Gene Expression Studies of Hematologic Malignacies

Andersson, Anna

2005

Gene Expression Studies of Hematologic Malignancies

Anna Andersson

Department of Clinical Genetics

Lund University

2006

ISBN 91-85481-35-1

PRINTED IN SWEDEN 2005 BY MEDIA-TRYCK, LUND

ORIGINAL ARTICLES

This thesis is based on the following articles, referred to in the text by their standard citation designation or their Roman numerals (I-IV).

I. Andersson A, Edén P, Lindgren D, Nilsson J, Lassen C, Heldrup J, Fontes M, Borg Å, Mitelman F, Johansson B, Höglund M, and Fioretos T. Gene expression profiling of leukemic cell lines reveals conserved molecular signatures among subtypes with specific genetic aberrations. *Leukemia* 2005;19:1042-1050.

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

III. Andersson A, Ritz C, Lindgren D, Edén P, Lassen C, Heldrup J, Olofsson T, Råde J, Fontes M, Porwit-McDonald A, Behrendtz M, Höglund M, Johansson B, and Fioretos T. 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. Manuscript.

IV. Karrman K, Andersson A, Björgvinsdóttir H, Strömbeck B, Lassen C, Olofsson T, Nguyen-Khac F, Berger R, Bernard O, Fioretos T, and Johansson B. 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. Submitted.

TABLE OF CONTENTS

ABBREVIATIONS

ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
ANN	Artificial neural network
BASE	BioArray software environment
BM	Bone marrow
BP	Breakpoint
BC	Blast crisis
cDNA	Complementary DNA
CLP	Common lymphoid progenitor
CML	Chronic myeloid leukemia
CMP	Common myeloid progenitor
CR	Complete remission
DNA	Deoxyribonucleic acid
FACS	Fluorescence-activated cell sorter
FISH	Fluorescence in situ hybridization
GMP	Myelomonocytic progenitor
GO	Gene ontology
HCA	Hierarchical clustering analysis
HSC	Hematopoietic stem cell
IGH	Immunoglobulin heavy chain
k-NN	*k*-nearest neighbor
LT-HSC	Long-term HSCs
LMPP	Lymphoid primed multipotent progenitor
Mb	Mega base
MDS	Multidimensional scaling
MEP	Megakaryocytic/erythroid progenitor
MLL	Mixed lineage leukemia
MPP	Multipotent progenitor
MRD	Minimal residual disease
mRNA	Messenger ribonucleic acid
NBM	Normal bone marrow
PB	Peripheral blood
PCA	Principal component analysis
PCR	Polymerase chain reaction
Ph	Philadelphia chromosome
PTD	Partial tandem duplication
RNA	Ribonucleic acid
RT-PCR	Reverse transcription PCR
SAM	Significance analysis of microarray
ST-HSC	Short-term HSC
SVM	Support vector machine
TCR	T-cell receptor
TSG	Tumor suppressor gene
UPD	Uniparental disomy

PREFACE

Hematologic malignancies are characterized by recurrent balanced chromosomal abnormalities that lead to deregulated expression of genes located in the proximity of the breakpoints or result in tumor-specific fusion genes. In acute leukemia, the rearranged genes often involve conserved transcription factors of importance for normal hematopoiesis, whereas chronic leukemias typically are characterized by rearrangements of protein-tyrosine kinase encoding genes. The genetic rearrangements present at diagnosis provide important clinical and pathogenetic information.

Although leukemias have been extensively characterized, resulting in improved risk stratification and better outcomes, there is still a need for a refined risk classification to identify patients with a favourable or adverse prognosis, who would benefit from alternative treatment modalities. In addition, our knowledge of how individual fusion genes elicit their leukemogeneic properties still remains quite limited. In this context, gene expression profiling, as determined by microarray analyses, has proved to be a powerful tool for identifying clinically and biologically important variables. A refined risk-assessment of leukemias will hopefully lead to identification of patient subgroups that would benefit from either more or less intensive treatment. In addition, improved knowledge about deregulated genetic networks in leukemias will provide important pathogenetic information and help identifying genes that can serve as molecular targets for novel therapies.

The general aim of this thesis was to characterize hematologic malignancies using gene expression profiling in order to obtain an improved classification and an increased understanding of the complex genetic networks that are deregulated in acute leukemia. This thesis is divided into three sections; the first part provides an overview of hematologic malignancies and the microarray technology, giving a general introduction to the field on which the original articles are based. In the second section, the specific aims of the thesis, a summary of materials and methods, and the results are given with a short discussion, followed by a general discussion. The third and final section contains the original articles on which this thesis is based.

Lund, December 2005

INTRODUCTION

Hematopoietic Malignancies

Molecularly, leukemias are a heterogeneous disease entity with different rearrangements and dysregulations of genes with important functions in cellular growth, differentiation, and death (apoptosis). At the cellular level, acute leukemias are characterized by an expansion of immature white blood cells (blasts) in the bone marrow and blood, where a lack of mature blood cells together with a suppression of normal residual hematopoiesis, eventually leads to anemia, thromobocytopenia, and leukopenia, which result in fatigue, bleeding, and infections.

Hematopoietic malignancies comprise acute and chronic leukemias, myeloproliferative disorders, and myelodysplastic syndromes. In Sweden, there are approximately 460 cases of acute leukemia per year (www.socialstyrelsen.se). In adults, acute myeloid leukemia (AML) predominates, with an incidence that increases with age, whereas in childhood, acute lymphoblastic leukemia (ALL) is more common. In fact, ALL is the most common malignancy in childhood with an age peak around 3-5 years of age and an incidence of about 5 cases per 100 000 and year (Hjalgrim et al., 2003)

Leukemias are characterized by the presence of specific genetic alterations at diagnosis that are intimately associated with leukemogenesis, clinical and morphologic subtypes, and outcome. For example, it is well known that t(12;21)(p13;q22) [*ETV6/RUNX1*] high hyperdiploidy (>50 chromosomes), t(8;21)(q22;q22) [*RUNX1/RUNX1T1*], and t(15;17)(q22;q21) [*PML/RARA*], are associated with a favorable prognosis, whereas t(1;19)(q23;p13) [*TCF3/PBX1*], t(9;22)(q34;q22) [*BCR/ABL1*], and 11q23/*MLL* rearrangements confer an adverse prognosis, unless intensively treated (Grimwade, 2001; Johansson et al., 2004). However, although genetic alterations in hematologic malignancies have been extensively studied, much remains to be known about how these genetic lesions cause leukemia. This notwithstanding, our present knowledge has recently led to the development of alternative treatment strategies. For example, the BCR/ABL1 chimeric protein is now targeted with imatinib mesylate (Gleevec) in patients with chronic myeloid leukemia (CML) (Druker et al., 1996; Deininger et al., 2005) and there are ongoing clinical trials with inhibitors targeting FLT3 in patients with AML (Fiedler et al., 2005; Stone et al., 2005).

Normal Hematopoiesis

In fetal development, hematopoiesis takes place in the yolk sac during the first eight weeks of life, after which the yolk sac diminishes, and then in the liver and spleen until close to term, after which hematopoiesis is mainly restricted to the bone marrow. During childhood, blood cells are produced in all bones of the body, but with increasing age, the peripheral parts of the bones are replaced with inactive marrow (yellow marrow); in the adult, blood cells are only developed in the central parts of the skeleton (Jandle, 1996).

The blood is composed of several different cells types, each of which has important functions necessary for survival. For example, the erythroid cells transport oxygen, platelets are involved in blood clotting, granulocytes and monocytes are responsible for the immunological defense response against fungi, parasites, and viruses, B-cells produce antibodies as an immunological response against bacteria and other microorganisms, and T-cells participate in the activation of B-cells as well as in the elimination of virus-infected cells (Kawamoto and Minato, 2004; Chen-Kiang, 2005). Hematopoiesis is a continuous process and dying cells need to be replaced in order to maintain a steady state. Subtle abnormalities affecting hematopoietic proliferation, differentiation, and/or apoptosis may eventually result in leukemia.

Hematopoietic Development and Differentiation

The generation of mature blood cells throughout life is governed by hematopoietic stem cells (HSC), which are rare cells characterized by their potential to self-renew and their capacity to differentiate and form cells of all blood lineages. The term self-renewal refers to the ability to produce daughter cells with identical characteristics as the original stem cell (Ogawa, 1993; Herzog et al., 2003). HSCs can be further subdivided into long-term HSCs (LT-HSC) with the capacity of indefinite self-renewal and short-term HSC (ST-HSC) that self-renew only for a defined period of time.

The differentiation hierarchy of blood cells is tightly regulated by cytokines and transcription factors. Ordered expression or downregulation of these regulatory molecules drive maturation and lineage commitment (Metcalf, 1993; Zhu and Emerson, 2002; Hoang, 2004). One generally accepted model of hematopoietic development starts with the LT-HSC, which gives rise to a ST-HSC that differentiates into a multipotent progenitor (MPP) with restricted, or no capacity for, self-renewal. The MPP may differentiate into a common lymphoid progenitor (CLP) or a common myeloid progenitor (CMP), both of which will give rise to lineage-restricted cells (Morrison et al., 1997). The CLP is committed to form cells of the B- and T-lineages and the CMP differentiate into a myelomonocytic progenitor (GMP) that gives rise to monocytes and granulocytes and a megakaryocytic/erythroid progenitor (MEP),

which differentiates to megakaryocytes and erythrocytes (Kondo et al., 1997; Akashi et al., 2000). Recently, an alternative model has been suggested where the pluripotent HSC loses the potential to differentiate to a megakaryocytic and erythroid progenitor and subsequently turns into a lymphoid primed multipotent progenitor (LMPP). The LMPP express FLT3 and when it loses the potential to differentiate to a granulocytic/monocytic progenitor it will generate the CLP (Adolfsson et al., 2005).

Hematopoietic cells express specific cell surface markers that are characteristic for their lineage and maturation, e.g., $CD19^+$ for B-lineage and $CD33^+$ for myeloid lineage. The detailed knowledge about the ordered expression of cell surface markers can be used to sort cells using a fluorescence-activated cell sorter (FACS). FACS produces a highly enriched cell population of a desired maturation.

Malignant Hematopoiesis

It has become generally accepted that cancer is a multistep process, where the accumulation of somatically acquired genetic changes disturbs the normal homeotic balance of controlled cell differentiation, proliferation, and death (Hanahan and Weinberg, 2000). The target cell for malignant transformation is in most cancers unknown, but the similarities between stem cells and cancer cells, both of which harbor the potential of self-renewal, indicate that the stem cell may be the target cell of transformation (Reya et al., 2001; Passegue et al., 2003). This is an attractive hypothesis since the stem cell already has self-renewal capacity; hence, only a limited number of genetic changes would be needed to give rise to a leukemic clone. In addition, stem cells are long-lived and therefore more likely to accumulate additional genetic changes. However, it has also been suggested that the first genetic hit could take place in a more committed progenitor cell, which would then reacquire self-renewal potential, accumulate genetic changes, and give rise to a malignant clone (Reya et al., 2001; Passegue et al., 2003).

Recent data suggest that the target cell for transformation may vary and be dependent on the specific genetic rearrangement. For example, in cases with the t(9;22)(q34;q11) [*BCR/ABL1*] rearrangement, the target cell for transformation is most likely an early HSC since the fusion gene has been found in myeloid, erythroid, B-, and sometimes also in T-cells (Fialkow et al., 1977; MacKinney et al., 1993). In addition, the *BCR/ABL1* fusion gene has been found in endothelial cells, indicating that the target cell for transformation may even be the very primitive and putative hemangioblast (Gunsilius et al., 2000; Fang et al., 2005). In contrast, the *ETV6/RUNX1* fusion gene, generated through a t(12;21)(p13;q22) and found in 25% of B-cell precursor pediatric ALLs, has been found in a more mature $CD34^+$, $CD38^-$, and $CD19^+$ population. The expression of CD19 indicates

that the target cell for transformation may be a cell already committed to the B-cell lineage (Castor et al., 2005).

Genetic Alterations in Leukemia

A Historical Perspective

During recent years, it has become increasingly clear that leukemias are characterized by recurrent chromosomal rearrangements that are closely associated with leukemic subtype and also, in many instances, with prognosis (Johansson et al., 2004; Mrozek et al., 2004). Today, more than 350 recurrent chromosomal abnormalities have been identified in hematologic malignancies (Mitelman et al., 2004), providing important clinical and pathogenetic information. The success in the identification of chromosomal rearrangements has primarily been a result of detailed chromosome banding analyses. However, it took more than ten years from the detection of the first chromosomal abnormality until the true nature of this aberration was revealed. It was in 1960 that Nowell and Hungerford studied the chromosomes in bone marrow samples from patients with CML and discovered the presence of a small marker chromosome (Nowell and Hungerford, 1960). It was soon evident that this was a recurrent cytogenetic alteration in CML and it was called the Philadelphia chromosome (Ph^1 or Ph) in honor of the city in which it was first discovered (Figure 1). However, it was not until the banding techniques were introduced in 1970 (Caspersson et al., 1970), that researchers could perform detailed analyses of the chromosomes and it could be shown that the Ph chromosome in CML in fact was a result of a translocation between chromosomes 9 and 22 (Rowley, 1973). In 1985, breakpoint characterization revealed that the t(9;22)(q34;q11) leads to the fusion of the *BCR* gene at 22q11 with the *ABL1* gene, translocated from 9q34 (Heisterkamp et al., 1985; Shtivelman et al., 1985). Molecularly, this fusion gene has been shown to lead to a constitutive activation of the tyrosine kinase-encoding gene *ABL1*, which subsequently activates intrasignalling pathways resulting in enhanced proliferation, inhibition of apoptosis, and altered adhesive properties of the leukemic cells (Salesse and Verfaillie, 2002; Melo and Deininger, 2004). The molecular understanding of the *BCR/ABL1* fusion recently led to the revolutionary development of Imatinib (Gleevec), a tyrosine kinase inhibitor of the BCR/ABL1 protein, which today is used to treat patients with CML (Melo and Deininger, 2004).

The discovery of chromosomal alterations that were intimately associated with clinically important variables paved the way for the identification of a large

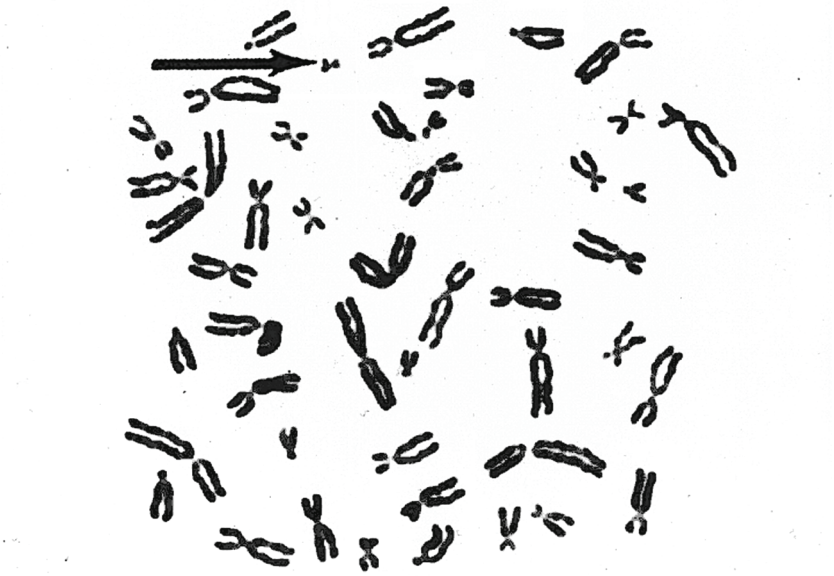

Figure 1. Unbanded metaphase of a bone marrow cell from a patient with CML. The arrow points at the Ph chromosome.

number of leukemia-associated genetic changes. Today, these specific abnormalities are used clinically for the classification of patients to different risk groups, receiving different therapies.

Genetic Alterations in Leukemia

Recurrent chromosomal aberrations, including translocations, inversions, deletions, duplications, monosomies, and trisomies are common in leukemia. Translocations result in the exchange of genetic material between two chromosomes and may have, at least, two different molecular consequences (Figure 2) (Rabbitts, 2001). Genes may be juxtaposed to the vicinity of strong regulatory elements, such as the T-cell receptor (*TCR*) or the immunoglobulin heavy chain (*IGH*) genes. In these cases, a translocated and structurally intact "oncogene" becomes activated by strong regulatory elements, resulting in inappropriate level and timing of expression. A prototypic example is the t(8;14)(q24;q32) in Burkitt lymphoma where *MYC* comes under the control of *IGH@* regulatory elements (Rabbitts, 2001). More commonly, a translocation results in the fusion of genetic material from two chromosomes resulting in the formation of a chimeric gene, a so-called fusion gene. Typically, the two genes break at intronic sequences with subsequent ligation. After splicing, a chimeric mRNA and protein with altered function as compared to the normal counterparts is produced (Rabbitts, 2001). So far, more than 200 fusion genes have

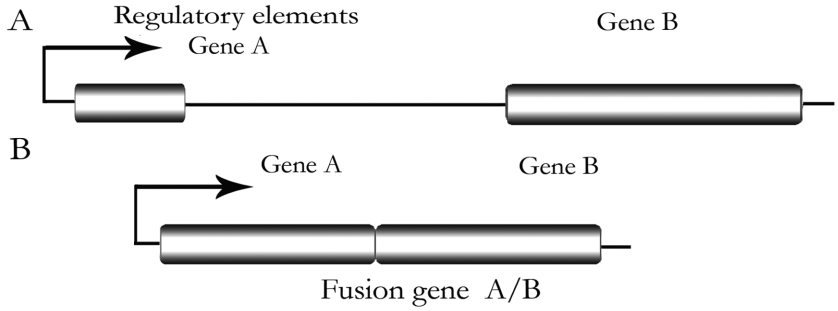

Figure 2. Chromosomal translocations may lead to two different molecular consequences. A. The structurally intact gene B comes under the influence of strong regulatory elements of, e.g., the *TCR* gene (gene A), resulting in a deregulated expression of gene B. **B.** If two genes break they may fuse and create a fusion gene (A/B), which gives rise to a chimeric protein with transforming properties.

been described in hematologic malignancies (Mitelman et al., 2004).

