Genetic Characterization of Pediatric T-cell Acute Lymphoblastic Leukemia

Karrman, Kristina

2014

Genetic Characterization of Pediatric T-cell Acute Lymphoblastic Leukemia

Kristina Karrman

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.

To be defended in Belfragesalen, BMC, Lund.

On Friday November 14, 2014, at 9:30 AM.

Faculty opponent

Professor Frank Speleman

Organization	Document name
LUND UNIVERSITY	Doctoral Dissertation
Division of Clinical Genetics	Date of issue November 14, 2014
Author(s) Kristina Karrman	Sponsoring organization

Title and subtitle **Genetic characterization of pediatric T-cell acute lymphoblastic leukemia**

Abstract

The aim of my thesis has been to characterize genetically pediatric T-cell acute lymphoblastic leukemia (T-ALL).
Articles I and **II** focus on molecular characterization of translocations involving T-cell receptor (TCR) loci. These types of aberration are characteristic for T-ALL and have previously proved pivotal in the identification of genes implicated in leukemogenesis. The translocation t(12;14)(p13;q11) was shown to result in overexpression of *CCND2*. The t(12;14) is the first neoplasia-associated translocation shown to result in overexpression of *CCND2* and the first example of a targeted deregulation of a member of the cyclin-encoding gene family in T-ALL. Cyclin D proteins are crucial to the cell cycle machinery and hence potential oncogenes. The second translocation cloned, t(X;7)(q22;q34), had not previously been reported in a neoplastic disorder. Breakpoint analysis revealed *IRS4* as a novel translocation partner to a TCR locus, resulting in deregulated IRS4 expression, both at the gene and protein level. IRS4 plays an important part in several intracellular signalling cascades, including PI3K-AKT, known to be activated in T-ALL. In a subsequent work, I showed that *IRS4* can also be targeted by alternative mechanisms in T-ALL, apart from TCR translocations, namely by mutations (**Article IV**).
In **Article III**, clinical characteristics and cytogenetic aberrations were ascertained and reviewed in a large, population-based Nordic series of 285 pediatric T-ALLs. Survival analyses revealed a correlation between rare TCR translocations and inferior outcome, an association that awaits confirmation in a separate study.
Finally, I used several different techniques – fluorescence *in situ* hybridization, single nucleotide polymorphism (SNP) array, and deep sequencing of 75 selected candidate genes – to characterize co-operative genetic aberrations in a consecutive series of paediatric T-ALL (**Article V**). One common change identified by SNP array was segemtal uniparental isodisomy (sUPID). This aberration was seen in 44% of the investigated cases, with most being sUPID9p that always were associated with homozygous *CDKN2A* deletions, with a heterozygous deletion occurring prior to the sUPID9p in all instances. Among the 75 genes investigated by deep sequencing, 14 were mutated in 28 cases. The genes targeted are involved in signalling transduction, epigenetic regulation, and transcription. In some cases, *NOTCH1* mutations were seen in minor subclones and lost at relapse, showing that such mutations also can be secondary events. These findings support a multistep leukemogenic process in pediatric T-ALL.
In summary, through different approaches and by various methods, the articles included in this thesis have deciphered genetic aberrations in pediatric T-ALL, contributing to a better understanding of leukemogenesis.

Key words pediatric, T-cell, acute lymphoblastic leukemia, genetic characterization

Classification system and/or index terms (if any)

Supplementary bibliographical information	Language
	English

ISSN and key title: 1652-8220		ISBN 978-91-7619-049-4
Recipient's notes	Number of pages 170	Price
	Security classification	

Signature _____ Date October 3, 2014_____

Genetic Characterization of Pediatric T-cell Acute Lymphoblastic Leukemia

Kristina Karrman

LUND
UNIVERSITY

Lund University, Faculty of Medicine, and Division of Clinical Genetics
ISBN 978-91-7619-049-4
ISSN 1652-8220
Faculty of Medicine Doctoral Dissertation Series 2014:120

Printed in Sweden by Media-Tryck, Lund University, Lund 2014

Contents

Preface

One hundred years ago, in 1914, Boveri's article "Concerning the Origin of Malignant Tumors" was published, in which he proposed that tumors arise as a consequence of incorrect chromosome combinations. Boveri's prescient ideas have since been confirmed, and the efforts of many subsequent scientists have resulted in major contributions to the cancer genetic landscape. However, the map still contains unexplored territories. Today, novel molecular methods give us unprecedented possibilities to investigate the genetic code and further explore and explain the field of cancer genetics.

Leukemia is the most common childhood malignancy, of which T-cell acute lymphoblastic leukemia (T-ALL) comprises a minor part. Genetic characterization has revealed several characteristic, leukemia-associated aberrations in T-ALL. Nevertheless, not all T-ALL cases can be classified by our current genetic knowledge. This implies that there are still unrecognized genetic changes important for T-cell oncogenesis. The aim of my thesis has been to characterize genetically pediatric T-ALL.

This thesis is divided into three sections. The first part provides a biological background to T-ALL, introducing the thymus and the T-cell. The second part is mainly focused on the clinical features, but will also address the epidemiological and etiological aspects of pediatric T-ALL. In the third and final part, I review the current cancer genetic landscape of T-ALL and I point out the fields to which my research work has contributed.

Lund, October 2014

Articles included in the thesis

This thesis is based on the following articles, which will be referred to in the text by their Roman numerals (Articles I-V):

I. Karrman K, Andersson A, Björgvinsdóttir H, Strömbeck B, Lassen C, Olofsson T, Nguyen-Khac F, Berger R, Bernard O, Fioretos T, and Johansson B. (2006) Deregulation of cyclin D2 by juxtaposition with T-cell receptor alpha/delta locus in t(12;14)(p13;q11)-positive childhood T-cell acute lymphoblastic leukemia. *Eur J Haematol* 77:27-34.

II. Karrman K, Kjeldsen E, Lassen C, Isaksson M, Davidsson J, Andersson A, Hasle H, Fioretos T, and Johansson B. (2009) The t(X;7)(q22;q34) in paediatric T-cell acute lymphoblastic leukaemia results in overexpression of the insulin receptor substrate 4 gene through illegitimate recombination with the T-cell receptor beta locus. *Br J Haematol* 144:546-551.

III. Karrman K, Forestier E, Heyman M, Andersen MK, Autio K, Blennow E, Borgström G, Ehrencrona H, Golovleva I, Heim S, Heinonen K, Hovland R, Johannsson JH, Kerndrup G, Nordgren A, Palmqvist L, and Johansson B. (2009) Clinical and cytogenetic features of a population-based consecutive series of 285 pediatric T-cell acute lymphoblastic leukemias: rare T-cell receptor gene rearrangements are associated with poor outcome. *Genes Chromosomes Cancer* 48:795-805.

IV. Karrman K, Isaksson M, Paulsson K, and Johansson B. (2011) The insulin receptor substrate 4 gene (*IRS4*) is mutated in paediatric T-cell acute lymphoblastic leukaemia. *Br J Haematol* 155:516-519.

V. Karrman K, Castor A, Behrendtz M, Forestier E, Olsson L, Ehinger M, Biloglav A, Fioretos T, Paulsson K, and Johansson B. (2014) Comprehensive genetic characterisation of paediatric T-cell acute lymphoblastic leukaemia. *Submitted*.

Abbreviations

ALL	Acute lymphoblastic leukemia
BAC	Bacterial artificial chromosome
BCP	B-cell precursor
bHLH	Basic helix-loop-helix
BM	Bone marrow
C	Constant
CD	Cluster of differentiation
CIMP	CpG island methylator phenotype
CGH	Comparative genomic hybridization
CNAs	Copy number alterations
CNS	Central nervous system
CpG	Cytosine-phosphate-guanine
D	Diversity
DN	Double negative
DNMTs	DNA methyltransferases
DP	Double positive
DSBs	Double strand breaks
ETP	Early T-cell precursor
ETPs	Early thymic progenitors
FISH	Fluorescence *in situ* hybridization
HD	Heterodimerization domain
HLA	Human leukocyte antigen
HR	High risk
iAMP21	Intrachromosomal amplification of chromosome 21

Ig	Immunoglobulin
IR	Intermediate risk
J	Joining
LMO	LIM-domain-only
LOH	Loss of heterozygosity
MHC	Major histocompatibility complex
MRD	Minimal residual disease
NGS	Next-generation sequencing
NOPHO	Nordic Society of Paediatric Haematology and Oncology
pEFS	Probability of event-free survival
PEST	Proline (P), glutamic acid (E), serine (S), and threonine (T) rich
pOS	Probability of overall survival
PRC2	Polycomb repressive complex 2
RNA-seq	RNA sequencing
RQ-PCR	Real-time quantitative polymerase chain reaction
RT-PCR	Reverse transcriptase polymerase chain reaction
SNP	Single nucleotide polymorphism
SP	Single positive
SR	Standard risk
sUPID	Segmental uniparental isodisomy
T-ALL	T-cell acute lymphoblastic leukemia
TCR	T-cell receptor
TdT	Terminal deoxynucleotidyl transferases
T-LBL	T-cell lymphoblastic lymphoma
UPID	Uniparental isodisomy
V	Variable
WBC	White blood cell
wUPID	Whole chromosome uniparental isodisomy

Biological background

The thymus

The thymus is an organ in the mediastinum, located right behind the sternum. During the early years of life, the thymus is prominent, but with increasing age a loss of tissue mass and structure takes place, a process called involution (Shanley *et al.*, 2009). The ancient Greeks knew about this organ – the name is believed to derive either from the Greek word thumos, roughly meaning spirit, soul, and courage, or from the plant Thymus vulgaris, possibly due to a vague resemblance of the thymus to the leaf of the thyme (Lavini *et al.*, 2008). However, the function of the thymus remained elusive for centuries. It was not until the 1960s that it was shown that the thymus plays a major role in lymphopoiesis and that it has important immunological functions (Miller, 1961). It is now well known that the thymus orchestrates the complex symphony of the T-cell differentiation that results in functional, self-tolerant T-cells (Miller, 2011).

The T-cell

T-cell differentiation

The thymus is dependent upon input of progenitor cells from the bone marrow (BM). The intrathymic differentiation process is depicted in Figure 1. Early thymic progenitors (ETPs) are the most immature thymocytes, retaining both B-cell and myeloid lineage potential; however, this potential is gradually lost as differentiation proceeds (Luc *et al.*, 2012). The presence or absence of specific surface markers, termed cluster of differentiation (CD), characterizes the maturing cells. The earliest thymocytes are double negative (DN) with regard to the absence of both CD4 and CD8. The DN cells proliferate and migrate from the cortico-medullary junction and make a loop through the outer cortex during which the microenvironment is crucial for the maturation process. One major input to sustain T-cell development and early differentiation is the interaction between NOTCH1, a transmembrane receptor, and its cell bound ligands (Delta-like 1, 3, and 4; and Jagged 1 and 2).

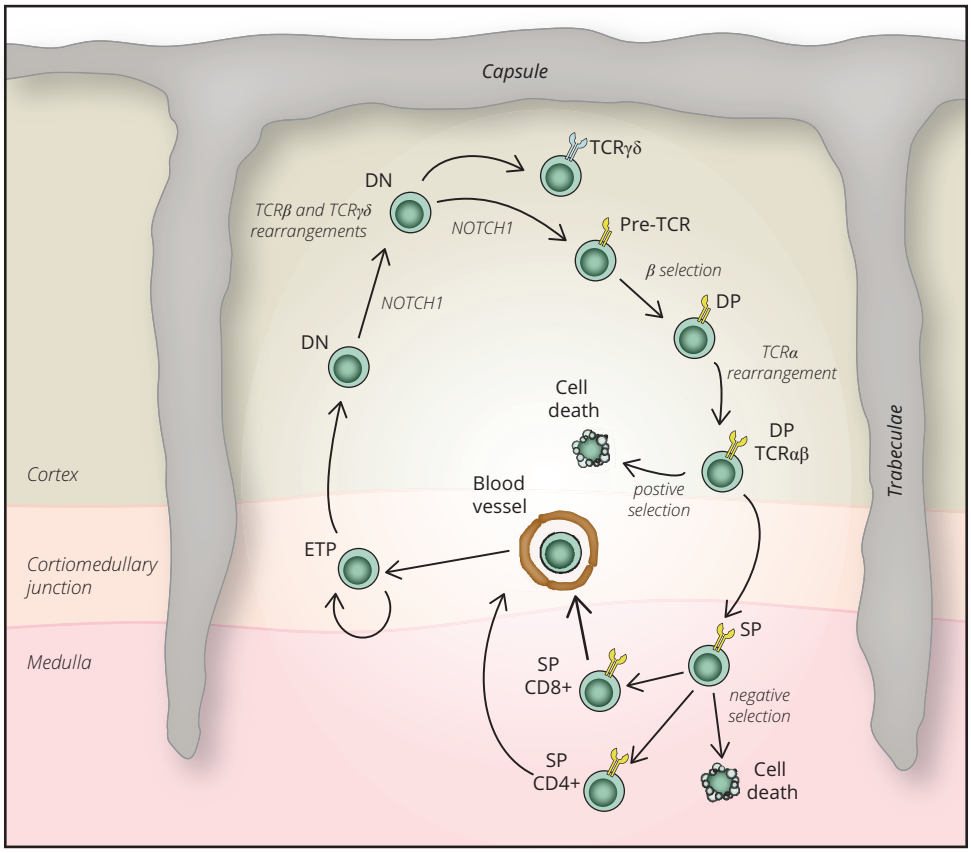


Figure 1. The intrathymic pathway followed by T-cells during the maturation process.
Early thymic progenitors (ETPs) enter the thymus through vessels in the corticomedullary junction. Through their thymic circuit, T-cells will undergo distinct developmental stages, depicted by expression of characteristic sets of surface markers and T-cell receptor (TCR) molecules. As immigrants move outward towards the cortex, the double negative (DN) cells will initiate TCR rearrangements. Expression of pre-TCR takes place in the outmost periphery of the cortex. T-cells that pass the so called β-seletion will transform to double positive (DP) cells and subsequently express the final TCR αβ at the cell surface. Stringent criteria are applied to ensure sensitivity and specificity of the newly assembled TCRs. The T-cells are exposed to the individual's unique major histocompatibility complex (MHC) proteins and only T-cells with TCRs that bind to self MHC with an appropriate affinity will survive, *i.e.*, positive selection. Cells that pass this check point will transform to single postive (SP) cells and migrate to the thymic medulla. In this part of the thymus, there are cells that display self antigens. Negative selection eliminates cells that bind to a self antigen. Eventually, after a strenuous differentiation process eliminating 95% of the cells, mature T-cells are ready to leave the thymus and to start their immunological mission. Graphics adapted from Aifantis *et al.* (2008) and Rothenberg *et al.* (2008).

The T-cell receptor (TCR) genes are then rearranged in the DN cells in a strictly ordered manner, resulting in either αβ or γδ TCRs (see below). The majority of the cells express αβ TCRs, which are assembled in a stepwise fashion. First, the TCR β

protein is expressed and incorporated in a prototype receptor, *i.e.*, pre-TCR. Thymocytes with a functional pre-TCR pass the so called β-selection that enables them to differentiate further. Second, the TCR α protein is transcribed, after which mature αβ TCR complexes can be assembled and presented on the cell surface. While the cells express pre-TCR, upregulation of both CD4 and CD8 results in double positive (DP) cells. The αβ T-cells can only recognize antigens when they are presented by a cell bound to major histocompatibility complex (MHC) proteins. MHC proteins, called human leukocyte antigen (HLA) in humans, are able to present a very large number of peptides on the cell surface. The TCRs associate with CD3 in complexes optimal for MHC recognition, in conjunction with the co-receptors CD4 or CD8. MHC classes I and II are structurally very similar but are expressed differentially: class I is present on virtually all nucleated cells and associates with CD8, whereas class II is expressed mainly by macrophages, dendritic cells, and B-cells and is associated with CD4. Cortical thymic epithelial cells express the individual's unique MHC proteins together with peptides. Only the DP cells with TCRs that bind to a peptide-MHC complex with an appropriate affinity will survive, a process called positive selection. Depending on if the TCR recognizes MHC class I or class II proteins, the T-cell either retains CD8 or CD4, respectively, and hence transforms to a single positive (SP) cell. The SP cells migrate towards the central medulla where they are exposed to epithelial cells expressing MHC-self-antigen complexes. Negative selection functions to destroy SP cells that bind with high affinity to self-antigens; this process reduces autoreactivity. In short, the development of ETPs into mature T-cells is an intricate process where complicated somatic rearrangements of the TCR genes have to be completed and the functionality and tolerance of the receptor scrutinized by positive and negative selection. Less than 5% of the thymocytes generated in the thymus leave as mature T-cells; this loss of more than 95% of cells is the price to pay for an effective, self-tolerant immune defence (Egerton *et al.*, 1990; Aifantis *et al.*, 2008; Rothenberg *et al.*, 2008; Miller, 2011).

T-cell receptors

The function of the adaptive immune system depends on the T-cells' ability to produce an astonishing variety of antigen receptors – the TCRs (Tonegawa, 1983; Hedrick *et al.*, 1984) (Figure 2). There are four TCR genes: *TRA* (located at 14q11, α protein), *TRB* (7q34, β protein), *TRD* (14q11, δ protein), and *TRG* (7p14, γ protein); the *TRD* locus maps within the *TRA* locus. A functional TCR is a heterodimer consisting either of αβ or γδ chains, and each T-cell expresses approximately 30,000 identical TCRs on the cell surface. The TCR genes are composed of several variable (V), diversity (D; only present in *TRB* and *TRD*), joining (J), and constant (C) segments. Together, *TRA* and *TRB* contain around 190 functional gene segments, whereas *TRD* and *TRG* consists of considerably fewer segments (Nishana & Raghavan, 2012). During T-cell differentiation, V(D)J

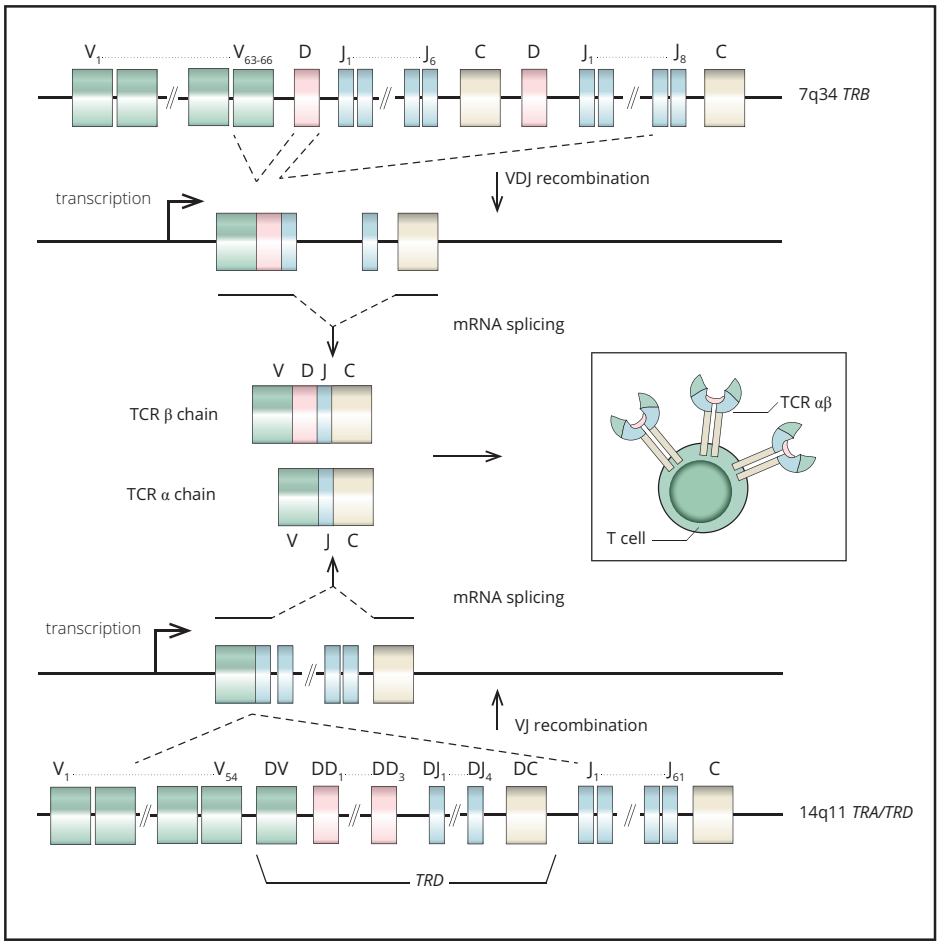


Figure 2. T-cell receptor (TCR) structures are assembled from multiple separate gene segments. The TCR genes *TRB*, *TRA,* and *TRD* (nestled within *TRA*) consist of several different segments, denoted variable (V), diversity (D), joining (J), and constant (C). *TRG* is not grapically depicted but is organized as the other TCR genes. In the process of *TRA* recombination, *TRD* will be excised due to its intragenic position. The combinatorial possibilities created by V(D)J recombination are astonishing and are a cornerstone in our immune defence. The numbers of the different segments were ascertained from Atlas of Genetics and Cytogenetics in Oncology and Haematology (http://atlasgeneticsoncology.org/). Graphics adapted from Nemazee (2006).

recombination takes place, a process strictly regulated as to lineage and developmental stage. In addition, the recombination follows a hierarchical order with regard to the TCR genes and their segments. The recombination process results in the joining of a specific V to a J segment and, in *TRB* and *TRD*, to a D segment (Figure 2). The combinatorial possibilities of V(D)J joining are the backbone of the diversity of the TCRs. V(D)J recombination is dependent upon the recombination activating genes *RAG1* and *RAG2* that encode the RAG proteins. These proteins bind to specific

recombination signal sequences immediately adjacent to the different TCR segments and introduce double strand breaks (DSBs) of the DNA. The RAG proteins will then cooperate with non-homologous end joining proteins to join the DNA ends. In the process of joining the fragments, non-templated nucleotides may be introduced by terminal deoxynucleotidyl transferases (TdT) or lost, a process known as junctional diversification. This adds not only to the variability of the TCRs but also to the risk of the rearrangements being out-of-frame. There are two *TRB* alleles but the vast majority of T-cells will only express one functional β TCR chain; this monospecificity is ensured by allelic exclusion that functions through asynchronous recombination and negative feedback mechanisms. Asynchronous recombination is accomplished by affecting the recombination efficacy in multiple ways, including the localisation of gene loci to repressive nuclear compartments and chromatin conformations affecting the proximity of TCR gene segments. These mechanisms, in addition to the error prone VDJ process with a high built in failure rate, make it unlikely that both *TRB* alleles undergo functional rearrangements at the same time. Negative feedback mechanisms are initiated once an in-frame VDJ rearrangement is produced, manifested by the pre-TCR. The critical signalling pathways that regulate feedback inhibition are only partially understood; however, the pre-TCR takes part in the initiation. Decontraction and reduced chromatin accessibility will function to prevent further recombination at the other *TRB* allele (Schlissel, 2003; Skok *et al.*, 2007; Schlimgen *et al.*, 2008; Kondilis-Mangum *et al.*, 2011; Schatz & Ji, 2011).

T-cell functions

B and T lymphocytes are derived from the same progenitor cell and are morphologically indistinguishable. They nevertheless have distinct, albeit interconnected, functions in the adaptive immune defence, *i.e.,* antibody- and T-cell-mediated immune responses, respectively (Miller, 2002). Mature T-cells exit from the thymus and migrate via the blood to peripheral lymphoid tissues. A minority of T-cells displays γδ TCRs and these cells mainly function in epithelia, such as the skin and gut mucosa. Based on the expression of the surface receptors CD4 or CD8, two major subclasses of αβ T-cells have been identified: the helper and the cytotoxic T-cells, respectively. Cytotoxic, CD8-positive T-cells kill infected cells or tumor cells, while helper, CD4-positive T-cells aid in the activation of cells involved in the immune response, such as B-cells and cytotoxic T-cells (Cantor & Boyse, 1975). A minority of CD4-positive T-cells develops into regulatory T-cells in the thymus and they play an important role in immunological self-tolerance. Self-reactive T-cells that have escaped thymic elimination are suppressed in the periphery by regulatory T-cells (Hori *et al.*, 2003). Mature, naïve T-cells have a lifespan of several years. The immune repertoire, *i.e.*, the TCR diversity, is acquired in childhood after which the thymus involutes. The quantity of T-cells in adults mainly relies on peripheral T-cell division; however, some thymic activity is retained into adulthood (den Braber *et al.*, 2012).

T-cell acute lymphoblastic leukemia (T-ALL)

Early descriptions of leukemia

> "Several cases exist with a great excess of white blood cells (…) Blood of such patients contains so many white blood cells that at first glance I thought they contained purulent matter. In fact I believe the excess of white blood cells is due to an arrest of maturation of blood. From my theory on the origin of blood cells, the overabundance of white blood cells should be the result of an arrest of the development of intermediate cells."

