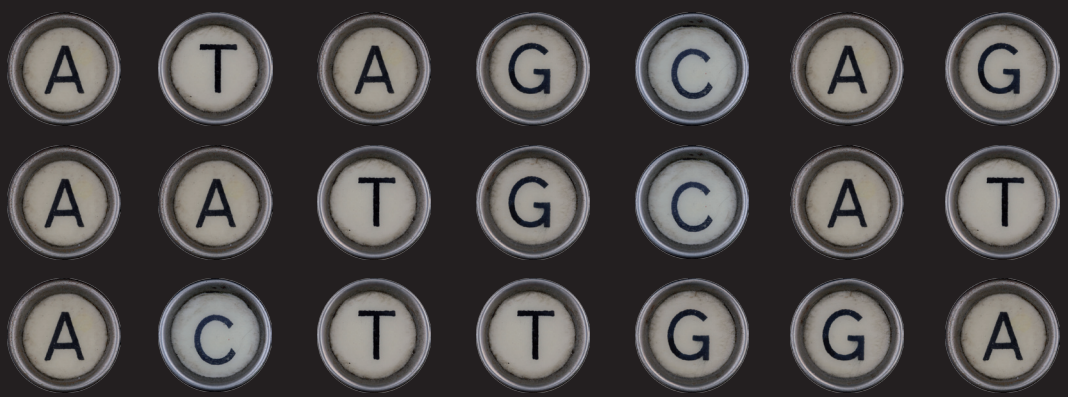
Genomic characterization of ETV6/RUNX1-positive acute lymphoblastic leukemia

Lilljebjörn, Henrik

2011

Genomic Characterization of *ETV6/RUNX1*-Positive Acute Lymphoblastic Leukemia



Henrik Lilljebjörn

Department of Clinical Genetics
Faculty of Medicine
Lund University, 2011

Genomic Characterization of *ETV6/RUNX1*-Positive Acute Lymphoblastic Leukemia

Henrik Lilljebjörn

Akademisk avhandling som med vederbörligt tillstånd av Medicinska Fakulteten vid Lunds universitet för avläggande av doktorsexamen i medicinsk vetenskap kommer att offentligen försvaras i föreläsningssal F3, Universitetssjukhuset i Lund, fredagen den 11 mars, kl 09.00.

Fakultetsopponent:
Professor Kjeld Schmiegelow
Department of Paediatrics,
The University Hospital Rigshospitalet,
Copenhagen, Denmark

Organization
LUND UNIVERSITY

Department of Clinical Genetics
University Hospital
SE-22185 Lund
Sweden

Document name
DOCTORAL DISSERTATION

Date of issue
March 11, 2011

Sponsoring organization
The Swedish Cancer Society, the Swedish Childhood Cancer Foundation, the Swedish Research Council, the Inga-Britt and Arne Lundberg Foundation, the Gunnar Nilsson Cancer Foundation, and the Medical Faculty of Lund University

Author(s) Henrik Lilljebjörn

Title and subtitle Genomic characterization of ETV6/RUNX1-positive acute lymphoblastic leukemia

Abstract

The t(12;21) translocation generates the ETV6/RUNX1 fusion gene, present in 25% of childhood acute lymphoblastic leukemia. This fusion gene is important for leukemia development but is not sufficient for leukemia to arise. Hence, the additional genetic changes present in leukemic ETV6/RUNX1-positive cells give important clues regarding the history of the leukemia. The aim of this thesis has been to characterize thoroughly the genetic changes present in ETV6/RUNX1-positive ALL. In Article I, seventeen ETV6/RUNX1-positive ALLs were characterized using array CGH. This revealed that gain of Xq material, present in six cases, was the most common copy number aberration (CNA). A large number of genes was present in the commonly gained region but the high and specific expression of SPANXB identified this gene as a likely target of the gain. In Article II, 24 ALLs were analyzed using 500K single nucleotide polymorphism arrays. The data from these ALLs were combined with previously published external SNP array data from 140 ETV6/RUNX1-positive ALLs. A high number of recurrent CNAs could be identified in this dataset. The recurrent CNAs were further analyzed using hierarchical clustering, connected pair analysis, and oncogenetic tree models. These analyzes revealed that the majority of cases had acquired a unique set of recurrent CNAs, indicating that the process of acquiring CNAs is unique for each ALL. In Article III, exome sequencing was used to sequence all protein coding genes in two ETV6/RUNX1-positive ALLs. Seven somatic single nucleotide mutations were present in each ALL, six of these were in genes previously implicated in cancer.

Key words: ETV6/RUNX1-positive ALL, childhood ALL, array CGH, SNP array, exome sequencing

Classification system and/or index termes (if any):

Supplementary bibliographical information:

Language

ISSN and key title:
1652-8220

ISBN
978-91-86671-67-9

Recipient's notes

Number of pages
107

Price

Security classification

Genomic Characterization of *ETV6/RUNX1*-Positive Acute Lymphoblastic Leukemia

Henrik Lilljebjörn

DEPARTMENT OF CLINICAL GENETICS
FACULTY OF MEDICINE

2011

Lund University

Printed in Sweden by Media Tryck, Lund 2011
ISSN 1652-8220
ISBN 978-91-86671-67-9
Lund University, Faculty of Medicine Doctoral Dissertation Series 2011:18

Table of Contents

Original Articles

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

I. **Lilljebjörn H, Heidenblad M, Nilsson B, Lassen C, Horvat A, Heldrup J, Behrendtz M, Johansson B, Andersson A, Fioretos T.** 2007. Combined high-resolution array-based comparative genomic hybridization and expression profiling of *ETV6/RUNX1*-positive acute lymphoblastic leukemias reveal a high incidence of cryptic Xq duplications and identify several putative target genes within the commonly gained region. *Leukemia* **21**:2137-2144.

II. **Lilljebjörn H, Soneson C, Andersson A, Heldrup J, Behrendtz M, Kawamata N, Ogawa S, Koeffler HP, Mitelman F, Johansson B, Fontes M, Fioretos T.** 2010. The correlation pattern of acquired copy number changes in 164 *ETV6/RUNX1*-positive childhood acute lymphoblastic leukemias. *Hum Mol Genet* **19**:3150-3158.

III. **Lilljebjörn H, Rissler M, Lassen C, Heldrup J, Behrendtz M, Mitelman F, Johansson B, Fioretos T.** Whole exome sequencing of pediatric acute lymphoblastic leukemia. *Submitted.*

Abbreviations

ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
array CGH	array comparative genomic hybridization
BCP ALL	B-cell precursor acute lymphoblastic leukemia
CBF	core-binding factor
CML	chronic myeloid leukemia
CNA	copy number aberration
COSMIC	Catalogue of Somatic Mutations in Cancer
CPA	connected pair analysis
FISH	fluorescence *in situ* hybridization
PTD	pointed domain
RD	repression domain
RHD	runt homology domain
RT-PCR	reverse transcriptase polymerase chain reaction
SNP	single nucleotide polymorphism
TA	transactivation domain
WHO	World Health Organization

PREFACE

Acquired genetic aberrations are the hallmark of neoplastic malignancies and are believed to have a direct causative role in the development of these diseases. The successful identification of the underlying genetic aberrations has, for many neoplasms, resulted in better classification into prognostic subgroups, improved treatment regimens and, for aberrations affecting tyrosine kinases, even molecularly targeted treatment. Several recurrent genetic aberrations have been identified in acute lymphoblastic leukemia (ALL). Of these, balanced translocations in particular have proved important for risk stratification and treatment guidance. However, several aspects regarding the contribution of additional genetic changes in these subtypes remain to be determined.

The topic of this thesis is genetic aberrations in *ETV6/RUNX1*-positive ALL, one of the most common subtypes of childhood ALL. The thesis contains three sections. The first part provides a general overview of ALL, with a special focus on the *ETV6/RUNX1*-positive subtype. The second part describes the studies that were performed during the work of this thesis. The third and final section contains the original articles upon which this thesis is based.

Lund, January 2011

INTRODUCTION

Leukemia

The term 'leukemia' was coined in 1856 by the german pathologist Rudolph Virchow to describe a disease characterized by unusual quantities of white cells visible in the blood when studied under a microscope (Virchow 1856). The name is derived from the Ancient Greek words λευκός (*leukos* – 'white') and αἷμα (*aima* – 'blood'). However, leukemia is no longer regarded as a single disease, but rather a collection of disorders with overlapping symptoms. The first division is usually made between acute and chronic leukemias, where the acute leukemias are associated with a rapid expansion of cells with an immature phenotype and the chronic leukemias are associated with a gradual buildup of more mature cells. Leukemias are also divided according to the morphologic and immunophenotypic features of the leukemic cells, *i.e.* if they resemble normal myeloid or lymphoid cells. The lymphoid leukemias can be further divided into B-cell or T-cell leukemias depending on the immunophenotype of the leukemic cells. Hence, acute lymphoblastic leukemia (ALL), the focus of the present thesis, is associated with a rapid expansion of immature lymphoid cells; these are termed 'lymphoblasts'.

Information regarding genetic alterations is increasingly used for classifying leukemias. In fact, genetic information was included for the first time in the third edition of the World Health Organization (WHO) classification of Tumours of the Haematopoietic and Lymphoid Tissues, published in 2001. The aim of this classification is to define clinically relevant tumor subgroups to facilitate diagnosis and treatment decisions. In the latest (the fourth) edition of the WHO classifications, several new entities defined by genetic features were included (Vardiman et al. 2009). Among the new categories were seven distinct entities of B-cell precursor acute lymphoblastic leukemia (BCP ALL), further discussed below.

Leukemia is the most common type of childhood cancer and ALL is the most common type of childhood leukemia. In Sweden, 85% of pediatric

leukemias are ALLs, while acute myeloid leukemia (AML) accounts for 10% and chronic myeloid leukemia (CML) together with other non-acute leukemias comprise 5% of the childhood leukemias. For adults, AML and CML are more common than ALL, but since more adults have leukemia approximately half of the Swedish ALL patients are children and the other half are adults. The annual incidence for childhood ALL in the Nordic countries is 3.9/100 000 children (Gustafsson et al. 2000); for Sweden this corresponds to ~60 new cases each year (www.socialstyrelsen.se).

Acute lymphoblastic leukemia

Acute lymphoblastic leukemia is, like all cancers, a genetic disease. It is the result of an accumulation of mutations which leads to deregulation of important cellular processes and ultimately, the leukemic phenotype. It should, however, be stressed that it is a somatically acquired genetic disease, and only in a small minority of cases is it possible to identify a genetic predisposing factor (Pui et al. 2008). Hence, the mutations are only present in the leukemic cells.

It has been found that ALL, especially BCP ALL, can be divided into non-overlapping, prognostically meaningful, subtypes, according to specific cytogenetic aberrations. The information regarding cytogenetic subtype is then used for risk stratification of patients to different treatment regimens. While adult and childhood ALLs are classified according to the same sets of genetic aberrations, their frequency differs substantially among the different age groups. The most common recurrent genetic abnormalities used within the WHO classifications for BCP ALL and their respective frequencies in childhood and adult ALL are listed in Table 1. These genetic aberrations are thought to constitute primary mutations that are responsible for important biological and clinical aspects of the leukemic phenotype. That the genetic subgroups are biologically relevant has been illustrated by murine models for several of the primary mutations (Kennedy and Barabé 2008) and by gene expression profiling of ALLs, revealing that the genetic subtypes are associated with characteristic expression signatures (Yeoh et al. 2002; Ross et al. 2003; Andersson et al. 2005). For a number of primary genetic aberrations there is

Table 1. The most common genetic aberrations used for WHO classification of BCP ALL[a]

Genetic aberration	Frequency in childhood ALL	Frequency in adult ALL
High hyperdiploidy (51-67 chromosomes)	25%	7%
t(12;21)(p13;q22)	22%	2%
11q23/MLL-rearrangements	8%	10%
t(1;19)(q23;p13.3)	5%	3%
t(9;22)(q34;q11.2)	3%	25%
Hypodiploidy (<45 chromosomes)	1%	2%

[a]Adapted from Pui et al. (2004) and Vardiman et al. (2009).

evidence that the specific change can precede the leukemia by several years (Greaves et al. 2003).