The major targets of chromosomal translocations in acute leukemia are conserved transcription factors that function as master regulators of normal hematopoiesis where they control blood development (Look, 1997; Rabbitts, 2001; Scandura et al., 2002). Typically, such rearrangements alter or interrupt the normal function of genetic programs controlled by the rearranged transcription factors. Recently, it was suggested that acute leukemia is a result of cooperating mutations, referred to as class I and class II mutations, respectively, of genes that cause a survival advantage and impaired differentiation (Speck and Gilliland, 2002). In this model, class I mutations are characterized by mutations in genes encoding tyrosine kinases, such as *FLT3*, *KIT*, or *RAS*, whereas class II mutations involve genes encoding transcription factors. A transcription factor often rearranged in both ALL and AML is *RUNX1* – a master regulator of hematopoiesis (Blyth et al., 2005). When *RUNX1* is rearranged by chromosomal translocations, the fusion gene interferes with the normal function of *RUNX1* causing impaired differentiation (Speck and Gilliland, 2002). In chronic leukemias, genes targeted by translocations often involve genes encoding tyrosine kinases, the prototypic example being the *BCR/ABL1* fusion in CML.

In leukemias, genes may also become altered through the gain or loss of chromosomes. Alternatively, amplification, i.e., the existence of a specific gene or sets of genes in multiple copies, may be present. Cytogenetically, amplification is seen as extra-chromosomal double minute chromosomes or as intra-chromosomal homogeneously staining regions. Typically, the amplified gene(s) provides the cell with a growth advantage and has the capacity to accelerate tumor formation, e.g., the cell cycle regulator *CCND1* (Donnellan and Chetty, 1998). Known examples of amplified oncogenes in leukemias include *RUNX1* and *MLL*, although they occur at relatively low frequencies (Harewood et al., 2003; Poppe et al., 2004).

More frequently, one or several chromosomes are gained in leukemia and if the modal number exceeds 50, it is referred to as high hyperdiploidy, which is the most common cytogenetic abnormality pattern in childhood ALL (Johansson et al., 2004). Next to nothing is known about the pathogenetic effect of chromosome gain, but it is likely that it contributes to leukemogenesis through a general gene dosage effect (Gruszka-Westwood et al., 2004).

Loss of chromosomal material may involve small deletions or whole losses of one or several chromosomes. Loss of chromosomal material may contribute to tumorigenesis through the loss of genes with the capacity to prevent tumor formation. Such genes are called tumor suppressor genes (TSGs) and known examples include *TP53* and *RB1.* Loss of a TSG is a recessive genetic event where both chromosomal copies have to be lost or inactivated in order for a gene to lose its function (Knudson, 1971). TSGs are tightly linked to programmed cell death and the cell cycle machinery, and upon DNA damage they may induce cell cycle arrest and subsequently DNA repair through the activation of DNA repair genes. However, if the DNA damage is too severe, TSGs instead induce cell death (Macleod, 2000). Hence, loss of a TSG may result in escape of apoptosis and eventually in tumor formation.

Losses of TSGs have been considered the main mechanism by which genetic changes result in solid tumor formation. However, it has been suggested that fusion genes may be more frequent in solid tumors than previously anticipated (Mitelman et al., 2004). Indeed, two recurrent chromosomal rearrangements, resulting in the creation of the *TMPRSS2/ERG* or *TMPRSS2/ETV1* fusion genes, were recently detected in a large proportion of cases with prostate cancer (Tomlins et al., 2005). During recent years, it has also become evident that some genes show haploinsufficiency, that is, the loss of one allele is sufficient to cause a phenotypic effect (Santarosa and Ashworth, 2004).

Genes may be altered through a point mutation, i.e., the change of a single base pair in the DNA sequence. The molecular consequences of a point mutation depend on which amino acid is affected. Potential effects include loss of function of an allele through a mutation that induces a premature stop codon or a gain-of-function, where the mutation results in oncogeneic activation and constitutive signalling, as exemplified by mutations in *FLT3* and *RAS* (Ehrhardt et al., 2002; Stirewalt and Radich, 2003).

Epigenetic alterations may also contribute to malignant transformation through gene silencing or activation of genes with important function in tumor formation. For example, gene silencing due to methylation of the cell cycle regulators *CDKN2A/B* genes is observed in about 30% of childhood B-lineage ALLs (Zhou et al., 1997). In addition, acquired segmental uniparental disomy (UPD) has been detected in 20% of AML cases with a normal karyotype, and in several

of the cases with UPD, homozygous mutations in leukemia-associated genes, i.e., *WT1*, *FLT3*, *CEBPA*, and *RUNX1* were seen (Fitzgibbon et al., 2005; Raghavan et al., 2005). UPD may also result in altered expression of imprinted genes, i.e., genes that are selectively expressed depending of their parental origin.

MicroRNAs, which are noncoding genes thought to be involved in tissue-specific gene regulation, have been suggested to play a role in leukemia as well as in other cancers (Calin et al., 2002; Chen, 2005). Two microRNA genes, *miR15a* and *miR16*, are deleted in a high proportion of cases with chronic lymphocytic leukemia (Calin et al., 2005) and were recently shown to regulate postranscriptionally the expression of BCL2, resulting in induction of apoptosis in hematopoietic cells (Cimmino et al., 2005). The exact role of microRNAs in leukemia, however, remains to be elucidated.

Common Genetic Changes in Acute Leukemia

The genetic changes found in hematopoietic malignancies serve as hallmarks for the leukemic subtype and provide important clinical information. Below, a short summary of characteristic genetic alterations in childhood leukemia, being a particular focus of the present study, is presented.

t(1;19)(q23;p13) TCF3/PBX1

The t(1;19)(q23;p13), which occurs in approximately 3% of pediatric ALL (Johansson et al., 2004), was cloned in 1990 by two groups (Kamps et al., 1990; Nourse et al., 1990) and shown to result in the *TCF3/PBX1* fusion gene. When the t(1;19) was first reported, it was considered to be associated with a high risk leukemia that presented with leucocytosis, central nervous system involvement, and an increased risk of relapse (Crist et al., 1990; Hunger, 1996). However, with intensified treatment protocols, the prognosis of patients with this translocation/ fusion gene has improved, but it is still considered a high risk genetic feature (Uckun et al., 1998). The translocation occurs both in a balanced and unbalanced form (Paulsson et al., 2005a) and it has been suggested that the unbalanced variant confer a better prognosis (Secker-Walker et al., 1992; Uckun et al., 1998), although this remains controversial (Pui et al., 1994).

The t(1;19)(q23;p13) targets the basic-loop-helix transcription factor *TCF3* (*E2A*) on chromosome 19 and the homeobox containing gene *PBX1* on chromosome 1. *TCF*3 gives rise to two protein products; E12 and E47 (Murre et al., 1989; Murre, 2005), whose expression is critical for B-cell development as demonstrated in knock-out mice, where deficiency of *Tcf3* causes arrest of B-cell

development at an early pro-B cell stage (Bain et al., 1994). In addition, ectopic expression of *E12* induces the expression of *Ebf1* as well as other genes of importance for B-cell development, e.g., *Il7rα* and *Rag1* (Kee and Murre, 1998). *PBX1* is a homeobox-containing transcription factor that normally is not expressed in the lymphoid lineages. In mice, it has been shown that *Pbx1* is required for the maintenance, but not the initiation of definitive hematopoiesis (DiMartino et al., 2001). *PBX1* can bind directly to *HOX* genes, or to *MEIS1*, another HOX-cofactor, thus interacting with HOX proteins in trimeric complexes (Shanmugam et al., 1999).

The *TCF3/PBX1* fusion gene retains the transactivation domain of *TCF3* and the homeodomain of *PBX1*, and unlike normal *PBX1*, the fusion protein is a transcriptional activator (LeBrun and Cleary, 1994; Sykes and Kamps, 2004). The fusion protein retains the capability to bind HOX proteins, but can no longer bind MEIS1, and it is likely that the fusion gene in complex with HOX genes results in deregulated expression of HOX/PBX1 target genes (Knoepfler et al., 1997; Lu and Kamps, 1997). In addition, the disruption of *TCF3*, which is critical for B-cell development, is likely to contribute to leukemia development. The *TCF3/PBX1* fusion gene rapidly induces leukemia in mice, but intriguingly, so far, the leukemia developing in these mice is of myeloid or T-cell type and never a B-cell leukemia (Sykes and Kamps, 2004).

In rare cases with the (1;19)(q23;p13), the *TCF3/PBX1* fusion transcript is absent (Hunger et al., 1991). Recently, a novel translocation involving *MEF2D* at 1q23 and *DAZAP1* at 19p13, resulting in the *MEF2D/DAZAP1* fusion gene, was cloned in a pre-B cell line with the t(1;19) rearrangement but that lacked the *TCF3/PBX1* fusion gene (Yuki et al., 2004).

RUNX1 Rearrangements

RUNX1 is frequently involved in chromosomal rearrangements in hematologic malignancies. In pediatric B-precursor ALL, *RUNX1* is rearranged by the t(12;21)(p13;q22), seen in about 25% of the cases (Harrison et al., 2005). This translocation fuses the *ETV6* (*TEL*) gene on chromosome 12 with *RUNX1* (*AML1*) on chromosome 21 (Golub et al., 1995; Romana et al., 1995). This fusion gene has been reported to correlate with a good prognosis (Shurtleff et al., 1995), although some recent studies have reported a high incidence of this fusion transcript in relapsed ALLs (Seeger et al., 1998). *RUNX1* alterations are also frequent in AML, e.g., through the t(8;21)(q22;q22) where *RUNX1* is fused to the *RUNX1T1* (*ETO*) gene on chromosome 8 (Erickson et al., 1992).

The transcription factor *RUNX1* encodes the heterodimeric partner of *CBFB*, which enhances the DNA-binding properties of *RUNX1* and protects it from ubiquitin-mediated proteolysis. Together, *CBFB* and *RUNX1* constitute a

component of the core binding factor (CBF) transcription factor complex (Blyth et al., 2005) and regulate transcription of a large number of genes with pivotal roles in all lineages of hematopoiesis, e.g., *IL3* (Uchida et al., 1997) and *Sfpi1 (Pu1)* (Okada et al., 1998). *RUNX1* may also function as a transcriptional repressor through the binding of Groucho-related co-repressors (Imai et al., 1998) and through interaction with mSin3A (Lutterbach et al., 2000).

RUNX1 has a strong DNA binding domain (Runt), which is retained in all fusion genes. In the t(12;21), the 5′part of *ETV6* is fused to almost the entire *RUNX1* (Golub et al., 1995; Romana et al., 1995). *ETV6* is widely expressed in most normal tissues and when fused to *RUNX1*, the expression of *RUNX1* will be driven by the *ETV6* promoter. The leukemogenic potential of *RUNX1* fusion proteins is probably coupled to the interaction and inhibition of the normal function of the *CBF*-transcription factor complex (Lutterbach and Hiebert, 2000; Speck and Gilliland, 2002). Moreover, the fusion protein has an increased affinity for *CBFB* as compared to the wild-type allele, resulting in repression of *RUNX1*-target genes. Mouse models with *ETV6/RUNX1* have shown that expression of the fusion protein causes impaired differentiation, mainly in the pro-B-cell compartment, but does not result in a complete differentiation block (Fischer et al., 2005).

Amplification of *RUNX1* is uncommon in ALL (1.5%), but has recently been shown to be associated with a poor prognosis (Harewood et al., 2003; Robinson et al., 2003; Harrison et al., 2005). Inactivating mutations of *RUNX1* occur in 10% of AMLs, being particularly frequent in AML M0 (22%) (Roumier et al., 2003). Interestingly, heterozygous mutations of *RUNX1* are responsible for the autosomal dominant genetic disease, familial platelet disorder, characterized by platelet defects and a predisposition to develop AML (Song et al., 1999).

11q23/MLL Rearrangements

Rearrangements of the *MLL* (Mixed Lineage Leukemia) gene at chromosome band 11q23 are common in leukemia, in particular among infants where 80% harbor such a rearrangement (Rubnitz et al., 1994). In older children, the frequency is much lower (4-8%) (Rubnitz et al., 1997; Forestier et al., 2000a). *MLL* abnormalities are also frequently seen in AML, in particular therapy-related AML arising after previous treatment with drugs targeting topoisomerase II (Felix, 1998). Leukemias with *MLL* rearrangements are quite often bilineage/biphenotypic expressing both lymphoid and myeloid surface antigens, and are seen in all hematopoietic lineages (Gregorini et al., 1998). Today, more than 40 different cloned partner genes are known to be involved in rearrangements with *MLL* (Mitelman et al., 2005). Generally, abnormalities of *MLL* correlate with a poor outcome, although recent data indicate that prognosis varies among the different fusion genes generated, with t(9;11)(p21;q23)

conferring a better prognosis in children with AML (Rubnitz et al., 2002; Pui et al., 2003). In addition, infants with *MLL* rearrangements have an adverse prognosis irrespective of the specific *MLL* abnormality (Pui et al., 2002, 2003).

The normal MLL protein is cleaved post-translationally into an N-terminal fragment with repressor activity and a C-terminal part with strong transcriptional activity. The C- and N-terminal parts of MLL dimerize and constitute the basic part of a large multiprotein complex that regulates and maintain the transcription of *HOX* genes, which are critical regulators of hematopoiesis (Yu et al., 1998; Nakamura et al., 2002; Hsieh et al., 2003). In addition, *HOX* genes regulate hematopoietic development and are expressed in hematopoietic cells during distinct stages of differentiation (Magli et al., 1991; Sauvageau et al., 1994). In line with the fundamental roles of *HOX* genes in hematopoiesis and the role of *MLL* to maintain *HOX* gene expression, it was recently shown that *Mll* is required for definite hematopoiesis (Ernst et al., 2004).

MLL fuses to a wide range of partner genes that may be subdivided into two types; nuclear genes (e.g., *AFF1* (*AF4*), *MLLT3* (*AF9*), *MLLT1* (*ENL*), *CREBBP*) or cytoplasmic genes (e.g., *MLLT4*, *ARHGEF12* (*LARG*), *GAS7*, *CBL*). Most of the nuclear genes are transcriptional activators, and there are now several studies reporting that the 3′ partner gene contributes to the oncogeneic property of *MLL* fusion genes, providing *MLL* chimeras with a gain-of-function rather than a loss-of-function (Ayton and Cleary, 2001). Among the cytoplasmic genes that fuse to *MLL*, none have roles in transcriptional regulation and little is known about the oncogeneic properties of these fusion genes. However, it was recently shown that the leukemogenic potential of cytoplasmic *MLL* fusion genes is contributed by oligomerization, that results in transcriptional activation of *MLL* target genes (So et al., 2003). Gene expression studies have shown that *MLL* chimeras give rise to a common gene signature with a high expression of genes, such as *HOXA9*, *HOXA10*, and *MEIS1* (Armstrong et al., 2002; Ross et al., 2004; Andersson et al., 2005a, b; Kohlmann et al., 2005). The finding that also cytoplasmic *MLL* fusion genes result in transcriptional activation of *MLL* target genes fits well with the finding of a common gene expression signature. Intriguingly, however, cases with partial tandem duplication (PTD) of *MLL* have a different gene expression profile as compared to cases where *MLL* is rearranged through a translocation, suggesting that alternative mechanisms contribute to malignant transformation in cases with *MLL* PTD (Ross et al., 2004).

High Hyperdiploidy

In childhood leukemias, high hyperdiploidy (>50 chromosomes) occurs in approximately 30-45% of B-cell precursor ALL (Forestier et al., 2000a; Moorman et al., 2003). This is the most common genetic abnormality pattern in pediatric

ALLs and is associated with a favorable prognosis with a 5-year event free survival of 70-80% (Forestier et al., 2000b; Moorman et al., 2003). Cytogenetically, high hyperdiploidy is characterized by a nonrandom gain of chromosomes with trisomy or tetrasomy 21 being the most frequent (Heerema et al., 2000). Other chromosomes commonly gained include X, 4, 6, 8, 10, 14, 17, and 18 (Heerema et al., 2000; Paulsson et al., 2005b). Attempts have been made to identify cytogenetic subgroups among the high hyperdiploid ALLs that correlate with outcome, suggesting that gain of chromosome 4, 10, 17, and 18 (Harris et al., 1992; Heerema et al., 2000; Moorman et al., 2003; Sutcliffe et al., 2005) are associated with a favorable prognosis. The presence of structural rearrangements has also been suggested to have a negative prognostic impact (Pui et al., 1989; Forestier et al., 2000b); however, this has been questioned (Raimondi et al., 1996; Moorman et al., 2003).

Little is known about the molecular consequences of hyperdiploidy, but it has been suggested that a general gene dosage effect of certain loci on the gained chromosomes contribute to leukemic development (Gruszka-Westwood et al., 2004). Imprinting - selective expression of a gene dependent on its parental origin – has also been suggested to play a role in the pathogenesis of hyperdiploidy (Haas, 1996). However, recent studies addressing this possibility have not found any evidence of preferential gain of a chromosome depending on the parental origin (Paulsson et al., 2003, 2005b). Gene expression studies of trisomies have shown that there is a general dose effect, but, in addition, some genes located on the duplicated chromosomes display either a substantially higher or a lower expression than expected, suggesting that alternative mutational mechanisms exist that cause deregulatated gene expression (Gruszka-Westwood et al., 2004; Andersson et al., 2005b).