The French physician Alfred Donné wrote this in 1844 (Donné, 1844; Degos, 2001; Xavier, 2013). Donné was a pioneer in microscopy and when he examined the blood from patients with splenomegaly he noted that what had previously been thought of as "pus in the blood" was in fact leukocytosis, and he suggested that the excess of white blood cells (WBC) could be due to a maturation arrest, a fact now well recognized in leukemia. The first reports recognizing leukemia as a distinct clinical entity are, however, often attributed to John Hughes Bennett and Rudolf Virchow. In 1845, Bennett, a professor at the Institutes of Medicine in Edinburgh, performed an autopsy on a 28-year-old man with hepatosplenomegaly. An extensive post mortem report was published, in which Bennett stated that the aberrant blood finding was independent of inflammation, concluding that it represented a systemic blood disorder and called the disease leucocythemia (Bennett, 1845; Kampen, 2012). Just a few months later, Virchow, a 24-year-old German pathologist at the Charité Hospital in Berlin, reported similar findings in an autopsy on a 50-year-old woman with edema, splenomegaly, and nose bleeding. When examining the blood, he found a skewed balance between red and white blood cells in comparison with normal blood. His publication was entitled "Weisses Blut" and in a paper a few years later, Virchow named the disease "Leukämie" (Virchow, 1845, 1848). Thus, the early understanding of leukemia in the mid-nineteenth century was a gradual process, involving several physicians. The first reported case of childhood leukemia was provided by Henry William Fuller and concerned a 9-year-old girl who presented at the St George's Hospital in London in 1850. The child had frequent hemorrhages, enlarged spleen, and leukocytosis, and died after a few months. Recent reviews of Fuller's detailed

notes make it reasonably to believe, however, that the girl had an unusual form of childhood leukemia, namely chronic myeloid leukemia (Fuller, 1850; Piller, 2001). The distinctions between acute and chronic leukemia as well as between leukemias originating from myeloid or lymphoid lineages were made in reports by Nikolaus Friedrich and Paul Ehrlich, respectively (Friedrich, 1857; Ehrlich, 1879). In the late 1860s, the BM was recognized as the site for hematopoiesis and a few years later, leukemic cells were also shown to derive from the BM (Neumann, 1872). The relatively high prevalence of acute leukemia in children was first recognized in 1917 through an epidemiological survey on 1457 cases of leukemia collected by Gordon Ward while on war service with the Royal Army Medical Corps (Ward, 1917). More than 50 years later, in 1973, an article was published on a case of pediatric acute leukemia where the lymphoblasts displayed T-cell surface markers, and in a paper a few years later, T-ALL was established as a clinically as well as biologically distinct disease entity (Borella & Sen, 1973; Sen & Borella, 1975). Thus, T-ALL is a surprisingly late addition to the acute leukemia group.

Characteristics of pediatric T-ALL

Definitions

Acute lymphoblastic leukemia (ALL) comprises three major subclasses: B-cell precursor (BCP) ALL, mature B-cell (Burkitt) leukemia, and T-ALL. When referring to childhood ALL in general, one usually means BCP ALL because that entity comprises approximately 85% of all pediatric ALL cases (Swerdlow *et al.*, 2008). Pediatric T-cell malignancies include T-ALL and T-cell lymphoblastic lymphoma (T-LBL), both included as T lymphoblastic leukemia/lymphoma in the current World Health Organization classification (Swerdlow *et al.*, 2008). T-ALL and T-LBL are usually considered to be different manifestations of the same disease, with the extent of BM involvement being the main feature to distinguish T-ALL (>25% BM blasts) from T-LBL (minimal BM involvement) (Swerdlow *et al.*, 2008). However, even though the entities display overlapping clinical, morphologic, and immunophenotypic features, some studies have identified distinct gene expression profiles, indicating a possibly greater biological difference than previously believed (Burkhardt, 2010; Basso *et al.*, 2011). T-ALL is a malignancy originating from an immature thymocyte in the thymus. The malignant clone will proliferate and spread throughout the body, preferentially in the bone marrow, peripheral blood, lymph nodes, and central nervous system (CNS).

Epidemiology

Leukemia is the most common childhood malignancy, accounting for one-third of all pediatric cancers in developed countries (Steliarova-Foucher *et al.*, 2004). ALL constitutes approximately 80% of all cases of childhood leukemia.

Patient characteristics are distinctly different between BCP ALL and T-ALL. BCP ALL has a marked incidence peak between 2 to 6 years of age, and the sex ratio shows only a slight male predominance with a male/female ratio of 1.2. In contrast, T-ALL has a flatter incidence curve (the median age at diagnosis is 9 years) and there is a marked male predominance – three times as many boys as girls are diagnosed with T-ALL (Forestier & Schmiegelow, 2006; Article III).

The incidence of pediatric ALL in Sweden is one of the highest in the world, with ~4 cases per 100 000 children per year (aged 0-14 years). In absolute numbers this translates into approximately 70 children and adolescents being diagnosed with ALL each year in Sweden. T-ALL comprises 10-15% of childhood ALL; the annual incidence is 0.36 cases per 100 000/year (Hjalgrim *et al.*, 2003). Some studies have reported an increased rate of ALL over time (Steliarova-Foucher *et al.*, 2004); however, the number of pediatric cases diagnosed in Sweden during the last decades has been quite constant (Hjalgrim *et al.*, 2003). Geography, mirroring ethnicity and socioeconomic status, influences the frequency of ALL; the lowest ALL incidences are reported in the sub-Saharan African countries (Stiller & Parkin, 1996). Differences in recorded number of childhood ALL cases, with regard to time trends and geographic origins, may be due to a number of factors, some of which more attributable to administrative factors and health service availability and quality (Adamson *et al.*, 2005). This notwithstanding, if reliable incidence differences were to be observed, that could provide important clues as to etiological factors (Hrušák *et al.*, 2002; Lim *et al.*, 2014).

Etiology

Why is ALL a common childhood malignancy? Etiological studies have mainly addressed risk factors for the more common form of ALL, namely BCP ALL. Pediatric T-ALL is a rare disease which makes it difficult to obtain robust results in etiological studies. However, investigations of causative mechanisms for BCP ALL may, at least to some extent, be relevant also to T-ALL, given that both diseases originate from hematopoietic progenitor cells.

Prenatally and during early life, lymphopoiesis is very active, with progenitor cells proliferating at a high rate and somatic rearrangements creating the immune repertoire. The intense kinetics makes the cells vulnerable to environmental insults that could result in leukemia-associated genetic aberrations. In the majority of childhood ALL cases, including T-ALL, leukemogenesis requires several genetic

events in a single cell, as illustrated by: i) mouse models; ii) the long latency between the first initiating event and overt leukemia in children; and iii) the presence of leukemia-associated anomalies in healthy individuals (Larson *et al.*, 1996; Greaves & Wiemels, 2003; Yasuda *et al.*, 2014). This is in line with, at minimum, a two-step model for cancer development (Knudson, 1971). In BCP ALL, the initiating event often occurs prenatally. This has been suggested also for some T-ALLs, although the origin of most T-ALL cases seems to be postnatal (Ford *et al.*, 1997; Gale *et al.*, 1997; Fischer *et al.*, 2007; Eguchi-Ishimae *et al.*, 2008). A genetic hallmark of T-ALL is TCR translocations (see section "Genetic aberrations in T-ALL" below). Conversely, *IGH* translocations are relatively rare in BCP ALL. However, the age profile for BCP ALL with *IGH* translocations matches the one for T-ALL, *i.e.*, older children and adolescents (Russell *et al.*, 2014). Thus, it seems that aberrations involving illegitimate rearrangements of immunoglobulin (Ig)/TCR loci could arise postnatally and might share a common etiological factor.

The only confirmed external causative mechanism for ALL is ionizing radiation, but because this is now a well-known risk factor, pre- and postnatal exposure is held to a minimum and irradiation should hence not contribute significantly to cases occurring today. Several other causative agents have been proposed, such as parental smoking, paint and household chemicals, pesticides, maternal diet, traffic fumes, and electric fields, but so far none has been shown to add convincingly to the incidence of childhood ALL (Eden, 2010; Wiemels, 2012).

That infections and/or the body's response to them may promote leukemia development has been put forward in two separate, albeit somewhat overlapping, theories: "population mixing" and "delayed-infection" (Kinlen, 1988; Greaves, 1988). Kinlen's hypothesis of population mixing postulates that leukemia may be the rare result of a common infection in a non-immune individual, a phenomenon most evident when there is an influx of new residents to a previously isolated community. However, in contrast to leukemia in animals and certain human malignancies, such as Burkitt lymphoma and adult T-cell leukemia, there is little evidence for direct viral transformation in childhood ALL (Eden, 2010). Greaves' theory of delayed-infection postulates that a delayed exposure to common infections may trigger an aberrant immune response, thus creating an immunological milieu favoring leukemic development; however, the exact underlying mechanism is unclear (Greaves, 2006). Some circumstantial support for the "delayed-infection" theory comes from investigations showing that children who are exposed to common infections at an early age, measured as day-care attendance, have a reduced risk of ALL (Urayama *et al.*, 2010).

Apart from external factors, could there be an inherent difference between children as to how susceptible they are to develop ALL? A few syndromes due to germline mutations in high-penetrance genes confer an increased risk of ALL, such as ataxia-telangiectasia that is associated with childhood lymphoid malignancies, in particular T-ALL (Seif, 2011). However, hereditary syndromes are a rare cause of ALL. This

notwithstanding, a study of late T-ALL relapses revealed that the leukemia- and clone-specific markers were completely different between diagnosis and recurrence in eight out of the 22 (36%) investigated cases, suggestive of new primary T-ALLs rather than relapses. One explanation could be that patients with two suspectedly separate T-ALLs are predisposed to ALL development; indeed, one of the patients in that study had a constitutional genetic aberration, del(11)(p12p13), previously associated with T-ALL (Szczepański *et al.*, 2011). However, the existence of a pre-leukemic, ancestral clone could potentially give rise to two seemingly different leukemias (Shlush *et al.*, 2014).

Furthermore, genome-wide association studies have identified five common polymorphisms at 7p12.2 (*IKZF1*), 9p21.3 (*CDKN2A*), 10p12.2-12.31 (*BMI1-PIP4K2A*), 10q21.2 (*ARID5B*), and 14q11.2 (*CEBPE*) that influence the risk, albeit with low impact, of childhood ALL. The polymorphism in *CDKN2A* is significantly associated also with T-ALL (Papaemmanuil *et al.*, 2009; Sherborne *et al.*, 2010; Xu *et al.*, 2013). If genetic and/or environmental risk factors play a role in the etiology of ALL, one would expect that siblings of children with ALL have an increased risk of this disease. In fact, in a recent study, a high ALL-subtype concordance rate was shown within sibships affected by multiple cases of pediatric ALL. Notably, in five of the six sibships presented, where the first case was T-ALL, the subsequent case (or cases) was also T-ALL. This indicates that etiologic factors, and a possibly increased sibling risk, may be subtype-dependent (Schmiegelow *et al.*, 2012). In conclusion, there is not a single cause of childhood ALL. This is not surprising considering how biologically and genetically diverse the disease is. Most likely, a combination of constitutional polymorphisms and different exposures will influence the ALL risk, potentially in a subtype-specific manner.

Symptoms, findings, and patient characteristics

The characteristic symptoms and findings in patients with ALL are caused by the BM failure due to maturation arrest. The lymphoblasts expand and hence normal hematopoiesis is repressed, resulting in a deficiency of normal blood cells. Fever, recurrent infections, fatigue, paleness, and skin/mucosal bleeding are frequent. Bone pain and arthralgia are also common. In a review of 189 childhood ALL patients, 96% had at least one classic leukemia symptom/finding; however, 4% were diagnosed accidentally (Bernbeck *et al.*, 2009). Mediastinal mass and CNS involvement is particularly common in T-ALL, seen in approximately 60% and 10% of the patients, respectively, and may lead to dyspnea, headache, nausea, and visual impairment. The term “acute” in acute leukemia implies a rather short medical history, between days and weeks; however, in a proportion of patients (~10%), a duration of months has been noted (Bernbeck *et al.*, 2009).

Diagnostic modalities

Clinical evaluation

The clinical evaluation assesses the patient's general condition, including signs of the above-mentioned symptoms and findings. Neurological examination and palpation of the lymph nodes, abdomen, and testes are performed in order to identify lymphadenopathy, hepatosplenomegaly, and testicular and/or CNS involvement. Echocardiography, chest X-ray, and analysis of spinal fluid samples are also part of the diagnostic work-up.

Blood samples

Laboratory findings at diagnosis often include anemia, thrombocytopenia, and neutropenia. Although the term leukemia means "white blood", implying leukocytosis, the WBC counts may be low, normal, or high. However, T-ALL patients most often display a high WBC count (median 66 x 10^9/l) and almost 50% have hyperleukocytosis >100 x 10^9/l. The differential blood count reveals the presence of peripheral lymphoblasts in most instances (Bernbeck *et al.*, 2009; Article III)

BM samples

The BM aspirates are used for morphologic, cytochemical, immunophenotypic, and genetic analyses, with the latter comprising Ig/TCR monoclonality assays and cytogenetic/molecular genetic screening for acquired chromosomal changes. A BM infiltration of more than 25% lymphoblasts is a prerequisite for the diagnosis of ALL – T-ALL cases most often display a much higher percentage of BM infiltration. Because T-ALL and BCP ALL lymphoblasts are morphologically indistinguishable, cytochemical and immunophenotypic analyses are required to identify the lineage involved (Swerdlow *et al.*, 2008).

Immunophenotype

As mentioned above, immunophenotyping is mandatory in order to distinguish between BCP ALL and T-ALL. CD3, expressed in the cytoplasm or on the cell surface, is T-lineage specific. In addition, TdT is usually positive as is a variable expression of CD1a, CD2, CD4, CD5, CD7, and CD8. Aberrant expression of one or both of the myeloid-associated markers CD13 and CD33 is seen in 20-30% of T-ALL cases.

As normal thymocytes mature, they are characterized by a changing pattern of CD markers, and T-ALL lymphoblasts will, to some extent, mirror the normal maturation process. The constellation of expressed CD markers in lymphoblasts can hence suggest the stage of the leukemic differentiation block. Four maturation subgroups of T-ALL have been defined according to the CD patterns: 1) pro-T: cCD3+/CD7+/CD2-/CD1a-/CD34+/-; 2) pre-T: cCD3+/CD7+/CD2+/CD1a-

/CD34+/-; 3) cortical T: cCD3+/CD7+/CD2+/CD1a+/CD34-; and 4) medullary T: cCD3+/CD7+/CD2+/CD1a-/CD34-/membrane CD3+. The immature pro- and pre-T stages are DN for CD4 and CD8 and the cortical T group is DP for CD4 and CD8, whereas the more mature medullary T-cell is SP for CD4 or CD8 (Bene *et al.*, 1995; Swerdlow *et al.*, 2008). A novel subtype of T-ALL – the early T-cell precursor ALL (ETP ALL) – is thought to originate from the earliest thymic immigrants (however, the cell of origin is debated; see section "Initiating event" below). ETP ALL displays a unique constellation of cell surface markers: CD1a-/CD8-/CD5(+) and expression of one or more myeloid- or stem cell-associated markers. Initially, ETP ALL was associated with a very dismal outcome; however, more recent studies allocate this subtype to intermediate risk (Coustan-Smith *et al.*, 2009; Patrick *et al.*, 2014).

TCR rearrangements

The TCR genes in normal T-cells are rearranged in a specific order during the differentiation process, starting with *TRD* followed by *TRG*, *TRB*, and *TRA*. Recombination of *TRA* will lead to the deletion of *TRD* since this gene is located within *TRA*. Each T-cell harbors a unique TCR rearrangement, like a DNA fingerprint, due to combinatorial and junctional diversification during the V(D)J recombination process (Figure 2). T-ALL is a clonal disease, *i.e.*, all leukemic cells descend from a single ancestor. Thus, if the TCR genes are rearranged in the ancestor cell, identical rearrangements will be present in all the lymphoblasts. Hence, in contrast to an inflammatory process that gives rise to a polyclonal, diverse T-cell response, the malignant T-cells display TCR monoclonality. More than 90% of T-ALL cases display at least one monoclonal rearrangement, most often involving *TRG* and *TRB*. In a small subset of cases, biallelic *TRG* deletions are absent; this represents an immature form of T-ALL associated with a poor outcome (Gutierrez *et al.*, 2010a). Cross-lineage immunoglobulin (Ig) rearrangements are found in 10-20% of T-ALL, whereas TCR rearrangements are found in up to 70% of BCP ALL. Thus, neither Ig nor TCR rearrangements are lineage-specific. They are hence not useful for lineage assignment. It should be emphasized that although TCR and Ig monoclonality is a marker for leukemic proliferation and useful for diagnostic purposes as well as for assessing the maturation stage of the leukemic clone and measuring minimal residual disease (MRD) (see below), they do not participate in the pathogenesis of T-ALL (Szczepański *et al.*, 2000; Brüggemann *et al.*, 2004; Kraszewska *et al.*, 2012a).

Genetic analyses of acquired chromosome changes

At the time of diagnosis, BM and blood samples are sent for cytogenetic and molecular genetic investigations. In the Nordic countries, mandatory genetic analyses are, at present, specified by the Nordic Society of Paediatric Haematology and Oncology (NOPHO) ALL-2008 protocol. The genetic investigations comprise chromosome banding analysis as a screening method and targeted analyses for the

translocations/fusion genes t(1;19)(q23;p13) [*TCF3/PBX1*], t(9;22)(q34;q11) [*BCR/ABL1*], and t(12;21)(p13;q22) [*ETV6/RUNX1*] using fluorescence *in situ* hybridization (FISH) or reverse transcriptase polymerase chain reaction (RT-PCR), intrachromosomal amplification of chromosome 21 (iAMP21) and dic(9;20)(p13;q11) using FISH, and 11q23/*MLL* rearrangements using FISH or Southern blot analyses (Toft *et al.*, 2013). These targeted analyses are mainly directed towards BCP ALL-associated aberrations that are either rarely or never seen in T-ALL (Mansur *et al.*, 2010).

In T-ALL, approximately 50-60% of the cases are cytogenetically abnormal, a frequency slightly lower than for pediatric ALL (60%) as a group (Heerema *et al.*, 1998; Forestier *et al.*, 2000; Schneider *et al.*, 2000; Articles III and V). Characteristic T-ALL-associated chromosome changes are described in section: "Genetic aberrations in T-ALL" below. In contrast to BCP ALL, in which the genetic changes are diagnostically as well as prognostically important, the acquired genetic changes in T-ALL, although occasionally of diagnostic importance, play no role in the risk stratification and treatment intensity of T-ALL in the NOPHO ALL-2008 protocol.

Treatment, MRD, and prognosis

Assessment of biological and clinical features associated with inferior outcome is essential in order to ensure that the patients are treated according to the aggressiveness of the individual leukemia. Stratifying patients according to risk ensures that the most intense, and hence most toxic, treatments are limited to the group of patients with the highest risk of relapse, thus sparing standard risk cases undue toxicities. Adjusted treatment can abolish the adverse impact of many high risk features once associated with poor prognosis. The understanding of different risk groups and the practice of combinatorial chemotherapy has gradually evolved since the 1970s, something that has played a major role for the improved outcome for children with ALL.

The NOPHO ALL-2008 protocol comprises three different treatment groups: standard risk (SR), intermediate risk (IR), and high risk (HR). The stratification is based on B or T-cell immunophenotype, WBC count (<100 or ≥100 x 10^9/l), CNS involvement, MRD status (days 15, 29, and 79), and the following specific genetic/chromosomal aberrations: hypodiploidy (<45 chromosomes or DNA index <0.85), t(1;19), dic(9;20), t(9;22), 11q23/*MLL* rearrangement, and iAMP21. Patients with t(9;22) are treated according to a separate protocol. In accordance to these criteria, T-ALL cases can be stratified either as IR or HR, never as SR (Toft *et al.*, 2013).

The treatment is divided into three different phases: i) induction phase, aiming at remission within approximately 4 weeks; ii) consolidation/delayed intensification phase, to eradicate residual leukemic blasts and strengthen remission; and iii) maintenance phase, to stabilize further remission (Stanulla & Schrappe, 2009). The

treatment duration in NOPHO ALL-2008 is 2.5 years and the therapy is based on a combination of cytotoxic drugs, mainly targeting cellular mechanisms crucial for rapidly dividing cells, such as DNA replication. The drugs are administrated in different ways, including intrathecally to eradicate lymphoblasts in the CNS (in the NOPHO ALL-2008 protocol, CNS irradiation is no longer used). Allogeneic stem cell transplantation is an option for persistent ALL, based on MRD findings (Toft *et al.*, 2013).

Monitoring MRD is of great importance in order to evaluate the effect of the given treatment as well as to assess the risk of relapse. MRD can be measured by three different approaches: 1) Real-time quantitative PCR (RQ-PCR) for monoclonal Ig/TCR rearrangements (see above); 2) flow cytometric detection of aberrant leukemic immunophenotypes (see above); and 3) RQ-PCR for fusion genes. The latter approach is seldom used in T-ALL because only a minority of such cases displays a suitable fusion target. The methodological sensitivities are 10^{-3} – 10^{-4}, *i.e.*, one leukemic cell among 1000-10,000 normal cells can be detected (Teachey & Hunger, 2013). In the NOPHO ALL-2008 protocol, MRD analysis is requested at days 15, 29, and 79. In T-ALL, MRD is measured by RQ-PCR for the Ig/TCR genes (Toft *et al.*, 2013). Since antigen receptor genes might undergo continuous recombination during leukemic development, leading to Ig/TCR clonal evolution, it is advisable to monitor two rearrangements (Campana, 2010). It has been convincingly shown that MRD is a strong prognostic indicator – for T-ALL, MRD $\geq 10^{-3}$ at day 79 seems to constitute the most powerful predictor for relapse (Schrappe *et al.*, 2011).

The survival of children diagnosed with ALL has improved dramatically since the 1960s, when only 5% of the patients were cured. In the previous NOPHO protocol (ALL-2000), the probabilities of 5-year event-free survival (pEFS) and overall survival (pOS) for BCP ALL exceeded 80% and 90%, respectively, while for T-ALL, the outcome was generally poorer with pEFS and pOS just over 60% and 70%, respectively (Schmiegelow *et al.*, 2010). Relapsed pediatric T-ALL has a very dismal survival rate of only approximately 20%. Stagnation in the cure rate has been noted during the last 15 years. With increasing knowledge about the profiles of genetic aberrations in different ALL subtypes, the ultimate goal is to design treatments that target specific genetic mutations (Roti & Stegmaier, 2014).

Genetic aberrations in T-ALL

Historic landmarks in cancer genetics – at a glance

In the early 1900s, Theodor Boveri recognized chromosomes as the carriers of the Mendelian hereditary factors (Boveri, 1904). By observing dispermic sea urchin eggs he noted that the distribution of unequal number of chromosomes to daughter cells gave rise to specific characteristics due to the random combination of chromosomes inherited. Some sea urchin eggs with aberrant chromosomal compositions displayed abnormal growth, a trait shared with many tumors. In 1914, Boveri published "Concerning the Origin of Malignant Tumors" in which he proposed that tumors arise as a consequence of incorrect chromosome combinations. He also postulated several other cancer biology concepts, such as cell cycle checkpoints, oncogenes ("stimulatory chromosomes"), tumor suppressor genes ("inhibitory chromosomes"), inherited cancer predisposition, a multistep genetic process of malignant transformation, and a clonal origin of tumors (Boveri, 1914; Balmain, 2001; Harris, 2008). However, at that time, it was not technically feasible to prove (or disprove) Boveri's hypotheses. Improvements in chromosome analyses between the mid 1950s and the early 1970s made it possible to identify and characterize the first cancer-associated aberrant chromosome, the Philadelphia chromosome, clearly supporting the view that chromosome abnormalities are important for carcinogenesis (Tjio & Levan, 1956; Nowell & Hungerford, 1960; Caspersson *et al.*, 1970; Rowley, 1973). Furthermore, in the 1970s, the cellular origin of proto-oncogenes was described, showing that genes in normal cells can become deregulated and contribute to malignant development as oncogenes (Stehelin *et al.*, 1976), and ten years later the same mechanism was described for tumor suppressor genes (Friend *et al.*, 1986). Alfred Knudson demonstrated with the two-hit hypothesis that multiple mutations are necessary for the development of pediatric retinoblastoma (Knudson, 1971). So, many of Boveri's prescient ideas have, by the work and efforts of many subsequent scientists, been proved to be correct.

Transformation from thymocyte to malignant T-cell

T-ALL is the consequence of a normal thymocyte transforming into its malignant counterpart. Although we know that this malevolent makeover is the result of accumulated intracellular genetic aberrations, many questions remain to be fully answered: Which genes are targeted? What mechanisms are used to alter gene expression? How many genetic changes are needed to complete the transformation? What is the duration between the first genetic hit and the clinical presentation of T-ALL? Can genetic aberrations in childhood T-ALL be utilized for more effective treatment?