Fortunately, improvements in treatment strategies have lead to increased success in treating childhood ALL, with a five year event-free survival rate close to 80% (Gustafsson et al. 2000; Pui and Evans 2006). Adult ALL, however, has a more dismal prognosis with a five year event-free survival rate of only 40% (Pui and Evans 2006). These differences can partly be explained by differences in the underlying genetic aberrations present in adult and pediatric ALLs. Both of the two largest cytogenetic subtypes of childhood ALL, high hyperdiploidy and the 12;21 translocation, are associated with a good prognosis; together they constitute almost 50% of childhood ALL. In adult ALL, however, both high hyperdiploidy and t(12;21) are rare and comprise less than 10% of the cases. The two most common cytogenetic subtypes in adult ALL are instead the t(9;22) (25%), and 11q23/*MLL*-rearrangements (10%), both associated with poor prognosis. These differences do not, however, explain the entire difference in survival between adult and childhood ALL. Interestingly, it has been found that young adults treated according to pediatric protocols fared significantly better than those treated according to adult

protocols (Boissel et al. 2003; de Bont et al. 2004). If these discrepancies should be attributed to differences in treatment regimens, compliance with protocol, or other factors is, however, currently not known (Pui and Evans 2006).

High hyperdiploid ALL is the most common type of childhood ALL and is characterized by a non-random gain of chromosomes resulting in an abnormal cytogenetic pattern with 51-67 chromosomes. The number of chromosomal gains can vary, but most commonly involve the chromosomes X, 4, 6, 10, 14, 17, 18, and 21 (Paulsson and Johansson 2009). It is not known if the gained chromosomes themselves contribute functional properties to the leukemia or if they are byproducts of other oncogenetic events. High hyperdiploidy is associated with a favorable prognosis. There is a marked incidence peak for high hyperdiploid ALL at 2-4 years of age (Paulsson and Johansson 2009). In fact, the general incidence peak for childhood ALL that occurs at 2-7 years of age can almost exclusively be explained by incidence peaks for high hyperdiploid ALL and *ETV6/RUNX1*-positive ALL (Forestier and Schmiegelow 2006). In addition to the whole chromosome gains that defines this subtype, the high hyperdiploid ALLs are also associated with dup(1q), gain of material from 17q, deletions of 6q, and microdeletions targeting *ETV6*, *CDKN2A*, *PAX5*, and *IKZF1* (Paulsson et al. 2010).

The *MLL* gene, located at 11q23, can be perturbed in several ways, including recurrent translocations such as t(4;11)(q21;q23), t(9;11)(p21;q23), and t(11;19)(q23;p13.3), which produce the fusion genes *MLL/AFF1*, *MLL/MLLT3*, and *MLL/MLLT1*. The most common of these is *MLL/AFF1*. This subtype is particularly common in infant ALLs where *MLL* rearrangements can be detected in 70-80% of the cases (Pui et al. 1995). Rearrangements of the *MLL* gene are associated with a dismal prognosis, especially in the age group below 1 year of age (Pui et al. 2003). It has been suggested that a rearranged *MLL* gene may be sufficient for leukemia development, based on the following arguments: 1) the strong association between *MLL* rearrangements and infant leukemia, demonstrating that the disease usually develops with short latency; 2) murine models with *MLL* fusion genes develop leukemia with a high penetrance (Barabé et al. 2007); 3) few secondary changes are identified in ALLs with *MLL* rearrangements,

especially few secondary changes are observed in infant ALLs with the *MLL/AFF1* fusion gene (Bardini et al. 2009).

The *TCF3/PBX1* fusion gene is generated by either a balanced t(1;19)(q23;p13) or, more commonly, an unbalanced der(19)t(1;19). It was initially associated with a poor prognosis, but with intensified treatment protocols the outcome has improved considerably (Pui et al. 1993; Pui and Evans 1998). Cases carrying the der(19)t(1;19) do not display uniparental disomy for the two normal chromosomes 1. Hence, the translocation is most likely preceded by trisomy 1, followed by a loss of the derivative chromosome 1 (Paulsson et al. 2005). No high resolution studies focusing specifically on *TCF3/PBX1*-positive cases of ALL have been reported. To the extent it has been studied, this subtype seems to be associated with few genetic aberrations. However, some of the recurrent microdeletions present in other types of childhood ALL, such as deletion of *PAX5,* have been observed (Mullighan et al. 2007).

The t(9;22)(q34;q11) results in the *BCR/ABL1* fusion gene. This fusion gene is most commonly present in CML, but is also present in a subset of childhood ALL. Several transcript variants of *BCR/ABL1* have been described and the two most common are termed P190 and P210; of these, P190 is the more common variant in ALL (Melo 1996). *BCR/ABL1*-positive ALL is associated with poor prognosis. Unfortunately, the BCR/ABL-inhibitor imatinib mesylate, which can be used to treat CML with great success, only results in a transient response (Druker et al. 2001). The inclusion of imatinib mesylate in multidrug regimens could, however, be a more promising approach (Thomas et al. 2004; Towatari et al. 2004). Presently, the only curative treatment is stem cell transplantation (Pui and Evans 2006). A very strong association between the *BCR/ABL1* fusion gene and microdeletions affecting *IKZF1* has been observed in both adult and childhood ALL (Mullighan et al. 2008). Deletions of *IKZF1* were present in 84% of *BCR/ABL1*-positive ALLs and 2/3 of CMLs in lymphoid blast crisis, but never in chronic phase CML, indicating that these deletions contribute to the lymphoid phenotype of *BCR/ABL1*-positive ALL (Mullighan et al. 2008).

ETV6/RUNX1-positive ALL

The ETV6/RUNX1 fusion gene

Ironically, the most common translocation in childhood BCP ALL, the t(12;21)(p13;q22), was not identified until 1994 (Romana et al. 1994), ten years after the other common translocations in this disease such as t(1;19)(q23;p13), t(9;22)(q34;q11), and t(4;11)(q21;q23) had all been described (Third International Workshop on Chromosomes in Leukemia 1983; Carroll et al. 1984). The reason for this late discovery was that the banding patterns of the involved regions of chromosomes 12 and 21 are so similar that the translocation can not be detected using conventional chromosome banding techniques. Consequently, t(12;21) was not identified until fluorescence *in situ* hybridization (FISH) became available (Romana et al. 1994). The translocation results in the *ETV6/RUNX1* fusion gene (often referred to as *TEL/AML1*), which was cloned less than a year after the translocation had been described (Golub et al. 1995; Romana et al. 1995). The reciprocal *RUNX1/ETV6* fusion gene is only expressed sporadically and is not thought to be pathogenetically important. Both *ETV6* and *RUNX1* are rearranged in several other hematologic malignancies, for example resulting in the recurrent fusion genes *ETV6/PDGFRB*, *ETV6/JAK2*, and *RUNX1/RUNX1T1* (*AML1/ETO*) (Erickson et al. 1992; Golub et al. 1994; Lacronique et al. 1997; Peeters et al. 1997). Considering the frequent involvement of these genes in hematologic malignancies, it is not surprising that both *ETV6* and *RUNX1* have been shown to be important regulators of normal hematopoiesis (Wang et al. 1996a, 1998).

RUNX1 encodes the α subunit of the heterodimeric transcription factor core-binding factor (CBF); the β subunit is encoded by *CBFB*. Alternative α subunits of the CBF transcription factor are encoded by *RUNX2* and *RUNX3*, but *RUNX1* in particular has been found to be important for hematopoiesis. For example, knockout models have shown that murine *Runx1* is required for definitive hematopoiesis and the transition of hematopoietic stem cells from endothelial cells (Okuda et al. 1996; Wang et al. 1996a; Chen et al. 2009). In fact, *Runx1* knockout mice display a similar phenotype to *Cbfb* knockout mice (Sasaki et al. 1996; Wang et al. 1996b). The same phenotype is not observed

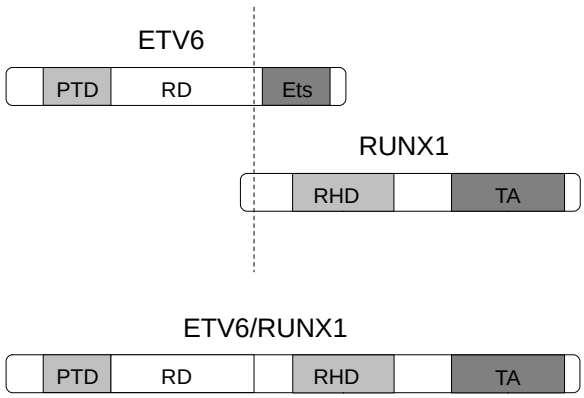


Figure 1. Schematic representation of ETV6, RUNX1, and ETV6/RUNX1 proteins. The ETV6/RUNX1 fusion protein retains the pointed domain (PTD) and the repression domain (RD) but not the Ets domain from native ETV6. Almost the entire RUNX1 protein is also retained, including the runt homology domain (RHD) and the transactivation domain (TA).

in murine *Runx2* and *Runx3* knockout models; instead, these genes seem to be important for bone formation and neural development (Otto et al. 1997; Komori et al. 1997; Inoue et al. 2002; Levanon et al. 2002). An indication that *RUNX1* plays an important role in human hematopoiesis is that constitutional haploinsufficiency for *RUNX1* leads to familial platelet disorder, a disease characterized by platelet abnormalities and a propensity to develop AML (Song et al. 1999). The most conserved region of the RUNX family of proteins is the 'runt homology domain' (RHD), which is sufficient for both DNA binding and interaction with CBFB (Meyers et al. 1993).

ETV6 is a transcriptional repressor that contains a helix-loop-helix dimerization domain termed the 'pointed domain' (PD), a repression domain (RD), and the DNA binding 'Ets' domain (Figure 1). Murine knockout models have revealed that loss of *Etv6* results in embryonic lethality (Wang et al. 1997). Chimeric knockout mice suggest that, while *Etv6* is not required for fetal liver hematopoiesis, it is required for later stage hematopoiesis in the bone marrow (Wang et al. 1998).

Both ETV6 and RUNX1 are, hence, transcription factors with important roles in normal hematopoiesis. The fusion between these proteins brings the repression domain of ETV6 together with the DNA binding domain of RUNX1 (Figure 1), which results in repression of genes normally activated by

RUNX1 (Hiebert et al. 1996; Fears et al. 1997; Fenrick et al. 1999; Uchida et al. 1999). It is, however, likely that the fusion protein also performs other important functions, as illustrated by the fact that loss of *Runx1* and expression of *ETV6/RUNX1* affect the number of long term repopulating hematopoietic stem cells in opposite directions (Schindler et al. 2009). The ETV6 pointed domain and corepressor domain together with the RUNX1 runt domain have been shown to be necessary for the function of the fusion protein (Morrow et al. 2007). In fact, as for the runt domain, both the DNA binding and CBFB interacting regions of this domain are required for ETV6/RUNX1 function (Morrow et al. 2007; Roudaia et al. 2009). The pointed domain of ETV6, retained in the ETV6/RUNX1 fusion protein, is necessary for normal ETV6 dimerization. This domain is functional in the fusion protein and allows it to dimerize both with ETV6/RUNX1 and normal ETV6 (McLean et al. 1996).

Additional genetic changes in ETV6/RUNX1-positive ALL

There are several lines of evidence suggesting that the *ETV6*/*RUNX1* fusion gene by itself is not sufficient to cause leukemia. Two of the most compelling ones are provided by retrospective studies identifying small *ETV6/RUNX1* positive cell populations in blood samples collected at birth from children that later developed *ETV6/RUNX1*-positive ALL (Wiemels et al. 1999b; Maia et al. 2001; Hjalgrim et al. 2002; McHale et al. 2003), and studies of monochorionic twins with concordant *ETV6/RUNX1*-positive ALLs showing that these have identical genomic breakpoints for *ETV6/RUNX1* (Ford et al. 1998; Wiemels et al. 1999a, 1999b; Maia et al. 2001; Teuffel et al. 2004; Bateman et al. 2010). In a recent study, an *ETV6/RUNX1*-positive 'preleukemic' clone was even detected in the healthy twin of a patient with *ETV6/RUNX1*-positive ALLs; also with identical genomic breakpoints to that of the fusion gene in the other twin (Hong et al. 2008). These examples show that the 'initial hit' of the leukemia can occur years before overt leukemia is detected, suggesting that additional events are necessary for leukemia development. In this context, it is interesting to note that in a recent study using single nucleotide polymorphism (SNP) arrays to characterize genetic changes in five pairs of *ETV6/RUNX1*-positive twin ALLs, none of the 32

secondary alterations characterized as 'driver' mutations overlapped between the twin ALLs (Bateman et al. 2010). Hence, the acquisition of these mutations is likely to constitute independent 'additional events' required for leukemogenesis.

