the whole genome of the patient with dic(9;20). The data are presented in log2 ratio format. Red arrows point on deletions of the short arm of chromosome 9 and the long arm of chromosome 20.
and it was located between genomic positions 14.5 and 46.5 Mb. In three out of seven patients, array-CGH allowed identification of a 0.9–4.2 Mb deletion of the most telomeric part of 21q, distal to the amplified region. In two cases we observed a loss of 9p and a gain of 9q with formation of an isochromosome 9q. In three diagnostic cases we observed a loss of 7q of variable size. In diagnostic and relapse samples from the same patient, the only aberration retained unchanged was the intrachromosomal amplification of 21q.

In two samples we observed structures looking like anaphase bridges that suggested BFB cycles as a possible mechanism of formation of this amplification.

**Characterization of dic(9;20)(p11-13;q11)(Paper III)**

In the array-CGH and FISH analyses of seven pediatric dic(9;20) positive BCP ALLs we observed a high frequency of cryptic deletions, consistent losses of 9p and 20q material, and clustering of breakpoints, both at 9p and at 20q (Figure 7). Although the breakpoints on chromosomes 9 and 20 clustered, they were not identical, being distributed in a 1.5 Mb segment in 9p13.2 and a 350 kb segment in 20q11.2. One of the ALLs, shown to have a complex dic(9;20), was further investigated by FISH. A rearrangement of the haemapoietic cell kinase isoform p61 (HCK) gene at 20q11 was revealed. The disruption of HCK may result in a fusion gene or in loss of function. Unfortunately, lack of material precluded further analyses of HCK.

**Characterization of patients with normal karyotype by G-banding (Paper IV)**

We characterized unbalanced changes in 28 patients with normal (n=23) or failed (n=5) karyotype by G-banding with a tiling-path 33 K BAC array. In 22 patients (79%), altogether 135 copy number alterations, including 69 gains and 66 losses, were detected. Most of the patients had CNAs below the resolution of G-banding. No specific pattern of gains or losses was identified in the whole cohort of patients. According to the molecular cytogenetic and array-CGH results the patients were divided into five groups: high hyperdiploidy (n=4), intrachromosomal amplification of 21q (n=5), ETV6/RUNX1 translocation (n=7), others (n=3), and no CNA (n=5). Samples with T-ALL
were analyzed separately. When aCGH was combined with interphase FISH for the most common balanced rearrangements, aberrations were found in 90% of the patients.
**DISCUSSION**

The results of this thesis stress the importance of a combination of well-established chromosome based methods with high-resolution molecular approaches, in order to shed new light onto genetic changes causing malignant transformation of hematopoietic cells. In spite of the improved outcome in patients suffering from ALL, the disease is still severe and all patients are not cured. Many relapse, and extended knowledge regarding events involved in the development of ALL are important for correct diagnosis, individualized treatment strategies and the possible development of new drugs.

**Application of whole genome expression arrays**

The oligonucleotide chips used in this study allowed us to compare the gene expression profiles in children and adults with ALL bearing different cytogenetic abnormalities. The platform does not require reference material to perform the analysis, and thus gene signatures from only leukemic blasts were compared. We observed that T- and B-cell immunophenotypes and prognostically important cytogenetic abnormalities guided gene expression profiles. This finding stresses the importance of cytogenetic changes in ALL as markers of genomic alterations underlying malignant transformation.

Our study has shown that samples with the same cytogenetic aberration have the most similar gene expression profiles, forcing such samples to cluster together on unsupervised hierarchical cluster analysis, and the results of such clustering may predict the presence of certain abnormalities in previously unclassified samples. This was proved by discovering the *ETV6/RUNX1* rearrangement in two samples (where presence of this aberration was not investigated earlier), based on their clustering together with other samples with the translocation. Studying peculiarities of gene expression in individual samples may aid to identify previously undetected aberrations. The increased expression of the *PBX1* gene guided the identification of a *PBX1/TCF3* rearrangement in one patient, later verified by interphase FISH. It has already been shown that samples within cytogenetic subgroups may be heterogeneous (Tsutsumi, et
Gene expression profiling has proved to be a useful tool to divide such samples into different groups, which later may be correlated with clinical or prognostical features. In our study we showed that BCP-ALL samples with BCR/ABL1 translocation, resulting from different breakpoints in the BCR gene cluster separately. The phenomenon of heterogeneous clustering of Philadelphia positive samples, as well as cell lines bearing translocations with different breakpoints, has been observed by others, stressing the necessity of a larger study on a selected group of patients bearing the translocation (Andersson, et al 2005, Fine, et al 2004, Juric, et al 2007).

In our study we also addressed differences between childhood and adult ALL. In spite of a huge difference in outcome between these age groups, no differences in the gene expression pattern were observed. Moreover, adults and children with similar cytogenetic abnormalities clustered together, but the size of the study was too small to perform age comparisons within cytogenetic subgroups. This observation supports recently published results where difference in outcome of children and young adults was shown to be protocol-induced (Hallbook, et al 2006, Jeha 2003, Sallan 2006), and showed better prognosis for young adults treated with pediatric protocols compared to those treated with adult protocols. Thus there may be no disease-specific age-dependent diversity in gene expression, and the difference in outcome between adults and children may at least partially be explained by the more prolonged and intensive protocols used in children, a higher rate of adverse prognostic abnormalities in adults and more strict compliance to the protocols in children (Boissel, et al 2003, Moricke, et al 2005, Nachman 2005, Ramanujachar, et al 2006).