Gene Expression Profiling

Historical Overview and Background

Gene expression profiling is a collective terminology for technologies that measure the expression of a large number of genes in a single or in a few experiments. In the past, gene expression analyses could only be performed on a gene-by-gene basis. Technological improvements made it possible, however, to obtain expression data of a large number of genes in a single experiment and to perform two-color hybridizations where the relative expression of genes could be measured in relation to a common reference (Lander, 1999). Currently, several different methods exist for monitoring gene expression; e.g., real-time quantitative PCR, filter based microarrays, and glass microarrays. Although all these methods measure the level

of expression of genes, they have intrinsic differences that should be taken into account when choosing a method for a particular experiment. For example, real-time PCR analysis can be both time-consuming and labor intensive when many genes are studied and are hence more suitable for investigating a smaller number of genes. With the introduction of filter based arrays (Lennon and Lehrach, 1991; Maier et al., 1994) it became possible to investigate the expression of a larger number of genes in a single experiment. However, it was not until the introduction of glass microarrays that the gene density increased considerably. In 1995, Schena and colleagues demonstrated that two-color cDNA microarrays could be used for high-throughput monitoring of gene expression changes in plants (Schena et al., 1995). Subsequently, in 1996, the same group investigated the expression of heat-shock induced genes in human T-cells using slides containing 1000 genes (Schena et al., 1996). The same year, the first microarray study of human cancer was performed on a melanoma cell line (DeRisi et al., 1996). In parallel with the cDNA array technology, microarray slides with short synthesized oligonucleotides also became available. Oligonucleotide slides offer the possibility of having several different sequence-specific oligonucleotides synthesized and are now used, e.g., for exon specific arrays. One of the first companies offering slides with synthesized oligonucleotides was Affymetrix, which uses a photolithography method to generate high density oligonucleotide slides (Lipshutz et al., 1999). In 1996, the density of the arrays was approximately 1000-2000 genes, a number that has increased dramatically to 30 000-100 000 elements on the arrays. With increased densities, gene expression profiling provides a unique possibility of high-throughput screening, and a large number of human cancers have now been analyzed using microarrays, providing important biological insights into the genetic pathways becoming deregulated in human malignancies. In addition, gene expression profiling of hematologic malignancies has shown that cytogenetic subclasses of leukemia show distinct and unique gene expression profiles that can be used to assign patients to genetic riskgroups at diagnosis. Below, an introduction to microarray analysis, with a special emphasis on cDNA microarrays, will follow.

Methodological Principles of cDNA Microarray

The microarray technology is based on the distinct nature of the DNA to base pair with a complementary sequence and form a double helix (Southern et al., 1999). This fundamental principle has been used for a long time in molecular genetic research, for example in northern blot analysis. Northern blot has been the standard method of choice for measuring gene expression and involves immobilization of the target RNA on a membrane and labeling of a probe from the gene of interest

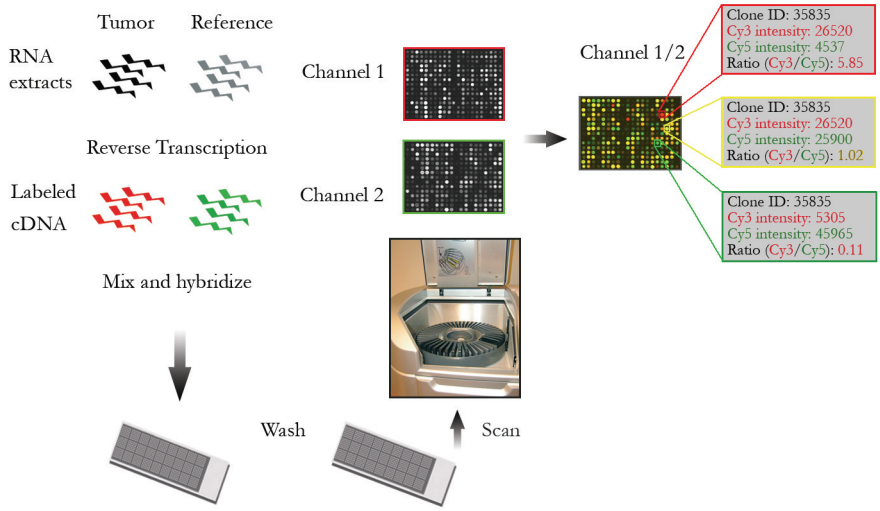

Figure 3. Methodological principles of the cDNA microarray technology. RNA extracts from the tumor and reference samples are reversely transcribed into cDNA, a reaction in which a fluorescent dye is incorporated. The samples are combined and hybridized to a microarray slide, which is then washed and subsequently scanned in a laser scanner. The laser excitates the fluorochromes and a detector measures the emission light. Two black and white images are created, one for each fluorochrome, and these are then combined to a pseudocolored image, in which red represents relative upregulation and green relative downregulation as compared to the reference. A gene with equal expression in the tumor and reference will be colored in yellow.

with radioactive nucleotides. The labeled probe is allowed to hybridize to the immobilized target RNA and the result is visualized using a phosphoimager. Northern blot is a robust method, but has the disadvantage that only a single gene can be investigated in each experiment. With the introduction of cDNA microarrays, where the probes are immobilized on a glass slide and the target cDNAs are hybridized to the slide, the number of genes that can be investigated at each experiment increases significantly. In cDNA microrrays, RNA from two targets, i.e., the sample of interest and a reference sample are labeled with two different dyes and hybridized to a slide containing the probes (Cheung et al., 1999; Duggan et al., 1999). Two-color microarray experiments provide the possibility to measure the relative expression of a large number of genes as compared to a common reference used in all hybridizations. In contrast to the more conventional northern blot, cDNA microarray is a high-throughput technique where several thousands of genes can be studied in only one single experiment, providing a snap-shot of the genes expressed in a tissue at a certain time point.

The generation of microarrays involves robotically printing of cDNA clones of 500-3000 base pair onto aminosilane-coated glass slides. Total RNA is

extracted from the tissue of interest, purified, and reversely transcribed into cDNA. RNA is also prepared from a reference sample, used in all hybridizations. The reference sample should show abundant gene expression, ideally expressing all genes on the microarray. To achieve this, it is common to use a pool of cell lines from different tissues, either made in-house or commercially available. When the RNA is reversely transcribed to cDNA, a fluorescent dye is incorporated, typically Cy3 for the test sample and Cy5 for the reference. The test and reference samples are then combined and allowed to hybridize to the cDNA clones on the microarray. Hybridization occurs in a competitive manner where the amounts of the samples bound to the probes depend upon their level of expression. The slides are washed in solutions with increased stringency to remove unbound cDNA and non-stringent binding between moderately matching sequences and are finally scanned at high resolution in a laser scanner where the fluorescent dyes are excitated and the emission light measured. The quantified emission is saved as two black and white images, one for each wavelength, which are then combined to a pseudocolored image (channel 1/channel 2) where red represents relative upregulation and green relative downregulation of the test sample as compared to the reference (Figure 3) (Cheung et al., 1999; Duggan et al., 1999; Harrington et al., 2000)

Data Management

Before the data generated from microarray experiments can be extracted and analyzed, they need to be normalized to compensate for technical and methodological biases. Several different methods exist for normalization; common to them all is that they rescale data to balance for potential differences in the amount of RNA labeled and to remove or dampen technical biases, such as differences in the detection of the dyes depending on the spatial location over the slide (Quackenbush, 2002; Yang et al., 2002). After normalization, a data set-specific filtering is performed to remove spots (genes) with a poor quality.

The data set-specific filtering can be performed using various quality cut-offs. For example, spots with a low intensity or with a diameter below a certain threshold may be filtered away. It is also common to consider only genes with a signal-to-noise or a log-ratio above a certain threshold. Such cut-offs are sensitive to threshholding, and genes with a high quality, but not fulfilling these criteria, may be filtered away. To avoid this, an alternative way of obtaining high quality data is to use an error-model (Andersson et al., 2005a, b, c). The error model uses signal-to-noise and fold change to correct for poor quality. Genes with a high uncertainty are moved towards the average of the gene and are hence more likely to be filtered away in a subsequent variation filter. After filtering for variance to remove genes with a low variation across experiments, a presence

filter is applied. The data are now considered to be of high quality and ready to be analyzed using sophisticated methods that compare gene expression data and explore the similarities and relations between samples.

Methods used for comparison of gene expression data can be divided into unsupervised and supervised methods. Unsupervised methods include algorithms that are used for exploration of gene expression data and where no previous knowledge about the distribution of data or group assignment is used for analysis. In particular, unsupervised methods can be used to propose novel hypotheses and to find novel subgroups (Quackenbush, 2001). Examples of traditional mathematical methods used for unsupervised analysis of microarray data include hierarchical clustering analysis (HCA) (Eisen et al., 1998), multidimensional scaling (MDS) (Khan et al., 1998), and principal component analysis (PCA) (Alter et al., 2000). HCA is a two-dimensional algorithm that, in the context of gene expression analysis, group samples with a similar gene expression pattern close to each other horizontally, with genes being organized vertically in a "heat map" reflecting their level of expression (Eisen et al., 1998). MDS and PCA are algorithms that reduce the high dimensionality of gene expression data into the two or three dimensions that contain most variance, with the distance between the samples reflecting their similarities at the gene expression level.

In supervised methods, on the other hand, previous knowledge of the data is used when class assignment is made, and they include discriminatory analyses (Ringnér et al., 2002). Discriminatory analysis requires that the data involves samples from at least two groups and include straightforward statistical analyses such as the T-test, but can also be performed using more sophisticated learning algorithms, e.g., support vector machines (SVM) (Brown et al., 2000), *k*-nearest neighbor (*k*-NN) (Dudoit and Fridlyand, 2002), and artificial neural networks (ANN) (Khan et al., 2001). Hence, supervised methods are used when the purpose of the data analysis is to construct a classifier for prediction of an unknown sample to an already defined class. Such methods have been successfully used in gene expression profiling of leukemia to predict the class of leukemic samples to clinically important variables, e.g., leukemia type, type of genetic change, and minimal residual disease (MRD) status (Yeoh et al., 2002; Ross et al., 2003, 2004; Valk et al., 2004; Andersson et al., 2005b; c; Cario et al., 2005; Haferlach et al., 2005; van Delft et al., 2005).

Gene Expression Profiling of Hematologic Malignancies

Since the introduction of microarray analysis, several investigations of gene expression profiles in hematologic malignancies have been performed, yielding insight into the genes dysregulated in leukemia. So far, most gene expression studies have been performed on adult leukemias (Virtaneva et al., 2001; Schoch et al.,

2002; Debernardi et al., 2003; Kohlmann et al., 2003; Bullinger et al., 2004; Chiaretti et al., 2004; Staber et al., 2004; Valk et al., 2004; Haferlach et al., 2005), and only a few large-scale gene expression studies have focoused on pediatric leukemias (Yeoh et al., 2002; Ross et al., 2003, 2004; Holleman et al., 2004; Andersson et al., 2005b, c; van Delft et al., 2005). Below, the most significant gene expression studies of hematologic malignancies are summarized, with a special emphasis on pediatric leukemia.

In 1999, the first article was published showing that leukemias can be classified, using supervised learning algorithms based on their gene expression profiles, into B-lineage ALL, T-cell ALL, and AML (Golub et al., 1999). Two years later, Armstrong and coworkers proposed that leukemias with 11q23/*MLL* abnormalities constitute a distinct and unique leukemia type (Armstrong et al., 2002). Subsequently, in 2002, the largest gene expression study of pediatric ALLs was published, investigating 360 pediatric ALLs and showing that such leukemias harbor distinct gene expression profiles and that classifiers could be built that predicted the class of an unknown sample with a high accuracy (Yeoh et al., 2002). In the same study, gene expression signatures associated with relapse in T-cell ALLs and in high hyperdiploid leukemias were identified. Furthermore, among leukemias with uncharacterized genetic changes, a novel subgroup was found. A smaller subset (132 cases) of the ALLs analyzed by Yeoh and collegues (2002) were reanalyzed the following year on higher density arrays, verifying that the expression profiles present at diagnosis can be used to classify leukemias into genetic risk groups with high accuracy (Ross et al., 2003). The following year, the largest study of pediatric AML was published (Ross et al., 2004), showing that distinct gene expression profiles associated with specific genetic abnormalities present at diagnosis could be identified in this subset of leukemia as well. These expression profiles were subsequently used to construct predictors that could assign an unknown sample to a known genetic class with an overall classification accuracy of 93%. By combining the AML and ALL data set previously analyzed by the same group (Ross et al., 2003), it was shown that 11q23/*MLL*-positive cases cluster primarily according to lineage. However, supervised analysis revealed the presence of a unique and common gene expression signature that was independent of lineage (Ross et al., 2004). Interestingly, it was also shown that AMLs with a partial tandem duplication (PTD) of *MLL* failed to cluster with the other *MLL*-positive cases, suggesting that such AMLs have a different mechanism of transformation (Ross et al., 2004). Several studies have subsequently verified that specific and distinct gene expression signatures correlate with lineages and genetic changes in hematologic malignancies (Moos et al., 2002; Schoch et al., 2002; Kohlmann et al., 2003; Yagi et al., 2003; Fine et al., 2004; Andersson et al., 2005a, b, c; Haferlach et al., 2005; Kohlmann et al., 2005; van Delft et al., 2005). Moreover, global gene expression analyses of hematopoietic cell lines of diverse origin, but with the same primary genetic changes,

have shown that such cell lines display similar gene expression profiles despite their diverse origin and numerous passages in vitro (Fine et al., 2004; Andersson et al., 2005a).

Although most microarray studies of hematologic malignancies have focused on the identification of genes associated with the specific genetic aberrations characteristically seen in leukemia, some have tried to use gene expression profiling to identify genes that are associated with response to treatment. For example, in a study investigating 173 pediatric ALLs for in vitro sensitivity to prednisolone, vincristine, asparaginase, and daunurubicine to identify genes associated with treatment resistance or sensitivity (Holleman et al., 2004), the genes associacted with drug resistance correlated with patient outcome; several of the genes identified had not previously been implicated in resistance for the drugs tested. In ALL, it has previously been reported that a high tumor load at day 29 of treatment significantly increases the risk of relapse (van Dongen et al., 1998; Björklund et al., 2003). Only two expression studies have tried to predict MRD status among childhood ALLs (Willenbrock et al., 2004; Cario et al., 2005). In the largest one (Cario et al., 2005), cases with no detectable MRD were compared to cases with a high MRD. A classifier of 62 clones was identified that could predict MRD status with a high accuracy. It was also recently shown that MRD status could be predicted in T-cell ALLs at the time of diagnosis (Andersson et al., 2005c).

Collectively, microarray analyses have successfully been utilized for classification purposes, both as regards leukemia type and specific genetic lesion present at diagnosis, with a high accuracy. It has, however, proved more difficult to identify gene expression profiles that could, already at the time of diagnosis, predict which patients who will relapse and to find novel subgroups of leukemia. Gene expression studies of hematologic malignancies have also resulted in important biological insights into the genetic programs becoming deregulated in leukemia and in an increased understanding of leukemia development and progression.

THE PRESENT STUDY

This section includes the specific aims of the thesis, a summary of the materials and methods, and the results together with a short discussion. A general discussion with emphasis on acute pediatric leukemia and gene expression profiling is given at the end of this section.

Specific Aims of the Study

The general aim of this thesis has been to characterize hematologic malignancies using gene expression profiling to improve the classification and to increase our understanding of the genetic mechanisms that control and contribute to leukemia development and progression. More specifically, the aims were:

- to investigate if immortalized hematologic cell lines with the same specific genetic alterations maintain a characteristic gene expression pattern despite their diverse origin and numerous passages in vitro (*Article I*),

- to study the gene expression patterns in pediatric acute leukemias and to investigate the expression pattern of these genes in normal hematopoietic cell subpopulations (*Article II*),

- to construct gene expression classifiers that can predict the class of an unknown leukemia sample to clinically important subgroups (*Article III*), and

- to use gene expression profiling to identify the genes deregulated as a consequence of the t(12;14)(p13;q11) in T-cell ALLs (*Article IV*).

MATERIALS AND METHODS

Patient Material, Purified Hematopoietic Subpopulations, and Cell Lines

The childhood acute leukemias analyzed in the present study (*Articles I*, *II*, *III*, and *IV*) were all diagnosed at Lund or Linköping University Hospitals, and the studies were reviewed and approved by the Research Ethics Committees of Lund and Linköping Universities, Sweden. Bone marrow (BM) or peripheral blood (PB) were collected at the time of diagnosis and put in TRIzol (Invitrogen, Carlsbad, CA). All samples were, as part of routine diagnostic procedures, analyzed cytogenetically and molecularly at the Department of Clinical Genetics, Lund, Sweden. The ALLs were analyzed for the presence of 11q23/*MLL* rearrangements, *BCR/ABL1*, *ETV6/RUNX1*, and *TCF3/PBX1* fusions. The AMLs were screened for 11q23/*MLL* rearrangements. Fluorescence in situ hybridization (FISH) investigations, using probe cocktails for the chromosomes commonly gained in high hyperdiploid ALLs, were performed on cases with either normal karyotypes or without analyzable metaphases. All cell lines used were cultured according the manufacturers' instructions and harvested 24 hours after medium exchange.

Article I

Forty hematologic cell lines were analyzed, 30 of which harbored the following primary genetic changes t(4;11)(q21;q23) [*MLL/AFF1*], t(6;11)(q27;q23) [*MLL/MLLT4*], t(9;11)(p21;q23) [*MLL/MLLT3*], t(11;19)(q23;p13) [*MLL/MLLT1*], dup(11)(q23q23) [PTD of *MLL*], t(X;11)(q13;q23) [*MLL/MLLT7*], t(1;19)(q23;p13) [*TCF3/PBX1*], del(4)(q12q12) [*FIP1L1/PDGFRA*], t(8;21)(q22;q22) [*RUNX1/RUNXT1*], t(8;14)(q24;q32) [*IGH@/MYC*], t(8;14)(q24;q11) [*TRA@/MYC*], t(9;22)(q34;q11) [P190 and P210 *BCR/ABL1*], t(10;11)(p12;q14) [*PICALM/MLLT10*], t(12;21)(p13;q22) [*ETV6/RUNX1*], t(15;17)(q22;q21) [*PML/RARA*], and inv(16)(p13q22) [*CBFB/MYH11*]. In addition, BMs from 11 children with AML or ALL, harboring *MLL/MLLT1*, *MLL/AFF1*, *TCF3/PBX1*, P190 *BCR/ABL1*, *ETV6/RUNX1*, *MLL/MLLT3*, or *RUNX1/RUNX1T1*, were investigated.