During the last decade, our knowledge about genetic abnormalities in T-ALL has increased dramatically. As outlined below, studies of T-ALL have implicated numerous genes, targeted by various biological mechanisms, as drivers of leukemogenesis. However, the cellular effect of these seemingly diverse aberrations in multiple genes may well condense to a limited set of signaling pathways, controlling a few crucial biological features, *i.e.*, cell fate, cell survival, and genome maintenance (Vogelstein *et al.*, 2013). The number of genetic aberrations required for malignancy still remains to be fully elucidated. Recently, it has been demonstrated that a pediatric T-ALLs harbors a mean of approximately ten protein-affecting mutations (Zhang *et al.*, 2012; De Keersmaecker *et al.*, 2013; Tzoneva *et al.*, 2013). For a single cell to acquire these crucial aberrations may, however, take several years. The duration between initiation and clinical presentation of pediatric T-ALL can be illustrated by the seemingly rare cases initiated prenatally or after gene therapy: the time span to overt T-ALL in these cases ranges from 1.3 to 11 years (Ford *et al.*, 1997; Hacein-Bey-Abina *et al.*, 2003; Eguchi-Ishimae *et al.*, 2008; Braun *et al.*, 2014). However, the median age for T-ALL in the three reported cases initiated prenatally was 9 years, whereas the median time between gene therapy and disease was 3 years. This discrepancy could possibly be explained by the extensive insertional mutagenesis resulting from gene therapy, which might accelerate the leukemic transformation and hence not fully correspond to the normal duration between initiation and disease. Among the genetic aberrations characterized in T-ALL, a few therapeutic targets have been identified, such as *NOTCH1,* cell cycle regulators including cyclin D, and proteins involved in PI3K-AKT-mTOR pathway. This may in the future result in a more effective T-ALL treatment strategy (Roti & Stegmaier, 2014).

In this thesis I review the genetics of T-ALL, a map which has been drawn by the effort of many scientists. While journeying through the cancer genetic landscape I will point out the fields to which my research work has contributed. I will also briefly mention some of the techniques used in order to reach the various destinations.

Chromosomal aberrations

Cytogenetic studies

Chromosome banding analyses have long been the gold standard in cancer genetic research and many important leukemia-associated chromosomal rearrangements have been characterized by this technique since Rowley (1973) described the t(9;22)(q34;q11) in chronic myeloid leukemia. Cells need to be cultured in order to obtain condensed chromosomes for analysis; the number of metaphases is enriched by adding a mitotic inhibitor to the cell culture. Slides are prepared and the chromosomes stained, in most instances by Giemsa resulting in a G-banding pattern. The strength of chromosome banding analyses is the global overview over all chromosomes, allowing detection of numerical as well as structural changes. However, the technique has a relatively low resolution (~5 Mb) and is labor-intensive (Smeets, 2004).

As mentioned above, G-banding analyses reveal chromosomal abnormalities in 50-60% of pediatric T-ALL cases. Most abnormal karyotypes are pseudodiploid, with structural aberrations being more common than numerical changes. The most frequent cytogenetic aberrations are, in order of decreasing frequency, translocations involving the TCR loci, del(9p), +8, del(6q), and 11q23 rearrangements (Heerema *et al.*, 1998; Schneider *et al.*, 2000; Articles III and V).

TCR translocations

Translocations involving the TCR loci are the oncogenic hallmark of T-ALL. Structural aberrations affecting 14q11 [*TRA/D*] or 7q34 [*TRB*] are detected in approximately 20-30% of pediatric T-ALL cases by G-banding and/or FISH analyses (Cauwelier *et al.*, 2006; Le Noir *et al.*, 2012; Articles III and V).

Why are translocations targeting the TCR genes frequent in T-ALL? Factors triggering structural aberrations include DNA DSBs, spatial proximity between translocation partners, and inappropriate DNA repair. The somatic process of creating a vast TCR repertoire is dependent on the induction of multiple DSBs, facilitated by the RAG proteins. If there are foreign DNA DSBs present close by during the process of somatic TCR rearrangements, the repair system may misalign the strands. The repair system involved, non-homologous end joining, links DSBs without extensive sequence homology, hence facilitating illegitimate joining and chromosomal rearrangements (Bunting & Nussenzweig, 2013). It has been postulated that DSBs introduced in non-TCR loci are the result of RAG proteins recognizing cryptic recombination signal sequences elsewhere in the DNA. However, recent studies characterizing the breakpoints outside TCR have revealed that they are not, in the majority of cases, induced by RAG proteins (Le Noir *et al.*, 2012; Larmonie *et al.*, 2013). Alternative mechanisms facilitating DSBs include specific DNA features, such as cytosine-phosphate-guanine (CpG) dinucleotides (see section

"Epigenetic alterations" below) and repetitive elements, transcriptional activity, and cellular stress (Mani & Chinnaiyan, 2010). The majority of TCR translocations generated will probably not afford any advantages to the cell and will hence not be selected for. However, if DSBs occur in the vicinity of a proto-oncogene, a translocation may place the target gene under the control of TCR enhancers or promoters, resulting in a deregulated expression of the oncogene. Rarely, a TCR translocation can also function to silence a tumor suppressor gene (Le Noir *et al.*, 2012). Consequently, translocations involving TCR loci may potentially be the spark that initiates tumor development (McCormack *et al.*, 2010). So, to answer the question above, translocations involving TCR loci are common in T-ALL because the mechanistic requirements are met and the product has an oncogenic potential, conferring a selective edge.

TRD and *TRB* are the TCR loci most often illegitimately recombined; *TRA* is only rarely targeted and aberrations affecting *TRG* are virtually non-existent (Cauwelier *et al.*, 2006; Le Noir *et al.*, 2012). To date, 19 genes have been reported as partners in TCR translocations in T-ALL (Mitelman *et al.*, 2014). The gene categories mainly targeted are class II basic helix-loop-helix (bHLH) and homeobox genes, encoding transcription factor proteins, and LIM-domain-only (LMO) genes, which modulate transcriptional complexes by protein-protein interactions. Other gene categories are more rarely implicated (Table 1). Specific genes, representing these categories, such as *TAL1*, *HOXA*, *TLX1*, *TLX3*, and *LMO1/2,* have been proposed to delineate separate molecular-cytogenetic T-ALL subgroups, characterized by distinct gene expression patterns (Van Vlierberghe *et al.*, 2008a; Graux et al., 2006).

It has been shown that many TCR translocations are not detected by chromosome banding analyses (Cauwelier *et al.*, 2006; Le Noir *et al.*, 2012). There are several possible reasons for this: i) telomeric breakpoints (*TRB;* 7q34) resulting in submicroscopic alterations; ii) the translocations/inversions may be masked in more complex rearrangements; and iii) outgrowth of only normal cells in culture, resulting in a false normal karyotype. To overcome these shortcomings, TCR rearrangements can be identified using FISH, which is not dependent on metaphase spreads and that has a relatively high resolution (~100 kb). FISH is based on the hybridization of DNA probes to complementary target sequences. The probes are directly or indirectly labelled with a fluorochrome and can hence be detected in a fluorescence microscope. Depending on the probe design and the hybridization patterns, various chromosomal aberrations, such as numerical, structural, and copy number alterations, can be detected and characterized by FISH. However, the technique is limited to answer the question asked by the specific probe design. Hence, FISH analysis is mainly hypothesis-driven in contrary to chromosome banding analysis in which all chromosomes are screened (Bishop, 2010).

Table 1. Characteristic TCR rearrangements in pediatric T-ALL (Van Vlierberghe & Ferrando, 2012; Mitelman *et al.*, 2014)

Gene category/ function	Partner gene	Abnormality	Oncogenic function	Normal thymic expression/ differentiation block	Frequency*
Class II bHLH/ transcription factor	*TAL1*	t(1;7)(p33;q34) t(1;14)(p33;q11)	Inhibits class I bHLH protein (E2A)	Very early stages/ late cortical, CD8+/CD4+	3-6%
LMO/ protein-protein interactions	*LMO1*	t(7;11)(q34;p15) t(11;14)(p15;q11)	Interacts with TAL1 to inhibit E2A. Self-renewal capacity†	*LMO2* very early stages, *LMO1* not expressed/ late cortical, CD8+/CD4+	1%
	LMO2	t(7;11)(q34;p13) t(11;14)(p13;q11)			6-12%
Class I homebox/ transcription factor	*HOXA* cluster	inv(7)(p15q34) t(7;7)(p15;q34)	Impair T-cell differentiation	Some HOXA at early stages/ mature, CD2-/CD4+/CD8-	3-6%
Class II homebox/ transcription factor	*TLX1*	t(7;10)(q34;q24) t(10;14)(q24;q11)	Induction of aneuploidy§	No/ early cortical CD1a+/CD4+/CD8+	5-10%
	TLX3	t(5;14)(q35;q11)			<1%
NOTCH receptor family/ transcription complex	*NOTCH1*	t(7;9)(q34;q34) t(9;14)(q34;q11)	T-cell determination, self-renewal, differentiation	Yes/ not applicable	<1%
D-type cyclin/ cell cycle	*CCND2*	t(7;12)(q34;p13) t(12;14)(p13;q11)	Proliferation	Yes/ not applicable	<1%
Insulin receptor substrate family/ signal transduction	*IRS4*	t(X;7)(q22;q34)	Proliferation, cellular growth, survival	No/ not applicable	<1%

*The frequency with which the gene is targeted by TCR rearrangements. The gene may well be deregulated in a much higher percentage of T-ALLs by other mechanisms, such as *TAL1* and *LMO2* by interstitial deletions, *TLX3* by translocation to *BCL11B* instead of TCR genes, *HOXA* by chimeric oncogenes, *e.g.*, *SET/NUP214*, *PICALM/MLLT10*, and *MLL/MLLT1*, and *NOTCH1* by activating mutations.

†Refers to the oncogenetic potential of *LMO2* (McCormack *et al.*, 2010)

§Refers to the suggested oncogenetic potential of *TLX1* (De Keersmaecker *et al.*, 2010).

In one study reporting a high incidence of submicroscopic TCR translocations in T-ALL, it was also noted that in some cases the translocation partner to TCR remained unidentified (Cauwelier *et al.*, 2006). Although most of the translocations that remain to be cloned are likely to be individually rare, the partner gene may play a greater role in leukemogenesis than suggested based on incidence alone. A very pertinent example of this is *NOTCH1*, a gene which is involved in TCR translocations in only a handful of T-ALL cases (Ellisen *et al.,* 1991; Mitelman *et al.*, 2014) (Table 1). However, it is now known that a majority of T-ALLs harbors activating mutations in *NOTCH1* (Weng *et al.*, 2004). Hence, investigating seemingly quite rare TCR abnormalities in T-ALL may well provide insights into pathogenetically important cellular pathways. As part of my thesis work, two rare TCR translocations in T-ALL were molecular characterized (Articles I and II), namely t(12;14)(p13;q11) and t(X;7)(q22;q34), as detailed below.

Cytogenetic analyses of two pediatric T-ALL cases identified a translocation between the short arm of chromosome 12 and the long arm of chromosome 14; t(12;14)(p13;q11). FISH analyses revealed breakpoints in the *TRA/D* locus (14q11) and close to the cyclin D2 (*CCND2*) gene at 12p13. Gene expression analyses were performed by cDNA microarray and RQ-PCR. The two samples harboring t(12;14) displayed a marked overexpression of *CCND2* compared with pediatric T-ALL cases without t(12;14). In conclusion, the recurrent t(12;14)(p13;q11) results in overexpression of *CCND2* by juxtaposition of *TRA/D* regulatory elements to the vicinity of *CCND2* (Clappier *et al.*, 2006; Article I) (Figure 3). Prior to the cloning of the t(12;14), the D-cyclins, comprising CCND1, CCND2, and CCND3, had been implicated in T-cell malignancies. Knock-down of *CCND3* inhibits proliferation of human T-ALL cells and cells infected with the HTLV-1 retrovirus, which is associated with adult T-cell leukemia/lymphoma, overexpress *CCND1* and *CCND2* (Mori *et al.*, 2002; Sicinska *et al.*, 2003). The t(12;14) was, however, the first neoplasia-associated translocation shown to result in overexpression of *CCND2* and the first example of a targeted deregulation of a member of the cyclin-encoding gene family in a T-cell neoplasm. Subsequently, a t(6;14)(p21;q32) was shown to deregulate *CCND3* by juxtaposition with the *IGH* locus in T-ALL (Nguyen-Khac *et al.*, 2010).The cyclin D family is an essential part of the cell-cycle machinery (Figure 3). D-cyclins interact with the cyclin-dependent kinases 4 and 6 (CDK4 and CDK6). This complex phosphorylates various substrates, such as the retinoblastoma 1 (RB1) protein, and promotes subsequent transition through the cell cycle. Aberrant cell cycle control is a common feature in human cancer. *CCND1* is the most well-established oncogene within the cyclin D family, with deregulated expression by copy number changes, translocations, and mutations having been reported in a variety of cancers, including B-cell neoplasms (Musgrove *et al.*, 2011).

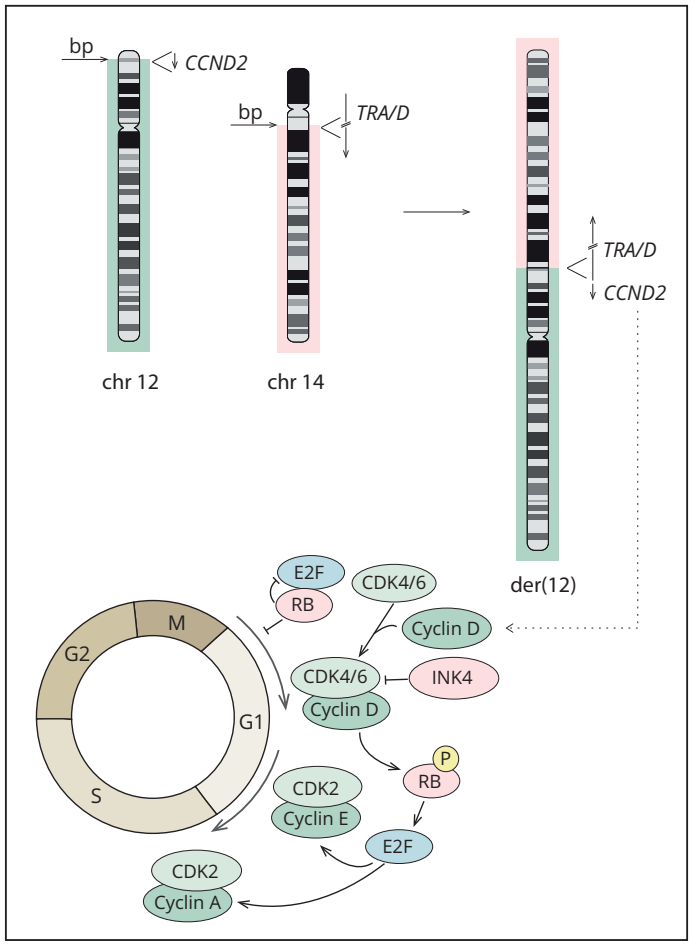


Figure 3. Characterization of t(12;14)(p13;q11) resulting in overexpression of the cyclin D2 (*CCND2*) gene in pediatric T-ALL.

Ideogram for chromosomes 12 and 14 (depicted as at the UCSC Genome Browser: http://genome-euro.ucsc.edu/). The chromosomal breakpoints (bp) identified by FISH analyses are indicated by arrows. Arrows next to the genes indicate their transcriptional direction. The colored bars behind the ideograms highlight the translocated segments. Only one of the two derivate (der) chromosomes resulting from the translocation is shown, namely der(12)t(12;14)(p13;q11). *CCND2* is retained on the der(12) and deregulated by *TRA/D* regulatory elements. Aberrant expression of *CCND2* will affect the cell cycle machinery. The cell cycle is divided into four successive phases: S phase (DNA synthesis) in which the DNA is replicated; M phase (mitosis), in which the cell divides; and the two gap phases G1 and G2, which are periods of cell growth and mitosis preparations, respectively, including diverse control mechanisms to ensure that the cell is ready for division. The transition from one phase of the cell cycle to the next is controlled by cyclin-dependant kinases (CDKs) in complex with cyclin proteins. The INK4 family, including CDKN2A and CDKN2B, inhibits the mitogenic effect of CDK4 and CDK6. The CDK4/6-cyclin D complex phosphorylates the RB1 protein, resulting in subsequent release of the transcription factor E2F. Transcription of several genes, including cyclins A and E, is mediated by E2F. Cell cycle graphics adapted from Strauss *et al.* (2012).

As shown by us and others, D-cyclins are involved also in T-ALL and hence inhibition of their activity may compromise T-ALL progression. Indeed, therapeutic targeting of cyclin D3:CDK4/6 in T-ALL has yielded some promising results (Sawai *et al.*, 2012).

The second TCR translocation in pediatric T-ALL characterized as part of my thesis work was t(X;7)(q22;q34). To map the breakpoints in Xq22 and 7q34 several FISH probes were used. The FISH analyses revealed a breakpoint at the *TRB* locus (7q34) and mapped the corresponding breakpoint to Xq22.3 (Figure 4). Gene expression analysis by RQ-PCR in the t(X;7)-positive case revealed a marked overexpression of the gene located closest to the Xq22.3 breakpoint, namely the insulin receptor substrate 4 (*IRS4*) gene. By Western blot analysis we could also show that IRS4 was expressed at the protein level. In addition, several cooperative genetic aberrations were identified: del(6q), *STIL/TAL1* fusion, and a *NOTCH1* mutation, in accordance with the fact that multiple genetic hits are required for T-ALL to occur (Kraszewska *et al.*, 2012b). In conclusion, molecular genetic characterization of the t(X;7), a translocation not previously reported in a neoplastic disorder, revealed *IRS4* as a novel translocation partner to a TCR locus (Article II) (Figure 4). *IRS4* has subsequently been shown to be a recurrent target of TCR translocations in T-ALL. (Kang *et al.*, 2012). The IRS family comprises four members (IRS1-4) that play a central role in maintaining basic cellular functions, *e.g.*, growth and metabolism. The IRS proteins act as cytoplasmic mediators between multiple growth factor receptors that possess tyrosine kinase activity, such as the insulin and insulin-like growth factor-1 receptors. When phosphorylated, the IRS proteins, which do not contain any intrinsic enzymatic activity, will act as docking sites for several intracellular effector molecules and result in activation of multiple signalling pathways, including the PI3K-AKT and RAS/ERK (Mardilovich *et al.*, 2009) (Figure 4). In humans, IRS1 and IRS2 are the most frequently expressed members, while IRS4, initially detected in human embryonic kidney cells, has only been detected in muscle and liver tissue (Lavan *et al.*, 1997; Schreyer *et al.*, 2003; Cuevas *et al.*, 2007). Mice lacking Irs4 exhibit mild defects in growth, reproduction, and glucose homeostasis; mice without Irs1 or Irs2 display the same phenotype but with more pronounced features (Fantin *et al.*, 2000). Although relatively little is known about the tumorigenic potential of IRS4, expression of this gene has been demonstrated in a human hepatoblastoma cell line and Irs4 has been implicated in proliferation of a murine T-cell lymphoma cell line (Ursø *et al.*, 2003; Cuevas *et al.*, 2007). Cancer cell lines with high IRS4 levels can drive signalling via the PI3K pathway, even in the absence of growth factors, and knockdown of IRS4 inhibits cell proliferation (Hoxhaj *et al.*, 2013). *IRS4* mutations have been detected in melanoma and both *IRS2* and *IRS4* mutations have been demonstrated in T-ALL (Kalender Atak *et al.*, 2012; Shull *et al.*, 2012; Article IV). Hence, by characterizing a rare TCR translocation, a potentially new oncogene in T-ALL was identified.

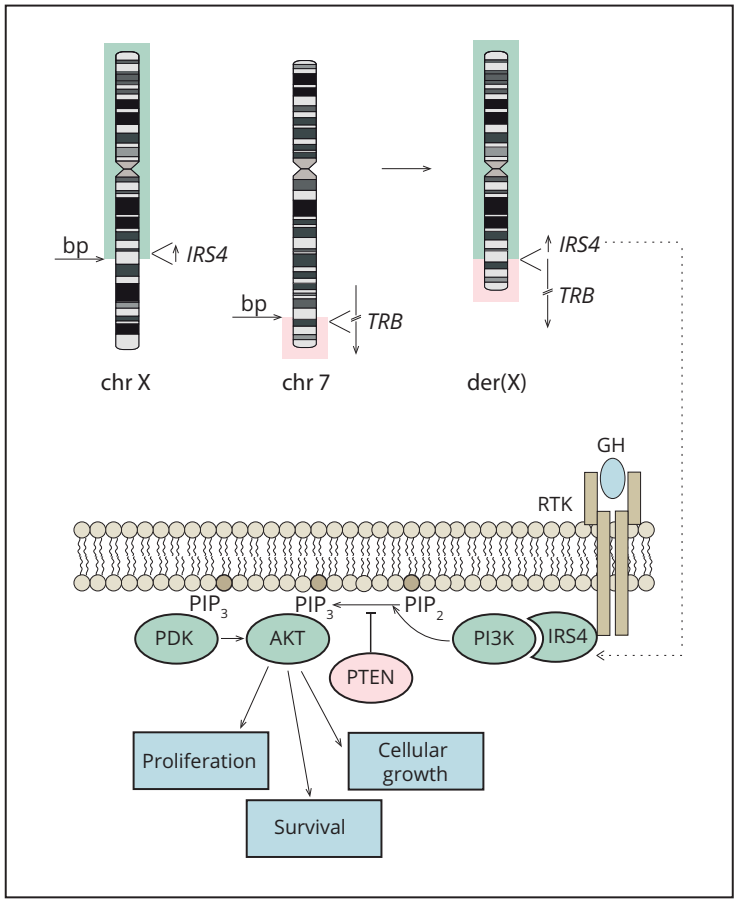


Figure 4. T-ALL-associated TCR translocation targeting IRS4, with a possible impact on the PI3K-AKT signaling pathway.

Ideogram for chromosomes X and 7 (depicted as at the UCSC Genome Browser: http://genome-euro.ucsc.edu/). FISH analyses were used to map the chromosomal breakpoints (bp), indicated by arrows. The positions of the TRB and IRS4 genes are specified and arrows show their transcriptional direction. Colored bars behind the ideograms highlight the translocated segments. Only the der(X)t(X;7)(q22;q34) is depicted, on which the created proximity of IRS4 and TRB is outlined. Aberrant expression of IRS4 may potentially affect several intracellular signaling pathways. However, the PI3K-AKT pathway has been delineated as one of the more important in leukemogenesis and is hence graphically outlined. Stimulation of receptor tyrosine kinases (RTK) by a growth hormone (GH) results in recruitment and phosphorylation of IRS4. Activated IRS4 will function as a docking site for phosphatidylinositol 3-kinase (PI3K). PI3K will subsequently convert phosphatidylinositol (4,5) biphosphate (PIP2) into phosphatidylinositol (3,4,5) triphosphate (PIP3). The phosphatase and tensin homolog (PTEN) dephosphorylates PIP3 and thereby terminates PI3K signaling. Deregulation, mainly through mutations, of PTEN is common in T-ALL. PIP3 recruits phosphatidylinositol-dependent kinase (PDK), which helps to activate the v-akt murine thymoma viral oncogene homolog (AKT). AKT mediates the activation and inhibition of several targets, resulting in proliferation, cellular growth, and survival (Engelman, 2009)

Submicroscopic copy number alterations and uniparental isodisomy

Array CGH analysis

Array-based technologies can detect copy number alterations (CNAs) at a higher resolution than conventional chromosome banding/FISH analyses and without the requirement for dividing cells. Two main types of platform are available: comparative genomic hybridization (CGH) and single nucleotide polymorphism (SNP) arrays (discussed below). Array CGH techniques have in the past used two different types of probe set: bacterial artificial chromosomes (BAC) or oligonucleotides. The method relies on competitive hybridization of differentially labeled tumor and normal samples to the probes on the array. Differences in hybridization signals will convey information regarding CNAs. The minimal size of an aberration that can be detected by the platform is referred to as resolution and is dependent on probe size and distribution as well as on array density. Oligonucleotides are short fragments with high coverage; thus, array CGH based on oligonucleotides has a superior resolution compared with BAC arrays (Maciejewski & Mufti, 2008; Maciejewski *et al.*, 2009).

SNP array analysis

SNPs are common interindividual DNA variations. By definition, a SNP is a single nucleotide substitution, leading to allele differences with the least common allele having a prevalence of 1% or more. There are over 63 million SNPs reported in the human genome (http://www.ncbi.nlm.nih.gov/SNP/; build 141). The effect of a SNP ranges from no phenotypic impact at all to disease-predisposing or even disease-causing, depending on its localization and effect on protein translation. Information about SNPs can be used in several research areas, such as genome-wide association studies, population genetics, and large-scale genome analysis. SNP array analysis has been utilized in research for almost a decade, providing valuable information on, *e.g.*, genetic aberrations in leukemias (Mullighan *et al.*, 2007).

