Genome-Wide Analysis of Severe Congenital Neutropenia and Leukemia

Implications for leukemogenesis

Renée Beekman

Genome-Wide Analysis of Severe Congenital Neutropenia and Leukemia; Implications for Leukemogenesis

ISBN: 978-94-6169-311-2

Cover design: Renée Beekman and Optima Grafische Communicatie
Layout and Printing: Optima Grafische Communicatie

The work presented in this thesis was performed at the Department of Hematology at the Erasmus Medical Center in Rotterdam and was financially supported by the Center for Translational Molecular Medicine (CTMM), the Dutch Cancer Society "Koningin Wilhelmina Fonds" (KWF Kankerbestrijding), the Netherlands Genomics Initiative (NGI) and the ERA-Net for Research Programmes on Rare Diseases (E-RARE).

Printing of this thesis was financially supported by de MPN Stichting - Fonds Wetenschappelijk Onderzoek, the Erasmus University Rotterdam, de J.E. Jurriaanse Stichting, Biozym TC B.V. and Integrated DNA Technologies.

Genome-Wide Analysis of Severe Congenital Neutropenia and Leukemia

Implications for leukemogenesis

Genoombrede analyse van ernstige congenitale neutropenie en leukemie

Implicaties voor leukemogenese

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.G. Schmidt

en volgens het besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
vrijdag 18 januari 2013 om 13:30 uur

door

Renée Beekman

geboren te Vlaardingen

PROMOTIECOMMISSIE

Promotor:
Prof.dr. I.P. Touw

Overige leden:
Prof.dr. H.R. Delwel
Prof.dr. R. Fodde
Dr. J.H. Jansen

Wees niet bang
Je mag opnieuw beginnen
Vastberaden, doelgericht
of aarzelend op de tast
Houd je aan de regels
of volg je eigen zinnen
Laat die hand maar los
of pak er juist één vast

Wees niet bang
voor al te grote dromen
Ga als je het zeker weet
en als je aarzelt wacht
Hoe ijdel zijn de dingen
die je je hebt voorgenomen
Het mooiste overkomt je
het minste is bedacht

Wees niet bang
voor wat ze van je vinden
Wat weet je van de ander
als je jezelf niet kent
Verlies je oorsprong niet
door je snel te binden
Het leven lijkt afwisselend
maar zelfs de liefde went

Wees niet bang
je bent één van de velen
en tegelijk
is er maar één als jij
Dat betekent dat
je vaak zal moeten delen
en soms zal moeten zeggen:
laat me vrij!

Geschreven door: Freek de Jonge
Uit: "Wees niet bang. Mijn 101 mooiste liedjes"
Uitgeverij: Uitgeverij Augustus

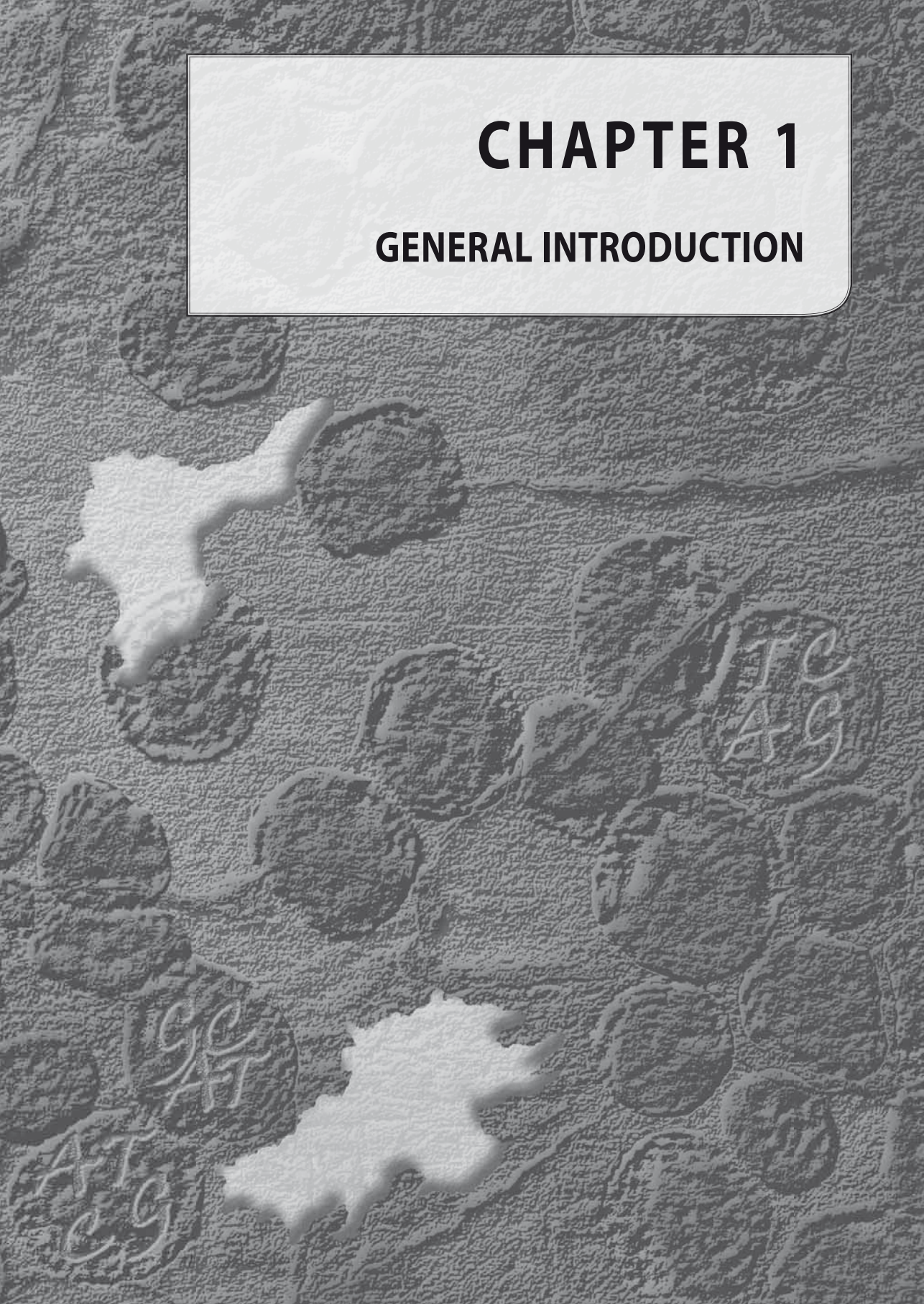
Voor jullie, lieve pap en mam
Jullie hebben mij geleerd niet bang te zijn voor grote dromen
En laten mij vrij mijn dromen te volgen