Articles II and *III*

Samples from BM (n=108) or PB (n=13) were obtained at the time of diagnosis from 121 children with ALL (87 B-lineage and 11 T-cell) or AML (n=23). In addition, six normal bone marrows (NBMs) and 10 selected purified hematopoietic subpopulations, collected from healthy donors, were included in the analysis. In *Article II*, all normal cells were obtained from the Department of Hematology, Lund, Sweden. $CD34^+$ cells were isolated from the mononuclear cell fraction by immunomagnetic beads (MACS, Miltenyi Biotec, Bergisch Gladbach, Germany). Mononuclear cells and $CD34^+$ cells were further fractionated by cell

sorting on a FACS Aria flow cytometer (Becton Dickinson Biosciences, San Jose, CA). From mononuclear cells, $CD3^+/CD4^+$ [helper T-lymphocytes], $CD3^+/CD8^+$ [suppressor T-lymphocytes], $CD19^+$ [mature B-lymphocytes], $CD15^{(+)}/CD33^+$ [neutrophils intermediate maturation], $CD15^{++}/CD33^{(+)}$ [mature neutrophils], and $CD71/GPA^+$ [erythroblasts] were isolated. From the $CD34^+$ cells, $CD34^+/CD19^+$ [early pre-B-cells], $CD34^+/CD117^+/CD45RA^-$ [common myeloid progenitors], and $CD34^+/CD117^+/CD45RA^+$ [granulocyte/macrophage progenitors], were obtained.

Article IV

Eight childhood T-cell ALLs, two of which harbored the t(12;14)(p13;q11), were analyzed. Seven of the cases were also included in *Articles II* and *III* and were diagnosed and treated at the Department of Pediatrics, Lund, Sweden. The remaining case was obtained from Saint Louis Hospital, Paris, and details about this case have been published elsewhere (Le Coniat et al., 1997).

Minimal Residual Disease Status

A subset of the pediatric leukemias investigated in *Articles II* and *III* was, as part of routine analyses, investigated for MRD status at day 0, 29, 50, and 100 of treatment. The ALLs from Lund (n=61) were monitored for MRD status by real-time quantitative PCR of patient-specific immunoglobulin and/or T-cell receptor gene rearrangements, as described in van Dongen *et al* (1998). The ALLs from Linköping (n=16) were monitored using flow cytometry as described previously (Björklund et al., 2003). The MRD status was translated to a scale from 1-6, where MRD of 1 corresponds to 0.001%, 2 to 0.01%, 3 to 0.1%, 4 to 1%, 5 to 10%, and 6 to 100% leukemic cells. In *Article III*, cases were classified based on MRD status at day 29 and were arbitrarily divided into two groups designated "Low" (MRD of 1-2) or "High" (MRD of 3-6).

RNA Isolation and Amplification

In *Articles I-IV*, total RNA was extracted using the TRIzol reagent (Invitrogen) and further purified using RNeasy columns (Qiagen, Valencia, CA). The Universal Human Reference RNA (Stratagene, La Jolla, CA), used as a reference for all microarray hybridizations, was prepared according to the instructions provided by the manufacturer. The patient material in *Articles I-IV* and the reference used for all hybridizations were linearly amplified using the RiboAmp - RNA Amplification kit (Arcturus, Mountain View, CA).

cDNA Array Production and Microarray Slides

The cDNA microarray slides were generated by growing bacterial clones containing cDNA clones, which were then purified and amplified using reverse-transcription PCR (RT-PCR). Purified and concentrated PCR products were robotically deposited on aminosilane-coated slides using the MicroGrid II (BioRobotics, Genomic Solutions, Huntingdon, UK). Slides were generated as part of the activites at the Swegene DNA Microarray Resource Center at Lund University, Sweden (http://swegene.onk.lu.se). In *Articles I-IV*, all patient samples were hybridized to 27K slides. The cell lines (*Article I*) were hybridized to 32K slides using the same clone set and design as the 27K slides.

cDNA Synthesis, Labeling, and Hybridization

In *Article I*, cDNA synthesis and labeling of the poly(A) RNA obtained from the cell lines were performed using the CyScribe Post-labeling Kit according to the manufacturer's instructions (Amersham Biosciences, Uppsala, Sweden). The Cy5 and Cy3 targets were pooled, and 12 μg pd(A) (Amersham Biosciences), 20 μg Cot-1 DNA (Invitrogen), 6 mg yeast tRNA (Invitrogen), and 1.5 μl 50 x Denhardt´s solution (Nalgene, Cleveland, Ohio) were added. The labeled targets were dried and resuspended in 40 μl DIG Easy Hyb (Roche, Mannheim, Germany), and pre-hybridization of the slides was performed in 5×SSC, 0.1% SDS, and 1% BSA in 42°. Targets from the reference and cell lines were hybridized simultaneously for 18 hours at 42° in a humidified chamber (Corning, Acton, MS). The slides were washed and scanned in the G2565AA Agilent DNA Microarray Scanner (Agilent Technologies, Palo Alto, CA). In *Articles II-IV*, the samples were labeled as above but for prehybridization, hybridization, and posthybridization washes, the Pronto Universal Microarray Reagent System (Corning) was used. Images were analyzed using the GenePix4.0 software (GenePix, Foster City, CA) and the obtained data matrix was uploaded onto the BioArray Software Environment (BASE) (Saal et al., 2002).

Microarray Data Analyses

Quality Filtering

In *Articles I-IV*, reporters flagged as bad or absent in the GenePix software were filtered away within BASE, and normalization was performed using the Lowess algorithm (Yang et al., 2002). To correct for poor quality spots, an error model was used. Briefly, the error model moves uncertain measurements towards the

mean across assays for the position. The effect is two-fold. First, since expression values close to the mean are less important when finding correlation to classes, the correction reduces the risk that a gene with a few uncertain outlier measurements is incorrectly ranked as highly relevant for a class. Second, if the measurement of a position is uncertain in several assays, many values are moved towards the mean, with a subsequent decrease of the variance for that position. The poorly measured reporter is then more likely to be excluded in the following variation filter. After error correction, the data were filtered for variation and presence.

In *Article I*, a standard deviation of 0.3 and 100% presence was required. To analyze the cell line and primary leukemia data sets together, which were hybridized on slides of two different designs (32K and 27K, respectively), the cell lines and acute leukemias were mean-centered individually, and for the few reporters that occurred in duplicate, measurements were merged.

In *Articles II* and *III*, the data were analyzed as above, but a 95% presence was required and reporter multiplets were merged before analysis. To correct for an initially observed deviation of the gene expression values with regard to sample referral site, the data were mean-centered with respect to hospital (Lund vs Linköping). In *Article III*, a standard deviation of 0.5 was required before subsequent analyses.

In *Article IV*, no further filtering of the data with regard to variation and presence was performed after normalization and error model correction. Molecular characterization using FISH analyses of two cases with t(12;14)(p13;q11) revealed that the breakpoints on chromosome 14 were located within the T-cell receptor alpha/delta locus and in the vicinity of the *CCND2* gene on 12p. Because the molecular consequence of this rearrangement most likely was deregulation and activation of an oncogene at 12p by the strong regulatory elements of the T-cell receptor, cDNA microarray analyses were used to investigate the expression of genes within a 5 Mb region on 12p spanning the breakpoints.

Normalization

Before gene expression data can be compared and analyzed, the data must be normalized in order to minimize systematic variations in the measured gene expression levels. By reducing such variations, biological differences between two samples can be more easily distinguished and reliably analyzed (Quackenbush, 2002; Yang et al., 2002; Smyth and Speed, 2003).

Normalization adjusts for differences in the labeling and detection efficiencies of the two fluorochromes. In addition, the data are adjusted to compensate for differences in the quantity of the RNA used for labeling. During normalization, ratios are log transformed to obtain a continuous spectrum of values (Quackenbush,

2001). Log transformed values treat up- and downregulated genes equally, and a gene that is upregulated by a factor of 2 will have a log2(ratio) of 1 and a gene downregulated by a factor of 2 will have a log2(ratio) of -1. Naturally, a gene with an equal expression in both the query and the reference sample will have a log2(ratio) equal to zero. After normalization, the data for each gene are given as a gene expression ratio, i.e., the normalized value of the tumor sample divided by the normalized value for the reference sample. Several different algorithms are available for normalization of gene expression data, most of which assume that all genes or a smaller set of house-keeping genes on the array have an average expression level equal to one. Two widely used normalization methods for gene expression analyses are total intensity and the Lowess algorithms (Quackenbush, 2002; Yang et al., 2002). Briefly, in the total intensity algorithm, the assumption is made that the amount of RNA used for labeling is the same both for the query sample and the reference. Moreover, an equal number of genes are assumed to be over- and underexpressed in the query sample relative to the reference. The normalization factor is then used to rescale the intensities for each gene (Quackenbush, 2002).

In all four articles, normalization was performed using the Lowess algorithm, which is a regression algorithm that can compensate for non-linear relations between sample and reference. Ideally, in a microarray experiment, the scatter plot of the intensities of test and reference sample should cluster along a straight line and have a slope of 1. However, due to technical issues, detection differences of the fluorochromes exist that may be intensity-dependent. In addition, local intensity differences on the slides may be present due to the spatial location of the clones on the slide, related to variations in the printing process. The Lowess algorithm has been shown to correct efficiently for such biases (Yang et al., 2002).

Hierarchical Clustering Analysis

HCA is one of the most commonly used methods for analysis and visualization of microarray data. This is a two-way dimensional technique that groups samples and genes based on their similarity. Hence, samples or genes which are similar to each other cluster next to or close to each other. The generated data matrix is presented as a "heat-map" where each sample is represented by a column and each gene as a row (Eisen et al., 1998). Typically, a gene that shows a relative over-expression compared to the reference is colored in red and a gene that show a relative underexpression in green. Genes with values close to zero are colored in black.

There are different clustering methods, which are designated divisive and agglomerative clustering. Divisive methods start with all cases in one cluster, which is then broken down into smaller clusters until each case is in a unique branch. In contrast to divisive clustering, agglomerative methods start with a single

member of a cluster and then gradually join different members together to form new clusters. Several different algorithms can be used for clustering; two commonly used include average- and complete-linkage (Quackenbush, 2001; Shannon et al., 2003).

HCA has for years been one of the most popular methods used for analysis and visualization of microarray data. It is important to realize, however, that the clustering method and distance used for similarity measurements influences the size and shape of the clusters formed. Moreover, clustering algorithms will produce clusters from any data, and it is difficult to validate the strength of cluster membership, and hence hard to estimate the stability of the clusters formed (Shannon et al., 2003). In addition, preprocessing of data, such as normalization, quality filtering, and number of genes used for input, also influences cluster formation. Despite these potential drawbacks, clustering analysis remains a valuable and commonly used method for the analysis and visualization of gene expression data.

Principal Component Analysis and Isomap

PCA is a well-known algorithm for studying "relationships between variables" and has during recent years been successfully applied to gene expression data for exploration of similarities between samples or genes. Microarray analyses generate high dimensional data and PCA is a powerful method for reducing dimensionality. In PCA, only the two or three dimensions containing the largest variance are kept, and samples or genes are then plotted using the two or three principal components, thereby efficiently visualizing relationships between samples/genes. As in hierarchical clustering, samples with similar gene expression profiles cluster closer together than samples that are different from each other. In *Article II*, PCA was used on a set of purified hematopoietic cells of different lineages and maturation stages, revealing that cells clustered with respect to lineage and maturation. The pediatric ALLs with specific primary genetic changes were studied together with the normal cells by projecting the centroid (the mean of all samples in a given group) for each genetic change onto the three principal components determined from PCA on the normal cells only. This analysis revealed similarities between the normal and malignant cells and provided information on the degree of maturation of the malignant cells.

In *Article I*, MDS using geodesic distances (Isomap) was used to visualize the similarities between samples. This method has previously been shown to efficiently visualize relationships between samples in microarray data (Nilsson et al., 2004).

Gene Discriminatory Analyses

In *Articles I* and *II*, genes associated with lineage and genetic subtype were ascertained using signal-to-noise statistics (Golub et al., 1999). This algorithm emphasizes that the correlation (r) of a gene to a class is given by: $r = [\mu_1(g) - \mu_2(g)]/[\sigma_1(g) + \sigma_2(g)]$, where μ denotes the mean expression; g the gene; σ the standard deviation of gene g, and the number the class. The larger value of r, the higher the correlation of a gene to a given class. In *Articles I*, *II*, and *III*, the *P*-values were estimated using permutation testing; typically, 1000-5000 permutations were performed.

Supervised Learning Algorithms

In *Articles II* and *III*, the supervised learning algorithm *k*-NN was utilized to build classifiers that could predict a class of an unknown leukemic sample. As briefly mentioned earlier, supervised methods use *a priori* knowledge of the samples to identify genes correlating with a known class. Typically, the class may be a cancer subtype or a prognostic marker. When the classifier is built, it is common to divide the data set into two groups, one serving as a training data set and the other as a validation set. It is important to split the data set before gene selection, since the samples in the validation data set otherwise will influence the gene selection process, which may result in overfitting. Typically, overfitting results in non-reproducibility of the results obtained (Ransohoff, 2004). Instead of dividing the data set into a training and validation set, a crossvalidation procedure may be used, in which one or several samples are withheld from the building process of the classifier; then, the class of the withheld samples are predicted (Radmacher et al., 2002).

k-NN is a widespread supervised method used for classification of gene expression data. In brief, the distances between a test sample and samples of known classes are calculated, typically using Euclidian distances. The class of the test sample is determined by the number of closest neighbors (k) of a known class. To evaluate the performance of the classifiers, a leave-one-out crossvalidation procedure was used. In *Article II*, gene expression classifiers were built that could predict the genetic subtype among the B-lineage ALLs (e.g., *TCF3/PBX1*, *ETV6/RUNX1*, or high hyperdiploidy) with a high accuracy. In *Article III*, gene expression predictors were constructed for lineage, all genetic subtypes among the B-lineage ALLs, and for the AMLs. In *Article III*, we also attempted to classify the leukemias based on clinical parameters, such as risk group assignment, risk of relapse, and prediction of MRD at diagnosis.

RESULTS AND DISCUSSION

GENE EXPRESSION PROFILING OF LEUKEMIC CELL LINES REVEALS CONSERVED MOLECULAR SIGNATURES AMONG SUBTYPES WITH SPECIFIC GENETIC ABERRATIONS (*ARTICLE I*)

Immortalized hematopoietic cell lines with characteristic fusion genes are widely used to model different aspects of leukemogenesis and have been fundamental tools in the cloning, characterization, and functional analyses of a large number of leukemia-associated fusion genes (Drexler et al., 2000). It is known that cell lines acquire additional genetic aberrations in culture and little is known to what extent pathogenetically important transcriptional programs remain conserved upon establishment and passaging of individual cell lines. Using cDNA microarrays, we determined the gene expression profiles of 40 cell lines, with various primary genetic abnormalities, as well as of primary leukemias harboring the same changes. Unsupervised analysis revealed that hematopoietic cell lines of diverse origin, but with the same primary genetic changes, co-segregated, strongly suggesting that pathogenetically important regulatory networks remain conserved. Moreover, primary leukemias co-segregated with cell lines carrying identical genetic rearrangements, further supporting that critical regulatory pathways remain intact in hematopoietic cell lines despite their diverse origin and numerous passages in vitro. Recently, Fine et al (2004) studied a set of hematopoietic cell lines and B-lineage ALLs, and showed that cell lines and leukemias with identical genetic aberrations co-segregate upon cluster analyses. In this study, we confirmed and further extended these observations by investigating a larger number of cell lines, derived from all hematopoietic lineages, and by analyzing a larger number of specific genetic aberrations. Transcriptional signatures correlating with clinical subtypes/primary genetic changes were identified and annotated based on their biological/ molecular properties and chromosomal localization, providing biological insights into the downstream genes deregulated by the primary genetic change.

Surprisingly, the myeloid cell lines HL-60, SIG-M5, and OCI-AML2 segregated with cell lines harboring 11q23/*MLL* rearrangements. Although this, in principle, could be a reflection of their lineage specificity, Southern blot analysis was performed to investigate if the *MLL* gene was rearranged in these cell lines. Interestingly, one cell line, OCI-AML2, established from a patient with AML M4, had a rearranged *MLL* gene, not previously reported. The similarity in gene expression pattern of HL-60 and SIG-M5 with *MLL*-positive cell lines indicates

that they have activated downstream genes similar to *MLL* target genes, warranting further investigations.

Genes encoding tyrosine kinases are frequently implicated in human cancer (Futreal et al., 2004), and given the development of selective tyrosine kinase inhibitors (Gschwind et al., 2004), such genes have been targeted and extensively studied in vitro as well as in vivo. The expression profile of tyrosine kinase-encoding genes was therefore selectively investigated, identifying several differentially expressed members, segregating with primary genetic changes, which may be targeted with tyrosine kinase inhibitors. For example, *FGFR3* was found to be upregulated in CML cell lines in blast crisis (BC). Interestingly, *FGFR3* was recently shown to be highly expressed in transplanted CML patients in relapse (Dvorakova et al., 2001) and in $CD34^+$ CML cells (Dvorak et al., 2003).

The identified conserved signatures are likely to reflect regulatory networks of importance for the transforming abilities of the primary genetic changes and offer important pathogenetic insights as well as a number of targets for future rational drug design.

Molecular Signatures in Childhood Acute Leukemia and Their Correlations to Expression Patterns in Normal Hematopoietic Subpopulations (*Article II*)

Global gene expression profiles of a consecutive series of 121 childhood acute leukemias (87 B-lineage ALL, 11 T-cell ALL, and 23 AML), six NBMs, and ten normal hematopoietic subpopulations of different lineages and maturations were ascertained using 27K cDNA microarrays. The inclusion of normal hematopoietic cells provided a unique possibility to compare the gene expression patterns in primary leukemias with those seen in distinct normal hematopoietic cell subpopulations.

Unsupervised analysis using HCA revealed that 64 (85%) of the 75 acute leukemias with primary genetic aberrations segregated according to their genetic changes, i.e., *TCF3/PBX1*, *IGH@/MYC*, *ETV6/RUNX1*, 11q23/*MLL*, and high hyperdiploidy. Two leukemias with a constitutional chromosome 21 co-segregated with cases harboring the *ETV6/RUNX1* fusion gene, indicating that leukemias in children with Down syndrome and *ETV6/RUNX1*-positive cases are similar at the global gene expression level.