The main advantage of SNP array compared with array CGH is the former method's ability to detect copy number neutral loss of heterozygosity (LOH) in addition to CNAs. The SNP array technique is based on hybridization of fragmented genomic DNA to probes on an array. According to the Illumina platform, used in my thesis work, each probe targets a specific SNP in the genome, binding to the complementary DNA sequence and stopping one base before the SNP. Allele specificity can be conferred by single base extension using fluorescently labeled nucleotides. The specific signal emitted by the extended nucleotide when laser excited is recorded and the signal intensity is normalized against the average signal intensity and converted to a log2 ratio, which acts as a proxy for copy number. The signal also conveys data regarding genotype, denoting the SNP as hetero- or homozygous. Genotype information facilitates the interpretation of CNAs and, maybe more importantly, add data regarding copy number neutral LOH. When large regions contain only homozygous SNPs with no concurrent copy number alteration they may

represent uniparental isodisomy (UPID), and if only encompassing part of a chromosome, it is denoted a segmental UPID (sUPID) (Maciejewski & Mufti, 2008; Maciejewski *et al.*, 2009) (Figure 5). The sensitivity of the method refers to the minimal proportion of cells allowed to be detected, *i.e.*, the method's ability to detect minor clones, constituting either tumor heterogeneity or normal/tumor admixture. In Article V, an Illumina SNP array (HumanOmni1-Quad BeadChip) containing >1 million markers with a median marker spacing of 1.5 kb was used, generating a sensitivity of approximately 10-20% and a resolution of approximately 20 kb. To distinguish between constitutional variants and cancer-associated aberrations it is important to compare diagnostic and matched normal DNA samples and to use databases containing reported genomic variants (Heinrichs *et al.*, 2010). Because of the great advantages of SNP arrays, despite their inability to detect balanced translocations, they are currently implemented in many centers for routine genetic analysis of pediatric leukemias (Wikhager *et al.*, 2012).

Submicroscopic CNAs

Pediatric T-ALLs display a mean of 3-7 CNAs/case when investigated by SNP arrays. The vast majority of the CNAs (>75%) are deletions (Mullighan *et al.*, 2008; Yu *et al.*, 2011; Article V) and it has been suggested that RAG-mediated deletions contribute to the acquisition of some of these aberrations (Raschke *et al.*, 2005; Mendes *et al.*, 2014). The main oncogenetic impact afforded by deletions is inactivation of tumor suppressor genes. In T-ALL, the chromosomal segment most commonly affected by deletions is 9p21 (Figure 5), encompassing the *CDKN2A* and *CDKN2B* genes at 9p21.3. Deletions of *CDKN2A* are seen in >70% of pediatric cases, with a concurrent *CDKN2B* deletion in a slightly lower frequency (Mullighan *et al.*, 2008; Article V). *CDKN2A* encodes two separate proteins, CDKN2A and P14ARF, regulating the cell cycle through inactivation of CDK4/6 and stabilization of the TP53 protein, respectively. *CDKN2B* also codes for a CDK4/6 inhibitory protein, CDKN2B. Several other tumor suppressor genes have been shown to be deleted in T-ALL, although to a lesser extent, such as *BCL11B* (14q32.2), *LEF1* (4q25), *NF1* (17q11.2), *PTPN2* (18p11.21), and *WT1* (11p13) (Balgobind *et al.*, 2008; Tosello *et al.*, 2009; Gutierrez *et al.*, 2010b; Kleppe *et al.*, 2010; Gutierrez *et al.*, 2011; Article V). Recently, recurrent deletions have been shown to target genes involved in epigenetic regulation, including *EZH2* (7q36.1), *PHF6* (Xq26.2), and *SUZ12* (17q11.2) (Van Vlierberghe *et al.*, 2010; Ntziachristos *et al.*, 2012; Article V). Many of the above-mentioned tumor suppressor genes, with the possible exception of *CDKN2A/2B*, have also been shown to be targeted by inactivating mutations (Sulong *et al.*, 2009; Van Vlierberghe & Ferrando, 2012).

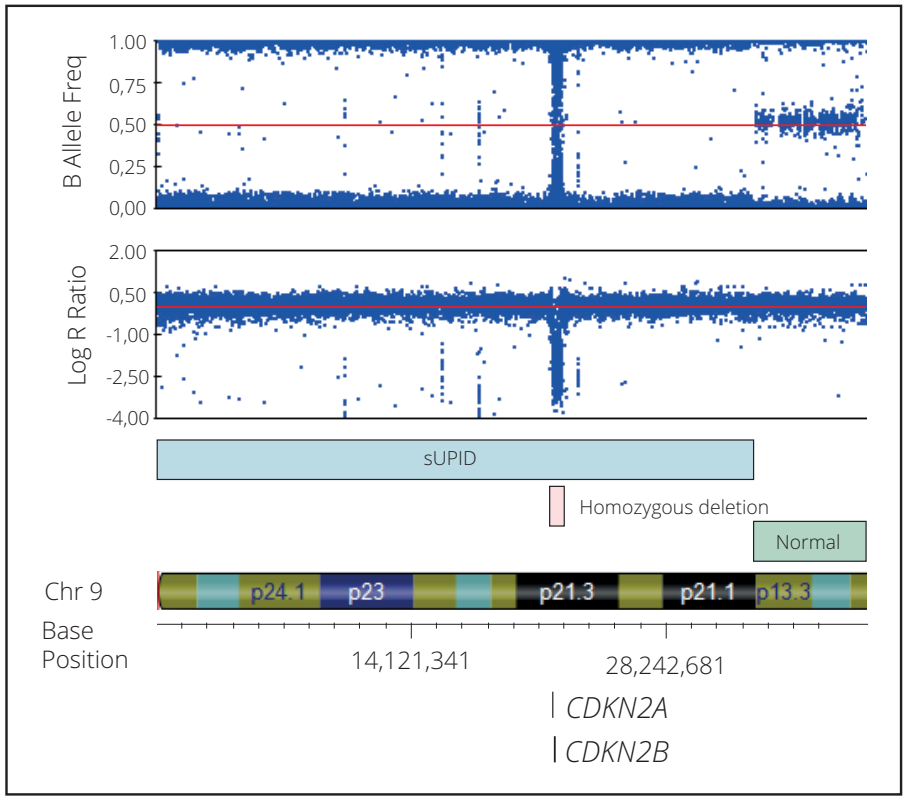


Figure 5. Graphical output for chromosome segment 9p13-pter from a SNP array analysis of a T-ALL case. Each dot represents one marker, in most instances a SNP. The top panel depicts genotype information presented as a B allele frequency (BAF), calculated as the signal intensity for allele B/(signal intensities for alleles A and B). If the SNP is heterozygous, AB, the BAF is 0.5 (indicated by the red line), while if it is homo- or hemizygous, BB/B- or AA/A-, it will be either 1 or 0, respectively. The BAF allows for the detection of copy number neutral aberrations, such as segmental UPIDs (sUPIDs). The lower panel presents signal intensity data (R), calculated as the log2 ratio (log R ratio), that can be used to ascertain copy number alterations. If the SNP is present in two copies (normal diploid copy number), the log R ratio is 0 (indicated by the red line). Increased and decreased log R ratios correspond to gained and deleted regions, respectively. The T-ALL case in this figure displays an aberrant BAF for the chromosomal segment 9p21.1-pter. This area does not contain any heterozygous SNPs, *i.e.*, no SNPs have a BAF value of 0.5 between 9p21.1 and pter. This BAF pattern could be the result of either an sUPID or a heterozygous deletion. To distinguish between these two aberration types information from the log R ratio is necessary. An sUPID does not affect the log R ratio, while a heterozygous deletion would result in a decreased ratio. The main part between 9p21.1 and pter does not demonstrate an alteration in the log R ratio; hence, the aberration affecting this segment is an sUPID (indicated by the blue bar). However, it is obvious that a second aberration affects band 9p21.3. The BAF pattern indicates that there are no SNPs to genotype, and the log R ratio is decreased. Thus, a homozygous deletion (indicated by the pink bar) is present within the sUPID. It should be noted that both the *CDKN2A* and *CDKN2B* genes will be deleted as a consequence of this aberration. In fact, sUPID for 9p with concurrent homozygous deletion of 9p21.3 is a recurrent mechanism for silencing *CDKN2A* and *CDKN2B* in T-ALL (Article V).

However, although tumor suppressor genes might be the most obvious targets for deletions, it has been shown that oncogenes can be activated by chromosomal losses through several different mechanisms. *TAL1* (1p33) is often deregulated in T-ALL, occasionally as the result of a translocation between *TAL1* and *TRA/D* or *TRB* (Table 1). However, in 10 to 40% of the cases, an interstitial 1p33 deletion places *TAL1* under the control of the promoter of the neighboring gene *STIL*, which is highly expressed in T-cells (Janssen *et al.*, 1993; Yu *et al.*, 2011; Article V). This promoter swapping results in overexpression of *TAL1* (Ferrando *et al.*, 2002). The oncogene *LMO2* (11p13) is often deregulated in T-ALL through illegitimate TCR translocations (Table 1). Array analyses have revealed an alternative way of *LMO2* activation, namely deletions encompassing a negative regulatory element in 11p12-p13, resulting in elevated *LMO2* expression (Van Vlierberghe *et al.*, 2006). Finally, submicroscopic deletions of 9q34 have been shown to result in the fusion gene *SET/NUP214*, which encodes a novel transcription factor that upregulates the *HOXA* genes (Van Vlierberghe *et al.*, 2008b).

In spite of the fact that the vast majority of CNAs detected by arrays are deletions a few recurrent chromosomal gains have been identified in T-ALL. A fusion between *NUP214* and *ABL1* at 9q34 is amplified on episomes in ~5% of T-ALL cases (Graux *et al.*, 2004). *NUP214/ABL1* codes for a constitutively active tyrosine kinase. Although the clinical experience is limited, a few, but definitely not all, T-ALL patients with this fusion treated with tyrosine kinase inhibitors have responded well (Clarke *et al.*, 2011). Another oncogene deregulated through gene duplication is *MYB* (6q23.3), which has also been shown to be targeted by TCR translocations (Clappier *et al.*, 2007; Lahortiga *et al.*, 2007).

Uniparental isodisomy

Copy number neutral stretches of LOH can arise through different mechanisms: i) as runs of homozygosity by descent, ii) meiotic UPID, and iii) mitotic UPID. A UPID is generated when both alleles are from the same parental origin and identical, hence giving rise to homozygosity (Engel, 1980). Homozygosity by descent is common in all populations and an increase in frequency and length is noted when there is close kinship between the parents (Kirin *et al.*, 2010). Although homozygosity by descent is a normal finding it may represent a risk for recessive diseases. Meiotic missegregation may result in uniparental heterodisomy (meiosis I error) or isodisomy (meiosis II error) that can give rise to different syndromes, sometimes associated with an increased risk of cancer (Lapunzina & Monk, 2011). Most cancer-related studies have, however, focused on somatically acquired, mitotic UPIDs. The notion that mitotic UPIDs can be involved in cancer development was first described in 1983 and, with the advent of the SNP array technology, ample data supporting this association have been forthcoming (Cavenee *et al.*, 1983; Fitzgibbon *et al.*, 2005). The mechanisms for mitotic UPID are not fully elucidated; whole chromosome UPIDs (wUPIDs) can occur due to mitotic missegregation or may reflect the

underlying mechanism by which certain subtypes, such as high hyperdiploid ALL, originate, whereas sUPIDs may result from mitotic homologous recombination (Fitzgibbon *et al.*, 2005; Paulsson & Johansson, 2009; Makishima & Maciejewski, 2011). Comparison between tumor and matched normal DNA is the most reliable method to distinguish between inherited runs of homozygosity and acquired mitotic UPIDs; however, size and localization may also be used for this discriminative purpose (Makishima & Maciejewski, 2011). The biological consequence of a sUPID is often duplication of a mutation/deletion in a tumor suppressor gene, resulting in homozygous silencing of the gene (Fitzgibbon *et al.*, 2005). However, sUPIDs could also potentiate activating mutations in oncogenes or result in changes in epigenetic patterns. In pediatric ALL, mitotic UPIDs are detected in approximately 25% of cases (Kawamata *et al.*, 2008); wUPIDs have been shown to be prevalent in high hyperdiploid ALL, with a frequency of close to 30% (Paulsson *et al.*, 2010), and in near-haploid/low hypodiploid ALL (Safavi *et al.*, 2013), whereas sUPIDs are seen in approximately 15% of pediatric ALL cases, with the most frequent gene segment targeted being 9p (Kawamata *et al.*, 2008). There is some discrepancy with regard to how often deletions of *CDKN2A* coincide with sUPID9p (Sulong *et al.*, 2009). In our T-ALL series, sUPIDs were seen in 44% of the cases, with the vast majority affecting 9p. We could show that all cases with sUPID9p had a concurrent *CDKN2A* deletion and because all homozygous *CDKN2A* deletions associated with sUPID9p had identical breakpoints this strongly suggests that a heterozygous deletion occurred prior to the sUPID (Article V). Hence, in T-ALL, homozygous *CDKN2A* deletions are frequently the result of a heterozygous *CDKN2A* deletion with a subsequent mitotic recombination resulting in an sUPID9p (Figure 5).

Gene mutations

Sanger sequencing

G-banding, FISH, and SNP array analyses, as described above, detect numerical and structural aberrations as well as small deletions, duplications, and UPIDs, with a resolution down to less than 100 kb. However, to detect aberrations affecting only a few nucleotides, such as base pair substitutions (point mutations) or small-scale insertions or deletions ("indels"), the genetic code needs to be deciphered base by base through sequencing. A method that has dominated DNA sequencing for nearly thirty years was presented in 1977 and was named Sanger sequencing after its inventor (Sanger *et al.*, 1977). Basically, the method relies on chain terminating nucleotides (ddNTPs). Sequencing products may be radioactively labeled or, most often nowadays, attached to a fluorescent dye (Smith *et al.*, 1986). The automated application separates the fragmented sequencing products by capillary electrophoresis, after which a laser detector will record the order of the four differentially fluorescently end-labeled ddNTPs. This information is converted to a chromatogram that displays

the nucleotide sequence. Sanger sequencing results in relatively long read-lengths with a high accuracy (Shendure & Ji, 2008).

Next-generation sequencing

During the last decade, the sequencing technology has been revolutionized with the development of methods that have a much higher throughput to a fraction of the cost compared with Sanger sequencing. These new techniques, referred to as next-generation sequencing (NGS), enable massively parallel sequencing of not just one DNA sequence, as by the Sanger method, but millions of different sequences at the same time.

There are several different NGS methods which all consist of three common steps: i) template preparation; ii) sequencing and imaging; and iii) alignment and bioinformatics. The DNA sequenced may represent the entire genome (whole genome sequencing), all exons (exome sequencing), a gene panel (targeted sequencing), or cDNA derived from RNA (transcriptome or RNA sequencing, RNA-seq). In the Illumina platform, applied in Article V, templates are created by fragmentation of DNA and ligation of two different adaptors, needed for subsequent sequencing reactions, to each end of the fragments. The templates are attached to a solid surface and then clonally amplified by bridge PCR that generates clusters, each consisting of numerous templates created from a single DNA strand. The amplification serves to enhance subsequent sequencing signals. One of the two adaptors is cleaved from the surface and the DNA is denatured to make it single-stranded, giving one single sequence in each cluster. Millions of spatially separated clusters enable the next step, *i.e.*, sequencing, to occur in a massively parallel fashion. The sequencing is based on the assembly of DNA strands complementary to the templates, referred to as reads. The reads will be sequenced as they are assembled, so called sequencing by synthesis. The sequencing is performed in a cyclic fashion consisting of base incorporation, imaging, and cleavage. For this purpose, the nucleotides used in the reactions are modified with reversible terminators, which allow only single-base incorporation in each cycle, and one of four fluorescent labels, corresponding to the identity of each nucleotide. After the single-base incorporation, imaging captures the fluorescent signals from all clusters on the surface at the same time. The signal recorded from each cluster is a consensus of all the individual fluorescent signals transmitted from every template constituting that specific cluster. Next, the fluorochrome molecule and terminator are cleaved from the incorporated nucleotide, thereby enabling a new modified nucleotide to be added to the read. Thus, the cycle of incorporation, imaging, and cleavage can be repeated numerous times. Incomplete extension or addition of more than one nucleotide in a given cycle will eventually give rise to signaling dephasing within the cluster, causing base-calling errors and limiting the read length. Hence, the fragments sequenced by NGS experiments are generally shorter than those analyzed by Sanger sequencing. In our experiments, 100 base pairs from both ends of the DNA fragments, paired-end, were

sequenced (Article V). For paired-end sequencing, after the first adaptor has been used for sequencing, a second round of bridge PCR is performed and the second adaptor is used to bind the primer. Each nucleotide will be sequenced several times, usually between 10-1000 times, as the same sequence will be present in multiple clusters. This is referred to as depth of coverage and will determine the strength of the method to identify small clones. The advantage of the NGS technology is its capacity to generate enormous amounts of data; however, advanced bioinformatic pipelines are crucially important in order to extract pathogenetically important findings from the wealth of sequences. The generated reads are aligned to a reference human genome and filtered according to different parameters, such as base quality, alignment score, mapping quality, and PCR duplicates. Next, variant calling analyses are performed, using confidence scores as well as read and variant coverage, to assign SNPs and other genomic variants. NGS can potentially detect the full range of genetic alterations (nucleotide substitutions, indels, CNAs, and structural aberrations as well as gene expression levels) depending on the input DNA (Shendure & Ji, 2008; Metzker, 2010; Meyerson *et al.*, 2010; Rizzo & Buck, 2012; Article V).

There are several sources of errors in NGS technology. False positive results can arise as a consequence of errors introduced during template amplification, sequencing, or because of incorrect alignment. False negative results may be due to the fact that some sequences are difficult to capture or map resulting in low coverage. Mutations can also go undetected due to samples being heavily "contaminated" by normal cells (Rizzo & Buck, 2012). Our experience of inconsistent findings obtained by Sanger sequencing and deep sequencing can be demonstrated by the *IRS4* mutation (c.670C>A; p.Pro213Thr; NP_003595.1) detected in Article IV using Sanger sequencing. DNA from this patient was analyzed by deep sequencing in Article V, but in this setting we could not detect this *IRS4* mutation. The deep sequencing raw data for the sample in question displayed high *IRS4* coverage, approximately 100 reads, with none of the reads displaying the alteration. The case displayed a *NOTCH1* mutation with a variant allele frequency of 0.49, indicating a high number of malignant cells in the sequenced sample. We attempted to re-analyze *IRS4* with Sanger sequencing but, unfortunately, we were not successful. Hence, we cannot fully explain this discrepancy.

After the samples have been analyzed by NGS, the acquired data need to be interpreted according to the biological context (Lawrence *et al.*, 2013). Which of all the mutations identified are of pathogenetic value? Which genes and mutations will have cancer-initiating/propagating effect ("drivers") and which are merely passive bystanders ("passengers")? Matched normal DNA samples as well as information from databases, such as dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), COSMIC (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/), and PolyPhen (http://genetics.bwh.harvard.edu/pph2/), may aid the interpretation regarding the origin (constitutional/acquired) and expected biological impact (driver/passenger) of the variants (Vogelstein *et al.*, 2013). Most previous studies have focused on protein-

altering variants; however, also synonymous mutations, not affecting the protein sequence, have been shown to be selected for in malignant cells, possibly contributing to cancer development (Gartner *et al.*, 2013). It has also been reported that 80% of the genome has biochemical functions; hence, alterations outside the 1% exome may also affect the cellular phenotype (ENCODE Project Consortium, 2012). A very recent report illustrated the potential of intergenic impact by the identification of recurrent duplications in a long-range *MYC* enhancer, conferring an essential role in *NOTCH1*-induced T-ALL (Herranz *et al.*, 2014).

Gene mutations in T-ALL

Numerous genes are mutated in T-ALL at frequencies equivalent to minor hills in the cancer genetic landscape (Vogelstein *et al.*, 2013). However, the mutation frequency of one gene will reach the peak of Mount Everest – namely, *NOTCH1*. NOTCH1 is crucial for T-cell fate and differentiation (Radtke *et al.*, 1999). The first indication that *NOTCH1* could be of pathogenetic importance in T-ALL came when the translocation t(7;9)(q34;q34) was cloned in 1991 (Ellisen *et al.*, 1991). This translocation, present in <1% of T-ALL cases (Mitelman *et al.*, 2014), results in a truncated and hence constitutively activated form of NOTCH1. However, the central role of this gene was highlighted when activating *NOTCH1* mutations were found to occur in more than 50% of T-ALLs (Weng *et al.*, 2004). The oncogenetic potential of activated NOTCH1 has also been demonstrated in mouse models (Pear *et al.*, 1996). NOTCH1 is a transmembrane heterodimeric receptor composed of one extracellular and one transmembrane/intracellular subunit, which interact via the heterodimerization domain (HD). The extracellular unit will bind to ligand molecules (the transmembrane proteins Delta-like 1, 3, and 4; and Jagged 1 and 2) that induce conformational changes and subsequent cleavages of NOTCH1, resulting in the release of the intracellular domain into the cytosol. The NOTCH1 intracellular domain will then translocate into the nucleus where it associates with DNA-binding proteins and cofactors, constituting a complex which will act as a transcriptional activator. NOTCH1 regulates the expression of several genes, mainly those involved in proliferation, metabolism, and cell cycle such as *MYC*, *HES1* (affecting the PI3K-AKT-mTOR signaling pathway), and *CCND3* (Paganin & Ferrando, 2011). The phosphorylation of the PEST (proline (P), glutamic acid (E), serine (S), and threonine (T) rich) domain and recruitment of FBXW7 induce NOTCH1 degradation. T-ALL-associated *NOTCH1* mutations cluster in the HD, resulting in ligand-independent activation, and in the PEST domain, stabilizing NOTCH1 by impairing degradation. The latter effect can also be accomplished by inactivating mutations in *FBXW7,* found in 15-30% of T-ALL cases. A proportion, about 20%, of T-ALLs displays dual HD and PEST mutations, or HD and *FBXW7* mutations; these co-occuring mutations will work in synergy, optimizing the NOTCH1 signaling (Paganin & Ferrando, 2011; Tosello & Ferrando, 2013). The extensive involvement of aberrant *NOTCH1* signaling in T-ALL makes it an attractive target for therapy and, recently, intense research has focused on different ways to inhibit

NOTCH1 activity in T-ALL, something that may result in improved therapy of this disease in the future (Roti & Stegmaier, 2014).

As for *NOTCH1*, most mutation analyses have focused on one or a few candidate genes, such as *FLT3*, *IL7R*, *IRS4*, *JAK1*, and *NRAS* (Bar-Eli *et al.*, 1989; Van Vlierberghe *et al.*, 2005; Flex *et al.*, 2008; Zenatti *et al.*, 2011; Article IV). Although such an approach has been fruitful, it does not provide data on co-operative mutations in the leukemogenesis. More recently, studies applying whole genome, exome, and transcriptome sequencing by NGS have been forthcoming, identifying an average of ten protein-altering mutations per pediatric T-ALL case. These reports have indicated novel genes harboring driver mutations, such as *CNOT3*, *RUNX1*, *STAT5B*, and ribosomal genes. But equally important, the results show that the co-operative effects of mutations affect signaling, transcription, and epigenetics (Zhang *et al.*, 2012; De Keersmaecker *et al.*, 2013; Kalender Atak *et al.*, 2013; Tzoneva *et al.*, 2013; Bandapalli *et al.*, 2014). Although the cost for such analyses is decreasing, they are still quite expensive if a larger number of cases are investigated. Furthermore, to ascertain the relevant mutations, among the thousands of gene changes identified, is a bioinformatic challenge. An alternative is to apply NGS on a gene panel. By targeted deep sequencing it is possible to analyze a list of candidate genes at a greater depth and in a larger cohort than is usually feasible by whole genome NGS, although the genes selected obviously will influence the type of information gained (Fieuw *et al.*, 2012). Targeted sequencing could potentially render data on: i) frequency of known mutations in driver genes; ii) detect new mutations in known driver genes; and iii) implicate new cancer-associated genes in a candidate gene approach (Kalender Atak *et al.*, 2012). In Article V the targeted gene approach was applied on a cohort of pediatric T-ALL patients, revealing mutations affecting signaling transduction (*JAK1*, *JAK3*, *NRAS*, *PI3KCA*, and *PTEN*), epigenetic regulators (*CREBBP*, *DNMT3A*, *EZH2*, *PHF6*, and *SETD2*), and transcription (*BCL11B*, *FBXW7*, *NOTCH1*, and *TCF3*).

Genetic aberrations in pediatric T-ALL – take-home message

So, as reviewed above, it is evident that aberrations in numerous genes have been described in T-ALL. Is there any take-home message based on this plethora of alterations? Yes, actually three types of aberration stand out as the genetic backbone in T-ALL, constituting a condensed take-home message: i) TCR translocations: T-ALL characteristic, signature-creating aberrations resulting in deregulated expression of various oncogenes, mainly transcription factors; ii) *NOTCH1* mutations: play a critical oncogenic role in T-ALL; iii) *CDKN2A* deletions: the most common aberration in T-ALL.