The fact that the *ETV6/RUNX1* fusion gene can be detected in the blood of newborns at a much higher frequency than that of children developing *ETV6/RUNX1*-positive ALL (Mori et al. 2002) is usually also used as evidence for the importance of secondary events in the development of leukemia. It has been estimated that approximately ~1% of newborns are positive for *ETV6/RUNX1* (Mori et al. 2002); this is 100 times higher than the incidence of *ETV6/RUNX1*-positive ALL. The fraction of positive cells was estimated to be between 10^{-3}-10^{-4} in these cases. These data have, however, recently been challenged by data from a larger study (Lausten-Thomsen et al. 2011). In this study, none of the individual samples positive by real-time quantitative RT-PCR could be confirmed positive upon retesting, indicating sporadic positivity either due to contamination or because the frequency of positive cells was very low (Lausten-Thomsen et al. 2011). In response to this study, a third group reported ~2% positivity for *ETV6/RUNX1* in a screening of 253 cord blood samples (Zuna et al. 2011). Clearly this issue is not yet settled and further studies are needed before the frequency with which the *ETV6/RUNX1* fusion gene is generated during normal fetal development can be known with certainty.

Regardless of the normal frequency, it is evident that the *ETV6/RUNX1* fusion gene, when associated with ALL, often arises *in utero* and that the fusion gene triggers the development of a small preleukemic population that requires additional genetic events and/or other events during childhood to develop into leukemia. Hence, the additional genetic changes found in *ETV6/RUNX1*-positive leukemias can give important clues to the developmental history of each leukemia. With the increased resolution in techniques for studying genetic changes, a growing number of genetic changes recurrently associated with the *ETV6/RUNX1*-fusion gene has been found. Even with older techniques such as G-banding and FISH it was recognized that *ETV6/RUNX1*-positive leukemias contain additional somatic genetic changes in the majority of cases (Stams et al. 2006; Forestier et al. 2007). In

fact, three of the most common genetic changes associated with *ETV6/RUNX1*-positive ALL are readily discovered using the same FISH probes required for detecting the translocation; deletion of *ETV6* on 12p, +21, and +der(21)t(12;21) (Stams et al. 2006). Other common changes observed using G-banding include deletions on 6q, 8p, 9p, 11q, and 13q as well as whole chromosome loss of X and gains of 10 and 16 (Forestier et al. 2007). Recent studies using molecular techniques for detecting unbalanced genetic changes, *i.e.* array comparative genomic hybridization (array CGH) and SNP arrays, have largely confirmed the previous findings of G-banding and FISH (Kuiper et al. 2007; Mullighan et al. 2007; Tsuzuki et al. 2007; Kawamata et al. 2008; Parker et al. 2008; Article I; Article II), with the curious exception of the X chromosome loss that has not been detected to a high extent using these techniques. However, in addition to confirming previous findings, several previously undetected recurrent genetic changes have been found using these techniques. For example, a large region of Xq is gained in ~20% of *ETV6/RUNX1*-positive ALLs (Article I; Article II). Although this gain was common and involved a relatively large chromosomal region it had previously not been identified using standard G-banding, most likely due to the poor banding quality often associated with childhood ALL. Several genes within the gained region displayed an overexpression in *ETV6/RUNX1*-positive cases compared with other ALLs; the most overexpressed genes were *SPANXB*, *HMGB3*, *FAM50A*, *HTATSF1,* and *RAP2C*. Of these, *SPANXB* was considered the most likely target of the Xq gain. The recent identification of a fusion gene involving *HMGB3* (Petit et al. 2010), however, indicates that this gene might also be an attractive target within the gained Xq region.

Focal deletions affecting only one or a small number of genes have been found to be common in all subtypes of childhood ALL (Kuiper et al. 2007; Mullighan et al. 2007). Different types of genes are targeted by these focal deletions, but genes involved in B-cell development, such as *PAX5*, *EBF1*, *IKZF1*, *IKZF3, LEF1*, and *TCF4,* are particularly often perturbed (Kuiper et al. 2007; Mullighan et al. 2007). Of these, only *PAX5*, *EBF1,* and *TCF4* are recurrently deleted in *ETV6/RUNX1*-positive cases (Article II). Loss of any of these genes in murine models is associated with impaired B-cell development (Lin and Grosschedl 1995; Zhuang et al. 1996; Nutt et al. 1999); hence, these

Table 2. The most common CNAs in *ETV6/RUNX1*-positive ALL[a]

Copy number aberration	Frequency
ETV6 deletion	59%
Deletion of *CDKN2A* and *CDKN2B*	22%
PAX5 deletion	20%
Deletion of 6q	20%
Gain of Xq	16%
Deletion of *CD200* and *BTLA*	13%
TBL1XR1 deletion	12%
+21	12%
+der(21)t(12;21)	12%

[a]Adapted from Article II.

deletions are likely to contribute to the immature B-cell phenotype characterizing BCP ALL. Interestingly, several of the focal deletions specific for *ETV6/RUNX1*-positive ALL, such as those affecting *TBL1XR1*, *NR3C1*, or *NR3C2*, are involved in nuclear hormone receptor response (Article II); however, the mechanism by which these deletions contribute to leukemia remains to be determined. A summary of the most common secondary changes in *ETV6/RUNX1*-positive ALL as determined by SNP array is presented in Table 2.

While focal deletions and other mutations likely are important for leukemogenesis, not many studies have addressed how the presence of these affect treatment outcome in *ETV6/RUNX1*-positive ALL. However, a large study of CpG methylation differences in childhood ALL identified that CpG methylation information can be used to divide the *ETV6/RUNX1*-positive subtype into two groups with significantly different probabilities for relapse (Milani et al. 2010).

With the recent introduction of highly parallel sequencing, a technique that enables whole exome- and even whole genome-sequencing of individual tumors, the list of recurrently altered genes in *ETV6/RUNX1*-positive ALL will

presumably grow significantly. Very recently, exome sequencing of two *ETV6/RUNX1*-positive ALLs revealed seven somatic gene mutations in each case, which is similar to the number of somatic copy number changes detectable using SNP array (Article III). Several of the identified mutations affected genes known to be perturbed in other types of tumors, such as *RUNX1T1* and *GKN1,* or genes involved in tumor suppressor pathways or otherwise related to tumor development such as *TP53INP1*, *CSMD2*, and *MCAM*, indicating that these single nucleotide mutations are important for the development of *ETV6/RUNX1*-positive ALL.

Animal models

Several animal models have been generated for studying the effect of *ETV6/RUNX1*-expression on hematopoietic cells. The different models have sometimes exhibited contradictory phenotypes, possibly due to different experimental setups. The majority of the animal models have been murine (Andreasson et al. 2001; Bernardin et al. 2002; Morrow et al. 2004; Tsuzuki et al. 2004; Fischer et al. 2005; Schindler et al. 2009), but also zebrafish and human cord blood cells, the latter injected into immunodeficient mice, have been used (Sabaawy et al. 2006; Hong et al. 2008). The most common approach for transgene delivery has been retroviral transduction of the *ETV6/RUNX1* fusion gene into either murine adult bone marrow cells or fetal liver cells (Bernardin et al. 2002; Morrow et al. 2004; Tsuzuki et al. 2004; Fischer et al. 2005). Two of the studies using this approach reported that expression of *ETV6/RUNX1* results in impaired (but not completely blocked) B-cell differentiation *in vivo* (Tsuzuki et al. 2004; Fischer et al. 2005). In addition to this impairment, Fischer et al. (2005) described the presence of an expanded $Ly6a^{+}/Kit^{hi}$ population, resembling multipotential progenitors, in both blood and bone marrow of the transplanted mice. Morrow et al. (2004) described an overrepresentation of *ETV6/RUNX1*-positive B-cells following transplantation. While this might seem contradictory to the impaired B-cell differentiation described by other groups, the surface antigen used to identify B-cells in this study would include also immature B-cells; hence, this result is compatible with impaired B-cell differentiation.

A non-mammal model has been developed using transgenic zebrafish. After

long latency, 3% of the transgenic fish developed oligoclonal B-cell acute lymphoblastic leukemia. Impaired B-cell differentiation and an expansion of progenitor cells was also observed (Sabaawy et al. 2006).

Hong et al. (2008) used lentiviral delivery of *ETV6/RUNX1* into human cord blood cells that were allowed to reconstitute the bone marrow of immunodeficient NOD/SCID mice. Interestingly, this gave rise to a population of $CD34^+/CD38^-/CD19^+$ cells. Such cells are not present in normal human bone marrow and this subpopulation is believed to house a pool of leukemic or preleukemic stem cells in *ETV6/RUNX1*-positive ALLs (Castor et al. 2005; Hong et al. 2008). The $CD34^+/CD38^-/CD19^+$ subpopulation could be transplanted into secondary mice and reconstituted both the $CD34^+/CD38^-/CD19^+$ population and mature B-cells.

In a recent study by Schindler et al. (2009), *Etv6/RUNX1* was expressed from the endogenous murine *Etv6* locus by replacing exon six of this gene with human *RUNX1* cDNA. A loxP flanked STOP-signal prevented the expression of the fusion transcript unless the STOP-signal was excised using Cre. This model, arguably mimicking several aspects of the native fusion gene, revealed that *Etv6/RUNX1* blocks the emergence of early B-cells when expressed in hematopoietic stem cells. An increase in hematopoietic stem cells was also observed. Importantly, combining *Etv6/RUNX1* expression with a mutagenic DNA alkylating agent resulted in aggressive lymphoblastic leukemia when *Etv6/RUNX1* was expressed in hematopoietic stem cells, but not when expressed in early B-cell progenitors.

A common theme in all the described models is that *ETV6/RUNX1* expression by itself is insufficient for leukemia development (Andreasson et al. 2001; Morrow et al. 2004; Tsuzuki et al. 2004; Fischer et al. 2005; Hong et al. 2008; Schindler et al. 2009). The two models where leukemia apparently developed spontaneously following *ETV6/RUNX1* expression were both associated with long latencies and low frequencies of spontaneous leukemia, further illustrating the need for cooperating mutations (Bernardin et al. 2002; Sabaawy et al. 2006). Another common theme is the impaired B cell differentiation observed by several groups (Tsuzuki et al. 2004; Fischer et al. 2005; Schindler et al. 2009). The exact nature of this impairment was however different between these models. Tsuzuki et al. (2004) and Fischer et al. (2005)

observed that the transition from pro- to pre-B-cell was perturbed, which corresponds well to the maturation block observed in *ETV6/RUNX1*-positive BCP ALL. Because of this, they speculated that this impairment contributes to leukemia development. Schindler et al. (2009), however, observed a block in maturation that occurred earlier than the pro- to pre-B-cell transition. Moreover, as this block was only evident in adult bone marrow hematopoiesis and not in fetal liver hematopoiesis, they speculated that this block is more likely to explain the association between *ETV6/RUNX1* and childhood ALL. It was argued that the early block in B-cell development observed in adult bone marrow would prohibit the emergence of a B-lineage committed *ETV6/RUNX1*-positive preleukemia in adult bone marrow and, hence, that this preleukemic clone can only develop *in utero*.

In conclusion, the animal models all support the idea that *ETV6/RUNX1* perturbs hematopoietic stem cells and gives rise to a preleukemic clone that can persist in the blood and bone marrow for several years and that the acquisition of specific mutations within this preleukemic clone results in ALL.

The Present Study

Specific aims of the study

The general aim of this thesis has been to characterize acquired genetic changes that occur in *ETV6/RUNX1*-positive ALLs. More specifically the aims were:

- To determine the presence of acquired copy number aberrations in *ETV6/RUNX1*-positive ALLs and to study the impact of these changes on the expression of genes within the altered regions. (Article I)
- To identify recurrent copy number changes in a large series of *ETV6/RUNX1*-positive ALLs and investigate how these are correlated. (Article II)
- To identify all somatic mutations in the exomes of two *ETV6/RUNX1*-positive ALLs. (Article III)

Materials and Methods

This section contains a brief overview of the material and methods used in the study. For a more detailed description the reader is referred to the *Materials and Methods* section and the corresponding *supplement* of each article.

Patient material

All the patients included in this study were treated at either Lund or Linköping University Hospitals. The presence of *ETV6/RUNX1, TCF3/PBX1, BCR/ABL1* and *MLL* fusion genes in the bone marrow or blood was determined at the Department of Clinical Genetics, Lund, using reverse transcriptase polymerase chain reaction (RT-PCR), Southern blot, and FISH, as a part of routine diagnostic analysis. DNA and RNA were extracted using standard methods at diagnosis and stored at the Department of Clinical

Genetics, Lund. This study was approved by the ethical committees of Lund and Linköping Universities.