Supervised discriminatory analyses were used to identify differentially expressed genes correlating with lineage and primary genetic change, providing

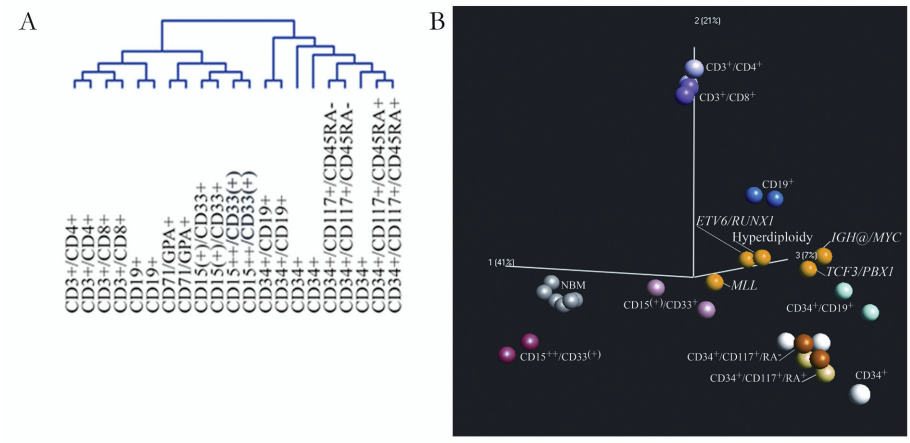

Figure 3. Unsupervised analyses of normal and malignant cells. A. Hierarchical dendogram using Pearson correlation and average linkage of the normal cells showing that samples cluster according to lineage and that the $CD34^+$ samples co-segregate and are separated from the more mature cell populations. **B.** The centroids of pediatric leukemias harboring specific genetic changes were projected into the principal components resulting from PCA of the normal cells only, thus enabling visualization of the similarities of the malignant cells in relation to the normal cell subpopulations.

biological insights into the different genetic subtypes of childhood leukemias. For example, in t(1;19)-positive ALLs – but not in the other genetic subtypes – an enrichment of cell cycle- and cell proliferation-associated genes was seen. Only two B-lineage ALLs harbored *IGH@/MYC*, but it is noteworthy that both displayed a distinct signature with elevated expression of *MYC* and deregulated expression of a number of genes known to be targeted by *MYC*, e.g., *BCL2* and *HMGA1* (Wood et al., 2000; Eischen et al., 2001). In ALLs with *ETV6/RUNX1*, pathway analysis of the upregulated genes showed enrichment of genes involved in the phosphatidylinositol signaling system and in beta-catenin signaling. A striking feature of the hyperdiploid ALLs was the significant correlation between the chromosomal location of the upregulated genes and the presence of trisomies/tetrasomies involving chromosomes X, 4, 6, 10, 14, 17, 18, and 21, likely reflecting a gene-dosage effect. Finally, among leukemias with 11q23/*MLL* rearrangements, a common transcriptional program was identified, irrespective of lineage, with elevated expression of *HOXA10*, *HOXA4*, *MEIS1*, and *PBX3*.

The gene expression profiles of normal hematopoietic cells were also studied. HCA revealed that the normal cells segregated according to lineage and that cells expressing the CD34 marker clustered together in a distinct branch (Figure 3A). Using PCA, a differentiation axis was exposed, reflecting lineages and maturation stages of normal hematopoietic cells. By applying the three principal components

obtained from PCA of the normal cells on the centroids of the leukemic samples, the degree of maturation of childhood leukemias with characteristic genetic changes could be visualized (Figure 3B). For example, the t(1;19)-positive ALLs clustered close to the $CD34^{+}/CD19^{+}$ cells, indicating an arrest at an early differentiation stage, despite their common lack of the CD34 marker. The t(12;21)-positive cases, on the other hand, clustered closer to the pre-B-cells, consistent with a recent report showing that this fusion gene arises in a committed B-cell progenitor (Castor et al., 2005).

To investigate the genes differentially expressed in the different genetic subtypes, the top 200 genes for each genetic subtype were investigated in the NBMs and the purified hematopoietic cells of different lineages and maturations. Interestingly, several of the genes were found to be preferentially expressed in the leukemias and not in the normal hematopoietic cells, suggesting that leukemic cells display a deregulated activation of transcriptional programs, not active in normal cells. However, some genes highly expressed by the leukemic cells were also highly expressed in normal cells but of a different lineage than the leukemic samples. Moreover, some clusters of genes, highly expressed by the leukemias, were found that were also highly expressed only in selected cell populations.

Apart from showing that leukemias segregate according to lineage and genetic subtype, this study also provided an extensive investigation of the genes correlating with primary genetic changes. For the first time, the expression patterns of these genes were also investigated in normal hematopoietic cells of different lineages and maturations, identifying genes preferentially expressed by the leukemic cells, suggesting an ectopic activation of a large number of genes. These genes likely reflect regulatory networks of pathogenetic importance and may also provide attractive targets for future directed therapies.

MICROARRAY-BASED CLASSIFICATION OF A CONSECUTIVE SERIES OF 121 CHILDHOOD LEUKEMIAS: PREDICTION OF LEUKEMIC AND GENETIC SUBTYPE AS WELL AS OF MINIMAL RESIDUAL DISEASE STATUS (*ARTICLE III*)

To build gene expression predictors that could classify the ALLs/AMLs analyzed in *Article II*, the supervised learning algorithm *k*-NN was utilized. Leukemic samples were classified according to lineage and genetic subtype, i.e., *TCF3/PBX1*, *ETV6/RUNX1*, high hyperdiploidy, and 11q23/*MLL* rearrangements, with high accuracies (97% and 98%, respectively). Validation experiments in an independent data set verified the high prediction accuracies of the classifiers. B-lineage ALLs

with uncharacteristic cytogenetic aberrations or with a normal karyotype displayed heterogeneous gene expression profiles, resulting in low prediction accuracies.

MRD status in T-cell ALLs with a high (>0.1%) MRD at day 29 (n = 9) could be predicted with 100% accuracy already at the time of diagnosis. Gene ontology (GO) analyses showed that cases having a high MRD showed a significant enrichment of genes involved in the JAK/STAT-cascade, e.g., *JAK3*, *STAT2*, and *STAT5A*, among the upregulated genes. Interestingly, *JAK3* is known to play an important role in T-cell development, and targeted deletion in mice leads to severe defects in thymocyte development (Eynon et al., 1999). In humans, *JAK3* mutations cause severe combined immunodeficiency with complete absence of T- and natural killer cells (Notarangelo et al., 2000). Furthermore, dysregulation of *Jak3* in mice causes increased resistance to apoptosis (Wen et al., 2001). In addition, many genes belonging to the cytochrome P450 superfamily, involved in drug metabolism, were highly expressed. It is tempting to speculate that these genes are responsible for the slow treatment response in T-cell ALL, although this remains to be elucidated.

A major problem in clinical practice is to identify patients with an increased risk of relapse. To detect such patients already at diagnosis would most likely be beneficial, making it possible to use a different treatment strategy already up-front. However, we failed to predict relapses in high hyperdiploid ALLs, in AMLs with *MLL* rearrangements, and in T-cell ALLs. The reasons for this are most likely manifold, including small patient groups, few relapses, and/or heterogeneous expression patterns associated with relapse in these subtypes. However, a gene expression profile distinguishing diagnostic samples from relapse samples in *ETV6/RUNX1*-positive ALL was identified.

In pediatric leukemias with uncharacteristic cytogenetic aberrations or a normal karytotype, HCA identified two subgroups: one consisting mainly of cases remaining in complete remission (CR) and one containing a few patients in CR and all but one of the patients who relapsed. Significance analysis of microarray (SAM) was used to identify genes associated with the two groups, identifying 72 genes. GO and pathway analyses revealed a highly significant enrichment of cell cycle-related genes among the lowly expressed genes in group 2, possibly indicating impaired cell proliferation in these samples which could be coupled to a decreased sensitivity towards drugs that are effective on proliferating cells. No distinct cytogenetic pattern was found to be characteristic for the two groups. Notably, however, three cases in group 1 harbored a dic(9;20)(p11-13;q11), a rearrangement that has been suggested to be associated with a favorable prognosis (Clark et al., 2000). Because the analyses were carried out on a rather small group of patients, further studies are clearly needed to validate these findings.

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 (*ARTICLE IV*)

In this study, two pediatric t(12;14)(p13;q11)-positive T-ALLs were characterized using FISH, cDNA microarray, and Real-Time PCR. FISH was used to map the region and revealed breakpoints (BPs) in the T-cell receptor alpha/delta locus (14q11) and in the vicinity of the *CCND2* gene at 12p13. To investigate the expression of genes in 12p13, cDNA microarray analysis was performed on one of the cases. The gene expression analysis was focused on a 2 Mb region spanning the BP and the relative level of expression of the genes in this region over the chromosome was investigated. Expression data for eight genes, including *CCND2*, surrounding the 12p BP were compared with those in other T-ALLs. The t(12;14)-positive T-ALL displayed a remarkably increased expression of *CCND2* compared to the controls, whereas the expression of the other genes was similar in all T-ALLs. Expression of *CCND2* and two additional genes (*PARP11* and *FGF23*), close to the 12p BP, was also investigated with real-time PCR of the two t(12;14)-positive cases and four controls. Neither *PARP11* nor *FGF23* displayed expression differences among the T-ALLs, whereas *CCND2* was clearly overexpressed in both t(12;14)-positive cases as compared to the mean expression level in the controls. CCND2 plays an important role in cell cycle progression and forms complexes with CDK4 and CDK6, whose activity is necessary for the G1/S-transition of the cell cycle. In addition, CCND2, has been shown to phosphorylate the RB1 protein (Ortega et al., 2002). While this manuscript was being prepared for submission, Clappier and coworkers (2005) reported a Real-Time quantitative PCR screening of T-ALLs, identifying three cases with *CCND2* overexpression, one harboring a t(12;14), further supporting the involvement of *CCND2* in this translocation.

The t(12;14) is the first neoplasia-associated translocation shown to result in overexpression of *CCND2* and it is also the first example of a T-cell neoplasm with a targeted deregulation of a member of a cyclin-encoding gene family.

GENERAL DISCUSSION

Since the discovery of the Ph chromosome in CML in 1960 (Nowell and Hungerford, 1960), close to 350 recurrent chromosomal aberrations have been described in hematologic malignancies (Mitelman et al., 2004). It has become increasingly clear that the specific genetic rearrangements detected in leukemias are intimately correlated with leukemic subtype and often also associated with prognosis. Today, the genetic abnormalities present at diagnosis are used together with other clinical parameters, such as age, white blood cell count, central nervous system involvement, and response to treatment, for stratification of patients to individual risk groups. Apart from providing important clinical information, the cytogenetic characterization of hematologic maliganancies has also been instrumental for the isolation of fusion genes of importance in leukemogenesis. The transforming properties of several fusion genes have been elegantly demonstrated using different mouse models (Rabbitts, 2001), but it still remains unclear how individual fusion genes elicit their leukemogenic properties.

The malignant state of cancer cells with their altered activity of pathways with important roles in cellular differentiation, proliferation, and death is suspected to be reflected in the global gene expression levels. In addition, the gene expression pattern in hematologic malignancies, characterized by the presence of specific translocations, is likely to mirror lineage-specific genes, reflecting the differentiation block seen in leukemia. There are now several gene expression studies of hematologic malignancies providing data showing that acute leukemias with specific genetic changes display unique and distinct gene expression profiles. Furthermore, a large number of genes have been identified that are associated with genetic subtypes, and sometimes also with prognostically important variables (Yeoh et al., 2002; Ross et al., 2003, 2004; Bullinger et al., 2004; Valk et al., 2004; Andersson et al., 2005a, b, c; Cario et al., 2005; Haferlach et al., 2005; van Delft et al., 2005).

Immortalized leukemic cell lines have been fundamental tools in the cloning and characterization of fusion genes in hematologic malignancies (Drexler et al., 2000). However, it has been unknown to what extent hematopoietic cell lines, despite their diverse origin and numerous passages in vitro, maintain the expression profiles associated with the primary genetic change. In *Article I*, it was shown that immortalized leukemic cell lines maintain a gene expression pattern characteristic of the primary genetic change, which is in line with a previous, but smaller, investigation (Fine et al., 2004). The finding of a maintained gene expression pattern in immortalized hematopoietic cell lines is important because they are likely to continue to serve as fundamental tools for investigating basic and applied aspects of leukemia cell biology. Given the recent developments of drugs targeting genes encoding tyrosine kinases in cancer, e.g., *BCR/ABL1*, *KIT*,

and *FLT3*, targeting of such genes in combination with gene expression profiling is likely to provide important biological insights into the pathways controlled by these tyrosine kinases. In addition, gene knock-down studies using RNA interference in combination with microarrays provide a powerful tool to monitor the regulatory networks deregulated in leukemia.

The molecular characterization of primary leukemia samples using gene expression profiling has focused on three main questions: (i) Is it possible to predict a class of an unknown sample to an already defined tumor class, referred to as "class prediction"? (ii) Can we find new tumor subtypes, referred to as "class discovery"? (iii) Can we increase our understanding of the underlying biology of the complex regulatory genetic networks and pathways deregulated in leukemia?

Class prediction has proved successful in leukemia, and several studies have constructed gene expression predictors using different supervised methods such as *k*-NN, SVM, and ANN (Yeoh et al., 2002; Ross et al., 2003, 2004; Andersson et al., 2005b, c; Haferlach et al., 2005; van Delft et al., 2005). In the present study, the supervised learning algorithm *k*-NN was used to build gene expression classifiers that could predict the leukemic and genetic subtype of an individual patient. It was shown that lineage and genetic change could be predicted with very high accuracies (97 and 98%, respectively), in line with recent reports (Yeoh et al., 2002; Ross et al., 2003, 2004; van Delft et al., 2005). However, when cases with uncharacteristic genetic changes were included, the classification accuracy dropped markedly, indicating that such cases have heterogeneous gene expression profiles. Classifiers were also built for different clinically important variables, most of which resulted in low prediction accuracies. The MRD status at day 29, could, however, be predicted at diagnosis in T-cell ALLs with 100% accuracy.

The high classification correctness of hematologic malignancies most probably reflects the strong influence of the primary genetic changes on the transcriptional networks in leukemia. However, although several large data sets of leukemia have been molecularly investigated, it has proved difficult to predict clinically important variables. The gene expression classifiers published so far for prediction of relapse or MRD have not been successful on independent data sets (Chiaretti et al., 2004; Andersson et al., 2005c). This may be a reflection of small patient subgroups resulting in identification of genes that cannot be reproducibly identified in independent studies. In addition, if the gene expression patterns are heterogeneous, or do not have a large impact on the expression profiles, it will be harder to find genes that can serve as good predictors across data sets (Ransohoff, 2004). One way to overcome the problem of small patient subgroups would be to perform meta-analytical studies and combine data sets generated from different groups.

As to the second question, addressing the possibility of finding new molecular subgroups using global expression profiling, only a few studies have identified novel subgroups, and their true nature remains to be verified in independent studies (Yeoh et al., 2002; Tsutsumi et al., 2003; Yagi et al., 2003; Bullinger et al., 2004; Valk et al., 2004; Andersson et al., 2005c). In *Article III*, HCA identified two groups of patients with uncharacteristic genetic changes, designated groups 1 and 2. GO analyses revealed enrichment of cell cycle-related genes, upregulated in group 1. This group mainly contained cases remaining in CR after treatment, whereas group 2 consisted of six diagnostic cases that remained in CR, one relapse sample, and two samples from diagnostic cases that later relapsed. It is tempting to speculate that a decreased expression of cell-cycle related genes in group 2 could be coupled to a decreased sensitivity to drugs targeting proliferating cells, a finding which could have important clinical implications. This, however, remains to be confirmed. The identified gene signature was applied on the samples designated "others" in the data set from Ross et al (2003); however, our set of genes was not able to separate the novel subgroup described by Ross and collegues (2003).

The final question addresses the possibility to improve our knowledge about the transcriptional programs becoming deregulated in leukemia using gene expression profiling. It has proved more difficult than initially hoped to draw biological conclusions out of the vast data generated from such studies (Ebert and Golub, 2004). This may in part be due to the fact that microarray data are subjected to substantial technical and biological noise. Moreover, it is likely that subtle gene expression patterns are hidden among the dominant patterns associated with lineage and differentiation, making it difficult to uncover important regulatory pathways. In addition, there has been a lack of gene expression studies of normal hematopoietic cells, which are fundamental in order to understand the perturbed transcriptional profiles that are deregulated in leukemia. Without detailed knowledge about the normal regulation of blood cells, it is difficult, perhaps even impossible to interpret the gene expression profiles found in malignant cells. Hence, gene expression studies of normal hematopoietic cells manipulated in vitro to differentiate and/or proliferate, and a careful monitoring of the gene expression changes during these processess would probably provide important information of the genetic regulatory mechanisms of normal and leukemic cells.