Epigenetic alterations

Leukemogenesis is dependent on the co-operative effect of several genetic events targeting the DNA molecule through different mechanisms, as described above. However, it has been known for more than 30 years that epigenetic modifications, *i.e.*, alterations affecting gene expression without altering the DNA sequence, also contribute to tumorigenesis (Feinberg & Vogelstein, 1983). Epigenetic alterations may take place through different mechanisms, such as DNA methylation, histone modifications, and microRNA (Jiang *et al.*, 2013).

DNA methylation occurs by DNA methyltransferases (DNMTs) at cytosine (C) bases located 5′ to guanine (G) bases, referred to as CpG dinucleotides. These dinucleotides are often clustered in so called CpG islands, frequently located within gene promoter regions (Burke & Bhatla, 2014). *DNMT3A* mutations have recently, at a low frequency, been reported in T-ALL (Van Vlierberghe *et al.*, 2011; Article V). The cancer genome is globally hypomethylated, while certain CpG islands are hypermethylated. Patterns of methylation have been shown to be disease- and subtype-specific and to discriminate between BCP ALL and T-ALL (Figueroa *et al.*, 2013). Promoter methylation often results in gene silencing (Davidsson *et al.*, 2009). Hence, identification of hypermethylated promoters can point to tumor suppressor genes. In T-ALL, some of the genes thus identified are *ABCG2*, *CDKN1A*, and *PAX5* (Van der Meulen *et al.*, 2014). Epigenetic data on a number of promoters differentially methylated have been applied in order to propose a CpG island methylator phenotype (CIMP) delineating cases as hypermethylated (CIMP+) or hypomethylated (CIMP-) (Toyota *et al.*, 1999; Kraszewska *et al.*, 2012c). Indeed, characterizing T-ALLs according to the CIMP phenotype may have prognostic and therapeutic implications (Borssén *et al.*, 2013).

The polycomb repressive complex 2 (PRC2) effectuates histone methylation, resulting in transcriptional silencing. The PRC2 consists of EZH2, EED, and SUZ12 and loss-of-function mutations and deletions in the corresponding genes have been reported in T-ALL (Zhang *et al.*, 2012; Article V). Similarly, closed chromatin configuration can be achieved by a PHF6 complex and recurrent *PHF6* inactivating mutations and deletions occur in T-ALL (Van Vlierberghe *et al.*, 2010; Article V). Conversely, histone acetylation is often associated with a more open chromatin conformation, facilitating transcription. Mutations affecting genes involved in histone acetyltransferase, *EP300* and *CREBBP*, are also present in T-ALL (Zhang *et al.*, 2012; Article V). It is noteworthy that promoters differentially methylated in T-ALL are enriched for genes with histone modification capacity, such as PRC2 and *PHF6* (Kraszewska *et al.*, 2012c; Borssén *et al.*, 2013).

MicroRNAs are short, single-stranded, non-coding RNA molecules that interfere with mRNA and negatively affect protein translation. MicroRNA can function as both tumor suppressors and oncogenes, depending on their target genes (Mets *et al.*, 2014a, b). Since each microRNA may target several different mRNAs due to

incomplete complementarity, it is a challenging and arduous task to predict target genes for individual microRNAs (Esquela-Kerscher & Slack, 2006). A network of microRNAs (miR-20a, miR26a, miR-92, and miR223), displaying overlapping effects on tumor suppressor genes, has been implicated in T-ALL pathogenesis. Genes targeted, and hence deregulated, by these microRNAs have previously been implicated in T-ALL pathogenesis, namely *FBXW7*, *IKZF1*, *NF1*, *PHF6*, and *PTEN* (Mavrakis *et al.*, 2011). Interestingly, the expression of microRNAs may be regulated by epigenetic alterations, while at the same time microRNAs can give rise to an aberrant pattern of epigenetic regulators; hence, these factors are intertwined in the control of gene expression (Fabbri & Calin, 2010; Mets *et al.*, 2014).

Gene expression

One way to evaluate how the genotype (DNA) will affect the phenotype (protein) is to characterize gene expression (RNA). However, various subsequent post-transcriptional regulations may affect the accuracy of this approximation (Vogel & Marcotte, 2012). RNA can be analyzed by different methods, such as Northern blot, RT-PCR, RQ-PCR, microarray, and RNA-seq. Northern blot and RT-PCR analyses will mainly show the presence or absence of specific RNAs, while the other methods will add quantitative measurements as well as a potentially global perspective. RT-PCR amplifies cDNA, resulting from RNA reverse transcription, in a cyclic fashion through different temperature stages, enabling denaturation, primer hybridization, and elongation to take place. The efficacy of the PCR is enhanced by the high temperatures at which amplification is accomplished, made possible by heat stable DNA polymerases (Saiki *et al.*, 1988). The end product of RT-PCR will be measured and a positive result signifies that the target RNA was present in the examined sample. RQ-PCR incorporates RT-PCR but is designed to measure also quantitatively the product during PCR progression. The quantification process relies on fluorescent reporter molecules, such as Taqman® probes (Articles I and II). The Taqman® probe is designed with two flanking elements: a fluorescent reporter dye and a quencher. Once the probe has hybridized to the target sequence and the PCR proceeds, the probe will be cleaved. The fragmentation of the probe will separate the quencher from the dye, after which the fluorescent signal can be detected. The intensity of the fluorescence acts as a proxy for the amount of target present in the sample at that specific time. During RQ-PCR there will be an exponential phase, in which the intensity of the fluorescent signal is well beyond background noise; a threshold fluorescence value is established within this phase. The number of PCR cycles necessary to reach the threshold, the C_t value, is dependent both on the total amount of RNA present in the initial reaction mixture and on the proportion of target RNA of the total RNA. Housekeeping genes, such as genes encoding rRNA or genes involved in basic metabolism, are utilized as a reference to normalize the C_t

result. Differences in normalized C_t values between samples may be used to calculate variation in gene expression (Bustin, 2000; Jensen, 2012; Articles I and II).

The microarray-based technology was reviewed above (see section "SNP array analysis"). Applications designed to investigate gene expression consist of arrays containing immobilized probes, cDNA or oligonucleotides, representing human genes. Different procedures are available for amplifying and labeling nucleic acids subsequently hybridized to the microarrays. According to the method applied in Article I, RNA is extracted from the investigated sample, amplified, and converted to fluorescently labeled cDNA. In parallel, cDNA from a reference sample is differentially labeled compared with the investigated sample. The reference and the investigated cDNA samples are competitively hybridized to the microarray. The signals are normalized and analyzed, relating signal intensity for both dyes from all probes on the array. The ratio of fluorescence signals between the sample and the control at each array feature conveys the relative difference in gene expression. This difference can be referred to as fold-difference or transformed to a base 2 logarithm of the intensity ratio. This value can then be used for comparison between samples (Sinicropi *et al.*, 2007; Article I).

The technology applied for cDNA sequencing, or RNA-seq, was addressed above (see section "Next-generation sequencing"). The results obtained from RNA-seq overlap with those from microarray analysis in the sense that both methods are able to quantify the global gene expression pattern. However, RNA-seq also provides data on nucleotide substitutions, gene fusions, and alternative splicing. In addition, and in contrast to microarray-based technology, RNA-seq is not limited to investigating the expression of only genes for which there are probes on the array (Ozsolak & Milos, 2011).

The microarray technology has been used for comparison of global gene expression data in samples with known annotations, so called supervised methods. These methods are valuable to construct classifiers, *i.e.*, genes differentially expressed between distinct groups, for prediction of unknown samples to an already defined class. By the use of these methods it has been possible to create reliable classifiers to distinguish T-ALL from BCP ALL (Yeoh *et al.*, 2002). The MRD status in T-ALL can also be predicted already at diagnosis based on gene expression patterns (Andersson *et al.*, 2007). Furthermore, classifiers for known T-ALL-associated oncogenes, such as *TAL1*, *TLX1*, and *TLX3*, have been recognized (Ferrando *et al.*, 2002; Van Vlierberghe *et al.*, 2008b). Some of these gene expression signatures are indicative of leukemic arrest at specific stages of normal thymocyte development; cases with high levels of *TLX1* or *TAL1* expression seem to be stalled at the early cortical and late cortical stage, respectively (Ferrando *et al.*, 2002). In unsupervised methods no previous knowledge about the samples is used for analysis of gene expression profiles. Unsupervised analyses have revealed novel oncogenes, such as *HOXA*, *SET/NUP214*, *NKX2-1*, and *MEF2C* in T-ALL (Soulier *et al.*, 2005; Van Vlierberghe *et al.*, 2008b; Homminga *et al.*, 2011). The global property of microarray

analysis is ideal for a discovery phase, with new oncogenes being verified by RQ-PCR (Soulier *et al.*, 2005; Van Vlierberghe *et al.*, 2008b; Homminga *et al.*, 2011). RNA-seq analyses of T-ALL have identified new candidate driver genes, such as *H3FA* and *PTK2B*, classified cases to T-ALL subgroups based on gene expression, and have revealed novel fusion genes encoding activated kinases *e.g.*, *SSBP2*/*FER* and *TPM3*/*JAK2* (Kalender Atak *et al.*, 2013).

Tumor evolution

Initiating event

It has been shown that specific genetic aberrations have the potential to give rise to so called pre-leukemic clones. These clones have been ascribed several characteristics: i) the ability to differentiate into several lineages, indicating a hematopoietic stem cell phenotype; ii) the capacity, after acquisition of additional aberrations, to evolve into leukemia-initiating cells; iii) a propensity to survive chemotherapy, hence acting as a potential reservoir for relapses; and iiii) exist in both remisson samples as well as in samples from healthy individuals (Shlush *et al.*, 2014). Of all the aberrations and mutations detected in T-ALL, which have the capacity to initiate the leukemogenic process? Oncogenes deregulated by, *e.g.*, TCR translocations, often representing transcription factors or proteins involved in transcription complexes, such as *TAL1*, *LMO1/2*, *TLX1/3*, and *HOXA*, have been suggested to possess pre-leukemic potentials (Tremblay & Curtis, 2014). In fact, mouse models have been able to substantiate that *LMO2* expression induces aberrant self-renewal in committed T-cells, hence generating a pre-leukemic clone (McCormack *et al.*, 2010). *NOTCH1* activation has also been implicated as a potential mechanism to generate a pre-leukemic or leukemia-initiating phenotype (Armstrong *et al.*, 2009; Blackburn *et al.*, 2012). Nevertheless, low-level mutation frequencies of *NOTCH1* and T-ALLs being positive for *NOTCH1* at diagnosis but *NOTCH1*-negative at relapse indicate that *NOTCH1* mutations also can be secondary events (Mansour *et al.*, 2007; Clappier *et al.*, 2011; Article V). The cell of origin is also debated. T-ALL-initiating cells, identified through serial mouse transplantation assays, can be traced by the same, original, clonal TCR rearrangement, indicating a probable residue in the committed T-cell progenitor rather than in the hematopoietic stem cell compartment (Armstrong *et al.*, 2009). It has been shown that the majority of TCR translocations will, due to the order of somatic TCR rearrangement during T-cell differentiation, take place in immature, CD1a-negative, DN thymocytes. However, many subgroups of T-ALL have a much later differentiation arrest, at the cortical CD1a-positive stage. Hence, in some cases, the oncogene activation and differentiation arrest may be uncoupled, which would be expected given the necessity for the leukemic clone to accumulate additional aberrations (Le Noir *et al.*, 2012). Conversely, some data indicate that T-ALLs with an immature gene expression profile could evolve from more mature T-cells (Berquam-Vrieze *et al.*, 2011).

Paired diagnostic-relapse samples

Relapsed T-ALL may display one of four different genetic relationships to the major clone at diagnosis: i) identical clones; ii) clonal evolution; iii) evolution from an ancestral clone; and iiii) genetically "distinct" leukemia (Mullighan *et al.*, 2008). The most common clonal origin of a relapse appears to be evolution from an ancestral clone (Tosello *et al.*, 2009; Tzoneva *et al.*, 2013; Article V), often present as a minor clone at diagnosis (Mullighan *et al.*, 2008; Clappier *et al.*, 2011). However, even though the relapse clones frequently lose chromosomal markers present at diagnosis and acquire new aberrations, a remarkable genomic stability is noted during the disease progression (Tosello *et al.*, 2009; Article V). Ancestral clones responsible for relapses could in part overlap with the pre-leukemic and leukemia-initiating clones discussed above. It could well be that the relapses considered to be genetically distinct, *i.e.*, with no detected genetic similarities to the diagnostic clone, in fact stem from a common pre-leukemic clone where the initiating event remains unidentified. Relapses evolving from minor subclones could have clinical implications: eradicating the bulk of leukemic cells at diagnosis might not result in cure if therapy-resistant ancestral clones persist.

Tumor heterogeneity in a single leukemia sample

Various kinds of tumor heterogeneity may be considered: i) across different cancer types; ii) between patients with a specific cancer type; and iii) within a single tumor sample (Lawrence *et al.*, 2013; Vogelstein *et al.*, 2013); the latter kind of heterogeneity will be addressed here. As mentioned above, relapses often evolve from a minor clone present at the time of diagnosis. This, thus, indicates a level of clonal heterogeneity at diagnosis. Given that a leukemia originates from a single cell, evolution over time would naturally generate clonal heterogeneity. The clonal composition at diagnosis reflects the different clones proliferate power, and hence selective edge, in response to the biological circumstances at diagnosis. When external circumstances change, *e.g.*, by treatment, a clone different from the major clone at diagnosis could be benefitted and hence evolve to a relapse (Paulsson, 2013; Landau *et al.*, 2014). The possibility to detect genomic heterogeneity has dramatically improved by FISH, SNP array, and, in particular, NGS analyses. For example, in Article V, targeted deep sequencing detected *NOTCH1*, *FBXW7*, and *PTEN* mutations in minor subclones.

Prognostic relevance of T-ALL-associated aberrations

Approximately one in five pediatric T-ALL patients relapses, and for those that do, the prognosis is very dismal; only around 20% survive (Nguyen *et al.*, 2008). The optimal stratification strategy is to predict risk for relapse already at the time of

diagnosis and hence adjust treatment to prevent this from occurring (see section "Treatment, MRD, and prognosis" above). As outlined above, a wealth of data about T-ALL-associated genetic aberrations is now available, elucidating cellular features, such as specific signalling pathways, cell cycle regulators, and transcription factors, important for leukemogenesis. A crucial application of this genetic insight is to seek to improve treatment stratification protocols for T-ALL.

The genes most frequently targeted in T-ALL are *CDKN2A* and *NOTCH1*/*FBXW7*. Aberrations affecting these genes are detected in >70% and >50% of cases, respectively. *CDKN2A* deletions have not conclusively been shown to confer any clinical impact on survival (Ramakers-van Woerden *et al.*, 2001; Krieger *et al.*, 2010; Article V), whereas there are conflicting results with regard to the prognostic significance of *NOTCH1* mutations. Several studies have identified a favorable early therapy response in cases with *NOTCH1* mutations, something that, however, did only translate into improved outcome in some studies (Kox *et al.*, 2010; Zuurbier *et al.*, 2010). This could of course reflect differences in therapy between treatment protocols (Zuurbier *et al.*, 2010). A recent report found a significantly improved overall survival of pediatric T-ALLs with co-occurring *FBXW7* and *NOTCH1* mutations (Jenkinson *et al.*, 2013). However, a positive effect of double *NOTCH1* and *FBXW7* mutations could not be confirmed in Swedish T-ALL patients (Fogelstrand *et al.*, 2014; Article V); however, increased *NOTCH1* activity, as reflected by aberrant *HES1* expression, was associated with improved outcome in one of these cohorts (Fogelstrand *et al.*, 2014). Furthermore, a meta-analysis of 711 pediatric T-ALL patients did not find any correlation between *NOTCH1* mutations and event-free survival (Ma & Wu, 2012).

It has been suggested that specific oncogenes, such as *TAL1*, *LMO1/2*, *TLX1/3*, and *HOXA*, deregulated by TCR translocations, define mutually exclusive subgroups in T-ALL characterized by specific expression patterns (see section "Gene expression" above) (Van Vlierberghe *et al.*, 2008a). Hence, it is possible that the prognosis is mainly associated with these signature-generating genes and not with *CDKN2A* or *NOTCH1*/*FBXW7,* which are present in all T-ALL subgroups (Jenkinson *et al.*, 2013). Aberrant expression of *TAL1* and *TLX1* has been suggested to confer a more favorable prognosis, while *TLX3* has been associated with a dismal outcome in some, but not all, studies (Van Vlierberghe & Ferrando, 2012). Finally, we found that rare TCR translocations correlated with an inferior prognosis (Article III); however, this has, as of yet, not been confirmed in a separate study. In conclusion, no genetic aberration in T-ALL has so far been clearly shown to be an independent prognostic marker.

Synopsis of the articles included in the thesis

Articles I and **II** focus on molecular characterization of rare translocations involving TCR loci. These types of aberration are characteristic for T-ALL and have previously proved pivotal in the identification of genes implicated in T-cell oncogenesis. The first investigated translocation, t(12;14)(p13;q11), was shown to result in overexpression of *CCND2*. The t(12;14) is the first neoplasia-associated translocation shown to result in overexpression of *CCND2* and the first example of a targeted deregulation of a member of a cyclin-encoding gene family in T-ALL. Cyclin D proteins are crucial to the cell cycle machinery and hence potential oncogenes. The second translocation cloned, t(X;7)(q22;q34), had not previously been reported in a neoplastic disorder. Breakpoint analysis revealed *IRS4* as a novel translocation partner to a TCR locus, resulting in deregulated IRS4 expression, both at the gene and protein level. IRS4 plays an important role in several intracellular signalling cascades, including PI3K-AKT, known to be activated in T-ALL. In a subsequent work, I could show that *IRS4* is also targeted by alternative mechanisms in T-ALL, apart from TCR translocations, namely by mutations (**Article IV**).

In **Article III**, clinical characteristics and cytogenetic aberrations were ascertained and reviewed in a large, population-based Nordic series of 285 pediatric T-ALLs. The clinical characteristics of our Nordic cohort agreed well with previous larger series: a median age of 9 years, male predominance (male/female ratio 3.1), median WBC count of 66.5 x 10^9/l, and a high incidence of mediastinal mass and CNS involvement. Survival analyses revealed a correlation between rare TCR translocations and inferior outcome, an association that still awaits confirmation in a separate study.

Finally, I used several different techniques – such as FISH, SNP array, and deep sequencing of 75 selected candidate genes – to characterize co-operative genetic aberrations in a consecutive series of pediatric T-ALL (**Article V**). One common change identified by SNP array was sUPID. This aberration was seen in 44% of the investigated cases; most were sUPID9p that always were associated with homozygous *CDKN2A* deletions, with a heterozygous deletion occurring prior to the sUPID9p in all instances. Among the 75 genes investigated by deep sequencing, 14 were mutated in 28 cases. The genes targeted are involved in signalling transduction, epigenetic regulation, and transcription. In some cases, *NOTCH1* mutations were seen in minor

subclones and lost at relapse, showing that such mutations also can be secondary events.

In summary, through different approaches and by various methods, the articles included in this thesis have deciphered genetic aberrations in pediatric T-ALL, thus contributing to a better understanding of leukemogenesis.

Future Perspectives

The mountains representing genes altered frequently in T-ALL have been explored, and it is unlikely that any major peak has gone undiscovered. Numerous hills, *i.e.*, genes targeted infrequently, are known, but some still await exploration. However, the identification of genes rarely involved in T-ALL will probably not dramatically change the understanding of the cancer genetic landscape. So, which genetic roads are less travelled by and may hence make all the difference in the future?

There is currently a massive tide of sequence data sweeping across the genetic landscape. The majority of NGS studies has focused on protein altering aberrations. When this approach has reached its potential, it may be fruitful to learn more about alterations affecting the genetic sequence but not the amino acid code, so called synonymous or silent mutations. Even though these variants are not protein-altering, they may affect protein confirmation and expression, hence potentially contributing to oncogenesis. Whole genome NGS generates extensive data also on the 99% of the genome that is non-protein coding. It is now well recognized that these extensive DNA stretches are not, as it was once considered, "junk DNA"; instead the majority of the genome participates in biochemical RNA- and/or chromatin-associated events. Thus, also genetic alterations in non-coding regions could affect cellular functions and hence be of interest to address in cancer genomes.

What impact do DNA alterations have on the RNA level? Transcriptome analyses in T-ALL are so far limited in number, but will probably add valuable information on the phenotype of the leukemic cells in the future. In addition, are all the hills seen in the global genetic analyses part of a greater picture? Is it possible to transform the seemingly high level of mutation heterogeneity to a limited number of cellular pathways in the majority of T-ALL cases? If so, are there candidates in these pathways that are suitable to target for therapy? Already, a few crucial genes and signaling pathways in T-ALL are known and clinical trials with *NOTCH1* inhibitors are promising, although they have not proved to be the sole salvation. Hopefully, with increasing knowledge about aberrant signaling pathways, multiple specific genes may be targeted which could render synergistic effects against T-ALL. Alterations in genes involved in epigenetics have just recently been implicated in T-ALL. Still, the epigenetic patterns in T-ALL have been addressed in relatively few studies and to explore further this avenue might lead to new insights in T-ALL.

With new technical solutions constantly pushing the limit for what can be achieved and an ever growing knowledge to build on, there is good reason to believe that future efforts will dispel the haze that linger over the cancer genetic landscape in T-ALL.

Summary in Swedish

Den vanligaste cancerformen hos barn är leukemi, d.v.s. cancer i benmärg och blod. Det finns olika former av leukemi; den som oftast drabbar barn är akut lymfatisk leukemi (ALL). Cancercellerna vid ALL är en form av vita blodkroppar, s.k. lymfocyter. B-celler och T-celler är två typer av lymfocyter, vilka tillsammans utgör en viktig del av vårt immunförsvar. Då en omogen T-cell omvandlas till en malign cell kallas sjukdomen T-cells ALL (T-ALL) och pediatrisk T-ALL drabbar barn. Omkring 10 barn och ungdomar insjuknar i T-ALL varje år i Sverige. Vanliga symptom vid leukemi är trötthet, blekhet, blåmärken och återkommande infektioner. Cytostatika ges i drygt två år mot T-ALL och runt 70 % av patienterna blir botade.

Varför drabbas vissa barn av leukemi? Vi vet inte det säkert. Troligen är det så att en del barn har en medfödd genetisk sårbarhet för denna sjukdom och i kombination med olika omgivningsfaktorer, såsom infektioner, kan leukemi uppstå. Det vi vet är att all cancer, också leukemi, orsakas av genetiska förändringar i cellen. Celler ska dela sig, mogna till specialiserade celler (differentiera) och så småningom dö; dessa cellulära processer är strikt reglerade. Genetiska förändringar kan sätta cellens regelsystem ur funktion, något som kan leda till snabb celldelning, utebliven differentiering och förlängt liv av cellen; d.v.s. karakteristiska egenskaper för cancerceller. I de flesta fall behöver ett flertal specifika genetiska avvikelser ansamlas i en och samma cell för att den ska förändras till en cancercell.

Genetiska avvikelser kan påverka prognosen, d.v.s. sjukdomens förlopp, för barn med ALL. Möjligheten att kombinera flera typer av cytostatika och ge olika intensiv behandling beroende på vilka genetiska förändringar som ses vid diagnos har bidragit till att förbättra prognosen för barn med leukemi; på 1960-talet överlevde endast 5% av barnen, idag botas drygt 90%. Som framgår ovan har dock patienter med T-ALL en sämre överlevnad. Till viss del kan detta bero på att vi fortfarande inte säkert vet om det finns specifika genetiska förändringar som påverkar prognosen vid T-ALL. Detta innebär att det inte finns så stora möjligheter idag att justera behandlingen för barn med T-ALL.