Array CGH

Array CGH was performed using a bacterial artificial chromosome array platform with 32 433 features covering 98% of the genome at 100 kb resolution (Article I). The arrays were produced at the SCIBLU DNA microarray resource center at Lund University. In brief, the arrays were used as follows: patient DNA and reference DNA were labeled with Cy3 and Cy5 respectively; the patient and reference DNAs were competitively hybridized to the arrays, which were then washed and scanned. Intensity information was retrieved from the scanned images using GenePix Pro 4.0 software (Axon Instruments, Foster City, USA). The intensity information was uploaded into the web based software environment BASE (Saal et al. 2002). Low intensity features were removed from the data, after which the data were normalized using Lowess curve fitting (Yang et al. 2002). Gains and losses were identified after plotting the data using a BASE implementation of the MATLAB toolbox CGH-plotter. (Autio et al. 2003).

SNP arrays

Three types of SNP arrays were used in the current study: The mapping 500K array set (Affymetrix, Santa Clara, USA) in Article II and cytogenetics whole-genome 2.7M arrays (Affymetix) and HumanCNV370-quad arrays (Illumina, San Diego, USA) in Article III. All arrays were used according to the respective manufacturer's protocol. The data from the mapping 500K array set were prepared and visualized using dChip (Lin et al. 2004). The data were segmented using DNAcopy (Venkatraman and Olshen 2007) from the Bioconductor software package (Gentleman et al. 2004) and copy number changes were detected from these segments. The cytogenetics whole-genome 2.7M arrays were analyzed using Chromosome analysis suite (Affymetrix) to detect copy number changes and genotype information was extracted from the HumanCNV370-quad arrays using Genome Studio (Illumina).

External data

Previously published SNP array data from external research groups were used in Article II. This allowed a thorough correlation analysis of secondary copy number changes in *ETV6/RUNX1*-positive ALLs. Two external data sets were used, one comprising 47 *ETV6/RUNX1*-positive cases from the data set described by Mullighan et al. (2007), and one consisting of 93 *ETV6/RUNX1*-positive cases from a data set published by Kawamata et al. (2008). Scanned intensities were downloaded as CEL-files; these were converted into log2 copy numbers using dChip for both external data sets. The data were segmented using DNAcopy (Venkatraman and Olshen 2007).

Exome enrichment and highly parallel sequencing

Exome enrichment was performed using 2.1M sequence capture human exome arrays (Roche, Madison, USA) and highly parallel sequencing of the enriched material was performed on the Genome Analyzer II (Illumina). These steps were performed in collaboration with Ambry Genetics, Aliso Viejo, USA. The samples were prepared using Illumina's protocol for paired end sequencing, modified only with a hybridization step to perform the exome enrichment. The libraries were subjected to 2x55 cycles of paired end sequencing. However, due to a machine error, only 42 cycles of data were available for one of the paired reads from one flow cell. Base calling was performed using RTA (Illumina) and quality filtering of the sequence data was performed using Illumina pipeline software (Illumina). The software 'bwa' (Li and Durbin 2009) was used to align the filtered reads to the hg18 build of the human genome. Filtering of PCR duplicates was performed using an in-house modified version of 'Picard' (http://picard.sourceforge.net/). Genotyping was performed using the maq consensus model implemented in samtools (Li et al. 2009). Genotype comparisons with SNP array data, filtering against dbSNP and comparison with reference genotype data were performed using custom perl scripts.

Results and discussion

Article I

In order to identify copy number aberrations (CNAs) and to study their impact on gene expression, 17 *ETV6/RUNX*-positive ALLs were analyzed with array CGH and gene expression profiling. Most of the CNAs previously identified using G-banding and FISH could be confirmed. Furthermore, 31 additional CNAs were identified. Of these, 12 CNAs were from cases where the G-banding analysis had failed and six CNAs where too small for detection using G-banding. The remaining 13 imbalances where, however, most likely missed due to poor chromosome morphology or similar banding patterns of the involved chromosome arms. The CNAs were found to be associated with general changes in gene expression for genes located in the affected regions; as expected, deletions were associated with decreased gene expression and gains were associated with increased expression.

The most common copy number change was a previously undescribed gain of Xq material of varying size, but with a common 35 Mb minimally duplicated region. Curiously, the six cases that were positive for this aberration were all in ALLs from boys, indicating this aberration is quite specific for boys with *ETV6/RUNX1*-positive ALL. In the three cases where material was available for FISH, the gained Xq material was found to be part of an unbalanced translocation with either chromosome 6 or 11. Since the breakpoints on Xq were variable and since different partner chromosomes were involved, it was considered unlikely that the unbalanced translocations would results in a fusion gene. That the functional outcome of Xq duplications would be increased expression of a gene within the minimally gained region important for leukemia development was considered a more likely scenario. The 35 Mb minimally gained region on Xq was found to harbor several genes with a specifically high expression in *ETV6/RUNX1*-positive ALLs, identifying them as candidate targets of the duplication. Three genes in particular were highly expressed in this subtype as determined by both gene expression analysis and real time quantitative PCR: *SPANXB*, *HMGB3*, and *FAM50A*. The exceptionally specific overexpression of *SPANXB* in *ETV6/RUNX1*-positive ALL was interpreted as an indication that this gene is the main target of the

duplication. However, since *HMGB3* has recently been described to be fused with *NUP98* in a therapy related AML (Petit et al. 2010), this gene should also be regarded as a potential target of the Xq duplication.

The finding that gain of Xq material is common in male *ETV6/RUNX1*-positive cases was unexpected and intriguing. Only sporadic cases of *ETV6/RUNX1*-positive ALLs with partial gain of Xq had previously been described. In total, only eight such cases were known prior to this study (Ma et al. 2001; Lu et al. 2002; Nordgren et al. 2002; Jarošová et al. 2003; Kuchinskaya et al. 2005; Martineau et al. 2005). Additional cases were, however, evident from studying an independent gene expression data set (Ross et al. 2003); gain of Xq could be inferred from the characteristic overexpression of genes in the minimally gained region in eight cases from this data set. An in depth analysis of the data set of Mullighan et al. (2007), published when the present data were being prepared for submission, confirmed the presence of Xq gains in around 20% of *ETV6/RUNX1*-positive cases, although this intriguing finding was not discussed in that study (Mullighan et al. 2007). Just like our cases, a large majority (at least 15/18) of the external cases were male, indicating a very specific association between this aberration and this gender. A possible explanation for this specific association could be that females already carry two copies of the entire X chromosome and have the potential to reactivate important genes on Xq without the need for Xq duplication. The fact that both *SPANXB* and *HMGB3* were overexpressed also in girls could argue in favor of this.

In summary, we identified gain of Xq material in a high proportion of *ETV6/RUNX1*-positive ALL in boys. The gain was associated with an increased expression of several genes in the minimally gained region, of which overexpression of *SPANXB* was considered to be particularly important due to the high and very specific expression of this gene in *ETV6/RUNX1*-positive ALLs.

Article II

In this study, 23 primary *ETV6/RUNX1*-positive ALLs and one *ETV6/RUNX1*-positive cell line were analyzed locally using 500K SNP arrays. In addition, previously published data on 140 external *ETV6/RUNX1*-positive

cases were reanalyzed together with the local data in order to produce reliable frequency estimates for the different CNAs and to study their interdependencies. The 24 local cases included also the 17 cases studied in Article I. All CNAs identified using array CGH in Article I were confirmed using the higher resolution SNP arrays in this study. In addition to the previously identified changes, a number of deletions below the detection limit of array CGH (around 300 kb) were identified. In total, 29 recurrent CNAs were identified in the 24 cases and 16 of these were recurrent deletions below 300 kb in size (referred to as focal deletions).

SNP array data describing *ETV6/RUNX1*-positive cases from two external data sets (Mullighan et al. 2007; Kawamata et al. 2008) were combined with the local data. This permitted detailed analysis of CNAs in 164 *ETV6/RUNX1*-positive cases. Since the two external data sets were produced using different platforms, not all recurrent CNAs could be studied in the entire combined data set. In total, 55 recurrent CNAs were identified, of which 45 could be studied in all 164 cases. This large data set permitted the study of interdependencies among the CNAs; this was accomplished using hierarchical clustering analysis, connected pair analysis and oncogenetic tree analysis.

A hierarchical clustering analysis based on the presence of recurrent CNAs in the 164 cases revealed ten clusters, mainly defined by ten different CNAs. Forty two cases (26%) displayed a profile of recurrent CNAs identical to another case in the data set. These cases were, however, in minority as 74% of all cases had a unique combination of recurrent CNAs. This indicates that the acquisition of CNAs is a stochastic process, a proposition further highlighted by the fact that the cases with identical CNAs all had a small number of recurrent CNAs (i.e. three, two, one, or zero recurrent CNAs).

Despite the stochastic nature of CNA accumulation, there might exist interdependencies between CNAs, for example due to functional redundancies, *i.e.* similar leukemogenic contribution from different CNAs. Such correlations were highlighted using connected pair analysis (CPA). Two significant ($q < 0.05$) negative associations were found, deletion of *ETV6* was negatively correlated with +der(21)t(12;21) and deletion of *ETV6* was also negatively correlated with gain of Xq material. Functional redundancy can

explain the former association as existing data suggest that heterodimerization between ETV6 and the ETV6/RUNX1 fusion protein impairs the function of the fusion protein (McLean et al. 1996). In this case, an increased ETV6/RUNX1 to ETV6 ratio would lead to a more functional fusion gene; deletion of *ETV6* or duplication of the fusion gene are two different ways to increase this ratio. The negative connection between gain of Xq and deletion of *ETV6* is, however, more elusive. Possibly this association indicates that the target gene on Xq is related to ETV6 function.

A number of positive correlations between CNAs were also identified. The most prominent of these was the association between the gains of chromosomes 10 and 16. This association was most likely present due to a general tendency to acquire whole chromosome gains in a small number of cases since those with gain of these chromosomes exhibited significantly more gains of other chromosomes. There was also a significant association between gain of Xq and deletion of a large region on 11q which is most likely explained by the fact that the Xq gain can be part of an unbalanced X;11 translocation, resulting in simultaneous gain of Xq and deletion of 11q (Article I). Several additional positive correlations were also identified although the exact nature behind their connections could not be explained.

The data were also analyzed using two types of oncogenetic tree models; branching oncogenetic trees and distance based oncogenetic trees (Desper et al. 1999, 2000). These have been developed for inferring and visualizing the sequential acquisition of mutations and to subgroup mutations based on their co-occurrence. The two oncogenetic trees were in good agreement as three quite similar groups of CNAs were separated in both models.

The strong association between gain of chromosomes 10 and 16 revealed by CPA was also apparent from the oncogenetic trees as these were connected in the branching tree and present in the same cluster in the distance based tree. Another observation was that among the three groups of CNAs identified by the oncogenetic tree models there were two consisting mainly of focal deletions while the third consisted mainly of gains and deletions of large chromosomal regions and whole chromosome gains. Hence, large and small genetic changes commonly occur together with CNAs of similar size, which indicates that predisposing mutations could precede these CNAs. As for the

whole chromosome gains, these were present in cases with a tendency to acquire whole chromosome gains in general (as described above). As for focal deletions, these have, in single cases, been shown to be the result of aberrant RAG activity (Novara et al. 2009); an enzyme involved in the normal somatic rearrangement of immunoglobulin and T-cell receptor loci. Hence, it is plausible that mutations affecting RAG activity in a subset of cases could be responsible for the clusters of focal deletions. Oncogenetic trees also inferred a sequential order of appearance for the aberrations and according to the two models the early events in the leukemogenic process are del(6q), +16, and deletions of *ETV6*, *CDKN2A/B*, and *PAX5*.