Investigation of the gene expression profiles of normal cells was partly addressed in *Article II* where malignant gene signatures were studied in normal hematopoietic cells. In that study, the gene expression profiles of a large series of pediatric leukemias as well as of normal hematopoietic cell subpopulations were investigated. The inclusion of normal cells provided a unique possibility to compare the malignant gene expression profiles identified for the genetic subtypes

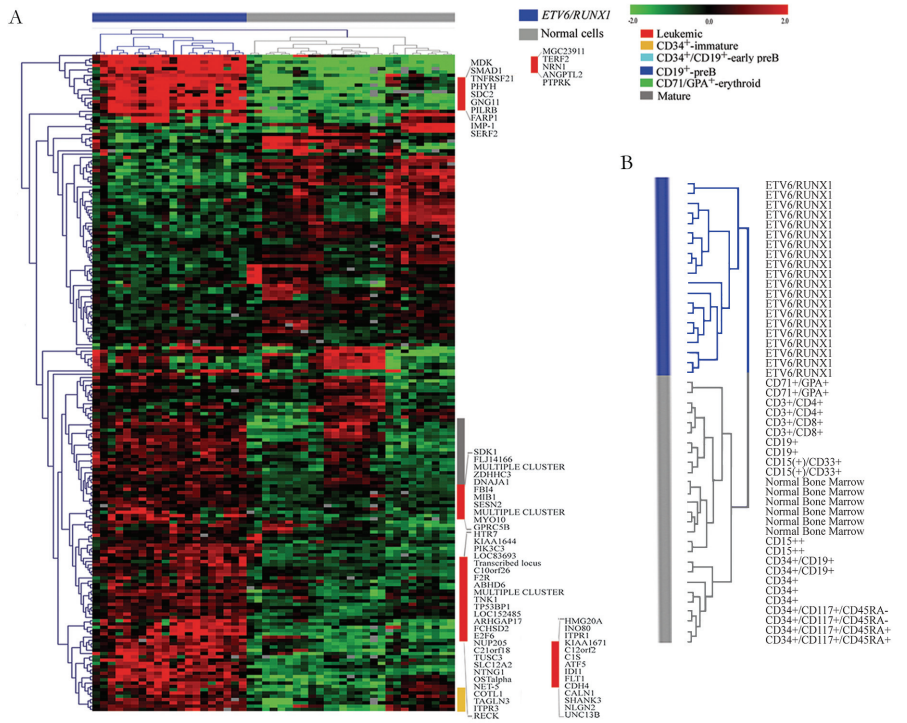

Figure 4. The expression of the top 200 genes in normal hematopoietic cells. A. The top 200 *ETV6/RUNX1*-associated genes applied on normal cells. The gene symbols for the genes highly expressed by the leukemic cells are given in the picture. **B.** Enhanced dendogram of the clustering of the top 200 genes.

in the normal cells. Unsupervised analyses showed that the leukemias segregated according to lineages and genetic changes, supporting previous reports (Yeoh et al., 2002; Ross et al., 2003, 2004; Andersson et al., 2005a). For the first time, malignant gene signatures were investigated in normal hematopoietic cells of different lineages and maturations. The main finding of this analysis was that leukemic cells express a large number of genes not found in the normal hematopoietic subpopulations, suggesting that leukemic cells display a deregulated activation of transcriptional programs not active in normal cells. In addition, genes were found highly expressed in the leukemias that also showed elevated expression in normal cells of a different lineage, suggesting an aberrant activation of genes not normally expressed in cells of that lineage (Figure 4). It is likely that the genes identified may not only reflect pathogenetically important regulatory pathways but that they also may provide attractive targets for future directed therapies.

In conclusion, the microarray technology is rapidly developing and with increasing numbers of probe sets and exon specific probes, it is likely that gene

expression profiling will improve our knowledge of the transcriptional programs that are altered in leukemia. In parallel with technical improvements, it is important that mathematical and bioinformatic tools are developed to allow improved data management and interpretation. Moreover, enhanced understanding of the transcriptional programs regulating the fine tuned expression of genes during normal cellular responses is needed before we can truly understand the underlying biology of the gene expression signatures of malignant cells. Although the enigma of the complex regulatory networks deregulated in leukemia cannot be solved solely by measuring mRNA expression levels, it is likely that the combined knowledge gathered through gene expression studies together with proteomic and other genomic tools available will provide valuable clues to the pathways that are involved in leukemia development and progression. Such analyses will hopefully also identify new molecular subgroups that may benefit from either more or less intensive treatment modalities and identify targets for future rational drug design.

CONCLUSIONS

In this thesis project, gene expression profiling was used to characterize molecularly immortalized hematopoietic cell lines, primary pediatric leukemias, and normal hematopoietic cells. The main conclusions of the present study may be summarized as follows:

Article I

- Immortalized hematopoietic cell lines with the same primary genetic changes, display similar gene expression profiles, despite their diverse origin and numerous passages in vitro, suggesting that pathogenetically important regulatory networks remain conserved.

Article II

- Based on their global gene expression patterns, pediatric leukemias segregate according to their lineage and primary genetic change.
- Several deregulated pathways were identified in the different genetic subtypes of the pediatric leukemias
- A large number of genes are preferentially expressed only by the leukemic cells, indicating ectopic activation of genes, likely to reflect regulatory networks of pathogenetic importance.
- Genes were highly expressed in malignant cells as well as in normal immature CD34-positive cells, possibly reflecting the cellular origin of the leukemia or, alternatively, an aberrant or maintained expression of genes that are active in immature hematopoietic cells.

Article III

- Gene expression classifiers could be built that with high accuracy predicted lineage and specific genetic change in childhood leukemias.
- The MRD status in T-cell ALLs at day 29 of treatment could be predicted with 100% accuracy at diagnosis.
- In leukemais with uncharacteristic genetic changes, unsupervised analyses identified two novel subgroups that differed with respect to the expression of cell cycle related genes.

Article IV

- In t(12;14)-positive T-ALLs, breakpoint mapping using FISH and gene expression analyses revealed that *CCND2* was juxtaposed to the vicinity of the TCR alpha/delta locus, resulting in aberrant expression of *CCND2.*

SUMMARY IN SWEDISH

Varje år insjuknar ca 460 personer i akut leukemi (blodcancer) i Sverige. Leukemi är en sjukdom som drabbar celler med viktiga funktioner i vårt immunförsvar och kännetecknas av att omogna celler ansamlas i framförallt benmärg och blod. Bristen på mogna och fungerande vita blodceller i benmärgen leder till en ökad infektionsbenägenhet, ökad risk för blödningar och blodbrist. Frekvensen av leukemi hos vuxna ökar med åldern och den vanligaste typen av leukemi hos äldre är akut myeloisk leukemi (AML). Hos barn ser mönstret annorlunda ut, med högst förekomst i 3-5 års ålder och barn drabbas framförallt av akut lymfatisk leukemi (ALL).

Vår arvsmassa (generna) finns lagrad som DNA i kromosomerna vilka finns i cellkärnan i varje cell i vår kropp. Cancer är en sjukdom som kännetecknas av förvärvade genetiska förändringar (mutationer) i DNA och där en ansamling av dessa vanligtvis sker vid cancerutveckling. Leukemier kännetecknas av specifika kromosomförändringar och sedan den första genetiska avvikelsen beskrevs 1960 av Nowell och Hungerford har nu fler än 350 återkommande genetiska förändringar identifierats vid leukemi. En vanlig genetisk avvikelse utgörs av så kallade translokationer, vilka innebär att genetiskt material från två olika kromosomer sätts samman och därigenom ger upphov till en ny sammanslagen gen, en sk fusionsgen. Fusionsgenen i sin tur bildar ett förändrat protein med cancer-omvandlande egenskaper. Eftersom fusionsgener många gånger består av sk transkriptionsfaktorer, det vill säga gener som styr andra geners uttryck, leder uttrycket av fusionsgenen till förändring i regulatoriska nätverk som styr viktiga funktioner för cellens tillväxt, utmognad och överlevnad. Förekomsten av en fusionsgen ger information om vilken typ av leukemi det rör sig om och många gånger erhålls viktig prognostisk information. Ibland kan den specifika fusions-genen dessutom ha betydelse för vilken typ av behandling som den enskilda patienten får. Trots att leukemier har studerats under lång tid är kunskapen om hur uttrycket av en fusionsgen ger upphov till leukemi relativt begränsad. I mitten på 1990-talet introducerades microarray-teknologin (på svenska ”mikromatris”), vilken har kommit att revolutionera vår möjlighet att studera hur cancer förändrar cellernas genetiska program. Microarraytekniken gör det möjligt att i ett enda försök studera uttrycket av tusentals gener och ger således ett molekylärt avtryck av de gener som är uttryckta i en vävnad vid ett visst tillfälle. Cancer är en sjukdom som kännetecknas av att genetiska program som styr cellens utmognad, delning och död är störda. I leukemier är det ofta gener som kodar för konserverade transkriptionsfaktorer med viktiga funktioner i cellernas utmognad som är förändrade, vilket karakteristiskt kan ses som en blockering vid

ett visst mognadsstadium. Genuttrycksmönstret i leukemicellerna kommer därför att reflekteras av vilka gener som är förändrade och vid vilken mognad cellerna blockerats. Eftersom arrayteknologin är så kraftfull finns det hopp om att den ska kunna användas diagnostiskt för att klassificera leukemier till kända grupper och för att identifiera undergrupper med bättre eller sämre prognos. Vidare ger den viktig information om de genetiska program som är förändrade till följd av leukemiuppkomsten.

Den övergripande målsättningen i denna avhandling har varit att använda microarrayteknologin för att utveckla diagnostiska klassningsverktyg samt studera de gener som förändrats vid leukemi. Vidare var målsättningen att finna nya undergrupper av leukemier med en bättre eller sämre prognos. Fyra delarbeten ingår i avhandlingen; i det första studerades genuttrycksprofilerna hos ett stort antal leukemicellinjer (*Artikel 1*). Cellinjer är viktiga modellsystem för att studera hur leukemier uppkommer och för att undersöka vilka genetiska program eller signalvägar som är störda vid leukemi. Det är känt att genetiska förändringar ansamlas när cellinjerna odlas i laboratoriet, men det har varit okänt huruvida de bibehåller det ursprungliga genuttrycksmönstret som kännetecknar den primära genetiska avvikelsen i leukemicellerna. Ett stort antal cellinjer och patienter med specifika genetiska förändringar undersöktes och visade att leukemicellinjer bibehåller ett genuttrycksmönster som speglar den primära genetiska förändringen trots att de har olika ursprung och har odlats under lång tid i skilda laboratorier. Denna kunskap är viktig eftersom leukemicellinjer är några av våra främsta verktyg för att studera de genetiska nätverk som förändrats vid leukemi.

I ett andra arbete studerades 121 barnleukemier och ett stort antal normala blodceller av olika typer och mognadsstadium (*Artikel II*). Denna undersökning visade att barnleukemier med karakteristiska genetiska avvikelser hade unika genuttrycksmönster. Vidare studerades uttrycksmönstret av gener som var associerade med primära genetiska förändringar i normala hematopoietiska celler av olika typer och mognadsgrad, vilket visade att flertalet gener var selektivt högt uttryckta i leukemicellerna. Detta kan tyda på att leukemierna aktiverar gener som inte är uttryckta under normal utmognad. Det är troligt att dessa gener är viktiga för leukemiuppkomsten och att de därmed också eventuellt kan fungera som mål för utveckling av framtida terapier.

I delarbete tre var syftet att använda matematiska/statistiska metoder för att undersöka om genuttrycksmönstret i leukemicellerna kunde användas för att förutsäga vilken typ av leukemi ett prov tillhörde samt om viktiga kliniska parametrar kunde förutspås redan vid diagnostillfället (*Artikel III*). Leukemierna kunde klassificeras med hög precision avseende vilken typ av leukemi och genetisk avvikelse det rörde sig om, något som visar att genuttrycksanalyser kan utgöra ett viktigt instrument för klinisk diagnostik av leukemier. Man vet idag att kvarvarande

leukemiceller dag 29 efter inledande behandling ökar risken för återfall och av den anledningen undersökte vi det om dessa patienter gick att identifiera redan vid diagnostillfället. På basis av genuttrycksmönstret vid diagnos var det möjligt att klassificera T-cellsleukemier beroende på om de skulle ha kvarvarande leukemiceller dag 29. En sådan prediktor kan i framtiden vara ett betydelsefullt kliniskt verktyg för att förbättra riskgrupperingen för denna leukemiform.

I det fjärde arbetet undersöktes brottspunkterna i en karakteristisk translokation – t(12;14)(p13:q11) – i två fall av T-cellsleukemi hos barn (*Artikel IV*). Fluorescent in situ hybridisering visade att brottspunkten i kromosom 14 låg i T-cellsreceptor alpha/delta lokus och i närheten av genen *CCND2* på kromosom 12. Microarray analys användes därefter för att undersöka uttrycket av 8 gener runt brottspunkten på kromosom 12. Denna analys visade att *CCND2*, en gen som är viktig för celltillväxt, var högt uttryckt i jämförelse med andra T-cellsleukemier, medan de andra generna uppvisade ett likartat uttryck i alla leukemierna. Detta fynd, som kunde bekräftas med sk realtids-PCR, är det första exemplet på dereglering av en cyclinrelaterad gen i T-cellsleukemi.

Sammanfattningsvis har studierna i denna avhandling bidragit till en ökad kunskap om de genuttrycksmönster och regulatoriska nätverk som förändras vid leukemi. Ett stort antal gener med ett selektivt högt uttryck i leukemiceller kunde identifieras, vilka ger viktig information om vilka genetiska program som är störda vid leukemi. Förhoppningsvis kan dessa gener i framtiden utgöra attraktiva mål för utveckling av nya behandlingsformer.

ACKNOWLEDGEMENTS

The work presented in this thesis would not have been possible without the help from a lot of people. In particular, I would like to express my sincere gratitude and appreciation to:

Thoas Fioretos, my excellent tutor and supervisor, for sharing with me your expertise in cancer genetics and for teaching me about the exciting world of research. In particular, I would like to thank you for your kindness, for your capacity to have time when there is none, for your ability to find solutions on problems I thought were impossible to solve, for your believe in me, and of course, for being so passionate about your work and for letting me be a part of it. These years have been extremely rewarding and you have certainly played a central role in making them so fun and inspiring – thank you!

Professor *Felix Mitelman*, Head of the Department of Clinical Genetics, for creating such a stimulating and creative environment for high quality research. In particular, I would like to thank you for giving me the opportunity to become a PhD student at the Department.

My co-supervisors, *Bertil Johansson* and *Mattias Höglund*. I would like to thank you *Bertil* for making me laugh and for the nice chats in the coffee room in the early mornings, we both know that "morgonstund har guld i mun". Your knowledge of cancer genetics and English grammar has resulted in many nice scientific discussions over the years and you have learnt me a lot. *Mattias*, I want to thank you for your excellent advises regarding lab work and for nice scientific discussions.

My friends and fellow PhD students: *Aikaterini*, *Anna*, *Björn*, *Charlotte*, *David G*, *David L*, *Helena*, *Josef*, *Kajsa*, *Karoline*, *Karin*, *Malin*, *Markus*, *Petra*, *Srinivas*, *Therese*, *Tord*, and *Ylva*, at the Department of Clinical Genetics. The years that we have spent together have resulted in a very good friendship, which is so important in order to be able to achieve new goals and produce good research. We have shared many laughs during our dinners, conferences, and Friday pubs. However, I think the most enjoyable times have been those everyday-moments in "doktorandbåsen" where it is impossible not to share life with each other. These years would not have been the same without you! In particular I want to thank: *Karin Broberg* - for friendship and nice discussions from science to life. *Kajsa Paulsson* and *Petra Håkansson* – for all the fun that we have shared and for the scientific and private discussions at our "leukemia research" dinners and during our travels. *Petra*, for being a terrific companion in the lab and for the pleasant morning chats. *Kajsa* for being an excellent company during our travels, for laughs, and for putting so much "glitter" to life, and of course for being a "skarping" in every aspects of life! *David Lindgren* for sharing with me the life of spots – colored in red and green – for our hilarious moments when sitting endless hours in front of the computers looking at scatter plots, hierarchical dendograms,

heat maps, PCAs, MDSs, SAMs, T-tests, QTCs, CASTs, COAs, PTMs, PAMs, and EASE analyses!

The senior research staff: *Catarina*, *Fredrik*, *Kristina*, *Karin*, *Ludmila*, *Mia*, *Nils*, *Nina*, *Ioannis*, *Samuel*, *Ulf*, and *Yuesheng*, for nice scientific discussions and for the highly appreciated out-of-the-lab activities.

The rest of the personnel at the Department of Clinical Genetics for nice coffee breaks, dinners, and pleasant company both in- and outside the lab. In particular, I want to thank *Carin Lassen* for your expertise and for teaching me so many things, not only about the lab work, but also about life. We have shared many laughs and thoughts during the years! *Yvonne Murray*, my first bench-mate and friend for life - for being so generous – for opening up your home and arms, you have become my second family.

Åke Borg and the people at the Department of Oncology. In particular I want to thank *Carina*, *Göran*, *Johan*, *Lao*, and *Sofia* for the sometimes tiring, but almost always very happy and cheerful work during the preparation of the clones and optimization of the technology during the early days of microarrays.

Patrik Edén and *Cecilia Ritz* at the Department of Complex System Division. In particular, I want to thank you *Patrik* for your expertise and for teaching me about the world or algorithms.

Magnus Fontes, *Johan Råd*e, and *Jens Nilsson* at the Department of Mathematical Sciences, for introducing me into the mysteries of mathematics and for the remarkable "spinning" projections of our data.

Jesper Heldrup at the Department of Pediatrics, for being so enthusiastic about your work and for tutoring me in clinical hematology.

To all the *co-authors* for stimulating collaborations.

My *Family*, my parents *Paul* and *Lisbeth*, for always believing in me and for being there for me, whatever has happened along my path of life. For sharing memorable moments and for the wonderful life we have together. My brothers, *Joakim* with family, and *Martin* for being the best brothers a sister can have - for putting up with me - and for always taking care of "syster yster".

Finally, I want to thank all my friends outside the lab for remembering me about "the world out there", for sharing so many laughs and for sharing life.

The original research presented in this thesis was supported by grants form the Swedish Cancer Society, the Swedish Children's Cancer Foundation, the Medical Faculty of Lund University, and the IngaBritt and Arne Lundberg Foundation.

REFERENCES

Adolfsson J, Mansson R, Buza-Vidas N, Hultquist A, Liuba K, Jensen CT, Bryder D, Yang L, Borge OJ, Thoren LA, Anderson K, Sitnicka E, Sasaki Y, Sigvardsson M, Jacobsen SE. 2005. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell 121:295-306.

Akashi K, Traver D, Miyamoto T, Weissman IL. 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 404:193-197.

Alter O, Brown PO, Botstein D. 2000. Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci U S A 97:10101-10106.

Andersson A, Edén P, Lindgren D, Nilsson J, Lassen C, Heldrup J, Fontes M, Borg A, Mitelman F, Johansson B, Höglund M, Fioretos T. 2005a. Gene expression profiling of leukemic cell lines reveals conserved molecular signatures among subtypes with specific genetic aberrations. Leukemia 19:1042-1050.