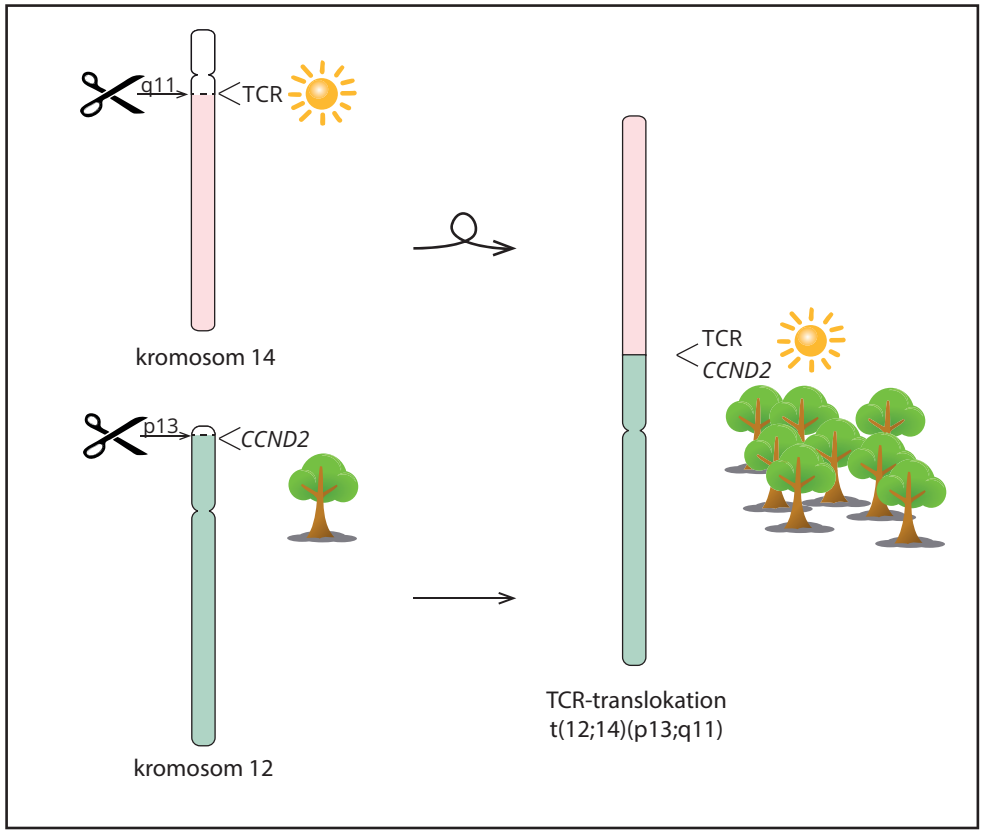
Mitt avhandlingsarbete har fokuserat på att beskriva och analysera genetiska avvikelser hos barn och ungdomar med T-ALL samt att undersöka om några genetiska förändringar kan användas för att bedöma prognosen.

T-ALL är en ovanlig sjukdom, men genom ett samarbete med Nordisk förening för pediatrisk hematologi och onkologi (NOPHO) har det varit möjligt att analysera data

från 285 barn och ungdomar som insjuknat i T-ALL mellan 1992 och 2006 i de Nordiska länderna. Fynden har dels lett till att vi kunde bekräfta att tidigare internationella resultat även gäller barn i Norden: T-ALL drabbar pojkar tre gånger så ofta som flickor, medianåldern vid insjuknandet är 9 år, patienterna har ofta ett högt antal vita blodkroppar i blodcirkulationen, många gånger återfinns en stor tumörmassa i mediastinum (utrymmet bakom bröstbenet) och leukemiceller har även ibland spridit sig till centrala nervsystemet. Studien har även givit ny kunskap om genetiska avvikelser och prognos. Hos hälften av patienterna ses genetiska förändringar i de sjuka cellerna vid en kromosomanalys. En återkommande avvikelse är en s.k. translokation. En translokation innebär att genetiskt material har bytt plats mellan två kromosomer (Figur 6). Vilket genetiskt material, d.v.s. vilka gener, som ingår i leukemiassocierade translokationer är inte slumpartat. Vid T-ALL är ofta gener som kodar för T-cellsreceptorer (TCR) involverade och dessa TCR-translokationer har betydelse för uppkomsten av T-ALL. I min studie fann vi att patienter med ovanliga TCR-translokationer tenderade att ha en sämre prognos (Artikel III).

Det som är speciellt intressant med TCR-translokationer är vilka gener, förutom TCR-generna, som är involverade. Den gen som via en translokation hamnar i närheten av en TCR-gen kommer ofta att få ett förhöjt uttryck, d.v.s. ett ökat antal gånger som genen används som mall för ett protein (Figur 6). En sådan gen kan genom ett förhöjt uttryck ofta ge cellen canceregenskaper, såsom snabb celldelning, utebliven differentiering och förlängt liv. Ett flertal gener som är involverade i TCR-translokationer är kända, men det finns fortfarande ovanliga TCR-translokationer där generna ännu inte är karakteriserade. Genom att beskriva tidigare okända gener som är involverade i T-ALL kan man få kunskap om biologiska mekanismer som är viktiga för att cancer ska uppstå samt förhoppningsvis även få ledtrådar till nya behandlingsformer. Som en del av mitt avhandlingsarbete har jag studerat två ovanliga TCR-translokationer.

Då en 14-årig flicka och en 1-årig pojke fick diagnosen T-ALL påvisade kromosomanalyserna en translokation mellan den korta armen, p, på kromosom 12, och den långa armen, q, på kromosom 14; t(12;14)(p13;q11) (Figur 6, Artikel I). Molekylärgenetiska analyser identifierade brottspunkter i TCR-genen samt i närheten av genen för cyklin D2 (*CCND2*). Analyser som undersöker geners uttryck kunde påvisa ett högt uttryck av *CCND2* i båda fallen med t(12;14) jämfört med T-ALL-fall som inte hade denna translokation. Detta är den första translokationen som har visats påverka uttrycket av cyklin D2. Cyklin D-proteiner medverkar till att cellen växer och delar sig. En ökad mängd av dessa proteiner har tidigare visats vara involverade i en rad olika cancerformer. Det pågår nu försök att behandla T-ALL genom att specifikt blockera cyklin D-proteiner.



Figur 6. Vad är en translokation?
Om två kromosomer klipps av och lagas med varandra bildas nya kromosomer, s.k. translokationer. Figuren illustrerar den TCR-translokation som beskrivs i Artikel I. Kromosomerna 14 och 12 klipps av (brottspunkter) i banden q11 respektive p13. En TCR-gen finns i brottspunkten på kromosom 14. Denna gen har en stark positiv inverkan på sin nära omgivning, symboliserat av en sol. *CCND2*, belägen nära brottspunkten på kromosom 12, är en gen som gör att cellen kan växa och dela sig, symboliserat av ett träd. Då TCR-translokationen t(12;14)(p13;q11) bildas hamnar TCR-genen och *CCND2* i närheten av varandra. Drivkraften, solen, som TCR-genen bär med sig kommer då att ha en positiv inverkan på, d.v.s. ge ett ökat uttryck av, *CCND2*. Det ökade uttrycket innebär att mer *CCND2* bildas vilket leder till att cellen växer och delar sig, illustrerat av skogen.

Den andra ovanliga TCR-translokationen som ingår i mitt avhandlingsarbet sågs vid diagnos hos en 12-årig pojke. Material mellan kromosomerna X och 7 hade bytt plats, t(X;7)(q22;q34), en förändring som inte tidigare hade rapporterats i T-ALL eller i någon annan cancerform. Brottspunkten i kromosom 7 visades vara i, eller i närheten av, en TCR-gen och X-brottspunkten var nära genen *IRS4* (Figur 4, Artikel II). Analyser visade att uttrycket av IRS4 var förhöjt på både gen- och proteinnivå i fallet med t(X;7) jämfört med andra T-ALL-fall utan denna translokation. Detta var första gången *IRS4* visats vara en partner i en TCR-translokation och överuttryckt i T-ALL.

Jag har i en kompletterande studie visat att *IRS4* även kan vara förändrad genom mutationer, d.v.s. skrivfel i DNA-strängen, i T-ALL (Artikel IV). IRS4 aktiveras i cellen av olika signalmolekyler och kan därefter aktivera ett flertal proteiner, såsom PI3K, vilka medverkar till ökad tillväxt och celldelning. Detta är egenskaper som selekteras för i cancerutvecklingen, något som talar för att IRS4 kan vara viktig för uppkomst av T-ALL.

Under de senaste tio åren har en dramatisk teknisk utveckling revolutionerat möjligheterna att undersöka vår arvsmassa. Två tekniker som gjort detta möjligt är SNP-array-analys och storskalig sekvensering. Gemensamt för dessa analysmetoder är att de ger stora mängder detaljrik data om hela vår arvsmassa. Jag har använt dessa samt ytterligare metoder för att genomföra en omfattande genetisk karakterisering av en serie med 47 pediatriska T-ALL-fall (Artikel V). Analyserna visade att leukemicellerna innehåller ett flertal genetiska avvikelser som påverkar viktiga reglersystem i cellen. De två gener som är förändrade i en majoritet av alla fall med T-ALL är *CDKN2A* och *NOTCH1*. Vi kunde påvisa en vanlig mekanism som ger upphov till förlust av *CDKN2A* i T-ALL hos barn samt ge mer data kring betydelsen av *NOTCH1*-förändringar för uppkomsten av T-ALL. Resultaten visar att cancerutvecklingen är en dynamisk process där genetiska förändringar kan uppstå genom en rad olika mekanismer samt att mutationer kan finnas i en större eller mindre andel av leukemicellerna och att återfall av sjukdomen ofta uppstår ur en forntida del av cancern.

Sammanfattningsvis har jag i mitt avhandlingsarbete genom en rad olika studier undersökt vilka genetiska förändringar som finns i leukemiceller hos barn med T-ALL. Mitt och många andra forskares arbete har lett till att sjukdomsassocierade förändringar i gener i T-ALL nu är tämligen väl kartlagda. Det som framöver behöver belysas är genetiska förändringar som finns i den stora del av vår arvsmassa som inte är gener och som därmed inte ger upphov till protein men som trots det skulle kunna påverka leukemiutvecklingen. Ytterligare undersökningar behöver också utföras av det s.k. epigenetiska mönstret i leukemiceller. Epigenetik innebär att genuttryck kan regleras utan att DNA-strängen förändras. Slutligen, även om vi idag har kunskap om avvikelser i gener vet vi inte så mycket om i vilken utsträckning dessa förändringar verkligen översätts till proteiner. Att fortsätta studera genuttryck och proteiner kan ge ledtrådar om vilka byggstenar som faktiskt finns i leukemiceller. Att finna genetiska avvikelser med prognostisk innebörd i T-ALL har visat sig vara svårt. Dock har den ökade kunskapen om genetiska avvikelser vid T-ALL lett till att man har identifierat några viktiga gener och signalvägar, t.ex. *NOTCH1* och gener som påverkar celldelningen såsom *CDKN2A* och cyklin D samt signalvägar involverande *PI3K*, som nästan alltid används av leukemicellen. Försök pågår nu att med olika läkemedel blockera effekten av förändringar i dessa gener, något som förhoppningsvis kan leda till effektivare behandling för patienter med T-ALL.

Acknowledgments

One of the things I enjoy most in life is to travel. In a way, the work with this thesis has been like a journey. I received my scientific passport in May 2007 when I was accepted as a PhD student. Since then I have learned a new language, in which words like illegitimate recombination and targeted re-sequencing are common. I have used many kinds of techniques, such as FISH, SNP array, Western blot, and NGS as means of transportation to reach genetic destinations. Taking the FISH trip was like sailing at night looking at the stars, Western blot was a bumpy ride, and NGS took me with the speed of a Concorde flight to completely new surroundings. Sometimes the destination was not what I expected or was already too crowded, but on a few occasions the journey took me to exciting places no one had ever seen before. I would then write letters to those left behind describing my discoveries, in the language I tried to master. But as always with travels, even though the sights are thrilling, the memories that linger are of the people you meet along the way. During my PhD expedition I have met many fantastic people who helped me when my luggage was heavy, joined me for a chat when the sights confused me, pointed me into the right direction when I was lost, and listened to my stories when I returned home. I want to mention a few who have been extra important to me on my road towards this thesis:

Bertil Johansson, my supervisor. I could not have wished for a better guide for this scientific expedition! You organized the itinerary, explained the route to me, but allowed me to test the path on my own. Sometimes I missed a connecting flight (I had a baby) or lost my passport (much work at the clinic), but you always waited patiently for me to return. When my destination turned out to be mentioned in the guidebook already, you encouraged me to keep looking for the unknown. And when I actually reached land unheard of, you shared my enthusiasm for the new discoveries. You were fantastic in phrasing the scientific letters, always finding the right words to elucidate the salient findings of our endeavors. You have been extremely generous with your time, giving me the opportunity to discuss with you reports from all parts of the cancer genetic landscape, teaching me how to learn from them. This journey would not have been the same without you. Thank you.

Kajsa Paulsson and **Thoas Fioretos**, my co-supervisors. I appreciate your interest in and support of my projects. Kajsa, I especially value your help with the SNP maps, deciphering the hills and valleys hidden there.

Felix Mitelman, founder of the Department of Clinical Genetics. Felix, you have more stamps in your scientific passport than most of us can dream of. Your amount of frequent flyer points entitles you to sit in the executive lounge and sip Champagne. But instead, you join me and my colleagues daily in the general waiting area, patienthotellet, and I greatly enjoy all our discussions: from science, to literature, to how to grow the perfect lawn.

Catarina, **Fredrik**, **Hans**, **Jacob**, **Kajsa**, **Maria**, **Nina**, **Samuel,** and **Ulf**, clinical colleagues. You have generously and with great knowledge guided me on my parallel but intertwined journey, in the field of clinical genetics. But maybe more important in this context, you made me forget about my travelling nerves. Over coffee we discuss small and big things, you make me laugh, we dance together at conference parties, and we have interesting lunch talks. I feel privileged to share my working days with you.

Linda and **Setareh**, PhD colleagues. You have been such generous and fun travel companions. The trip to San Francisco was amazing. Linda, when I got all confused in the sequencing landscape you always turned up to show me the way ahead. I also want to thank **Eleanor**, **Kristina**, and **Rebeqa**, in my research group, for encouragement along the way. Principle investigators, postdocs, PhD students, technicians, and other personnel at the **Department of Clinical Genetics**, you constitute a fantastic scientific homeland and it has been very educational to hear your reports from various foreign destinations during the Friday seminars. **Anette**, thank you for keeping track of all the logistics that comes with travelling.

Andrea, **Bodil**, **Carin**, and **Margareth**, laboratory engineer and technicians. It is an art to make the techniques run smoothly to the places you want to go. You master this art, thank you for all your assistance.

Anna, **Emma**, **Nisse**, **Petra**, and **Tord**, colleagues and hospital geneticists. Thank you for all the time you have taken to explain the wonders of chromosome banding, FISH, SNP array, and sequence analyses to me. Your expertise has been an important part of my guidebook to this adventure.

Everyone at the Clinic. In your competent care I have gradually grown into a clinical geneticist. Thank you for helping me along the way.

ST-ALF and **Barncancerfonden**, thank you for the financial contributions that made the new discoveries possible.

Mamma och **Pappa**. Min färd mot avhandlingen har tagit lång tid men utan all er hjälp hade den tagit ännu mycket längre tid. Tack för alla gånger ni hämtat och passat barnen och för att jag vet att de alltid har det så fint hos er. Det är också med er jag

gjort mina första resor, ofta på upptäcktsfärd i Europa. Er entusiasm för resande bär jag nu med mig.

Annika och **Ingrid**, mina systrar. Tillsammans med er och era fantastiska familjer; **Jakob**, **Jonas**, **Erik**, **Klara**, **Sofia** och **Nikolaj** hämtar jag krafter som behövs för vetenskaplig transit. Speciellt tack för alla sköna dagar vi njutit tillsammans i Steninge.

Malin, min svägerska. Att resa tar på krafterna och det är av stor vikt att vara i god fysisk form. Du tar ditt uppdrag som min PT på stort allvar. Jag vet att utgångsläget inte gett dig mycket att jobba med, men tack för att du inte ger upp.

Cissi, min bästa vän. Vi har rest tillsammans, till geografiska och litterära platser. Jag uppskattar alltid ditt sällskap.

Fredric, min man. Du är min främste resekamrat, i livet och runt jorden. Tack för att du har gett mig utrymme att göra denna forskningsfärd på egen hand. Och tack för att du har gjort bilderna från mina destinationer så bra.

Anton och **Emma**, mina fina barn. Den nya resan är tillsammans med er. Några ord på färden:

> Alla de försiktiga med långa håvar
> träffar havets jätteskratt.
> Vänner, vad söker ni på stranden?
> kunskap kan aldrig fångas,
> kan aldrig ägas.
>
> Men om du rak som en droppe
> faller i havet att upplösas,
> färdig för all förvandling –
> då skall du vakna med pärlemorhud
> och gröna ögon
> på ängar där havets hästar betar
> och vara kunskap.
>
> Karin Boye

References

Adamson P, Law G, and Roman E. (2005) Assessment of trends in childhood cancer incidence. *Lancet* 365: 753.

Aifantis I, Raetz E, and Buonamici S. (2008) Molecular pathogenesis of T-cell leukaemia and Lymphoma. *Nat Rev Immunol* 8: 380-390.

Andersson A, Ritz C, Lindgren D, Edén P, Lassen C, Heldrup J, Olofsson T, Råde J, Fontes M, Porwit-Macdonald A, Behrendtz M, Höglund M, Johansson B, and Fioretos T. (2007) Microarray-based classification of a consecutive series of 121 childhood acute leukemias: prediction of leukemic and genetic subtype as well as of minimal residual disease status. *Leukemia* 21: 1198-1203.

Armstrong F, Brunet de la Grange P, Gerby B, Rouyez MC, Calvo J, Fontenay M, Boissel N, Dombret H, Baruchel A, Landman-Parker J, Roméo PH, Ballerini P, and Pflumio F. (2009) NOTCH is a key regulator of human T-cell acute leukemia initiating cell activity. *Blood* 113: 1730-1740.

Balgobind BV, Van Vlierberghe P, van den Ouweland AM, Beverloo HB, Terlouw-Kromosoeto JN, van Wering ER, Reinhardt D, Horstmann M, Kaspers GJ, Pieters R, Zwaan CM, Van den Heuvel-Eibrink MM, and Meijerink JPP. (2008) Leukemia-associated *NF1* inactivation in patients with pediatric T-ALL and AML lacking evidence for neurofibromatosis. *Blood* 111: 4322-4328.

Balmain A. (2001) Cancer genetics: from Boveri and Mendel to microarrays. *Nat Rev Cancer* 1: 77-82.

Bandapalli OR, Schuessele S, Kunz JB, Rausch T, Stütz AM, Tal N, Geron I, Gershman N, Izraeli S, Eilers J, Vaezipour N, Kirschner-Schwabe R, Hof J, von Stackelberg A, Schrappe M, Stanulla M, Zimmermann M, Koehler R, Avigad S, Handgretinger R, Frismantas V, Bourquin JP, Bornhauser B, Korbel JO, Muckenthaler MU, and Kulozik AE. (2014) The activating STAT5B N642H mutation is a common abnormality in pediatric T-cell acute lymphoblastic leukemia and confers a higher risk of relapse. *Haematologica* Published online June 27, 2014.

Bar-Eli M, Ahuja H, Foti A, and Cline MJ. (1989) N-RAS mutations in T-cell acute lymphocytic leukaemia: analysis by direct sequencing detects a novel mutation. *Br J Haematol* 72: 36-39.

Basso K, Mussolin L, Lettieri A, Brahmachary M, Lim WK, Califano A, Basso G, Biondi A, Cazzaniga G, and Rosolen A. (2011) T-cell lymphoblastic lymphoma shows differences and similarities with T-cell acute lymphoblastic leukemia by genomic and gene expression analyses. *Genes Chromosomes Cancer* 50: 1063-1075.

Bene MC, Castoldi G, Knapp W, Ludwig WD, Matutes E, Orfao A, and van't Veer MB, European Group for the Immunological Characterization of Leukemias (EGIL). (1995) Proposals for the immunological classification of acute leukemias. *Leukemia* 9: 1783-1786.

Bennett JH. (1845) Case of hypertrophy of the spleen and liver, in which death took place from suppuration of the blood. *Edingburgh Medical and Surgical Journal* 64: 413-423.

Bernbeck B, Wuller D, Janssen G, Wessalowski R, Göbel U, and Schneider DT. (2009) Symptoms of childhood acute lymphoblastic leukemia: red flags to recognize leukemia in daily practice. *Klin Padiatr* 221: 369-373.

Berquam-Vrieze KE, Nannapaneni K, Brett BT, Holmfeldt L, Ma J, Zagorodna O, Jenkins NA, Copeland NG, Meyerholz DK, Knudson CM, Mullighan CG, Scheetz TE, and Dupuy AJ. (2011) Cell of origin strongly influences genetic selection in a mouse model of T-ALL. *Blood* 118: 4646-4656.

Bishop R. (2010) Application of fluorescence *in situ* hybridization (FISH) in detecting genetic aberrations of medical significance. *Bioscience Horizons* 3: 85-95.

Blackburn JS, Liu S, Raiser DM, Martinez SA, Feng H, Meeker ND, Gentry J, Neuberg D, Look AT, Ramaswamy S, Bernards A, Trede NS, and Langenau DM. (2012) Notch signaling expands a pre-malignant pool of T-cell acute lymphoblastic leukemia clones without affecting leukemia-propagating cell frequency. *Leukemia* 26: 2069-2078.

Borella L, and Sen L. (1973) T cell surface markers on lymphoblasts from acute lymphocytic leukemia. *J Immunol* 111: 1257-1260.

Borssén M, Palmqvist L, Karrman K, Abrahamsson J, Behrendtz M, Heldrup J, Forestier E, Roos G, and Degerman S. (2013) Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. *PLoS One* 8: e65373.

Boveri T. (1904) Ergebnisse über die Konstitution der Chromatischen Substanz des Zelkerns. *Gustav Fischer Verlag, Jena*.

Boveri T. (1914) Zur Frage der Entstehung maligner Tumoren. *Gustav Fischer Verlag, Jena*.

Braun CJ, Boztug K, Paruzynski A, Witzel M, Schwarzer A, Rothe M, Modlich U, Beier R, Göhring G, Steinemann D, Fronza R, Ball CR, Haemmerle R, Naundorf S, Kühlcke K, Rose M, Fraser C, Mathias L, Ferrari R, Abboud MR, Al-Herz W, Kondratenko I, Maródi L, Glimm H, Schlegelberger B, Schambach A, Albert MH, Schmidt M, von Kalle C, and Klein C. (2014) Gene therapy for Wiskott-Aldrich syndrome--long-term efficacy and genotoxicity. *Sci Transl Med* 6: 227ra233.

Brüggemann M, van der Velden VHJ, Raff T, Droese J, Ritgen M, Pott C, Wijkhuijs AJ, Gökbuget N, Hoelzer D, van Wering ER, van Dongen JJM, and Kneba M. (2004) Rearranged T-cell receptor beta genes represent powerful targets for quantification of minimal residual disease in childhood and adult T-cell acute lymphoblastic leukemia. *Leukemia* 18: 709-719.

Bunting SF, and Nussenzweig A. (2013) End-joining, translocations and cancer. *Nat Rev Cancer* 13: 443-454.

Burke MJ, and Bhatla T. (2014) Epigenetic modifications in pediatric acute lymphoblastic leukemia. *Front Pediatr* 2: 1-7.

Burkhardt B. (2010) Paediatric lymphoblastic T-cell leukaemia and lymphoma: one or two diseases? *Br J Haematol* 149: 653-668.

Bustin SA. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 25: 169-193.

Campana D. (2010) Minimal residual disease in acute lymphoblastic leukemia. *Hematology Am Soc Hematol Educ Program* 2010: 7-12.

Cantor H, and Boyse EA. (1975) Functional subclasses of T-lymphocytes bearing different Ly antigens. I. The generation of functionally distinct T-cell subclasses is a differentiative process independent of antigen. *J Exp Med* 141: 1376-1389.

Caspersson T, Zech L, and Johansson C. (1970) Differential binding of alkylating fluorochromes in human chromosomes. *Exp Cell Res* 60: 315-319.

Cauwelier B, Dastugue N, Cools J, Poppe B, Herens C, De Paepe A, Hagemeijer A, and Speleman F. (2006) Molecular cytogenetic study of 126 unselected T-ALL cases reveals high incidence of *TCRβ* locus rearrangements and putative new T-cell oncogenes. *Leukemia* 20: 1238-1244.

Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, Murphree AL, Strong LC, and White RL. (1983) Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 305: 779-784.

Clappier E, Cuccuini W, Cayuela JM, Vecchione D, Baruchel A, Dombret H, Sigaux F, and Soulier J. (2006) Cyclin D2 dysregulation by chromosomal translocations to TCR loci in T-cell acute lymphoblastic leukemias. *Leukemia* 20: 82-86.

Clappier E, Cuccuini W, Kalota A, Crinquette A, Cayuela JM, Dik WA, Langerak AW, Montpellier B, Nadel B, Walrafen P, Delattre O, Aurias A, Leblanc T, Dombret H, Gewirtz AM, Baruchel A, Sigaux F, and Soulier J. (2007) The *C-MYB* locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. *Blood* 110: 1251-1261.

Clappier E, Gerby B, Sigaux F, Delord M, Touzri F, Hernandez L, Ballerini P, Baruchel A, Pflumio F, and Soulier J. (2011) Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. *J Exp Med* 208: 653-661.

Clarke S, O'Reilly J, Romeo G, and Cooney J. (2011) NUP214-ABL1 positive T-cell acute lymphoblastic leukemia patient shows an initial favorable response to imatinib therapy post relapse. *Leuk Res* 35: e131-133.

Coustan-Smith E, Mullighan CG, Onciu M, Behm FG, Raimondi SC, Pei D, Cheng C, Su X, Rubnitz JE, Basso G, Biondi A, Pui CH, Downing JR, and Campana D. (2009) Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol* 10: 147-156.

Cuevas EP, Escribano O, Chiloeches A, Ramirez Rubio S, Román ID, Fernández-Moreno MD, and Guijarro LG. (2007) Role of insulin receptor substrate-4 in IGF-I-stimulated HEPG2 proliferation. *J Hepatol* 46: 1089-1098.