Article III

Newly developed methods for large scale sequencing have made it possible to sequence the whole genome of individual leukemias and tumor cell lines (Mardis 2010). However, whole genome sequencing is still relatively expensive. Another application also enabled by the new sequencing techniques is whole exome sequencing, where sequencing is preceded by an enrichment step in order to maximize the amount of sequenced material originating from coding regions. The latter application has enabled the identification of disease causing mutations in several rare Mendelian disorders (Choi et al. 2009; Ng et al. 2010). In this study, mutations affecting the coding regions of two *ETV6/RUNX1*-positive ALLs were characterized using SNP arrays and exome capture followed by highly parallel sequencing. In total, 80 and 85% of the coding regions were successfully genotyped by exome sequencing in the two ALLs. Comparing the genotypes with dbSNP (Sherry et al. 2001) and genotypes from a matched normal sample, followed by confirmation using traditional capillary sequencing, revealed seven single nucleotide mutations in the coding regions of each ALL. One of the mutations was in codon 835 of *FLT3*, which is well known to be mutated in both lymphoid and myeloid leukemias (Stirewalt and Radich 2003). This mutation was only present in 3 of 30 reads from the highly parallel sequencing and could at first not be confirmed using capillary sequencing. However, when a restriction enzyme assay was used in combination with capillary sequencing, the *FLT3* mutation was confirmed to be present in ~15% of the *FLT3* alleles, providing an

explanation for both the low number of reads in the highly parallel sequencing and the failure to detect the mutation using capillary sequencing. This example demonstrates the power of highly parallel sequencing over traditional capillary sequencing, especially at high read depths.

In addition to the single nucleotide mutations, the two ALLs contained eight and 14 CNAs, respectively. Of these, five CNAs in each ALL were from the subset of CNAs found to be recurrent in *ETV6/RUNX1*-positive ALL in Article II and, hence, most likely constitute mutations that contribute to the disease.

The single nucleotide mutations affected different types of genes. The *FLT3* mutation described above is well known to be associated with leukemia. The remaining 13 mutations, however, were in genes with no previously known involvement in leukemia. Therefore, the presence of all 14 mutations was analyzed in an extended series of 27 primary *ETV6/RUNX1*-positive ALLs and two *ETV6/RUNX1*-positive cell lines. These did, however, not contain additional mutations. The mutated genes were then compared with mutated genes in other types of cancer through the Catalogue of Somatic Mutations in Cancer (COSMIC) (Forbes et al. 2008). The effect of the mutations on the resulting protein was also modeled using PolyPhen-2 (Adzhubei et al. 2010) and SIFT (Ng and Henikoff 2003). These analyses revealed that five genes, in addition to *FLT3*, had previously been found to be mutated in a tumor. Two of the genes with mutations in other cancer types, *GKN1* and *RUNX1T1*, had protein altering mutations. Of these two, the amino acid change in *GKN1* was predicted to be tolerated by both PolyPhen-2 and SIFT. The mutation in *RUNX1T1*, however, affected a splice donor site and is very likely to affect the splicing and thereby change the final protein dramatically. Interestingly, *RUNX1T1* is commonly rearranged in acute myeloid leukemia as part of the fusion gene *RUNX1/RUNX1T1* (also known as *AML1/ETO*), further demonstrating that alterations affecting this gene can be oncogenic. In addition to the mutations affecting *FLT3*, *GKN1*, and *RUNX1T1*, the mutations in the genes *TP53INP1*, *CSMD2*, and *MCAM* were also considered candidate driver mutations. The *TP53INP1* mutation was included because this gene encodes a protein that is suggested to function as a tumor suppressor and acts downstream of TP53. Also, the amino acid change caused by this

mutation was predicted by both models to disrupt the protein. The mutation in *CSMD2* was considered possibly oncogenic due to 1) that the highly related *CSMD1* and *CSMD3* are commonly altered and suggested to be tumor suppressor genes (Ma et al. 2009; Gylfe et al. 2010); and 2) that only the mutated version of *CSMD2* could be detected by RT-PCR followed by sequencing in the mutated ALL in this study. The mutation in *MCAM*, predicted by PolyPhen-2 to be 'probably damaging', was also considered likely to be oncogenic due to the involvement of *MCAM* in melanoma progression and prostate and breast tumors (Ouhtit et al. 2009). Hence, although none of the 14 single nucleotide mutations identified by exome sequencing was recurrent in 29 *ETV6/RUNX1*-positive ALLs, several of them are likely to be important for leukemogenesis.

Conclusions

Article I

- Array CGH is a robust method for detecting CNAs in *ETV6/RUNX1*-positive ALL.
- Copy number changes have a great impact on gene expression in the affected regions.
- Gain of Xq is present in ~35% of *ETV6/RUNX1*-positive ALLs; this copy number change is specific for male cases.
- Several candidate target genes are located within the minimally gained Xq region, of which the gene *SPANXB* is the most likely target of the duplication.

Article II

- *ETV6/RUNX1*-positive ALLs display a high number of recurrent CNAs.
- The majority (74%) of *ETV6/RUNX1*-positive ALLs have acquired a unique set of recurrent CNAs, suggesting that the process of acquiring

CNAs is unique for each case.

- Genes involved in nuclear hormone response are frequently targeted by focal deletions in *ETV6/RUNX1*-positive ALLs.
- Oncogenetic tree models identify deletions of *ETV6*, *CDKN2A/B*, *PAX5*, and 6q and +16 as possible early events in the leukemogenesis of *ETV6/RUNX1*-positive ALLs.

Article III

- Using exome sequencing it was possible to produce high quality genotype information for 80 – 85% of the coding regions.
- Although highly parallel sequencing is associated with low error rates, the large amount of data result in high rates of false positives; hence, it is crucial to confirm the mutations in a separate experiment.
- *ETV6/RUNX1*-positive ALLs carry approximately equal numbers of single nucleotide mutations and CNAs.
- Six of the 14 single nucleotide mutations affected genes that were either present in tumor suppressor pathways, closely related to known tumor suppressors, or known to be perturbed in other tumors, strongly indicating that these mutations are important for leukemia development.

Concluding remarks and future perspectives

Although treatment of childhood ALL has become a tremendous success story, with cure rates improving from zero to ~80% in the last 60 years (Kersey 1997), many aspects of this disease remain elusive. It is known, especially from studies of *ETV6/RUNX1*-positive ALL, that the initial events preceding the leukemia can occur several years before the disease has its onset, often even *in utero*. The period from the initiating event to the debut of overt ALL is, however, shrouded in mystery. If the cellular path towards leukemia was known, and more importantly, could be interrupted, many children could be spared the intensive treatment and side effects associated with current ALL management protocols. One elusive aspect is the normal frequency for the occurrence of leukemia associated genetic changes. It has, for example, been reported that the *ETV6/RUNX1* fusion gene occurs in the cord blood of around 1% of newborns (Mori et al. 2002), a frequency approximately 100 times higher than expected if all newborns carrying the fusion gene would develop *ETV6/RUNX1*-positive leukemia. These data have, however, recently been challenged by a study indicating that a majority of the positive samples represent either contaminations or, possibly, samples with very small numbers of positive cells, below the actual detection limit of the assay (Lausten-Thomsen et al. 2011). If the data from the latter study prove correct it might be possible to devise an assay that identifies the preleukemic clones that will evolve into full blown leukemias, which could possibly allow earlier, less toxic, therapeutic interventions.

The cytogenetic subtype of childhood ALL has a strong influence on prognosis and this information is also considered when determining the intensity of treatment. During the present study it has become clear that although there is considerable overlap in secondary genetic changes between the cytogenetic subtypes, several of the secondary changes are subtype specific, such as the *ETV6/RUNX1*-specific gain of Xq described in Article I and the alterations affecting nuclear hormone response described in Article II. The fact that this study has focused exclusively on the *ETV6/RUNX1*-positive subtype

of ALL has enabled the discovery of these subtype-specific properties and has enabled us to appreciate their contribution to this specific subtype better. The gain of Xq, for example, is present in less than 5% of the ALLs in the data set of Mullighan et al. (2007) and could easily be discarded as rare and uninteresting if the connection to male *ETV6/RUNX1*-positive cases was not considered. However, since the publication of Article I, the high incidence of Xq gains in *ETV6/RUNX1*-positive ALLs has been confirmed further. Practically all high-resolution studies of copy number changes in *ETV6/RUNX1*-positive ALLs reported since the publication of Article I have described Xq gains, although the frequency of positive cases has varied between 16 and 26% (Parker et al. 2008; Horsley et al. 2008; Bateman et al. 2010; Article II). Interestingly, recent studies have also highlighted one of our original candidates, in addition to *SPANXB,* as a possible target for the Xq duplication. The fusion gene *NUP98/HMGB3* has been identified in a therapy-related AML, illustrating that *HMGB3* can be involved in leukemogenesis (Petit et al. 2010), consistent with its role in regulation of hematopoiesis (Nemeth et al. 2006). Hence, the individual contribution of each of the different overexpressed genes on Xq has not been settled. Functional studies where the most likely candidates are silenced and/or overexpressed in relevant model systems are probably required to settle this issue.

The overall aim of the current study has been to define the full complement of genetic changes present in *ETV6/RUNX1*-positive ALL. This has been accomplished using state of the art microarray and sequencing techniques. The results of the exome sequencing of Article III, however, illustrate that the picture is not yet complete. As whole exome and whole genome sequencing are becoming available, studies of larger series of *ETV6/RUNX1*-positive ALLs will undoubtedly identify additional recurrent genetic changes. So, what are the potential benefits of this increased knowledge? A number of important applications might emerge with the new information. First of all, as the *ETV6/RUNX1* fusion gene is not sufficient for inducing leukemia on its own in most of the models produced thus far, (Andreasson et al. 2001; Morrow et al. 2004; Tsuzuki et al. 2004; Fischer et al. 2005; Hong et al. 2008; Schindler et al. 2009) the addition of known secondary changes could enable the

establishment of more relevant models of this ALL subtype. Also, detailed knowledge of the secondary changes in individual *ETV6/RUNX1*-positive ALLs would permit studies addressing the sequential acquisition of mutations, if material was available from time points prior to leukemia, as demonstrated by Horsley et al. (2008). The information regarding which genes are perturbed during neoplastic hematopoiesis could also give important insights into the function of these genes during normal hematopoiesis. Importantly, the secondary genetic changes may constitute independent risk factors to consider for the overall prognosis. And lastly, the secondary genetic changes could, hopefully, represent possible drug targets for improved therapies.

Svensk Sammanfattning

Leukemi är ett samlingsbegrepp för sjukdomar där sjuka vita blodkroppar ansamlas i benmärgen. Leukemier indelas vanligen i akuta eller kroniska varianter beroende på de sjuka cellernas mognadsgrad och hastigheten på sjukdomsförloppet. Dessutom benämns leukemier efter vilken typ av celler som ansamlas – lymfatiska eller myeloiska. De lymfatiska leukemierna kan ytterligare delas in beroende på om de lymfatiska cellerna är B- eller T-celler. B-cells akut lymfatisk leukemi är således en snabb ansamling av omogna lymfatiska B-celler. För att framhäva att det just är omogna B-celler som ansamlas kallas denna leukemi även B-cells prekursor akut lymfatisk leukemi (BCP ALL). Denna leukemi är den vanligaste formen av cancer hos barn. Liksom andra typer av cancer orsakas BCP ALL av förvärvade genetiska förändringar i de sjuka cellerna. Närvaron av vissa genetiska förändringar har visat sig ha stor betydelse för prognosen vid BCP ALL. En translokation (utbyte av kromosomalt material) mellan kromosomerna 12 och 21 som ger upphov till fusionsgenen *ETV6/RUNX1* finns i cirka 25% av BCP ALL hos barn. BCP ALL:er med denna fusionsgen uppvisar vanligen en god prognos. Det finns goda belägg för att uppkomsten av denna fusionsgen är en tidig förändring som kräver ytterligare genetiska förändringar innan en leukemi uppstår. Att kartlägga vilka genetiska förändringar som förekommer tillsammans med *ETV6/RUNX1* kan alltså ge viktiga ledtrådar om hur denna leukemi uppstår. Möjligen kan det också ge svar på varför behandlingen av BCP ALL misslyckas i vissa fall. De genetiska förändringarna kan även, i bästa fall, användas som måltavlor för framtida läkemedel mot BCP ALL. Målet med studierna i avhandlingen har varit att med nyutvecklade tekniker kartlägga genetiska förändringar som förekommer tillsammans med *ETV6/RUNX1* i BCP ALL hos barn.

I den första studien undersöktes kopietalsförändringar i 17 *ETV6/RUNX1*-positiva BCP ALL:er från barn med hjälp av array CGH. Denna teknik kan påvisa kopietalsförändringar (förlust eller tillskott av genetiskt material) ned till en storlek av ca 300 000 baspar. Totalt identifierades 45 kopietalsförändringar.