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

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, Fioretos T. 2005c. 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. Manuscript.

Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR, Korsmeyer SJ. 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet 30:41-47.

Ayton PM, Cleary ML. 2001. Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. Oncogene 20:5695-5707.

Bain G, Maandag EC, Izon DJ, Amsen D, Kruisbeek AM, Weintraub BC, Krop I, Schlissel MS, Feeney AJ, van Roon M, van der Valk M, te Riele H, Berns A, Murre C. 1994. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell 79:885-892.

Björklund E, Mazur J, Söderhall S, Porwit-MacDonald A. 2003. Flow cytometric follow-up of minimal residual disease in bone marrow gives prognostic information in children with acute lymphoblastic leukemia. Leukemia 17:138-148.

Blyth K, Cameron ER, Neil JC. 2005. The RUNX genes: gain or loss of function in cancer. Nat Rev Cancer 5:376-387.

Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, Jr., Haussler D. 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci U S A 97:262-267.

Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, Dohner H, Pollack JR. 2004. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 350:1605-1616.

Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F, Croce CM. 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 99:15524-15529.

Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, Iorio MV, Visone R, Sever NI, Fabbri M, Iuliano R, Palumbo T, Pichiorri F, Roldo C, Garzon R, Sevignani C, Rassenti L, Alder H, Volinia S, Liu CG, Kipps TJ, Negrini M,

Croce CM. 2005. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. N Engl J Med 353:1793-1801.
Cario G, Stanulla M, Fine BM, Teuffel O, Neuhoff NV, Schrauder A, Flohr T, Schafer BW, Bartram CR, Welte K, Schlegelberger B, Schrappe M. 2005. Distinct gene expression profiles determine molecular treatment response in childhood acute lymphoblastic leukemia. Blood 105:821-826.
Caspersson T, Hulten M, Lindsten J, Zech L. 1970. Distinction between extra G-like chromosomes by quinacrine mustard fluorescence analysis. Exp Cell Res 63:240-243.
Castor A, Nilsson L, Åstrand-Grundström I, Buitenhuis M, Ramirez C, Anderson K, Strömbeck B, Garwicz S, Bekassy AN, Schmiegelow K, Lausen B, Hokland P, Lehmann S, Juliusson G, Johansson B, Jacobsen SE. 2005. Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia. Nat Med 11:630-637.
Chen CZ. 2005. MicroRNAs as oncogenes and tumor suppressors. N Engl J Med 353:1768-1771.
Chen-Kiang S. 2005. Biology of plasma cells. Best Pract Res Clin Haematol 18:493-507.
Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. 1999. Making and reading microarrays. Nat Genet 21:15-19.
Chiaretti S, Li X, Gentleman R, Vitale A, Vignetti M, Mandelli F, Ritz J, Foa R. 2004. Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. Blood 103:2771-2778.
Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M, Rassenti L, Alder H, Volinia S, Liu CG, Kipps TJ, Negrini M, Croce CM. 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A 102:13944-13949.
Clappier E, Cuccuini W, Cayuela JM, Vecchione D, Baruchel A, Dombret H, Sigaux F, Soulier J. 2005. Cyclin D2 dysregulation by chromosomal translocations to TCR loci in T-cell acute lymphoblastic leukemias. Leukemia Epub ahead of print.
Clark R, Byatt SA, Bennett CF, Brama M, Martineau M, Moorman AV, Roberts K, Secker-Walker LM, Richards S, Eden OB, Goldstone AH, Harrison CJ. 2000. Monosomy 20 as a pointer to dicentric (9;20) in acute lymphoblastic leukemia. Leukemia 14:241-246.
Crist WM, Carroll AJ, Shuster JJ, Behm FG, Whitehead M, Vietti TJ, Look AT, Mahoney D, Ragab A, Pullen DJ, et al. 1990. Poor prognosis of children with pre-B acute lymphoblastic leukemia is associated with the t(1;19)(q23;p13): a Pediatric Oncology Group study. Blood 76:117-122.
Debernardi S, Lillington DM, Chaplin T, Tomlinson S, Amess J, Rohatiner A, Lister TA, Young BD. 2003. Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events. Genes Chromosomes Cancer 37:149-158.
Deininger M, Buchdunger E, Druker BJ. 2005. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood 105:2640-2653.
DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 14:457-460.
DiMartino JF, Selleri L, Traver D, Firpo MT, Rhee J, Warnke R, O'Gorman S, Weissman IL, Cleary ML. 2001. The Hox cofactor and proto-oncogene Pbx1 is required for maintenance of definitive hematopoiesis in the fetal liver. Blood 98:618-626.
Donnellan R, Chetty R. 1998. Cyclin D1 and human neoplasia. Mol Pathol 51:1-7.
Drexler HG, Matsuo AY, MacLeod RA. 2000. Continuous hematopoietic cell lines as model systems for leukemia-lymphoma research. Leuk Res 24:881-911.

Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S, Zimmermann J, Lydon NB. 1996. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat Med 2:561-566.

Dudoit S, Fridlyand J. 2002. Introduction to classification in microarray experiments. In: Berrar DP, Dubitzky W, Granzow M, editors. A Practical Approach to Microarray Data Analysis: Kluwer Academic Publishers. p 201-215.

Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. 1999. Expression profiling using cDNA microarrays. Nat Genet 21:10-14.

Dvorak P, Dvorakova D, Doubek M, Faitova J, Pacholikova J, Hampl A, Mayer J. 2003. Increased expression of fibroblast growth factor receptor 3 in CD34+ BCR-ABL+ cells from patients with chronic myeloid leukemia. Leukemia 17:2418-2425.

Dvorakova D, Krejci P, Mayer J, Fajkus J, Hampl A, Dvorak P. 2001. Changes in the expression of FGFR3 in patients with chronic myeloid leukaemia receiving transplants of allogeneic peripheral blood stem cells. Br J Haematol 113:832-835.

Ebert BL, Golub TR. 2004. Genomic approaches to hematologic malignancies. Blood 104:923-932.

Ehrhardt A, Ehrhardt GR, Guo X, Schrader JW. 2002. Ras and relatives–job sharing and networking keep an old family together. Exp Hematol 30:1089-1106.

Eischen CM, Packham G, Nip J, Fee BE, Hiebert SW, Zambetti GP, Cleveland JL. 2001. Bcl-2 is an apoptotic target suppressed by both c-Myc and E2F-1. Oncogene 20:6983-6993.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863-14868.

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

Ernst P, Fisher JK, Avery W, Wade S, Foy D, Korsmeyer SJ. 2004. Definitive hematopoiesis requires the mixed-lineage leukemia gene. Dev Cell 6:437-443.

Eynon EE, Livak F, Kuida K, Schatz DG, Flavell RA. 1999. Distinct effects of Jak3 signaling on alphabeta and gammadelta thymocyte development. J Immunol 162:1448-1459.

Fang B, Zheng C, Liao L, Han Q, Sun Z, Jiang X, Zhao RC. 2005. Identification of human chronic myelogenous leukemia progenitor cells with hemangioblastic characteristics. Blood 105:2733-2740.

Felix CA. 1998. Secondary leukemias induced by topoisomerase-targeted drugs. Biochim Biophys Acta 1400:233-255.

Fialkow PJ, Jacobson RJ, Papayannopoulou T. 1977. Chronic myelocytic leukemia: clonal origin in a stem cell common to the granulocyte, erythrocyte, platelet and monocyte/macrophage. Am J Med 63:125-130.

Fiedler W, Serve H, Dohner H, Schwittay M, Ottmann OG, O'Farrell AM, Bello CL, Allred R, Manning WC, Cherrington JM, Louie SG, Hong W, Brega NM, Massimini G, Scigalla P, Berdel WE, Hossfeld DK. 2005. A phase 1 study of SU11248 in the treatment of patients with refractory or resistant acute myeloid leukemia (AML) or not amenable to conventional therapy for the disease. Blood 105:986-993.

Fine BM, Stanulla M, Schrappe M, Ho M, Viehmann S, Harbott J, Boxer LM. 2004. Gene expression patterns associated with recurrent chromosomal translocations in acute lymphoblastic leukemia. Blood 103:1043-1049.

Fischer M, Schwieger M, Horn S, Niebuhr B, Ford A, Roscher S, Bergholz U, Greaves M, Lohler J, Stocking C. 2005. Defining the oncogenic function of the TEL/AML1 (ETV6/RUNX1) fusion protein in a mouse model. Oncogene.

Fitzgibbon J, Smith LL, Raghavan M, Smith ML, Debernardi S, Skoulakis S, Lillington D, Lister TA, Young BD. 2005. Association between acquired uniparental disomy

and homozygous gene mutation in acute myeloid leukemias. Cancer Res 65:9152-9154.

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

Forestier E, Johansson B, Gustafsson G, Borgström G, Kerndrup G, Johannsson J, Heim S. 2000b. Prognostic impact of karyotypic findings in childhood acute lymphoblastic leukaemia: a Nordic series comparing two treatment periods. For the Nordic Society of Paediatric Haematology and Oncology (NOPHO) Leukaemia Cytogenetic Study Group. Br J Haematol 110:147-153.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. 2004. A census of human cancer genes. Nat Rev Cancer 4:177-183.

Golub TR, Barker GF, Bohlander SK, Hiebert SW, Ward DC, Bray-Ward P, Morgan E, Raimondi SC, Rowley JD, Gilliland DG. 1995. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. Proc Natl Acad Sci U S A 92:4917-4921.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531-537.

Gregorini A, Cinti C, Young BD. 1998. Molecular abnormalities in leukemia: the 11q23 story so far. J Biol Regul Homeost Agents 12:95-105.

Grimwade D. 2001. The clinical significance of cytogenetic abnormalities in acute myeloid leukaemia. Best Pract Res Clin Haematol 14:497-529.

Gruszka-Westwood AM, Horsley SW, Martinez-Ramirez A, Harrison CJ, Kempski H, Moorman AV, Ross FM, Griffiths M, Greaves MF, Kearney L. 2004. Comparative expressed sequence hybridization studies of high-hyperdiploid childhood acute lymphoblastic leukemia. Genes Chromosomes Cancer 41:191-202.

Gschwind A, Fischer OM, Ullrich A. 2004. Timeline: The discovery of receptor tyrosine kinases: targets for cancer therapy. Nat Rev Cancer 4:361-370.

Gunsilius E, Duba HC, Petzer AL, Kahler CM, Grunewald K, Stockhammer G, Gabl C, Dirnhofer S, Clausen J, Gastl G. 2000. Evidence from a leukaemia model for maintenance of vascular endothelium by bone-marrow-derived endothelial cells. Lancet 355:1688-1691.

Haas OA. 1996. Is genomic imprinting involved in the pathogenesis of hyperdiploid and haploid acute lymphoblastic leukemia of childhood? Acta Genet Med Gemellol (Roma) 45:239-242.

Haferlach T, Kohlmann A, Schnittger S, Dugas M, Hiddemann W, Kern W, Schoch C. 2005. A global approach to the diagnosis of leukemia using gene expression profiling. Blood.

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. Cell 100:57-70.

Harewood L, Robinson H, Harris R, Al-Obaidi MJ, Jalali GR, Martineau M, Moorman AV, Sumption N, Richards S, Mitchell C, Harrison CJ. 2003. Amplification of AML1 on a duplicated chromosome 21 in acute lymphoblastic leukemia: a study of 20 cases. Leukemia 17:547-553.

Harrington CA, Rosenow C, Retief J. 2000. Monitoring gene expression using DNA microarrays. Curr Opin Microbiol 3:285-291.

Harris MB, Shuster JJ, Carroll A, Look AT, Borowitz MJ, Crist WM, Nitschke R, Pullen J, Steuber CP, Land VJ. 1992. Trisomy of leukemic cell chromosomes 4 and 10 identifies children with B-progenitor cell acute lymphoblastic leukemia with a very low risk of treatment failure: a Pediatric Oncology Group study. Blood 79:3316-3324.

Harrison CJ, Moorman AV, Barber KE, Broadfield ZJ, Cheung KL, Harris RL, Jalali GR, Robinson HM, Strefford JC, Stewart A, Wright S, Griffiths M, Ross FM,

Harewood L, Martineau M. 2005. Interphase molecular cytogenetic screening for chromosomal abnormalities of prognostic significance in childhood acute lymphoblastic leukaemia: a UK Cancer Cytogenetics Group Study. Br J Haematol 129:520-530.

Heerema NA, Sather HN, Sensel MG, Zhang T, Hutchinson RJ, Nachman JB, Lange BJ, Steinherz PG, Bostrom BC, Reaman GH, Gaynon PS, Uckun FM. 2000. Prognostic impact of trisomies of chromosomes 10, 17, and 5 among children with acute lymphoblastic leukemia and high hyperdiploidy (> 50 chromosomes). J Clin Oncol 18:1876-1887.

Heisterkamp N, Stam K, Groffen J, de Klein A, Grosveld G. 1985. Structural organization of the bcr gene and its role in the Ph' translocation. Nature 315:758-761.

Herzog EL, Chai L, Krause DS. 2003. Plasticity of marrow-derived stem cells. Blood 102:3483-3493.

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

Hoang T. 2004. The origin of hematopoietic cell type diversity. Oncogene 23:7188-7198.

Holleman A, Cheok MH, den Boer ML, Yang W, Veerman AJ, Kazemier KM, Pei D, Cheng C, Pui CH, Relling MV, Janka-Schaub GE, Pieters R, Evans WE. 2004. Gene-expression patterns in drug-resistant acute lymphoblastic leukemia cells and response to treatment. N Engl J Med 351:533-542.

Hsieh JJ, Ernst P, Erdjument-Bromage H, Tempst P, Korsmeyer SJ. 2003. Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. Mol Cell Biol 23:186-194.

Hunger SP, Galili N, Carroll AJ, Crist WM, Link MP, Cleary ML. 1991. The t(1;19)(q23;p13) results in consistent fusion of E2A and PBX1 coding sequences in acute lymphoblastic leukemias. Blood 77:687-693.

Hunger SP. 1996. Chromosomal translocations involving the E2A gene in acute lymphoblastic leukemia: clinical features and molecular pathogenesis. Blood 87:1211-1224.

Imai Y, Kurokawa M, Tanaka K, Friedman AD, Ogawa S, Mitani K, Yazaki Y, Hirai H. 1998. TLE, the human homolog of groucho, interacts with AML1 and acts as a repressor of AML1-induced transactivation. Biochem Biophys Res Commun 252:582-589.

Jandle JH. 1996. Blood: Textbook of Hematology. In: Jandl JH, editor. Second ed: Lippincott Williams & Wilkins.

Johansson B, Mertens F, Mitelman F. 2004. Clinical and biological importance of cytogenetic abnormalities in childhood and adult acute lymphoblastic leukemia. Ann Med 36:492-503.

Kamps MP, Murre C, Sun XH, Baltimore D. 1990. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. Cell 60:547-555.

Kawamoto H, Minato N. 2004. Myeloid cells. Int J Biochem Cell Biol 36:1374-1379.

Kee BL, Murre C. 1998. Induction of early B cell factor (EBF) and multiple B lineage genes by the basic helix-loop-helix transcription factor E12. J Exp Med 188:699-713.

Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, Smith PD, Jiang Y, Gooden GC, Trent JM, Meltzer PS. 1998. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res 58:5009-5013.

Khan J, Wei JS, Ringnér M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS. 2001. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med 7:673-679.

Knoepfler PS, Calvo KR, Chen H, Antonarakis SE, Kamps MP. 1997. Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1. Proc Natl Acad Sci U S A 94:14553-14558.

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

Kohlmann A, Schoch C, Schnittger S, Dugas M, Hiddemann W, Kern W, Haferlach T. 2003. Molecular characterization of acute leukemias by use of microarray technology. Genes Chromosomes Cancer 37:396-405.

Kohlmann A, Schoch C, Dugas M, Schnittger S, Hiddemann W, Kern W, Haferlach T. 2005. New insights into MLL gene rearranged acute leukemias using gene expression profiling: shared pathways, lineage commitment, and partner genes. Leukemia 19:953-964.

Kondo M, Weissman IL, Akashi K. 1997. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell 91:661-672.

Lander ES. 1999. Array of hope. Nat Genet 21:3-4.

Le Coniat M, Della Valle V, Marynen P, Berger R. 1997. A new breakpoint, telomeric to TEL/ETV6, on the short arm of chromosome 12 in T cell acute lymphoblastic leukemia. Leukemia 11:1360-1363.

LeBrun DP, Cleary ML. 1994. Fusion with E2A alters the transcriptional properties of the homeodomain protein PBX1 in t(1;19) leukemias. Oncogene 9:1641-1647.

Lennon GG, Lehrach H. 1991. Hybridization analyses of arrayed cDNA libraries. Trends Genet 7:314-317.

Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. 1999. High density synthetic oligonucleotide arrays. Nat Genet 21:20-24.

Look AT. 1997. Oncogenic transcription factors in the human acute leukemias. Science 278:1059-1064.

Lu Q, Kamps MP. 1997. Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specifities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain–demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo. Oncogene 14:75-83.

Lutterbach B, Hiebert SW. 2000. Role of the transcription factor AML-1 in acute leukemia and hematopoietic differentiation. Gene 245:223-235.

Lutterbach B, Westendorf JJ, Linggi B, Isaac S, Seto E, Hiebert SW. 2000. A mechanism of repression by acute myeloid leukemia-1, the target of multiple chromosomal translocations in acute leukemia. J Biol Chem 275:651-656.

MacKinney AA, Jr., Clark SS, Borcherding W, Fizzotti M, Hong R. 1993. Simultaneous demonstration of the Philadelphia chromosome in T, B, and myeloid cells. Am J Hematol 44:48-52.

Macleod K. 2000. Tumor suppressor genes. Curr Opin Genet Dev 10:81-93.

Magli MC, Barba P, Celetti A, De Vita G, Cillo C, Boncinelli E. 1991. Coordinate regulation of HOX genes in human hematopoietic cells. Proc Natl Acad Sci U S A 88:6348-6352.

Maier E, Meier-Ewert S, Ahmadi AR, Curtis J, Lehrach H. 1994. Application of robotic technology to automated sequence fingerprint analysis by oligonucleotide hybridisation. J Biotechnol 35:191-203.