Davidsson J, Lilljebjörn H, Andersson A, Veerla S, Heldrup J, Behrendtz M, Fioretos T, and Johansson B. (2009) The DNA methylome of pediatric acute lymphoblastic leukemia. *Hum Mol Genet* 18: 4054-4065.

De Keersmaecker K, Real PJ, Gatta GD, Palomero T, Sulis ML, Tosello V, Van Vlierberghe P, Barnes K, Castillo M, Sole X, Hadler M, Lenz J, Aplan PD, Kelliher M, Kee BL, Pandolfi PP, Kappes D, Gounari F, Petrie H, Van der Meulen J, Speleman F, Paietta E, Racevskis J, Wiernik PH, Rowe JM, Soulier J, Avran D, Cavé H, Dastugue N, Raimondi S, Meijerink JPP, Cordon-Cardo C, Califano A, and Ferrando AA. (2010) The *TLX1* oncogene drives aneuploidy in T cell transformation. *Nat Med* 16: 1321-1327.

De Keersmaecker K, Atak ZK, Li N, Vicente C, Patchett S, Girardi T, Gianfelici V, Geerdens E, Clappier E, Porcu M, Lahortiga I, Lucá R, Yan J, Hulselmans G, Vranckx H, Vandepoel R, Sweron B, Jacobs K, Mentens N, Wlodarska I, Cauwelier B, Cloos J, Soulier J, Uyttebroeck A, Bagni C, Hassan BA, Vandenberghe P, Johnson AW, Aerts S, and Cools J. (2013) Exome sequencing identifies mutation in *CNOT3* and ribosomal genes *RPL5* and *RPL10* in T-cell acute lymphoblastic leukemia. *Nat Genet* 45: 186-190.

Degos L. (2001) John Hughes Bennett, Rudolph Virchow... and Alfred Donné: the first description of leukemia. *Hematol J* 2: 1.

den Braber I, Mugwagwa T, Vrisekoop N, Westera L, Mögling R, de Boer AB, Willems N, Schrijver EHR, Spierenburg G, Gaiser K, Mul E, Otto SA, Ruiter AFC, Ackermans MT, Miedema F, Borghans JAM, de Boer RJ, and Tesselaar K. (2012) Maintenance of peripheral naive T cells is sustained by thymus output in mice but not humans. *Immunity* 36: 288-297.

Donné A. (1844) Cours de microscopie complémentaire des études médicales. Anatomie microscopique et physiologie des fluides de l´Economie. *Paris Ballière* 135-136.

Eden T. (2010) Aetiology of childhood leukaemia. *Cancer Treat Rev* 36: 286-297.

Egerton M, Scollay R, and Shortman K. (1990) Kinetics of mature T-cell development in the thymus. *Proc Natl Acad Sci U S A* 87: 2579-2582.

Eguchi-Ishimae M, Eguchi M, Kempski H, and Greaves MF. (2008) *NOTCH*1 mutation can be an early, prenatal genetic event in T-ALL. *Blood* 111: 376-378.

Ehrlich P. (1879) Methodologische Beitrage zur Physiologie und Pathologie der verschisdenen Formen der Leukocyten. *Zeitschrift Klinische Medizinische* 1: 553-560.

Ellisen LW, Bird J, West DC, Soreng AL, Reynolds TC, Smith SD, and Sklar J. (1991) *TAN-1*, the human homolog of the Drosophila *Notch* gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell* 66: 649-661.

ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74.

Engel E. (1980) A new genetic concept: uniparental disomy and its potential effect, isodisomy. *Am J Med Genet* 6: 137-143.

Engelman JA. (2009) Targeting PI3K signalling in cancer: opportunities, challenges and limitations. *Nat Rev Cancer* 9: 550-562.

Esquela-Kerscher A, and Slack FJ. (2006) Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 6: 259-269.

Fabbri M, and Calin GA. (2010) Epigenetics and miRNAs in human cancer. *Adv Genet* 70: 87-99.

Fantin VR, Wang Q, Lienhard GE, and Keller SR. (2000) Mice lacking insulin receptor substrate 4 exhibit mild defects in growth, reproduction, and glucose homeostasis. *Am J Physiol Endocrinol Metab* 278: E127-133.

Feinberg AP, and Vogelstein B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301: 89-92.

Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, Behm FG, Pui CH, Downing JR, Gilliland DG, Lander ES, Golub TR, and Look AT. (2002) Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* 1: 75-87.

Fieuw A, De Wilde B, Speleman F, Vandesompele J, and De Preter K. (2012) Cancer gene prioritization for targeted resequencing using FitSNP scores. *PLoS One* 7: e31333.

Figueroa ME, Chen SC, Andersson AK, Phillips LA, Li Y, Sotzen J, Kundu M, Downing JR, Melnick A, and Mullighan CG. (2013) Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *J Clin Invest* 123: 3099-3111.

Fischer S, Mann G, Konrad M, Metzler M, Ebetsberger G, Jones N, Nadel B, Bodamer O, Haas OA, Schmitt K, and Panzer-Grümayer ER. (2007) Screening for leukemia- and clone-specific markers at birth in children with T-cell precursor ALL suggests a predominantly postnatal origin. *Blood* 110: 3036-3038.

Fitzgibbon J, Smith LL, Raghavan M, Smith ML, Debernardi S, Skoulakis S, Lillington D, Lister TA, and Young BD. (2005) Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Res* 65: 9152-9154.

Flex E, Petrangeli V, Stella L, Chiaretti S, Hornakova T, Knoops L, Ariola C, Fodale V, Clappier E, Paoloni F, Martinelli S, Fragale A, Sanchez M, Tavolaro S, Messina M, Cazzaniga G, Camera A, Pizzolo G, Tornesello A, Vignetti M, Battistini A, Cavé H, Gelb BD, Renauld JC, Biondi A, Constantinescu SN, Foà R, and Tartaglia M. (2008) Somatically acquired *JAK1* mutations in adult acute lymphoblastic leukemia. *J Exp Med* 205: 751-758.

Fogelstrand L, Staffas A, Wasslavik C, Sjögren H, Söderhäll S, Frost BM, Forestier E, Degerman S, Behrendtz M, Heldrup J, Karrman K, Johansson B, Heyman M, Abrahamsson J, and Palmqvist L. (2014) Prognostic implications of mutations in *NOTCH1* and *FBXW7* in childhood T-ALL treated according to the NOPHO ALL-1992 and ALL-2000 protocols. *Pediatr Blood Cancer* 61: 424-430.

Ford AM, Pombo-de-Oliveira MS, McCarthy KP, MacLean JM, Carrico KC, Vincent RF, and Greaves MF. (1997) Monoclonal origin of concordant T-cell malignancy in identical twins. *Blood* 89: 281-285.

Forestier E, Johansson B, Borgström G, Kerndrup G, Johansson J, and Heim S. (2000) Cytogenetic findings in a population-based series of 787 childhood acute lymphoblastic leukemias from the Nordic countries. *Eur J Haematol* 64: 194-200.

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.

Friedrich N. (1857) Ein neuer fall von Leukamie. *Virchows Arch path Anat* 12: 37-58.

Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, and Dryja TP. (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 323: 643-646.

Fuller HW. (1850) Encephaloid tumour of the abdomen. *Transactions of the Pathology Society of London* 4: 224-225.

Gale KB, Ford AM, Repp R, Borkhardt A, Keller C, Eden OB, and Greaves MF. (1997) Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc Natl Acad Sci U S A* 94: 13950-13954.

Gartner JJ, Parker SC, Prickett TD, Dutton-Regester K, Stitzel ML, Lin JC, Davis S, Simhadri VL, Jha S, Katagiri N, Gotea V, Teer JK, Wei X, Morken MA, Bhanot UK, Program NCS, Chen G, Elnitski LL, Davies MA, Gershenwald JE, Carter H, Karchin R, Robinson W, Robinson S, Rosenberg SA, Collins FS, Parmigiani G, Komar AA, Kimchi-Sarfaty C, Hayward NK, Margulies EH, and Samuels Y. (2013) Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proc Natl Acad Sci U S A* 110: 13481-13486.

Graux C, Cools J, Melotte C, Quentmeier H, Ferrando AA, Levine R, Vermeesch JR, Stul M, Dutta B, Boeckx N, Bosly A, Heimann P, Uyttebroeck A, Mentens N, Somers R, MacLeod RAF, Drexler HG, Look AT, Gilliland DG, Michaux L, Vandenberghe P, Wlodarska I, Marynen P, and Hagemeijer A. (2004) Fusion of *NUP214* to *ABL1* on amplified episomes in T-cell acute lymphoblastic leukemia. *Nat Genet* 36: 1084-1089.

Graux C, Cools J, Michaux L, Vandenberghe P, and Hagemeijer A. (2006) Cytogenetics and molecular genetics of T-cell acute lymphoblastic leukemia: from thymocyte to lymphoblast. *Leukemia* 20: 1496-1510.

Greaves MF. (1988) Speculation on the cause of childhood acute lymphoblastic leukemia. *Leukemia* 2: 120-125.

Greaves MF, and Wiemels J. (2003) Origins of chromosome translocations in childhood leukaemia. *Nat Rev Cancer* 3: 639-649.

Greaves MF. (2006) Infection, immune responses and the aetiology of childhood leukaemia. *Nat Rev Cancer* 6: 193-203.

Gutierrez A, Dahlberg SE, Neuberg DS, Zhang J, Grebliunaite R, Sanda T, Protopopov A, Tosello V, Kutok J, Larson RS, Borowitz MJ, Loh ML, Ferrando AA, Winter SS, Mullighan CG, Silverman LB, Chin L, Hunger SP, Sallan SE, and Look AT. (2010a) Absence of biallelic *TCRγ* deletion predicts early treatment failure in pediatric T-cell acute lymphoblastic leukemia. *J Clin Oncol* 28: 3816-3823.

Gutierrez A, Sanda T, Ma W, Zhang J, Grebliunaite R, Dahlberg S, Neuberg D, Protopopov A, Winter SS, Larson RS, Borowitz MJ, Silverman LB, Chin L, Hunger SP, Jamieson C, Sallan SE, and Look AT. (2010b) Inactivation of *LEF1* in T-cell acute lymphoblastic leukemia. *Blood* 115: 2845-2851.

Gutierrez A, Kentsis A, Sanda T, Holmfeldt L, Chen SC, Zhang J, Protopopov A, Chin L, Dahlberg SE, Neuberg DS, Silverman LB, Winter SS, Hunger SP, Sallan SE, Zha S, Alt FW, Downing JR, Mullighan CG, and Look AT. (2011) The *BCL11B* tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. *Blood* 118: 4169-4173.

Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, Sorensen R, Forster A, Fraser P, Cohen JI, de Saint Basile G, Alexander I, Wintergerst U, Frebourg T, Aurias A, Stoppa-Lyonnet D, Romana S, Radford-Weiss I, Gross F, Valensi F, Delabesse E, Macintyre E, Sigaux F, Soulier J, Leiva LE, Wissler M, Prinz C, Rabbitts TH, Le Deist F, Fischer A, and Cavazzana-Calvo M. (2003) *LMO2*-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302: 415-419.

Harris H. (2008) Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J Cell Sci* 121 Suppl 1: 1-84.

Hedrick SM, Cohen DI, Nielsen EA, and Davis MM. (1984) Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* 308: 149-153.

Heerema NA, Sather HN, Sensel MG, Kraft P, Nachman JB, Steinherz PG, Lange BJ, Hutchinson RS, Reaman GH, Trigg ME, Arthur DC, Gaynon PS, and Uckun FM. (1998) Frequency and clinical significance of cytogenetic abnormalities in pediatric T-lineage acute lymphoblastic leukemia: a report from the Children's Cancer Group. *J Clin Oncol* 16: 1270-1278.

Heinrichs S, Li C, and Look AT. (2010) SNP array analysis in hematologic malignancies: avoiding false discoveries. *Blood* 115: 4157-4161.

Herranz D, Ambesi-Impiombato A, Palomero T, Schnell SA, Belver L, Wendorff AA, Xu L, Castillo-Martin M, Llobet-Navas D, Cordon-Cardo C, Clappier E, Soulier J, and Ferrando AA. (2014) A NOTCH1-driven *MYC* enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat Med* Published online September 7, 2014.

Hjalgrim LL, Rostgaard K, Schmiegelow K, Söderhäll S, Kolmannskog S, Vettenranta K, Kristinsson J, Clausen N, Melbye M, Hjalgrim H, and Gustafsson G. (2003) Age- and sex-specific incidence of childhood leukemia by immunophenotype in the Nordic countries. *J Natl Cancer Inst* 95: 1539-1544.

Homminga I, Pieters R, Langerak AW, de Rooi JJ, Stubbs A, Verstegen M, Vuerhard M, Buijs-Gladdines J, Kooi C, Klous P, van Vlierberghe P, Ferrando AA, Cayuela JM, Verhaaf B, Beverloo HB, Horstmann M, de Haas V, Wiekmeijer AS, Pike-Overzet K, Staal FJT, de Laat W, Soulier J, Sigaux F, and Meijerink JPP. (2011) Integrated transcript and genome analyses reveal *NKX2-1* and *MEF2C* as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell* 19: 484-497.

Hori S, Nomura T, and Sakaguchi S. (2003) Control of regulatory T cell development by the transcription factor *Foxp3*. *Science* 299: 1057-1061.

Hoxhaj G, Dissanayake K, and MacKintosh C. (2013) Effect of IRS4 levels on PI 3-kinase signalling. *PLoS One* 8: e73327.

Hruŝák O, Trka J, Zuna J, Poloučková A, Kalina T, and Starý J. (2002) Acute lymphoblastic leukemia incidence during socioeconomic transition: selective increase in children from 1 to 4 years. *Leukemia* 16: 720-725.

Janssen JW, Ludwig WD, Sterry W, and Bartram CR. (1993) SIL-TAL1 deletion in T-cell acute lymphoblastic leukemia. *Leukemia* 7: 1204-1210.

Jenkinson S, Koo K, Mansour MR, Goulden N, Vora A, Mitchell C, Wade R, Richards S, Hancock J, Moorman AV, Linch DC, and Gale RE. (2013) Impact of NOTCH1/FBXW7 mutations on outcome in pediatric T-cell acute lymphoblastic leukemia patients treated on the MRC UKALL 2003 trial. *Leukemia* 27: 41-47.

Jensen EC. (2012) Real-time reverse transcription polymerase chain reaction to measure mRNA: use, limitations, and presentation of results. *Anat Rec* 295: 1-3.

Jiang Y, Hatzi K, and Shaknovich R. (2013) Mechanisms of epigenetic deregulation in lymphoid neoplasms. *Blood* 121: 4271-4279.

Kalender Atak Z, De Keersmaecker K, Gianfelici V, Geerdens E, Vandepoel R, Pauwels D, Porcu M, Lahortiga I, Brys V, Dirks WG, Quentmeier H, Cloos J, Cuppens H, Uyttebroeck A, Vandenberghe P, Cools J, and Aerts S. (2012) High accuracy mutation detection in leukemia on a selected panel of cancer genes. *PLoS One* 7: e38463.

Kalender Atak Z, Gianfelici V, Hulselmans G, De Keersmaecker K, Devasia AG, Geerdens E, Mentens N, Chiaretti S, Durinck K, Uyttebroeck A, Vandenberghe P, Wlodarska I, Cloos J, Foà R, Speleman F, Cools J, and Aerts S. (2013) Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. *PLoS Genet* 9: e1003997.

Kampen KR. (2012) The discovery and early understanding of leukemia. *Leuk Res* 36: 6-13.

Kang DH, Kim SH, Jun JW, Lee YW, Shin HB, Ahn JY, Hong DS, Lee YK, and Jeon BR. (2012) Simultaneous translocation of both *TCR* Loci (14q11) with rare partner loci (Xq22 and 12p13) in a case of T-lymphoblastic leukemia. *Ann Lab Med* 32: 220-224.

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 WD, 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.

Kinlen L. (1988) Evidence for an infective cause of childhood leukaemia: comparison of a Scottish New Town with nuclear reprocessing sites in Britain. *Lancet* 332: 1323-1327.

Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, and Wilson JF. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5: e13996.

Kleppe M, Lahortiga I, El Chaar T, De Keersmaecker K, Mentens N, Graux C, Van Roosbroeck K, Ferrando AA, Langerak AW, Meijerink JPP, Sigaux F, Haferlach T, Wlodarska I, Vandenberghe P, Soulier J, and Cools J. (2010) Deletion of the protein tyrosine phosphatase gene *PTPN2* in T-cell acute lymphoblastic leukemia. *Nat Genet* 42: 530-535.

Knudson AG. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68: 820-823.

Kondilis-Mangum HD, Shih HY, Mahowald G, Sleckman BP, and Krangel MS. (2011) Regulation of TCRβ allelic exclusion by gene segment proximity and accessibility. *J Immunol* 187: 6374-6381.

Kox C, Zimmermann M, Stanulla M, Leible S, Schrappe M, Ludwig WD, Koehler R, Tolle G, Bandapalli OR, Breit S, Muckenthaler MU, and Kulozik AE. (2010) The favorable effect of activating *NOTCH1* receptor mutations on long-term outcome in T-ALL patients treated on the ALL-BFM 2000 protocol can be separated from *FBXW7* loss of function. *Leukemia* 24: 2005-2013.

Kraszewska MD, Dawidowska M, Kosmalska M, Sędek L, Grzeszczak W, Szczepański T, and Witt M. (2012a) Immunoglobulin/T-cell receptor gene rearrangements in the diagnostic paradigm of pediatric patients with T-cell acute lymphoblastic leukemia. *Leuk Lymphoma* 53: 1425-1428.

Kraszewska MD, Dawidowska M, Szczepański T, and Witt M. (2012b) T-cell acute lymphoblastic leukaemia: recent molecular biology findings. *Br J Haematol* 156: 303-315.

Kraszewska MD, Dawidowska M, Larmonie NS, Kosmalska M, Sędek L, Szczepaniak M, Grzeszczak W, Langerak AW, Szczepański T, Witt M, and Polish Pediatric Leukemia Lymphoma Study G. (2012c) DNA methylation pattern is altered in childhood T-cell acute lymphoblastic leukemia patients as compared with normal thymic subsets: insights into CpG island methylator phenotype in T-ALL. *Leukemia* 26: 367-371.

Krieger D, Moericke A, Oschlies I, Zimmermann M, Schrappe M, Reiter A, and Burkhardt B. (2010) Frequency and clinical relevance of DNA microsatellite alterations of the *CDKN2A/B*, *ATM* and *p53* gene loci: a comparison between pediatric precursor T-cell lymphoblastic lymphoma and T-cell lymphoblastic leukemia. *Haematologica* 95: 158-162.

Lahortiga I, De Keersmaecker K, Van Vlierberghe P, Graux C, Cauwelier B, Lambert F, Mentens N, Beverloo HB, Pieters R, Speleman F, Odero MD, Bauters M, Froyen G, Marynen P, Vandenberghe P, Wlodarska I, Meijerink JPP, and Cools J. (2007) Duplication of the *MYB* oncogene in T cell acute lymphoblastic leukemia. *Nat Genet* 39: 593-595.

Landau DA, Carter SL, Getz G, and Wu CJ. (2014) Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia* 28: 34-43.

Lapunzina P, and Monk D. (2011) The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol Cell* 103: 303-317.

Larmonie NSD, Dik WA, Meijerink JPP, Homminga I, van Dongen JJM, and Langerak AW. (2013) Breakpoint sites disclose the role of the V(D)J recombination machinery in the formation of T-cell receptor (TCR) and non-TCR associated aberrations in T-cell acute lymphoblastic leukemia. *Haematologica* 98: 1173-1184.

Larson RC, Lavenir I, Larson TA, Baer R, Warren AJ, Wadman I, Nottage K, and Rabbitts TH. (1996) Protein dimerization between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. *EMBO J* 15: 1021-1027.

Lavan BE, Fantin VR, Chang ET, Lane WS, Keller SR, and Lienhard GE. (1997) A novel 160-kDa phosphotyrosine protein in insulin-treated embryonic kidney cells is a new member of the insulin receptor substrate family. *J Biol Chem* 272: 21403-21407.

Lavini C, Moran CA, Morandi U, and Schoenhuber R (Eds). (2008) Thymus Gland Pathology. Chapter 1, The thymus from antiquity to the present day: the history of a mysterious gland. *Springer* Verlag Italia 1-12.

Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortés ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau DA, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Lee RS, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CW, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES, and Getz G. (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499: 214-218.

Le Noir S, Ben Abdelali R, Lelorch M, Bergeron J, Sungalee S, Payet-Bornet D, Villarèse P, Petit A, Callens C, Lhermitte L, Baranger L, Radford-Weiss I, Grégoire MJ, Dombret H, Ifrah N, Spicuglia S, Romana S, Soulier J, Nadel B, Macintyre E, and Asnafi V. (2012) Extensive molecular mapping of TCRαδ- and TCRβ-involved chromosomal translocations reveals distinct mechanisms of oncogene activation in T-ALL. *Blood* 120: 3298-3309.

Lim JY, Bhatia S, Robison LL, and Yang JJ. (2014) Genomics of racial and ethnic disparities in childhood acute lymphoblastic leukemia. *Cancer* 120: 955-962.

Luc S, Luis TC, Boukarabila H, Macaulay IC, Buza-Vidas N, Bouriez-Jones T, Lutteropp M, Woll PS, Loughran SJ, Mead AJ, Hultquist A, Brown J, Mizukami T, Matsuoka S, Ferry H, Anderson K, Duarte S, Atkinson D, Soneji S, Domanski A, Farley A, Sanjuan-Pla A, Carella C, Patient R, de Bruijn M, Enver T, Nerlov C, Blackburn C, Godin I, and Jacobsen SEW. (2012) The earliest thymic T cell progenitors sustain B cell and myeloid lineage potential. *Nat Immunol* 13: 412-419.

Ma J, and Wu M. (2012) The indicative effect of Notch1 expression for the prognosis of T-cell acute lymphocytic leukemia: a systematic review. *Mol Biol Rep* 39: 6095-6100.

Maciejewski JP, and Mufti GJ. (2008) Whole genome scanning as a cytogenetic tool in hematologic malignancies. *Blood* 112: 965-974.

Maciejewski JP, Tiu RV, and O'Keefe C. (2009) Application of array-based whole genome scanning technologies as a cytogenetic tool in haematological malignancies. *Br J Haematol* 146: 479-488.

Makishima H, and Maciejewski JP. (2011) Pathogenesis and consequences of uniparental disomy in cancer. *Clin Cancer Res* 17: 3913-3923.

Mani RS, and Chinnaiyan AM. (2010) Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nat Rev Genet* 11: 819-829.

Mansour MR, Duke V, Foroni L, Patel B, Allen CG, Ancliff PJ, Gale RE, and Linch DC. (2007) Notch-1 mutations are secondary events in some patients with T-cell acute lymphoblastic leukemia. *Clin Cancer Res* 13: 6964-6969.

Mansur MB, Emerenciano M, Splendore A, Brewer L, Hassan R, Pombo-de-Oliveira MS. (2010) T-cell lymphoblastic leukemia in early childhood presents *NOTCH1* mutations and *MLL* rearrangements. *Leuk Res* 34: 483-486.

Mardilovich K, Pankratz SL, and Shaw LM. (2009) Expression and function of the insulin receptor substrate proteins in cancer. *Cell Commun Signal* 7: 14-29.

Mavrakis KJ, Van Der Meulen J, Wolfe AL, Liu X, Mets E, Taghon T, Khan AA, Setty M, Rondou P, Vandenberghe P, Delabesse E, Benoit Y, Socci NB, Leslie CS, Van Vlierberghe P, Speleman F, and Wendel HG. (2011) A cooperative microRNA-tumor suppressor gene network in acute T-cell lymphoblastic leukemia (T-ALL). *Nat Genet* 43: 673-678.

McCormack MP, Young LF, Vasudevan S, de Graaf CA, Codrington R, Rabbitts TH, Jane SM, and Curtis DJ. (2010) The *Lmo2* oncogene initiates leukemia in mice by inducing thymocyte self-renewal. *Science* 327: 879-883.

Mendes RD, Sarmento LM, Cante-Barrett K, Zuurbier L, Buijs-Gladdines JG, Póvoa V, Smits WK, Abecasis M, Yunes JA, Sonneveld E, Horstmann MA, Pieters R, Barata JT, and Meijerink JPP. (2014) *PTEN* micro-deletions in T-cell acute lymphoblastic leukemia are caused by illegitimate RAG-mediated recombination events. *Blood* 124: 567-578.

Mets E, Van Peer G, Van der Meulen J, Boice M, Taghon T, Goossens S, Mestdagh P, Benoit Y, De Moerloose B, Van Roy N, Poppe B, Vandesompele J, Wendel HG, Van Vlierberghe P, Speleman F, and Rondou P. (2014a) MicroRNA-128-3p is a novel oncomiR targeting *PHF6* in T-cell acute lymphoblastic leukemia. *Haematologica* 99: 1326-1333.