Framförallt identifierades ett tidigare ej beskrivet tillskott av material från den långa armen på kromosom X (benämnd Xq). Denna förändring fanns i sex (35%) av fallen, uteslutande pojkar. Hos pojkar förekom Xq-duplikationen i 55% av fallen. Jämförelser med genuttrycksdata visade att både tillkomst och förlust av genetiskt material ledde till allmänna ökningar och minskningar i genuttryck för gener med förändrat kopietal. I den duplicerade Xq-regionen fanns flera gener vars uttryck var signifikant förhöjt i *ETV6/RUNX1*-positiva BCP ALL:er, och därmed starka kandidater till att vara målgener för duplikationen. Den gen som hade det högsta och mest specifika uttrycket i *ETV6/RUNX1*-positiva BCP ALL:er var *SPANXB* som normalt endast uttrycks i testikelvävnad. Senare studier har visat att genen *HMGB3,* som även den hade ett högt och specifikt uttryck i *ETV6/RUNX1*-positiva BCP ALL:er, ingår i en fusionsgen i akut myeloisk leukemi. Därför bör även denna gen betraktas som en intressant kandidat i det duplicerade Xq-området.

I den andra studien undersöktes 164 *ETV6/RUNX1*-positiva BCP ALL:er med SNP-array, en teknik som kan detektera kopietalsförändringar ned till ca 10 000 baspars storlek. Bland de 164 BCP ALL:er som undersöktes fanns 24 BCP ALL:er för vilka hela SNP-array analysen gjordes av oss; resterande material utgjordes av tidigare publicerade fall från två stora studier gjorda av andra forskargrupper. Totalt kunde 55 förvärvade kopietalsförändringar som förekom i mer än ett fall, det vill säga var återkommande, identifieras. Eftersom en andel av fallen i studien endast studerats med en SNP-arrayplatform med lägre upplösning kunde dock endast 45 förändringar studeras i hela materialet. En majoritet av fallen (74%) hade en helt unik sammansättning av de återkommande kopietalsförändringarna vilket tyder på att processen som föregår leukemin ser olika ut för varje enskilt fall. Den vanligaste typen av förändring som identifierades var förluster av material mindre än 300 000 baspar i storlek. Dessa förluster resulterade oftast i att en eller två gener avlägsnades. Två grupper av gener som ofta drabbades av sådana genförluster var gener som styr B-cellsutvecklingen och gener som styr svaret på hormoner i cellkärnan. En modell för det tidsmässiga förloppet för tillkomsten av kopietalsförändringarna kunde göras med hjälp av en matematisk/statistisk metod benämnd ”onkogenetiska träd”. Enligt denna modell var förlusterna av *ETV6*, *CDKN2A/B*, *PAX5*, och 6q samt tillkomsten

av en extra kopia av kromosom 16 sannolikt tidiga förändringar.

I den tredje studien undersöktes två stycken *ETV6/RUNX1*-positiva BCP ALL:er med exomsekvensering, det vill säga storskalig sekvensering som möjliggör bestämning av varje enskild nukleotid i den del av arvsmassan som uttrycks till protein. Detta motsvarar ca 25 miljoner nukleotider, knappt en procent av DNA:t i en mänsklig cell. I denna studie beskrivs för första gången hur exomsekvenseringstekniken kan användas för att finna mutationer i tumörprov. Efter sekvenseringen jämfördes varje nukleotid med den mänskliga referenssekvensen, kända normalvarianter i en databas (dbSNP) och medfödda varianter hos patienten, för att identifiera förvärvade mutationer unika för de leukemiska cellerna. Mutationskandidaterna bekräftades slutligen med traditionell kapillärsekvensering. Sju nukleotidförändringar kunde identifieras i vardera BCP ALL, vilket innebär att nukleotidförändringar är ungefär lika vanliga som kopietalsförändringar i *ETV6/RUNX1*-positiv BCP ALL. I de två undersökta fallen fanns åtta respektive 14 kopietalsförändringar. Av nukleotidförändringarna visade sig 11 påverka aminosyrasammansättningen i genens proteinprodukt, vilket betyder att funktionen hos proteinet kan förändras eller förstöras. Närvaron av de 14 nukleotidförändringarna undersöktes i 29 *ETV6/RUNX1*-positiva BCP ALL:er, men ingen av dem återfanns i det materialet. En jämförelse med mutationer beskrivna i databasen COSMIC (”Catalogue of Somatic Mutations in Cancer”) visade att sex av de gener som påverkats av nukleotidförändringar tidigare visats vara förändrade i andra tumörtyper. Detta tillsammans med den normala funktionen av de drabbade generna och mutationernas förväntade påverkan på det uttryckta proteinet visade att nukleotidförändringarna sannolikt är viktiga för leukemiuppkomsten i de undersökta fallen.

Sammanfattningsvis har studierna som ingår i denna avhandling gett ny information om det komplexa mönster av genetiska förändringar som bidrar till utvecklingen av en av de vanligaste formerna av barnleukemi, *ETV6/RUNX1*-positiv BCP ALL. Förhoppningsvis kan denna kunskap leda till förbättrad framtida diagnostik och behandling av dessa sjukdomar.

Acknowledgements

I would like to express my sincere gratitude and appreciation to the following:

Thoas Fioretos, my excellent supervisor, for your brilliant scientific mind, your encouragement and your enthusiasm. Thank you for believing in me and for teaching me so much about science. Working with you is truly a pleasure!

My co-supervisors, **Anna Andersson** and **Bertil Johansson**.
Anna for being both an inspiration and a great co-supervisor. Thanks for always being positive and encouraging and for always taking time to read and comment my work and discuss science.
Bertil for generously sharing your great knowledge in everything from cancer genetics to historical physicists. Thank you for teaching me so much about the English language and for inspiring me to think both twice and trice about my wording. I promise to never again split an infinitive!

Jesper Heldrup and **Mikael Behrendtz**, for support and good collaboration.

Felix Mitelman, former head of the department, for creating the excellent and unique research environment at the Department of Clinical Genetics and for always being friendly and supportive.

Ulf Kristoffersson, head of the clinical side of the department, for trusting me with the challenge of establishing clinical protocols for QF-PCR and quantitative realtime-PCR, which ultimately lead to me becoming a PhD student.

Magnus Fontes and **Lotta Sonesson** for reintroducing me to the wonderful world of mathematics and for all the nice lunch/fika/sciencemeetings.

Carin Lassen and **Marianne Rissler**, both exceptionally skilled technicians,

for always being so helpful and positive and fun to be with. A special thanks also to **Margareth**, **Jenny N**, **Andrea B**, and **Linda M** for being so knowledgeable and helpful.

Past and present PhD students and postdocs at the department: **Björn Nilsson**, **Charles Walther**, **Daniel Sehic**, **David Lindgren**, **Emely Möller**, **Gisela Lundberg**, **Hammurabi Bartuma**, **Helena Ågerstam**, **Jenny Karlsson**, **Josef Davidsson**, **Kajsa Paulsson**, **Karolin Hansén Nord**, **Kristina Karrman**, **Linda Holmquist Mengelbier**, **Linda Olsson**, **Marcus Järås**, **Maria Askmyr**, **Nils Hansen**, **Panagiotis Tsapogas**, **Srinivas Veerla**, and **Ylva Stewénius** for all the good times, all the interesting discussions and for great companionship when traveling, both for science and otherwise.

Everybody in the 'DNA-lab' that I had the pleasure to work with during my time as technician and 'sjukhusgenetiker': **Ali A**, **Andrea K**, **Anna C**, **Anna H**, **Emelie**, **Helena S**, **Ingrid RvdW**, **Ingela L**, **Ioannis**, **Maria S**, **Mina**, **Mahtab**, **Petra**, and **Tord**.

The senior research and clinical staff: **Catarina**, **Charlotte**, **David**, **Emma**, **Figge**, **Hans**, **Nisse**, **Nina**, **Markus**, **Mia**, **Samuel**, and **Yuesheng**.

Anette Welin, for solving just about any problem with a smile on your face.

Eva Lindstedt, for always being kind and helpful.

The entire clinical staff.

Everyone at the oncology department and SCIBLU for nice floorball games and for being so generous with your equipment.

My parents, **Inger** and **Sven-Åke**, and my sisters **Charlotta**, **Elisabet**, and **Kristina**, for believing in me and for being so encouraging.

And **Elin**, for everything!

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, and Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* **7**: 248-249.

Andersson A, Olofsson T, Lindgren D, Nilsson B, Ritz C, Edén P, Lassen C, Råde J, Fontes M, Mörse H, Heldrup J, Behrendtz M, Mitelman F, Höglund M, Johansson B, and Fioretos T. 2005. Molecular signatures in childhood acute leukemia and their correlations to expression patterns in normal hematopoietic subpopulations. *Proc Natl Acad Sci U S A* **102**: 19069-19074.

Andreasson P, Schwaller J, Anastasiadou E, Aster J, and Gilliland D. 2001. The expression of *ETV6/CBFA2* (*TEL/AML1*) is not sufficient for the transformation of hematopoietic cell lines in vitro or the induction of hematologic disease in vivo. *Cancer Genet Cytogenet* **130**: 93-104.

Autio R, Hautaniemi S, Kauraniemi P, Yli-Harja O, Astola J, Wolf M, and Kallioniemi A. 2003. CGH-Plotter: MATLAB toolbox for CGH-data analysis. *Bioinformatics* **19**: 1714-1715.

Barabé F, Kennedy JA, Hope KJ, and Dick JE. 2007. Modeling the initiation and progression of human acute leukemia in mice. *Science* **316**: 600-604.

Bardini M, Spinelli R, Bungaro S, Mangano E, Corral L, Cifola I, Fazio G, Giordan M, Basso G, De Rossi G, Biondi A, Battaglia C, and Cazzaniga G. 2009. DNA copy-number abnormalities do not occur in infant ALL with t(4;11)/*MLL-AF4*. *Leukemia* **24**: 169-176.

Bateman CM, Colman SM, Chaplin T, Young BD, Eden TO, Bhakta M, Gratias EJ, van Wering ER, Cazzaniga G, Harrison CJ, Hain R, Ancliff P, Ford AM, Kearney L, and Greaves M. 2010. Acquisition of genome-wide copy

number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood* **115**: 3553-3558.

Bernardin F, Yang Y, Cleaves R, Zahurak M, Cheng L, Civin C, and Friedman A. 2002. TEL-AML1, expressed from t(12;21) in human acute lymphocytic leukemia, induces acute leukemia in mice. *Cancer Res* **62**: 3904-3908.

Boissel N, Auclerc M, Lhéritier V, Perel Y, Thomas X, Leblanc T, Rousselot P, Cayuela J, Gabert J, Fegueux N, Piguet C, Huguet-Rigal F, Berthou C, Boiron J, Pautas C, Michel G, Fière D, Leverger G, Dombret H, and Baruchel A. 2003. Should adolescents with acute lymphoblastic leukemia be treated as old children or young adults? Comparison of the French FRALLE-93 and LALA-94 trials. *J Clin Oncol* **21**: 774-780.

de Bont JM, Holt BVD, Dekker AW, van der Does-van den Berg A, Sonneveld P, and Pieters R. 2004. Significant difference in outcome for adolescents with acute lymphoblastic leukemia treated on pediatric vs adult protocols in the Netherlands. *Leukemia* **18**: 2032-2035.

Carroll AJ, Crist WM, Parmley RT, Roper M, Cooper MD, and Finley WH. 1984. Pre-B cell leukemia associated with chromosome translocation 1;19. *Blood* **63**: 721-724.

Castor A, Nilsson L, Astrand-Grundström I, Buitenhuis M, Ramirez C, Anderson K, Strömbeck B, Garwicz S, Békássy AN, Schmiegelow K, Lausen B, Hokland P, Lehmann S, Juliusson G, Johansson B, and Jacobsen SEW. 2005. Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia. *Nat Med* **11**: 630-637.

Chen MJ, Yokomizo T, Zeigler BM, Dzierzak E, and Speck NA. 2009. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**: 887-891.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu

A, Özen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, and Lifton RP. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* **106**: 19096-19101.

Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, and Schäffer AA. 2000. Distance-based reconstruction of tree models for oncogenesis. *J Comput Biol* **7**: 789-803.

Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, and Schäffer AA. 1999. Inferring tree models for oncogenesis from comparative genome hybridization data. *J Comput Biol* **6**: 37-51.

Druker BJ, Sawyers CL, Kantarjian H, Resta DJ, Reese SF, Ford JM, Capdeville R, and Talpaz M. 2001. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N Engl J Med* **344**: 1038-1042.

Erickson P, Gao J, Chang KS, Look T, Whisenant E, Raimondi S, Lasher R, Trujillo J, Rowley J, and Drabkin H. 1992. Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, *AML1/ETO*, with similarity to Drosophila segmentation gene, runt. *Blood* **80**: 1825-1831.