Melo JV, Deininger MW. 2004. Biology of chronic myelogenous leukemia–signaling pathways of initiation and transformation. Hematol Oncol Clin North Am 18:545-568, vii-viii.

Metcalf D. 1993. Hematopoietic regulators: redundancy or subtlety? Blood 82:3515-3523.

Mitelman F, Johansson B, Mertens F. 2004. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. Nat Genet 36:331-334.

Mitelman F, Johansson B, Mertens F. 2005. Mitelman Database of Chromosome Aberrations in Cancer. http://cgap.nci.nih.gov/Chromosomes/Mitelman.

Moorman AV, Richards SM, Martineau M, Cheung KL, Robinson HM, Jalali GR, Broadfield ZJ, Harris RL, Taylor KE, Gibson BE, Hann IM, Hill FG, Kinsey SE,

Eden TO, Mitchell CD, Harrison CJ. 2003. Outcome heterogeneity in childhood high-hyperdiploid acute lymphoblastic leukemia. Blood 102:2756-2762.

Moos PJ, Raetz EA, Carlson MA, Szabo A, Smith FE, Willman C, Wei Q, Hunger SP, Carroll WL. 2002. Identification of gene expression profiles that segregate patients with childhood leukemia. Clin Cancer Res 8:3118-3130.

Morrison SJ, Wandycz AM, Hemmati HD, Wright DE, Weissman IL. 1997. Identification of a lineage of multipotent hematopoietic progenitors. Development 124:1929-1939.

Mrozek K, Heerema NA, Bloomfield CD. 2004. Cytogenetics in acute leukemia. Blood Rev 18:115-136.

Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, et al. 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell 58:537-544.

Murre C. 2005. Helix-loop-helix proteins and lymphocyte development. Nat Immunol 6:1079-1086.

Nakamura T, Mori T, Tada S, Krajewski W, Rozovskaia T, Wassell R, Dubois G, Mazo A, Croce CM, Canaani E. 2002. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol Cell 10:1119-1128.

Nilsson J, Fioretos T, Höglund M, Fontes M. 2004. Approximate geodesic distances reveal biologically relevant structures in microarray data. Bioinformatics 20:874-880.

Notarangelo LD, Giliani S, Mazza C, Mella P, Savoldi G, Rodriguez-Perez C, Mazzolari E, Fiorini M, Duse M, Plebani A, Ugazio AG, Vihinen M, Candotti F, Schumacher RF. 2000. Of genes and phenotypes: the immunological and molecular spectrum of combined immune deficiency. Defects of the gamma(c)-JAK3 signaling pathway as a model. Immunol Rev 178:39-48.

Nourse J, Mellentin JD, Galili N, Wilkinson J, Stanbridge E, Smith SD, Cleary ML. 1990. Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell 60:535-545.

Nowell PC, Hungerford DA. 1960. A minute chromosome in human granulocytic leukemia. Science:1497.

Ogawa M. 1993. Differentiation and proliferation of hematopoietic stem cells. Blood 81:2844-2853.

Okada H, Watanabe T, Niki M, Takano H, Chiba N, Yanai N, Tani K, Hibino H, Asano S, Mucenski ML, Ito Y, Noda T, Satake M. 1998. AML1(-/-) embryos do not express certain hematopoiesis-related gene transcripts including those of the PU.1 gene. Oncogene 17:2287-2293.

Ortega S, Malumbres M, Barbacid M. 2002. Cyclin D-dependent kinases, INK4 inhibitors and cancer. Biochim Biophys Acta 1602:73-87.

Passegue E, Jamieson CH, Ailles LE, Weissman IL. 2003. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? Proc Natl Acad Sci U S A 100 Suppl 1:11842-11849.

Paulsson K, Panagopoulos I, Knuutila S, Jee KJ, Garwicz S, Fioretos T, Mitelman F, Johansson B. 2003. Formation of trisomies and their parental origin in hyperdiploid childhood acute lymphoblastic leukemia. Blood 102:3010-3015.

Paulsson K, Horvat A, Fioretos T, Mitelman F, Johansson B. 2005a. Formation of der(19)t(1;19)(q23;p13) in acute lymphoblastic leukemia. Genes Chromosomes Cancer 42:144-148.

Paulsson K, Mörse H, Fioretos T, Behrendtz M, Strömbeck B, Johansson B. 2005b. Evidence for a single-step mechanism in the origin of hyperdiploid childhood acute lymphoblastic leukemia. Genes Chromosomes Cancer 44:113-122.

Poppe B, Vandesompele J, Schoch C, Lindvall C, Mrozek K, Bloomfield CD, Beverloo HB, Michaux L, Dastugue N, Herens C, Yigit N, De Paepe A, Hagemeijer A, Speleman F. 2004. Expression analyses identify MLL as a prominent target of 11q23 amplification and support an etiologic role for MLL gain of function in myeloid malignancies. Blood 103:229-235.

Pui CH, Raimondi SC, Dodge RK, Rivera GK, Fuchs LA, Abromowitch M, Look AT, Furman WL, Crist WM, Williams DL. 1989. Prognostic importance of structural chromosomal abnormalities in children with hyperdiploid (greater than 50 chromosomes) acute lymphoblastic leukemia. Blood 73:1963-1967.

Pui CH, Raimondi SC, Hancock ML, Rivera GK, Ribeiro RC, Mahmoud HH, Sandlund JT, Crist WM, Behm FG. 1994. Immunologic, cytogenetic, and clinical characterization of childhood acute lymphoblastic leukemia with the t(1;19) (q23; p13) or its derivative. J Clin Oncol 12:2601-2606.

Pui CH, Gaynon PS, Boyett JM, Chessells JM, Baruchel A, Kamps W, Silverman LB, Biondi A, Harms DO, Vilmer E, Schrappe M, Camitta B. 2002. Outcome of treatment in childhood acute lymphoblastic leukaemia with rearrangements of the 11q23 chromosomal region. Lancet 359:1909-1915.

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

Quackenbush J. 2001. Computational analysis of microarray data. Nat Rev Genet 2:418-427.

Quackenbush J. 2002. Microarray data normalization and transformation. Nat Genet 32 Suppl:496-501.

Rabbitts TH. 2001. Chromosomal translocation master genes, mouse models and experimental therapeutics. Oncogene 20:5763-5777.

Radmacher MD, McShane LM, Simon R. 2002. A paradigm for class prediction using gene expression profiles. J Comput Biol 9:505-511.

Raghavan M, Lillington DM, Skoulakis S, Debernardi S, Chaplin T, Foot NJ, Lister TA, Young BD. 2005. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. Cancer Res 65:375-378.

Raimondi SC, Pui CH, Hancock ML, Behm FG, Filatov L, Rivera GK. 1996. Heterogeneity of hyperdiploid (51-67) childhood acute lymphoblastic leukemia. Leukemia 10:213-224.

Ransohoff DF. 2004. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer 4:309-314.

Reya T, Morrison SJ, Clarke MF, Weissman IL. 2001. Stem cells, cancer, and cancer stem cells. Nature 414:105-111.

Ringnér M, Peterson C, Khan J. 2002. Analyzing array data using supervised methods. Pharmacogenomics 3:403-415.

Robinson HM, Broadfield ZJ, Cheung KL, Harewood L, Harris RL, Jalali GR, Martineau M, Moorman AV, Taylor KE, Richards S, Mitchell C, Harrison CJ. 2003. Amplification of AML1 in acute lymphoblastic leukemia is associated with a poor outcome. Leukemia 17:2249-2250.

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

Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K, Williams WK, Liu HC, Mahfouz R, Raimondi SC, Lenny N, Patel A, Downing JR. 2003. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. Blood 102:2951-2959.

Ross ME, Mahfouz R, Onciu M, Liu HC, Zhou X, Song G, Shurtleff SA, Pounds S, Cheng C, Ma J, Ribeiro RC, Rubnitz JE, Girtman K, Williams WK, Raimondi SC, Liang DC, Shih LY, Pui CH, Downing JR. 2004. Gene expression profiling of pediatric acute myelogenous leukemia. Blood 104:3679-3687.

Roumier C, Fenaux P, Lafage M, Imbert M, Eclache V, Preudhomme C. 2003. New mechanisms of AML1 gene alteration in hematological malignancies. Leukemia 17:9-16.

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

Rubnitz JE, Link MP, Shuster JJ, Carroll AJ, Hakami N, Frankel LS, Pullen DJ, Cleary ML. 1994. Frequency and prognostic significance of HRX rearrangements in infant acute lymphoblastic leukemia: a Pediatric Oncology Group study. Blood 84:570-573.

Rubnitz JE, Behm FG, Pui CH, Evans WE, Relling MV, Raimondi SC, Harrison PL, Sandlund JT, Ribeiro RC, Grosveld G, Downing JR. 1997. Genetic studies of childhood acute lymphoblastic leukemia with emphasis on p16, MLL, and ETV6 gene abnormalities: results of St Jude Total Therapy Study XII. Leukemia 11:1201-1206.

Rubnitz JE, Raimondi SC, Tong X, Srivastava DK, Razzouk BI, Shurtleff SA, Downing JR, Pui CH, Ribeiro RC, Behm FG. 2002. Favorable impact of the t(9;11) in childhood acute myeloid leukemia. J Clin Oncol 20:2302-2309.

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

Salesse S, Verfaillie CM. 2002. Mechanisms underlying abnormal trafficking and expansion of malignant progenitors in CML: BCR/ABL-induced defects in integrin function in CML. Oncogene 21:8605-8611.

Santarosa M, Ashworth A. 2004. Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. Biochim Biophys Acta 1654:105-122.

Sauvageau G, Lansdorp PM, Eaves CJ, Hogge DE, Dragowska WH, Reid DS, Largman C, Lawrence HJ, Humphries RK. 1994. Differential expression of homeobox genes in functionally distinct CD34+ subpopulations of human bone marrow cells. Proc Natl Acad Sci U S A 91:12223-12227.

Scandura JM, Boccuni P, Cammenga J, Nimer SD. 2002. Transcription factor fusions in acute leukemia: variations on a theme. Oncogene 21:3422-3444.

Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467-470.

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci U S A 93:10614-10619.

Schoch C, Kohlmann A, Schnittger S, Brors B, Dugas M, Mergenthaler S, Kern W, Hiddemann W, Eils R, Haferlach T. 2002. Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. Proc Natl Acad Sci U S A 99:10008-10013.

Secker-Walker LM, Berger R, Fenaux P, Lai JL, Nelken B, Garson M, Michael PM, Hagemeijer A, Harrison CJ, Kaneko Y, et al. 1992. Prognostic significance of the balanced t(1;19) and unbalanced der(19)t(1;19) translocations in acute lymphoblastic leukemia. Leukemia 6:363-369.

Seeger K, Adams HP, Buchwald D, Beyermann B, Kremens B, Niemeyer C, Ritter J, Schwabe D, Harms D, Schrappe M, Henze G. 1998. TEL-AML1 fusion transcript in relapsed childhood acute lymphoblastic leukemia. The Berlin-Frankfurt-Munster Study Group. Blood 91:1716-1722.

Shanmugam K, Green NC, Rambaldi I, Saragovi HU, Featherstone MS. 1999. PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins. Mol Cell Biol 19:7577-7588.

Shannon W, Culverhouse R, Duncan J. 2003. Analyzing microarray data using cluster analysis. Pharmacogenomics 4:41-52.

Shtivelman E, Lifshitz B, Gale RP, Canaani E. 1985. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature 315:550-554.

Shurtleff SA, Buijs A, Behm FG, Rubnitz JE, Raimondi SC, Hancock ML, Chan GC, Pui CH, Grosveld G, Downing JR. 1995. TEL/AML1 fusion resulting from a cryptic

t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. Leukemia 9:1985-1989.
Smyth GK, Speed T. 2003. Normalization of cDNA microarray data. Methods 31:265-273.
So CW, Lin M, Ayton PM, Chen EH, Cleary ML. 2003. Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias. Cancer Cell 4:99-110.
Song WJ, Sullivan MG, Legare RD, Hutchings S, Tan X, Kufrin D, Ratajczak J, Resende IC, Haworth C, Hock R, Loh M, Felix C, Roy DC, Busque L, Kurnit D, Willman C, Gewirtz AM, Speck NA, Bushweller JH, Li FP, Gardiner K, Poncz M, Maris JM, Gilliland DG. 1999. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. Nat Genet 23:166-175.
Southern E, Mir K, Shchepinov M. 1999. Molecular interactions on microarrays. Nat Genet 21:5-9.
Speck NA, Gilliland DG. 2002. Core-binding factors in haematopoiesis and leukaemia. Nat Rev Cancer 2:502-513.
Staber PB, Linkesch W, Zauner D, Beham-Schmid C, Guelly C, Schauer S, Sill H, Hoefler G. 2004. Common alterations in gene expression and increased proliferation in recurrent acute myeloid leukemia. Oncogene 23:894-904.
Stirewalt DL, Radich JP. 2003. The role of FLT3 in haematopoietic malignancies. Nat Rev Cancer 3:650-665.
Stone RM, DeAngelo DJ, Klimek V, Galinsky I, Estey E, Nimer SD, Grandin W, Lebwohl D, Wang Y, Cohen P, Fox EA, Neuberg D, Clark J, Gilliland DG, Griffin JD. 2005. Patients with acute myeloid leukemia and an activating mutation in FLT3 respond to a small-molecule FLT3 tyrosine kinase inhibitor, PKC412. Blood 105:54-60.
Sutcliffe MJ, Shuster JJ, Sather HN, Camitta BM, Pullen J, Schultz KR, Borowitz MJ, Gaynon PS, Carroll AJ, Heerema NA. 2005. High concordance from independent studies by the Children's Cancer Group (CCG) and Pediatric Oncology Group (POG) associating favorable prognosis with combined trisomies 4, 10, and 17 in children with NCI Standard-Risk B-precursor Acute Lymphoblastic Leukemia: a Children's Oncology Group (COG) initiative. Leukemia 19:734-740.
Sykes DB, Kamps MP. 2004. E2a/Pbx1 induces the rapid proliferation of stem cell factor-dependent murine pro-T cells that cause acute T-lymphoid or myeloid leukemias in mice. Mol Cell Biol 24:1256-1269.
Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310:644-648.
Tsutsumi S, Taketani T, Nishimura K, Ge X, Taki T, Sugita K, Ishii E, Hanada R, Ohki M, Aburatani H, Hayashi Y. 2003. Two distinct gene expression signatures in pediatric acute lymphoblastic leukemia with MLL rearrangements. Cancer Res 63:4882-4887.
Uchida H, Zhang J, Nimer SD. 1997. AML1A and AML1B can transactivate the human IL-3 promoter. J Immunol 158:2251-2258.
Uckun FM, Sensel MG, Sather HN, Gaynon PS, Arthur DC, Lange BJ, Steinherz PG, Kraft P, Hutchinson R, Nachman JB, Reaman GH, Heerema NA. 1998. Clinical significance of translocation t(1;19) in childhood acute lymphoblastic leukemia in the context of contemporary therapies: a report from the Children's Cancer Group. J Clin Oncol 16:527-535.
Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, Beverloo HB, Moorhouse MJ, van der Spek PJ, Lowenberg B, Delwel R. 2004. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 350:1617-1628.
van Delft FW, Bellotti T, Luo Z, Jones LK, Patel N, Yiannikouris O, Hill AS, Hubank M, Kempski H, Fletcher D, Chaplin T, Foot N, Young BD, Hann IM, Gammerman

A, Saha V. 2005. Prospective gene expression analysis accurately subtypes acute leukaemia in children and establishes a commonality between hyperdiploidy and t(12;21) in acute lymphoblastic leukaemia. Br J Haematol 130:26-35.

van Dongen JJ, Seriu T, Panzer-Grumayer ER, Biondi A, Pongers-Willemse MJ, Corral L, Stolz F, Schrappe M, Masera G, Kamps WA, Gadner H, van Wering ER, Ludwig WD, Basso G, de Bruijn MA, Cazzaniga G, Hettinger K, van der Does-van den Berg A, Hop WC, Riehm H, Bartram CR. 1998. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. Lancet 352:1731-1738.

Virtaneva K, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, Bloomfield CD, de La Chapelle A, Krahe R. 2001. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc Natl Acad Sci U S A 98:1124-1129.

Wen R, Wang D, McKay C, Bunting KD, Marine JC, Vanin EF, Zambetti GP, Korsmeyer SJ, Ihle JN, Cleveland JL. 2001. Jak3 selectively regulates Bax and Bcl-2 expression to promote T-cell development. Mol Cell Biol 21:678-689.

Willenbrock H, Juncker AS, Schmiegelow K, Knudsen S, Ryder LP. 2004. Prediction of immunophenotype, treatment response, and relapse in childhood acute lymphoblastic leukemia using DNA microarrays. Leukemia 18:1270-1277.

Wood LJ, Mukherjee M, Dolde CE, Xu Y, Maher JF, Bunton TE, Williams JB, Resar LM. 2000. HMG-I/Y, a new c-Myc target gene and potential oncogene. Mol Cell Biol 20:5490-5502.

Yagi T, Morimoto A, Eguchi M, Hibi S, Sako M, Ishii E, Mizutani S, Imashuku S, Ohki M, Ichikawa H. 2003. Identification of a gene expression signature associated with pediatric AML prognosis. Blood 102:1849-1856.

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

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, 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 BD, Hanson RD, Hess JL, Horning SE, Korsmeyer SJ. 1998. MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proc Natl Acad Sci U S A 95:10632-10636.

Yuki Y, Imoto I, Imaizumi M, Hibi S, Kaneko Y, Amagasa T, Inazawa J. 2004. Identification of a novel fusion gene in a pre-B acute lymphoblastic leukemia with t(1;19)(q23;p13). Cancer Sci 95:503-507.

Zhou M, Gu L, Yeager AM, Findley HW. 1997. Incidence and clinical significance of CDKN2/MTS1/P16ink4A and MTS2/P15ink4B gene deletions in childhood acute lymphoblastic leukemia. Pediatr Hematol Oncol 14:141-150.

Zhu J, Emerson SG. 2002. Hematopoietic cytokines, transcription factors and lineage commitment. Oncogene 21:3295-3313.