**Array-CGH to study genomic imbalances in ALL**

**Intrachromosomal amplification of 21q**

We were able to identify the structure and size of the amplicon in patients with intrachromosomal amplification of 21q. This abnormality has earlier been associated with an adverse prognosis (Robinson, et al 2003). The diagnosis in our material was based on the identification of clusters of signals from RUNX1 on interphase FISH, when leukemic cells were analyzed for the presence of the ETV6/RUNX1 rearrangement.
Array CGH allowed characterization of the amplified region in eight patients where this abnormality was suspected, and we were also able to characterize additional genomic imbalances in the genome and follow them in a patient where material from two relapses was available. This approach allowed us to prove that amplification of 21q was the only aberration which remained unchanged in three consecutive samples. This observation suggests that intrachromosomal amplification of 21q is an important event, required for the development of ALL, probably occurring in an early hematopoietic precursor.

**Characterization of dic(9;20)**

The DNA content from eight patients with dic(9;20) was successfully analyzed with array-CGH. This subtle aberration is often missed on standard karyotyping and is often misclassified as monosomy 20. This recurrent rearrangement, with a favorable prognosis, occurs in around 2% of pediatric ALL and is extremely rare in adults (Song, *et al* 2007). The detailed characterization of the breakpoints was important to investigate the possible formation of a fusion gene. However, although we showed clustering of breakpoints on both chromosomes 9 and 20, no common breakpoint region was found. All breakpoints on 9p clustered in a 1.5 Mb segment of the sub-band 9p13.2 and in three of the cases, the 20q breakpoints mapped to three adjacent clones covering a distance of 350 kb at 20q11.2. These results have been confirmed by the observation of different breakpoints in two patients with dic(9;20) in another study, utilizing BAC array with a resolution of 1 Mb (Strefford, *et al* 2007). Thus, it remains to be elucidated whether dic(9;20)(p13.2;q11.2) leads to a chimeric gene or whether the functionally important outcome is loss of 9p and 20q material.

**Characterization of patients with a normal karyotype**

Array-CGH analysis of patients with normal or failed karyotype by G-banding revealed that 79% of the samples had gains and losses. When these results were combined with those obtained by FISH with probes for prognostically important rearrangements in ALL (*ETV6/RUNX1, BCR/ABL1, PBX1/ETV6, MLL* gene rearrangements) aberrations were found in 90% of the patients. No specific pattern of
gains and losses was obtained and cases were classified into groups based on a combination of array-CGH and FISH results. The number of gains was almost equal to the number of losses. Although these results are in contradiction with the prevalence of losses observed in other studies (Kuiper, et al 2007, Mullighan, et al 2007), the phenomenon may be explained by the fact that most of the gains were diagnosed in cases with high hyperdiploidy and intrachromosomal amplification of 21q. In conclusion, array-CGH is a valuable tool to refine karyotypes in patients with ALL, especially those with normal or failed karyotypes.
**FUTURE DEVELOPMENT**

The results of this thesis emphasize the importance of detailed analyses of genomic changes in ALL. The results of gene expression and array-CGH studies stress the importance of cytogenetic changes in ALL as key events in leukemogenesis. Array-CGH is a valuable research tool to identify genomic imbalances, which are hallmarks for the development of ALL and to aid in the classification and reclassification of cases into prognostically relevant subgroups. This technique is also an important diagnostic tool, especially in those patients where no information about the karyotype can be obtained through standard diagnostic approach using G-banding, SKY and FISH. Further employment of expression arrays based on a standardized approach in all stages of experiments will discover new molecular subgroups in ALL and determine molecular targets of new treatment modalities. Combination of whole gene expression analysis and results from tiling-resolution genomic arrays will lead to more precise definition of genes involved in pathophysiology of leukemia. Future work should be directed towards joint international efforts using a combination of modern diagnostic and research methods together to correlate all changes, occurring in the genome along with the development of ALL in a comprehensive picture. Assembling the information about normal variations in the genome may reveal genetic factors predisposing to the development of ALL.
ACKNOWLEDGEMENTS

I want to express my sincere gratitude to all people who surrounded me during my postgraduate studies for their great support. I would like especially thank:

Elisabeth Blennow, my main supervisor for accepting me as a PhD student and for being so immensely positive and patient. You introduced me to the field of molecular genetics and helped to understand how research works and how researchers think. I thank you for always encouraging me with my ideas, finding nice solutions in difficult situations, especially in my “last minute” problems. I have learnt a lot from you and with you. I just regret that I did not talk more often with you.

Magnus Nordenskjöld, my main “co-supervisor” for your endless support and fast decisions. Thank you for your generosity and creating stimulating and interesting work environment. Your short notes about historical events stimulated me to learn more about Russian and Swedish history in connection to each other.