Fears S, Gavin M, Zhang DE, Hetherington C, Ben-David Y, Rowley JD, and Nucifora G. 1997. Functional characterization of *ETV6* and *ETV6/CBFA2* in the regulation of the *MCSFR* proximal promoter. *Proc Natl Acad Sci U S A* **94**: 1949-1954.

Fenrick R, Amann J, Lutterbach B, Wang L, Westendorf J, Downing J, and Hiebert S. 1999. Both TEL and AML-1 contribute repression domains to the t(12;21) fusion protein. *Mol Cell Biol* **19**: 6566-6574.

Fischer M, Schwieger M, Horn S, Niebuhr B, Ford A, Roscher S, Bergholz U,

Greaves M, Lohler J, and Stocking C. 2005. Defining the oncogenic function of the TEL/AML1 (*ETV6/RUNX1*) fusion protein in a mouse model. *Oncogene* **24**: 7579-7591.

Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, and Stratton MR. 2008. The catalogue of somatic mutations in cancer (COSMIC). *Curr Protoc Hum Genet* **57**: 10.11.1–10.11.26.

Ford A, Bennett C, Price C, Bruin M, Van Wering E, and Greaves M. 1998. Fetal origins of the *TEL-AML1* fusion gene in identical twins with leukemia. *Proc Natl Acad Sci U S A* **95**: 4584-4588.

Forestier E, Andersen M, Autio K, Blennow E, Borgström G, Golovleva I, Heim S, Heinonen K, Hovland R, Johannsson J, Kerndrup G, Nordgren A, Rosenquist R, Swolin B, and Johansson B. 2007. Cytogenetic patterns in *ETV6/RUNX1*-positive pediatric B-cell precursor acute lymphoblastic leukemia: a Nordic series of 245 cases and review of the literature. *Genes Chromosomes Cancer* **46**: 440-450.

Forestier E, and Schmiegelow K. 2006. The incidence peaks of the childhood acute leukemias reflect specific cytogenetic aberrations. *J Pediatr Hematol Oncol* **28**: 486-495.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, and Zhang J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80.

Golub TR, Barker GF, Lovett M, and Gilliland DG. 1994. Fusion of PDGF receptor beta to a novel *ets*-like gene, *tel*, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell* **77**: 307-316.

Golub T, Barker G, Bohlander S, Hiebert S, Ward D, Bray-Ward P, Morgan E, Raimondi S, Rowley J, and Gilliland D. 1995. Fusion of the TEL gene on 12p13 to the *AML1* gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **92**: 4917-4921.

Greaves MF, Maia AT, Wiemels JL, and Ford AM. 2003. Leukemia in twins: lessons in natural history. *Blood* **102**: 2321-2333.

Gustafsson G, Schmiegelow K, Forestier E, Clausen N, Glomstein A, Jonmundsson G, Mellander L, Mäkipernaa A, Nygaard R, and Saarinen-Pihkala U. 2000. Improving outcome through two decades in childhood ALL in the Nordic countries: the impact of high-dose methotrexate in the reduction of CNS irradiation. Nordic Society of Pediatric Haematology and Oncology (NOPHO). *Leukemia* **14**: 2267-2275.

Gylfe AE, Sirkiä J, Ahlsten M, Järvinen H, Mecklin J, Karhu A, and Aaltonen LA. 2010. Somatic mutations and germline sequence variants in patients with familial colorectal cancer. *Int J Cancer* **127**: 2974-2980.

Hiebert S, Sun W, Davis J, Golub T, Shurtleff S, Buijs A, Downing J, Grosveld G, Roussell M, Gilliland D, Lenny N, and Meyers S. 1996. The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. *Mol Cell Biol* **16**: 1349-1355.

Hjalgrim LL, Madsen HO, Melbye M, Jørgensen P, Christiansen M, Andersen MT, Pallisgaard N, Hokland P, Clausen N, Ryder LP, Schmiegelow K, and Hjalgrim H. 2002. Presence of clone-specific markers at birth in children with acute lymphoblastic leukaemia. *Br J Cancer* **87**: 994-999.

Hong D, Gupta R, Ancliff P, Atzberger A, Brown J, Soneji S, Green J, Colman S, Piacibello W, Buckle V, Tsuzuki S, Greaves M, and Enver T. 2008. Initiating and cancer-propagating cells in *TEL-AML1*-associated childhood leukemia. *Science* **319**: 336-339.

Horsley SW, Colman S, McKinley M, Bateman CM, Jenney M, Chaplin T, Young BD, Greaves M, and Kearney L. 2008. Genetic lesions in a preleukemic aplasia phase in a child with acute lymphoblastic leukemia. *Genes Chromosomes Cancer* **47**: 333-340.

Inoue K, Ozaki S, Shiga T, Ito K, Masuda T, Okado N, Iseda T, Kawaguchi S, Ogawa M, Bae S, Yamashita N, Itohara S, Kudo N, and Ito Y. 2002. *Runx3* controls the axonal projection of proprioceptive dorsal root ganglion neurons. *Nat Neurosci* **5**: 946-954.

Jarošová M, Holzerová M, Mihál V, Lakomá I, Divoký V, Blažek B, Pospíšilová D, Hajdúch M, Novák Z, Dušek L, Koptíková J, Poulsen T, and Indrák K. 2003. Complex karyotypes in childhood acute lymphoblastic leukemia: cytogenetic and molecular cytogenetic study of 21 cases. *Cancer Genet Cytogenet* **145**: 161-168.

Kawamata N, Ogawa S, Zimmermann M, Kato M, Sanada M, Hemminki K, Yamatomo G, Nannya Y, Koehler R, Flohr T, Miller CW, Harbott J, Ludwig W, Stanulla M, Schrappe M, Bartram CR, and Koeffler HP. 2008. Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* **111**: 776-784.

Kennedy JA, and Barabé F. 2008. Investigating human leukemogenesis: from cell lines to *in vivo* models of human leukemia. *Leukemia* **22**: 2029-2040.

Kersey JH. 1997. Fifty years of studies of the biology and therapy of childhood leukemia. *Blood* **90**: 4243-4251.

Komori T, Yagi H, Nomura S, Yamaguchi A, Sasaki K, Deguchi K, Shimizu Y, Bronson RT, Gao YH, Inada M, Sato M, Okamoto R, Kitamura Y, Yoshiki S, and Kishimoto T. 1997. Targeted disruption of *Cbfa1* results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* **89**:

755-764.

Kuchinskaya E, Heyman M, Grandér D, Linderholm M, Söderhäll S, Zaritskey A, Nordgren A, Porwit-MacDonald A, Zueva E, Pawitan Y, Corcoran M, Nordenskjöld M, and Blennow E. 2005. Children and adults with acute lymphoblastic leukaemia have similar gene expression profiles. *Eur J Haematol* **74**: 466-480.

Kuiper RP, Schoenmakers EFPM, van Reijmersdal SV, Hehir-Kwa JY, van Kessel AG, van Leeuwen FN, and Hoogerbrugge PM. 2007. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia* **21**: 1258-1266.

Lacronique V, Boureux A, Valle VD, Poirel H, Quang CT, Mauchauffé M, Berthou C, Lessard M, Berger R, Ghysdael J, and Bernard OA. 1997. A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. *Science* **278**: 1309-1312.

Lausten-Thomsen U, Madsen HO, Vestergaard TR, Hjalgrim H, Nersting J, and Schmiegelow K. 2011. Prevalence of t(12;21)[*ETV6-RUNX1*]-positive cells in healthy neonates. *Blood* **117**: 186-189.

Levanon D, Bettoun D, Harris-Cerruti C, Woolf E, Negreanu V, Eilam R, Bernstein Y, Goldenberg D, Xiao C, Fliegauf M, Kremer E, Otto F, Brenner O, Lev-Tov A, and Groner Y. 2002. The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. *EMBO J* **21**: 3454-3463.

Li H, and Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R. 2009. The Sequence Alignment/Map format and

SAMtools. *Bioinformatics* **25**: 2078-2079.

Lin H, and Grosschedl R. 1995. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**: 263-267.

Lin M, Wei L, Sellers WR, Lieberfarb M, Wong WH, and Li C. 2004. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**: 1233-1240.

Lu X, Harris C, Cooley L, Margolin J, Steuber P, Sheldon M, Rao P, and Lau C. 2002. The utility of spectral karyotyping in the cytogenetic analysis of newly diagnosed pediatric acute lymphoblastic leukemia. *Leukemia* **16**: 2222-2227.

Ma C, Quesnelle KM, Sparano A, Rao S, Park MS, Cohen MA, Wang Y, Samanta M, Kumar MS, Aziz MU, Naylor TL, Weber BL, Fakharzadeh SS, Weinstein GS, Vachani A, Feldman MD, and Brose MS. 2009. Characterization *CSMD1* in a large set of primary lung, head and neck, breast and skin cancer tissues. *Cancer Biol Ther* **8**: 907-916.

Maia AT, Ford AM, Jalali GR, Harrison CJ, Taylor GM, Eden OB, and Greaves MF. 2001. Molecular tracking of leukemogenesis in a triplet pregnancy. *Blood* **98**: 478-482.

Mardis ER. 2010. Cancer genomics identifies determinants of tumor biology. *Genome Biol* **11**: 211.

Martineau M, Jalali G, Barber K, Broadfield Z, Cheung K, Lilleyman J, Moorman A, Richards S, Robinson H, Ross F, and Harrison C. 2005. *ETV6/RUNX1* fusion at diagnosis and relapse: some prognostic indications. *Genes Chromosomes Cancer* **43**: 54-71.

Ma S, Wan T, Cheuk A, Fung L, Chan G, Chan S, Ha S, and Chan L. 2001. Characterization of additional genetic events in childhood acute lymphoblastic

leukemia with *TEL/AML1* gene fusion: a molecular cytogenetics study. *Leukemia* **15**: 1442-1447.

McHale CM, Wiemels JL, Zhang L, Ma X, Buffler PA, Guo W, Loh ML, and Smith MT. 2003. Prenatal origin of *TEL-AML1*-positive acute lymphoblastic leukemia in children born in California. *Genes Chromosomes Cancer* **37**: 36-43.

McLean T, Ringold S, Neuberg D, Stegmaier K, Tantravahi R, Ritz J, Koeffler H, Takeuchi S, Janssen J, Seriu T, Bartram C, Sallan S, Gilliland D, and Golub T. 1996. TEL/AML-1 dimerizes and is associated with a favorable outcome in childhood acute lymphoblastic leukemia. *Blood* **88**: 4252-4258.

Melo JV. 1996. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. *Blood* **88**: 2375-2384.

Meyers S, Downing JR, and Hiebert SW. 1993. Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: the *runt* homology domain is required for DNA binding and protein-protein interactions. *Mol Cell Biol* **13**: 6336-6345.

Milani L, Lundmark A, Kiialainen A, Nordlund J, Flaegstad T, Forestier E, Heyman M, Jonmundsson G, Kanerva J, Schmiegelow K, Söderhäll S, Gustafsson MG, Lönnerholm G, and Syvänen A. 2010. DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood* **115**: 1214-1225.

Mori H, Colman S, Xiao Z, Ford A, Healy L, Donaldson C, Hows J, Navarrete C, and Greaves M. 2002. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A* **99**: 8242-8247.

Morrow M, Samanta A, Kioussis D, Brady HJM, and Williams O. 2007. TEL-AML1 preleukemic activity requires the DNA binding domain of AML1 and the dimerization and corepressor binding domains of TEL. *Oncogene* **26**:

4404-4414.

Morrow M, Horton S, Kioussis D, Brady HJM, and Williams O. 2004. TEL-AML1 promotes development of specific hematopoietic lineages consistent with preleukemic activity. *Blood* **103**: 3890-3896.

Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, Su X, Pui C, Relling MV, Evans WE, Shurtleff SA, and Downing JR. 2007. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**: 758-764.

Mullighan CG, Miller CB, Radtke I, Phillips LA, Dalton J, Ma J, White D, Hughes TP, Le Beau MM, Pui C, Relling MV, Shurtleff SA, and Downing JR. 2008. *BCR-ABL1* lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453**: 110-114.

Nemeth M, Kirby M, and Bodine D. 2006. Hmgb3 regulates the balance between hematopoietic stem cell self-renewal and differentiation. *Proc Natl Acad Sci U S A* **103**: 13783-13788.