Mets E, Van der Meulen J, Van Peer G, Boice M, Mestdagh P, Van de Walle I, Lammens T, Goossens S, De Moerloose B, Benoit Y, Van Roy N, Clappier E, Poppe B, Vandesompele J, Wendel HG, Taghon T, Rondou P, Soulier J, Van Vlierberghe P, and Speleman F. (2014b) MicroRNA-193b-3p acts as a tumor suppressor by targeting the *MYB* oncogene in T-cell acute lymphoblastic leukemia. *Leukemia* Published online September 18, 2014.

Metzker ML. (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11: 31-46.

Meyerson M, Gabriel S, and Getz G. (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 11: 685-696.

Miller JFAP. (1961) Immunological function of the thymus. *Lancet* 2: 748-749.

Miller JFAP. (2002) The discovery of thymus function and of thymus-derived lymphocytes. *Immunol Reviews* 185: 7-14

Miller JFAP. (2011) The golden anniversary of the thymus. *Nat Rev Immunol* 11: 489-495.

Mitelman F, Johansson B, and Mertens F (Eds). (2014) Mitelman database of chromosome aberrations and gene fusions in cancer, (accessed May 30, 2014) http://cgap.nci.nih.gov/Chromosomes/Mitelman.

Mori N, Fujii M, Hinz M, Nakayama K, Yamada Y, Ikeda S, Yamasaki Y, Kashanchi F, Tanaka Y, Tomonaga M, and Yamamoto N. (2002) Activation of cyclin D1 and D2 promoters by human T-cell leukemia virus type I tax protein is associated with IL-2-independent growth of T cells. *Int J Cancer* 99: 378-385.

Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, Su X, Pui CH, 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, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, and Downing JR. (2008) Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* 322: 1377-1380.

Musgrove EA, Caldon CE, Barraclough J, Stone A, and Sutherland RL. (2011) Cyclin D as a therapeutic target in cancer. *Nat Rev Cancer* 11: 558-572.

Nemazee D. (2006) Receptor editing in lymphocyte development and central tolerance. *Nat Rev Immunol* 6: 728-740.

Neumann E. (1872) Ein neuer Fall von Leukamie mit Erkrankung des Knochenmarks. *Archives Heilkunde* 13: 502-508.

Nguyen K, Devidas M, Cheng SC, La M, Raetz EA, Carroll WL, Winick NJ, Hunger SP, Gaynon PS, Loh ML, and Children's Oncology G. (2008) Factors influencing survival after relapse from acute lymphoblastic leukemia: a Children's Oncology Group study. *Leukemia* 22: 2142-2150.

Nguyen-Khac F, Barin C, Chapiro E, Macintyre EA, Romana S, and Bernard OA. (2010) Cyclin D3 deregulation by juxtaposition with *IGH* locus in a t(6;14)(p21;q32)-positive T-cell acute lymphoblastic leukemia. *Leuk Res* 34: e13-14.

Nishana M, and Raghavan SC. (2012) Role of recombination activating genes in the generation of antigen receptor diversity and beyond. *Immunology* 137: 271-281.

Nowell PC, and Hungerford DA. (1960) Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst* 25: 85-109.

Ntziachristos P, Tsirigos A, Van Vlierberghe P, Nedjic J, Trimarchi T, Flaherty MS, Ferres-Marco D, da Ros V, Tang Z, Siegle J, Asp P, Hadler M, Rigo I, De Keersmaecker K, Patel J, Huynh T, Utro F, Poglio S, Samon JB, Paietta E, Racevskis J, Rowe JM, Rabadan R, Levine RL, Brown S, Pflumio F, Dominguez M, Ferrando AA, and Aifantis I. (2012) Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat Med* 18: 298-301.

Ozsolak F, and Milos PM. (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12: 87-98.

Paganin M, and Ferrando AA. (2011) Molecular pathogenesis and targeted therapies for NOTCH1-induced T-cell acute lymphoblastic leukemia. *Blood Rev* 25: 83-90.

Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, Kinsey SE, Lightfoot T, Roman E, Irving JA, Allan JM, Tomlinson IP, Taylor M, Greaves MF, and Houlston RS. (2009) Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* 41: 1006-1010.

Patrick K, Wade R, Goulden N, Mitchell C, Moorman AV, Rowntree C, Jenkinson S, Hough R, and Vora A. (2014) Outcome for children and young people with Early T-cell precursor acute lymphoblastic leukaemia treated on a contemporary protocol, UKALL 2003. *Br J Haematol* 166: 421-424.

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

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. (2013) Genomic heterogeneity in acute leukemia. *Cytogenet Genome Res* 139: 174-180.

Pear WS, Aster JC, Scott ML, Hasserjian RP, Soffer B, Sklar J, and Baltimore D. (1996) Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated *Notch* alleles. *J Exp Med* 183: 2283-2291.

Piller G. (2001) Leukaemia - a brief historical review from ancient times to 1950. *Br J Haematol* 112: 282-292.

Radtke F, Wilson A, Stark G, Bauer M, van Meerwijk J, MacDonald HR, and Aguet M. (1999) Deficient T cell fate specification in mice with an induced inactivation of *Notch1*. *Immunity* 10: 547-558.

Ramakers-van Woerden NL, Pieters R, Slater RM, Loonen AH, Beverloo HB, van Drunen E, Heyman M, Calero Moreno T, Rots MG, van Wering ER, Kamps WA, Janka-Schaub GE, and Veerman AJP. (2001) *In vitro* drug resistance and prognositic impact of $p16^{INK4A}$/$p15^{INK4B}$ deletions in childhood T-cell acute lymphoblastic leukaemia. *Br J Haematol* 112: 680-690.

Raschke S, Balz V, Efferth T, Schulz WA, and Florl AR. (2005) Homozygous deletions of *CDKN2A* caused by alternative mechanisms in various human cancer cell lines. *Genes Chromosomes Cancer* 42: 58-67.

Rizzo JM, and Buck MJ. (2012) Key principles and clinical applications of "*next-generation*" DNA sequencing. *Cancer Prev Res* 5: 887-900.

Rothenberg EV, Moore JE, and Yui MA. (2008) Launching the T-cell-lineage developmental programme. *Nat Rev Immunol* 8: 9-21.

Roti G, and Stegmaier K. (2014) New approaches to target T-ALL. *Front Oncol* 4: 170.

Rowley JD. (1973) A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* 243: 290-293.

Russell LJ, Enshaei A, Jones L, Erhorn A, Masic D, Bentley H, Laczko KS, Fielding AK, Goldstone AH, Goulden N, Mitchell CD, Wade R, Vora A, Moorman AV, and Harrison CJ. (2014) *IGH@* translocations are prevalent in teenagers and young adults with acute lymphoblastic leukemia and are associated with a poor outcome. *J Clin Oncol* 32: 1453-1462.

Safavi S, Forestier E, Golovleva I, Barbany G, Nord KH, Moorman AV, Harrison CJ, Johansson B, and Paulsson K. (2013) Loss of chromosomes is the primary event in near-haploid and low-hypodiploid acute lymphoblastic leukemia. *Leukemia* 27: 248-250.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, and Erlich HA. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487-491.

Sanger F, Nicklen S, and Coulson AR. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-5467.

Sawai CM, Freund J, Oh P, Ndiaye-Lobry D, Bretz JC, Strikoudis A, Genesca L, Trimarchi T, Kelliher MA, Clark M, Soulier J, Chen-Kiang S, and Aifantis I. (2012) Therapeutic targeting of the cyclin D3:CDK4/6 complex in T cell leukemia. *Cancer Cell* 22: 452-465.

Schatz DG, and Ji Y. (2011) Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol* 11: 251-263.

Schlimgen RJ, Reddy KL, Singh H, and Krangel MS. (2008) Initiation of allelic exclusion by stochastic interaction of *Tcrb* alleles with repressive nuclear compartments. *Nat Immunol* 9: 802-809.

Schlissel MS. (2003) Regulating antigen-receptor gene assembly. *Nat Rev Immunol* 3: 890-899.

Schmiegelow K, Forestier E, Hellebostad M, Heyman M, Kristinsson J, Söderhäll S, and Taskinen M. (2010) Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia* 24: 345-354.

Schmiegelow K, Lausten Thomsen U, Baruchel A, Pacheco CE, Pieters R, Pombo-de-Oliveira MS, Andersen EW, Rostgaard K, Hjalgrim H, and Pui CH. (2012) High concordance of subtypes of childhood acute lymphoblastic leukemia within families: lessons from sibships with multiple cases of leukemia. *Leukemia* 26: 675-681.

Schneider NR, Carroll AJ, Shuster JJ, Pullen DJ, Link MP, Borowitz MJ, Camitta BM, Katz JA, and Amylon MD. (2000) New recurring cytogenetic abnormalities and association of blast cell karyotypes with prognosis in childhood T-cell acute lymphoblastic leukemia: a pediatric oncology group report of 343 cases. *Blood* 96: 2543-2549.

Schrappe M, Valsecchi MG, Bartram CR, Schrauder A, Panzer-Grümayer R, Möricke A, Parasole R, Zimmermann M, Dworzak M, Buldini B, Reiter A, Basso G, Klingebiel T, Messina C, Ratei R, Cazzaniga G, Koehler R, Locatelli F, Schäfer BW, Aricó M, Welte K, van Dongen JJM, Gadner H, Biondi A, and Conter V. (2011) Late MRD response determines relapse risk overall and in subsets of childhood T-cell ALL: results of the AIEOP-BFM-ALL 2000 study. *Blood* 118: 2077-2084.

Schreyer S, Ledwig D, Rakatzi I, Klöting I, and Eckel J. (2003) Insulin receptor substrate-4 is expressed in muscle tissue without acting as a substrate for the insulin receptor. *Endocrinology* 144: 1211-1218.

Seif AE. (2011) Pediatric leukemia predisposition syndromes: clues to understanding leukemogenesis. *Cancer Genet* 204: 227-244.

Sen L, and Borella L. (1975) Clinical importance of lymphoblasts with T markers in childhood acute leukemia. *N Engl J Med* 292: 828-832.

Shanley DP, Aw D, Manley NR, and Palmer DB. (2009) An evolutionary perspective on the mechanisms of immunosenescence. *Trends Immunol* 30: 374-381.

Shendure J, and Ji H. (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135-1145.

Sherborne AL, Hosking FJ, Prasad RB, Kumar R, Koehler R, Vijayakrishnan J, Papaemmanuil E, Bartram CR, Stanulla M, Schrappe M, Gast A, Dobbins SE, Ma Y, Sheridan E, Taylor M, Kinsey SE, Lightfoot T, Roman E, Irving JAE, Allan JM, Moorman AV, Harrison CJ, Tomlinson IP, Richards S, Zimmermann M, Szalai C, Semsei AF, Erdelyi DJ, Krajinovic M, Sinnett D, Healy J, Gonzalez Neira A, Kawamata N, Ogawa S, Koeffler HP, Hemminki K, Greaves MF, and Houlston RS. (2010) Variation in *CDKN2A* at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nat Genet* 42: 492-494.

Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, Kennedy JA, Schimmer AD, Schuh AC, Yee KW, McLeod JL, Doedens M, Medeiros JJF, Marke R, Kim HJ, Lee K, McPherson JD, Hudson TJ, The HALT Pan-Leukemia Gene Panel Consortium, Brown AMK, Yousif F, Trinh QM, Stein LD, Minden MD, Wang JCY, and Dick JE. (2014) Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 506: 328-333.

Shull AY, Latham-Schwark A, Ramasamy P, Leskoske K, Oroian D, Birtwistle MR, and Buckhaults PJ. (2012) Novel somatic mutations to PI3K pathway genes in metastatic melanoma. *PLoS One* 7: e43369.

Sicinska E, Aifantis I, Le Cam L, Swat W, Borowski C, Yu Q, Ferrando AA, Levin SD, Geng Y, von Boehmer H, and Sicinski P. (2003) Requirement for cyclin D3 in lymphocyte development and T cell leukemias. *Cancer Cell* 4: 451-461.

Sinicropi D, Cronin M, and Liu M-L. (2007) Gene expression profiling utilizing microarray technology and RT-PCR. *BioMEMS and Biomedical Nanotechnology (Springer)* 2: 23-45.

Skok JA, Gisler R, Novatchkova M, Farmer D, de Laat W, and Busslinger M. (2007) Reversible contraction by looping of the *Tcra* and *Tcrb* loci in rearranging thymocytes. *Nat Immunol* 8: 378-387.

Smeets DFCM. (2004) Historical prospective of human cytogenetics: from microscope to microarray. *Clin Biochem* 37: 439-446.

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, and Hood LE. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* 321: 674-679.

Soulier J, Clappier E, Cayuela JM, Regnault A, Garcia-Peydró M, Dombret H, Baruchel A, Toribio ML, and Sigaux F. (2005) *HOXA* genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood* 106: 274-286.

Stanulla M, and Schrappe M. (2009) Treatment of childhood acute lymphoblastic leukemia. *Semin Hematol* 46: 52-63.

Stehelin D, Varmus HE, Bishop JM, and Vogt PK. (1976) DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* 260: 170-173.

Steliarova-Foucher E, Stiller C, Kaatsch P, Berrino F, Coebergh JW, Lacour B, and Parkin M. (2004) Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since the 1970s (the ACCIS project): an epidemiological study. *Lancet* 364: 2097-2105.

Stiller CA, and Parkin DM. (1996) Geographic and ethnic variations in the incidence of childhood cancer. *Br Med Bull* 52: 682-703.

Strauss R, Hamerlik P, Lieber A, and Bartek J. (2012) Regulation of stem cell plasticity: mechanisms and relevance to tissue biology and cancer. *Mol Ther* 20: 887-897.

Sulong S, Moorman AV, Irving JAE, Strefford JC, Konn ZJ, Case MC, Minto L, Barber KE, Parker H, Wright SL, Stewart ARM, Bailey S, Bown NP, Hall AG, and Harrison CJ. (2009) A comprehensive analysis of the *CDKN2A* gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. *Blood* 113: 100-107.

Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, and Vardiman JW (Eds). (2008) WHO Classification of Tumours of Haematopoietic and Lymphiod Tissues. *IARC Lyon.*

Szczepański T, Langerak AW, Willemse MJ, Wolvers-Tettero ILM, van Wering ER, and van Dongen JJM. (2000) T cell receptor gamma (*TCRG*) gene rearrangements in T cell acute lymphoblastic leukemia refelct 'end-stage' recombinations: implications for minimal residual disease monitoring. *Leukemia* 14: 1208-1214.

Szczepański T, van der Velden VHJ, Waanders E, Kuiper RP, Van Vlierberghe P, Gruhn B, Eckert C, Panzer-Grümayer R, Basso G, Cavé H, Stadt UZ, Campana D, Schrauder A, Sutton R, van Wering E, Meijerink JPP, and van Dongen JJM. (2011) Late recurrence of childhood T-cell acute lymphoblastic leukemia frequently represents a second leukemia rather than a relapse: first evidence for genetic predisposition. *J Clin Oncol* 29: 1643-1649.

Teachey DT, and Hunger SP. (2013) Predicting relapse risk in childhood acute lymphoblastic leukaemia. *Br J Haematol* 162: 606-620.

Tjio JH, and Levan A. (1956) The chromosome number of man. *Hereditas* 42: 1-6.

Toft N, Birgens H, Abrahamsson J, Bernell P, Griŝkevičius L, Hallböök H, Heyman M, Holm MS, Hulegårdh E, Klausen TW, Marquart HV, Jónsson OG, Nielsen OJ, Quist-Paulsen P, Taskinen M, Vaitkeviciene G, Vettenranta K, Åsberg A, and Schmiegelow K. (2013) Risk group assignment differs for children and adults 1-45 yr with acute lymphoblastic leukemia treated by the NOPHO ALL-2008 protocol. *Eur J Haematol* 90: 404-412.

Tonegawa S. (1983) Somatic generation of antibody diversity. *Nature* 302: 575-581.

Tosello V, Mansour MR, Barnes K, Paganin M, Sulis ML, Jenkinson S, Allen CG, Gale RE, Linch DC, Palomero T, Real P, Murty V, Yao X, Richards SM, Goldstone A, Rowe J, Basso G, Wiernik PH, Paietta E, Pieters R, Horstmann M, Meijerink JPP, and Ferrando AA. (2009) *WT1* mutations in T-ALL. *Blood* 114: 1038-1045.

Tosello V, and Ferrando AA. (2013) The NOTCH signaling pathway: role in the pathogenesis of T-cell acute lymphoblastic leukemia and implication for therapy. *Ther Adv Hematol* 4: 199-210.

Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, and Issa JP. (1999) CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* 96: 8681-8686.

Tremblay CS, and Curtis DJ. (2014) The clonal evolution of leukemic stem cells in T-cell acute lymphoblastic leukemia. *Curr Opin Hematol* 21: 320-325.

Tzoneva G, Perez-Garcia A, Carpenter Z, Khiabanian H, Tosello V, Allegretta M, Paietta E, Racevskis J, Rowe JM, Tallman MS, Paganin M, Basso G, Hof J, Kirschner-Schwabe R, Palomero T, Rabadan R, and Ferrando AA. (2013) Activating mutations in the *NT5C2* nucleotidase gene drive chemotherapy resistance in relapsed ALL. *Nat Med* 19: 368-371.

Urayama KY, Buffler PA, Gallagher ER, Ayoob JM, and Ma X. (2010) A meta-analysis of the association between day-care attendance and childhood acute lymphoblastic leukaemia. *Int J Epidemiol* 39: 718-732.

Ursø B, Ilondo MM, Holst PA, Christoffersen CT, Ouwens M, Giorgetti S, Van Obberghen E, Naor D, Tornqvist H, and De Meyts P. (2003) IRS-4 mediated mitogenic signalling by insulin and growth hormone in LB cells, a murine T-cell lymphoma devoid of IGF-I receptors. *Cell Signal* 15: 385-394.

Van Vlierberghe P, Meijerink JPP, Stam RW, van der Smissen W, van Wering ER, Beverloo HB, and Pieters R. (2005) Activating *FLT3* mutations in CD4+/CD8- pediatric T-cell acute lymphoblastic leukemias. *Blood* 106: 4414-4415.

Van Vlierberghe P, van Grotel M, Beverloo HB, Lee C, Helgason T, Buijs-Gladdines J, Passier M, van Wering ER, Veerman AJP, Kamps WA, Meijerink JPP, and Pieters R. (2006) The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of *LMO2* in pediatric T-cell acute lymphoblastic leukemia. *Blood* 108: 3520-3529.

Van Vlierberghe P, Pieters R, Beverloo HB, and Meijerink JPP. (2008a) Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. *Br J Haematol* 143: 153-168.

Van Vlierberghe P, van Grotel M, Tchinda J, Lee C, Beverloo HB, van der Spek PJ, Stubbs A, Cools J, Nagata K, Fornerod M, Buijs-Gladdines J, Horstmann M, van Wering ER, Soulier J, Pieters R, and Meijerink JPP. (2008b) The recurrent *SET-NUP214* fusion as a new *HOXA* activation mechanism in pediatric T-cell acute lymphoblastic leukemia. *Blood* 111: 4668-4680.

Van Vlierberghe P, Palomero T, Khiabanian H, Van der Meulen J, Castillo M, Van Roy N, De Moerloose B, Philippe J, Gonzalez-Garcia S, Toribio ML, Taghon T, Zuurbier L, Cauwelier B, Harrison CJ, Schwab C, Pisecker M, Strehl S, Langerak AW, Gecz J, Sonneveld E, Pieters R, Paietta E, Rowe JM, Wiernik PH, Benoit Y, Soulier J, Poppe B, Yao X, Cordon-Cardo C, Meijerink JPP, Rabadan R, Speleman F, and Ferrando AA. (2010) *PHF6* mutations in T-cell acute lymphoblastic leukemia. *Nat Genet* 42: 338-342.

Van Vlierberghe P, Ambesi-Impiombato A, Perez-Garcia A, Haydu JE, Rigo I, Hadler M, Tosello V, Della Gatta G, Paietta E, Racevskis J, Wiernik PH, Luger SM, Rowe JM, Rue M, and Ferrando AA. (2011) *ETV6* mutations in early immature human T cell leukemias. *J Exp Med* 208: 2571-2579.

Van Vlierberghe P, and Ferrando AA. (2012) The molecular basis of T cell acute lymphoblastic leukemia. *J Clin Invest* 122: 3398-3406.

Van der Meulen J, Van Roy N, Van Vlierberghe P, and Speleman F. (2014) The epigenetic landscape of T-cell acute lymphoblastic leukemia. *Int J Biochem Cell Biol* 53: 547-557.

Ward GR. (1917) The infective theory of acute leukaemia. *Br J of Childhood Diseases* 14: 10-20.

Weng AP, Ferrando AA, Lee W, Morris JPt, Silverman LB, Sanchez-Irizarry C, Blacklow SC, Look AT, and Aster JC. (2004) Activating mutations of *NOTCH1* in human T cell acute lymphoblastic leukemia. *Science* 306: 269-271.

Wiemels J. (2012) Perspectives on the causes of childhood leukemia. *Chem Biol Interact* 196: 59-67.

Wikhager C, Ögärd I, Martinsson T, Abrahamsson J, Palmqvist L, and Sjögren H. (2012) High-resolution copy number array in the molecular cytogenetic diagnostics of pediatric malignant hematological disorders. *Oncol Rep* 27: 1429-1434.

Virchow R. (1845) Weisses Blut. *Neue Notizen aus dem Gebiete der Natur- und Heilkunde* 36: 151-156.

Virchow R. (1848) Zur patologischen Physologie des Bluts. II Weisses Blut (Leukämie). *Virchows Arch path Anat* 1: 563-572.

Vogel C, and Marcotte EM. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 13: 227-232.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., and Kinzler KW. (2013) Cancer genome landscapes. *Science* 339: 1546-1558.

Xavier T. (2013) First contributors in the history of leukemia. *World J Hematol* 2:62-70

Xu H, Yang W, Perez-Andreu V, Devidas M, Fan Y, Cheng C, Pei D, Scheet P, Burchard EG, Eng C, Huntsman S, Torgerson DG, Dean M, Winick NJ, Martin PL, Camitta BM, Bowman WP, Willman CL, Carroll WL, Mullighan CG, Bhojwani D, Hunger SP, Pui CH, Evans WE, Relling MV, Loh ML, and Yang JJ. (2013) Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. *J Natl Cancer Inst* 105: 733-742.

Yasuda T, Ueno T, Fukumura K, Yamato A, Ando M, Yamaguchi H, Soda M, Kawazu M, Sai E, Yamashita Y, Murata M, Kiyoi H, Naoe T, and Mano H. (2014) Leukemic evolution of donor-derived cells harboring *IDH2* and *DNMT3A* mutations after allogeneic stem cell transplantation. *Leukemia* 28: 426-428.

Yeoh EJ, 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 CH, 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.

Yu L, Slovak ML, Mannoor K, Chen C, Hunger SP, Carroll AJ, Schultz RA, Shaffer LG, Ballif BC, and Ning Y. (2011) Microarray detection of multiple recurring submicroscopic chromosomal aberrations in pediatric T-cell acute lymphoblastic leukemia. *Leukemia* 25: 1042-1046.

Zenatti PP, Ribeiro D, Li W, Zuurbier L, Silva MC, Paganin M, Tritapoe J, Hixon JA, Silveira AB, Cardoso BA, Sarmento LM, Correia N, Toribio ML, Kobarg J, Horstmann M, Pieters R, Brandalise SR, Ferrando AA, Meijerink JPP, Durum SK, Yunes JA, and Barata JT. (2011) Oncogenic *IL7R* gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia. *Nat Genet* 43: 932-939.

Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, Lu C, Chen SC, Wei L, Collins-Underwood JR, Ma J, Roberts KG, Pounds SB, Ulyanov A, Becksfort J, Gupta P, Huether R, Kriwacki RW, Parker M, McGoldrick DJ, Zhao D, Alford D, Espy S, Bobba KC, Song G, Pei D, Cheng C, Roberts S, Barbato MI, Campana D, Coustan-Smith E, Shurtleff SA, Raimondi SC, Kleppe M, Cools J, Shimano KA, Hermiston ML, Doulatov S, Eppert K, Laurenti E, Notta F, Dick JE, Basso G, Hunger SP, Loh ML, Devidas M, Wood B, Winter S, Dunsmore KP, Fulton RS, Fulton LL, Hong X, Harris CC, Dooling DJ, Ochoa K, Johnson KJ, Obenauer JC, Evans WE, Pui CH, Naeve CW, Ley TJ, Mardis ER, Wilson RK, Downing JR, and Mullighan CG. (2012) The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 481: 157-163.

Zuurbier L, Homminga I, Calvert V, te Winkel ML, Buijs-Gladdines JGCAM, Kooi C, Smits WK, Sonneveld E, Veerman AJP, Kamps WA, Horstmann M, Petricoin EF III, Pieters R, and Meijerink JPP. (2010) *NOTCH1* and/or *FBXW7* mutations predict for initial good prednisone response but not for improved outcome in pediatric T-cell acute lymphoblastic leukemia patients treated on DCOG or COALL protocols. *Leukemia* 24: 2014-2022.