Mats Heyman, my co-supervisor for your deep knowledge in hematology and interesting ideas. Working together with you was always exciting and learning to me. Thank you for correcting my English in papers and this thesis.

Andrey Yurievich Zaritskey, my supervisor in St-Petersburg for accepting me, a newborn physician at the department of hematology and introducing me to the wonderful world of hematology, medicine and science. Thank you for accepting me as a PhD student, for interesting discussions about science and all possible things (“a space rocket is still rotating and rotating”), and always supporting me. You are my first teacher in medicine!

Stefan Söderhäll for you fruitful discussions on our group meetings and always taking a sober view of things.

Dan Grander for your endless interest in our team-work, for your interesting ideas and deep insight into genes function.

Sigrid Sahlén, for introducing me to the lab, for teaching me all the things concerning work in the lab, always patiently answering my endless questions and helping with practical things. My lab experience will never be so smooth without you! And thanks to all technicians on the 2nd floor of CMM for always helping with practical issues, Anki Thelander, Anna-Lena Kastman, Christina Nyström.

Günther Weber for always helping with any scientific and practical question.

My former and present colleagues and friends of the second floor of CMM: Ann Nordgren for your never endless enthusiasm, Charlotta Lindvall for introducing me to
expression array field, **Erik Björk** for sharing Affymetrix and Gene Spring experience, as well as nice figures, **Jacqueline Schoumans** for teaching me to work with genomic arrays and fruitful discussions about our experiments, **Helena Malmgren, Britt-Marie Anderlid, Peter Gustavsson, Kim Ericson, Maria Bradley, Eva Rudd, Elisabeth Ekelund, Mai-Britt Giacobini, Zheng Chengyun, Anica Sääf, Anna Lindstrand** especially for lessons in Swedish, **Cilla Söderhäll, Johanna Skoglund, Tatiana Djureinovic, Mickela Barabro, Emma Tham, Agne Liedén, Johanna Rantala, Josephine Vincent, Josefine Edner, Clara Chamorro.** You all shared your knowledge and interests with me and I felt not lonely here.

**Ynonne Cowan, Gunilla Risberg, Majalena Granqvist** for administrative support and **Lennart Helledey and Rudolph Matousek** and **Dagmar Vesica** for help with computer issues.

Staff in the clinical lab for always helping me with finding samples and with my FISH and SKY experiments, especially **Irene White, Coralia Burgos, Margareta Lagerberg, Kerstin Holmberg, Karin Kindberg.**

My collaborators at CCK and Lund University, **Michael Corcoran, Katya Pokrovskaja, Alena Malukova, Johan Staaf and Johan Valon-Cristesen.**

**Tommy Linné** for your endless enthusiasm in Joint Research Training Program and for giving me an opportunity to participate in it. Rector of SPMU, **Nikolai Antonovich Yaitskii, professor Salman Hasunovich Al-Shukri, and the head of the faculty therapy department Eugeni Vladimirovich Shlyakhto** for supporting my participation in the program.

All present and former members for JRTP I and II: **Sasha and Sasha, Anna, Olga, Zainab, Alexei, Tatjana, Ksenia, Kristina** for our long evening teas and mornings coffees, for interesting talks and simply good time together. It was so good to meet you and make friends with you! Good luck with your scientific achievements!

**Irina** and **Olga** who opened me a world of sushi and for all nice times together in Stockholm and in St.-Petersburg.

**Anja,** my dear friend. It was so much shared together!! Thank you for sharing one room, one lunch, one bottle of vine on special occasions, as well as long walks, good talks and for moral support.

My friends **Alena** and **Tonja** for being together for almost 25 years, for short, but productive meetings and time together on vacations. I’m looking forward to travel together with you once again. MFC – forever, do you remember that? **Tonja,** we should meet all of us one day somewhere on the Earth. **Tanja and Roman** for being to
wonderful friends, Gelja and Alexei, and my goddaughter Katya for nice time we had together.

Here for help and support, and for teaching me to be strong and independent.

My parents, for your endless love and support and always being on my side. Thank you for encouraging me to start to work on these thesis, for always helping me and taking care of me even when I did not thought of doing so myself. Thank you for giving me a possibility to feel as a small daughter although being a mother already. Special thank to my sister Tanja, it so good to have a sister!! Good luck to you with all your projects. I love all of you!!! To my godmother Lidija Fedorovna and to djadja Gena for simply being a part of my family.

My beloved daughter, Дашеньке, моему маленькому солнышку. I live for you, I’m so happy to be your mother and I love you so much! My small wonder, you are the best result of these years!!!

This work was supported by the Swedish Institute, the Swedish Cancer Society, the Swedish Children Cancer Foundation, the Swedish Medical Research Council, the Mary Béve Foundation for paediatric cancer research, the Netpharma research fund.
REFERENCES


Maloney, K., McGavran, L., Murphy, J., Odom, L., Stork, L., Wei, Q. & Hunger, S. (1999) TEL-AML1 fusion identifies a subset of children with standard risk acute lymphoblastic leukemia who have an excellent prognosis when treated with therapy that includes a single delayed intensification. *Leukemia*, 13, 1708-1712.