Ng PC, and Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**: 3812-3814.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, and Bamshad MJ. 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**: 30-35.

Nordgren A, Heyman M, Sahlén S, Schoumans J, Söderhäll S, Nordenskjöld M, and Blennow E. 2002. Spectral karyotyping and interphase FISH reveal abnormalities not detected by conventional G-banding. Implications for treatment stratification of childhood acute lymphoblastic leukaemia: detailed analysis of 70 cases. *Eur J Haematol* **68**: 31-41.

Novara F, Beri S, Bernardo ME, Bellazzi R, Malovini A, Ciccone R, Cometa AM, Locatelli F, Giorda R, and Zuffardi O. 2009. Different molecular mechanisms causing 9p21 deletions in acute lymphoblastic leukemia of childhood. *Hum Genet* **126**: 511-520.

Nutt SL, Heavey B, Rolink AG, and Busslinger M. 1999. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**: 556-562.

Okuda T, van Deursen J, Hiebert SW, Grosveld G, and Downing JR. 1996. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* **84**: 321-330.

Otto F, Thornell AP, Crompton T, Denzel A, Gilmour KC, Rosewell IR, Stamp GW, Beddington RS, Mundlos S, Olsen BR, Selby PB, and Owen MJ. 1997. *Cbfa1*, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* **89**: 765-771.

Ouhtit A, Gaur RL, Abd Elmageed ZY, Fernando A, Thouta R, Trappey AK, Abdraboh ME, El-Sayyad HI, Rao P, and Raj MGH. 2009. Towards understanding the mode of action of the multifaceted cell adhesion receptor CD146. *Biochim Biophys Acta* **1795**: 130-136.

Parker H, An Q, Barber K, Case M, Davies T, Konn Z, Stewart A, Wright S, Griffiths M, Ross FM, Moorman AV, Hall AG, Irving JA, Harrison CJ, and Strefford JC. 2008. The complex genomic profile of *ETV6-RUNX1* positive acute lymphoblastic leukemia highlights a recurrent deletion of *TBL1XR1*. *Genes Chromosomes Cancer* **47**: 1118-1125.

Paulsson K, Forestier E, Lilljebjörn H, Heldrup J, Behrendtz M, Young BD, and Johansson B. 2010. Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **107**: 21719-21724.

Paulsson K, Horvat A, Fioretos T, Mitelman F, and Johansson B. 2005.

Formation of der(19)t(1;19)(q23;p13) in acute lymphoblastic leukemia. *Genes Chromosomes Cancer* **42**: 144-148.

Paulsson K, and Johansson B. 2009. High hyperdiploid childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer* **48**: 637-660.

Peeters P, Raynaud SD, Cools J, Wlodarska I, Grosgeorge J, Philip P, Monpoux F, Van Rompaey L, Baens M, Van den Berghe H, and Marynen P. 1997. Fusion of *TEL,* the ETS-variant gene 6 (*ETV6*), to the receptor-associated kinase *JAK2* as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia. *Blood* **90**: 2535-2540.

Petit A, Ragu C, Della-Valle V, Mozziconacci MJ, Lafage-Pochitaloff M, Soler G, Schluth C, Radford I, Ottolenghi C, Bernard OA, Penard-Lacronique V, and Romana SP. 2010. NUP98-HMGB3: a novel oncogenic fusion. *Leukemia* **24**: 654-658.

Pui CH, Behm FG, and Crist WM. 1993. Clinical and biologic relevance of immunologic marker studies in childhood acute lymphoblastic leukemia. *Blood* **82**: 343-362.

Pui CH, and Evans WE. 1998. Acute lymphoblastic leukemia. *N Engl J Med* **339**: 605-615.

Pui CH, Kane JR, and Crist WM. 1995. Biology and treatment of infant leukemias. *Leukemia* **9**: 762-769.

Pui C, Chessells JM, Camitta B, Baruchel A, Biondi A, Boyett JM, Carroll A, Eden OB, Evans WE, Gadner H, Harbott J, Harms DO, Harrison CJ, Harrison PL, Heerema N, Janka-Schaub G, Kamps W, Masera G, Pullen J, Raimondi SC, Richards S, Riehm H, Sallan S, Sather H, Shuster J, Silverman LB, Valsecchi MG, Vilmer E, Zhou Y, Gaynon PS, and Schrappe M. 2003. Clinical heterogeneity in childhood acute lymphoblastic leukemia with 11q23 rearrangements. *Leukemia* **17**: 700-706.

Pui C, and Evans WE. 2006. Treatment of acute lymphoblastic leukemia. *N Engl J Med* **354**: 166-178.

Pui C, Robison LL, and Look AT. 2008. Acute lymphoblastic leukaemia. *Lancet* **371**: 1030-1043.

Romana SP, Le Coniat M, and Berger R. 1994. t(12;21): a new recurrent translocation in acute lymphoblastic leukemia. *Genes Chromosomes Cancer* **9**: 186-191.

Romana S, Mauchauffé M, Le Coniat M, Chumakov I, Le Paslier D, Berger R, and Bernard O. 1995. The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. *Blood* **85**: 3662-3670.

Ross M, Zhou X, Song G, Shurtleff S, Girtman K, Williams W, Liu H, Mahfouz R, Raimondi S, Lenny N, Patel A, and Downing J. 2003. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* **102**: 2951-2959.

Roudaia L, Cheney MD, Manuylova E, Chen W, Morrow M, Park S, Lee C, Kaur P, Williams O, Bushweller JH, and Speck NA. 2009. CBF{beta} is critical for AML1-ETO and TEL-AML1 activity. *Blood* **113**: 3070-3079.

Saal L, Troein C, Vallon-Christersson J, Gruvberger S, Borg Å, and Peterson C. 2002. BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. *Genome Biol* **3**: SOFTWARE0003.

Sabaawy HE, Azuma M, Embree LJ, Tsai H, Starost MF, and Hickstein DD. 2006. *TEL-AML1* transgenic zebrafish model of precursor B cell acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **103**: 15166-15171.

Sasaki K, Yagi H, Bronson RT, Tominaga K, Matsunashi T, Deguchi K, Tani Y,

Kishimoto T, and Komori T. 1996. Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor β. *Proc Natl Acad Sci U S A* **93**: 12359-12363.

Schindler JW, Van Buren D, Foudi A, Krejci O, Qin J, Orkin SH, and Hock H. 2009. TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia. *Cell Stem Cell* **5**: 43-53.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, and Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**: 308-311.

Song WJ, Sullivan MG, Legare RD, Hutchings S, Tan X, Kufrin D, Ratajczak J, Resende IC, Haworth C, Hock R, Loh M, Felix C, Roy DC, Busque L, Kurnit D, Willman C, Gewirtz AM, Speck NA, Bushweller JH, Li FP, Gardiner K, Poncz M, Maris JM, and Gilliland DG. 1999. Haploinsufficiency of *CBFA2* causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet* **23**: 166-175.

Stams W, Beverloo H, den Boer M, de Menezes R, Stigter R, van Drunen E, Ramakers-van-Woerden N, Loonen A, van Wering E, Janka-Schaub G, and Pieters R. 2006. Incidence of additional genetic changes in the *TEL* and *AML1* genes in DCOG and COALL-treated t(12;21)-positive pediatric ALL, and their relation with drug sensitivity and clinical outcome. *Leukemia* **20**: 410-416.

Stirewalt DL, and Radich JP. 2003. The role of FLT3 in haematopoietic malignancies. *Nat Rev Cancer* **3**: 650-665.

Teuffel O, Betts DR, Dettling M, Schaub R, Schäfer BW, and Niggli FK. 2004. Prenatal origin of separate evolution of leukemia in identical twins. *Leukemia* **18**: 1624-1629.

Third International Workshop on Chromosomes in Leukemia. 1983.

Chromosomal abnormalities and their clinical significance in acute lymphoblastic leukemia. *Cancer Res* **43**: 868-873.

Thomas DA, Faderl S, Cortes J, O'Brien S, Giles FJ, Kornblau SM, Garcia-Manero G, Keating MJ, Andreeff M, Jeha S, Beran M, Verstovsek S, Pierce S, Letvak L, Salvado A, Champlin R, Talpaz M, and Kantarjian H. 2004. Treatment of Philadelphia chromosome-positive acute lymphocytic leukemia with hyper-CVAD and imatinib mesylate. *Blood* **103**: 4396-4407.

Towatari M, Yanada M, Usui N, Takeuchi J, Sugiura I, Takeuchi M, Yagasaki F, Kawai Y, Miyawaki S, Ohtake S, Jinnai I, Matsuo K, Naoe T, and Ohno R. 2004. Combination of intensive chemotherapy and imatinib can rapidly induce high-quality complete remission for a majority of patients with newly diagnosed *BCR-ABL*-positive acute lymphoblastic leukemia. *Blood* **104**: 3507-3512.

Tsuzuki S, Karnan S, Horibe K, Matsumoto K, Kato K, Inukai T, Goi K, Sugita K, Nakazawa S, Kasugai Y, Ueda R, and Seto M. 2007. Genetic abnormalities involved in t(12;21) *TEL-AML1* acute lymphoblastic leukemia: analysis by means of array-based comparative genomic hybridization. *Cancer Sci* **98**: 698-706.

Tsuzuki S, Seto M, Greaves M, and Enver T. 2004. Modeling first-hit functions of the t(12;21) *TEL-AML1* translocation in mice. *Proc Natl Acad Sci U S A* **101**: 8443-8448.

Uchida H, Downing JR, Miyazaki Y, Frank R, Zhang J, and Nimer SD. 1999. Three distinct domains in TEL-AML1 are required for transcriptional repression of the IL-3 promoter. *Oncogene* **18**: 1015-1022.

Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR, and Orkin SH. 1997. Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *EMBO J* **16**: 4374-4383.

Wang LC, Swat W, Fujiwara Y, Davidson L, Visvader J, Kuo F, Alt FW, Gilliland DG, Golub TR, and Orkin SH. 1998. The *TEL/ETV6* gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev* **12**: 2392-2402.

Wang Q, Stacy T, Binder M, Marin-Padilla M, Sharpe AH, and Speck NA. 1996a. Disruption of the *Cbfa2* gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc Natl Acad Sci U S A* **93**: 3444-3449.

Wang Q, Stacy T, Miller JD, Lewis AF, Gu T, Huang X, Bushweller JH, Bories J, Alt FW, Ryan G, Liu PP, Wynshaw-Boris A, Binder M, Marín-Padilla M, Sharpe AH, and Speck NA. 1996b. The CBFβ Subunit Is Essential for CBFα2 (AML1) Function In Vivo. *Cell* **87**: 697-708.

Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, Harris NL, Le Beau MM, Hellstrom-Lindberg E, Tefferi A, and Bloomfield CD. 2009. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**: 937-951.

Venkatraman ES, and Olshen AB. 2007. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**: 657-663.

Wiemels JL, Ford AM, Van Wering ER, Postma A, and Greaves M. 1999a. Protracted and variable latency of acute lymphoblastic leukemia after *TEL-AML1* gene fusion in utero. *Blood* **94**: 1057-1062.

Wiemels J, Cazzaniga G, Daniotti M, Eden O, Addison G, Masera G, Saha V, Biondi A, and Greaves M. 1999b. Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* **354**: 1499-1503.

Virchow R. 1856. *Die Leukämie. Die Gesammelte Abhandlungen zur Wissenschaftlichen Medizin*. Medinger Sohn, Frankfurt, Germany.

Yang Y, Dudoit S, Luu P, Lin D, Peng V, Ngai J, and Speed T. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**: e15.

Yeoh E, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling MV, Patel A, Cheng C, Campana D, Wilkins D, Zhou X, Li J, Liu H, Pui C, Evans WE, Naeve C, Wong L, and Downing JR. 2002. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1**: 133-143.

Zhuang Y, Cheng P, and Weintraub H. 1996. B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, *E2A*, *E2-2*, and *HEB*. *Mol Cell Biol* **16**: 2898-2905.

Zuna J, Madzo J, Krejci O, Zemanova Z, Kalinova M, Muzikova K, Zapotocky M, Starkova J, Hrusak O, Horak J, and Trka J. 2011. *ETV6/RUNX1* (*TEL/AML1*) is a frequent prenatal first hit in childhood leukemia. *Blood* **117**: 368-369.

LUND UNIVERSITY
Faculty of Medicine

ISSN 1652-8220
ISBN 978-91-86671-67-9
LUND UNIVERSITY, FACULTY OF MEDICINE DOCTORAL DISSERTATION SERIES 2011:18