Voor ons drietjes, lieve Uri en Ona
Het mooiste wat me kon overkomen

TABLE OF CONTENTS

CHAPTER 1

GENERAL INTRODUCTION



1. HEMATOPOIESIS

The process of blood cell formation is known as hematopoiesis. During this process mature blood cells are formed in the bone marrow, followed by their release in the peripheral blood. Once in the periphery, mature blood cells exert their functions; erythrocytes play a crucial role in gas transport from the lungs to the peripheral organs and vice versa, platelets are essential for blood clotting and leukocytes, comprising granulocytes, monocytes/macrophages, lymphocytes and natural killer cells, are indispensable for the defense against micro-organisms.

The mammalian blood system is self renewing and the number of blood cells is controlled by a tight balance of production and degradation. Throughout the entire life, mature blood cells derive from a small population of pluripotent hematopoietic stem cells (HSCs) within the bone marrow. HSCs have the exceptional ability to and self-renew and differentiate into progenitors of all hematopoietic lineages[1]. In this way, HSCs provide an unlimited source of blood cells. Hematopoietic progenitors on their turn develop through a number of differentiation stages into mature blood cells (Figure 1). The lifespan of mature blood cells is limited, consequently a constant production is necessary to maintain adequate blood cell numbers[2].

Hematopoiesis is largely controlled by external signals, e.g., hematopoietic growth factors and cell-cell interactions, within the bone marrow niche[2, 3]. These are essential for the production of sufficient numbers of blood cells under physiological conditions and provide opportunities to increase production if necessary, e.g., during blood loss or infections, a process known as "stress" or "emergency" hematopoiesis[4, 5]. Furthermore, extracellular signals can direct differentiation of hematopoietic progenitors into specific lineages. Growth factors that play a major role in this process are thrombopoeitin (TPO), essential for platelet production, erythropoietin (EPO), regulating the production of erythrocytes, and colony stimulating factors (CSFs), involved in myeloid cell development[2, 6-8].

The majority of leukocytes in the peripheral blood are neutrophilic granulocytes or neutrophils, which are essential for the early, aspecific immune response against micro-organisms. The process of neutrophilic development is known as myelopoiesis. During this process, HSCs develop through a number of differentiation stages, i.e., myeloblasts, promyelocytes, neutrophilic myelocytes, neutrophilic metamyelocytes and neutrophilic band cells into mature neutrophils (Figure 2).

2. ACUTE MYELOID LEUKEMIA

Hematopoiesis is tightly regulated to ensure the production of sufficient numbers of blood cells. Equally important however, is the prevention of uncontrolled growth of hematopoietic progenitors. In acute myeloid leukemia (AML) this regulation is disturbed, myeloid progenitors have an increased proliferation capacity and have lost their ability

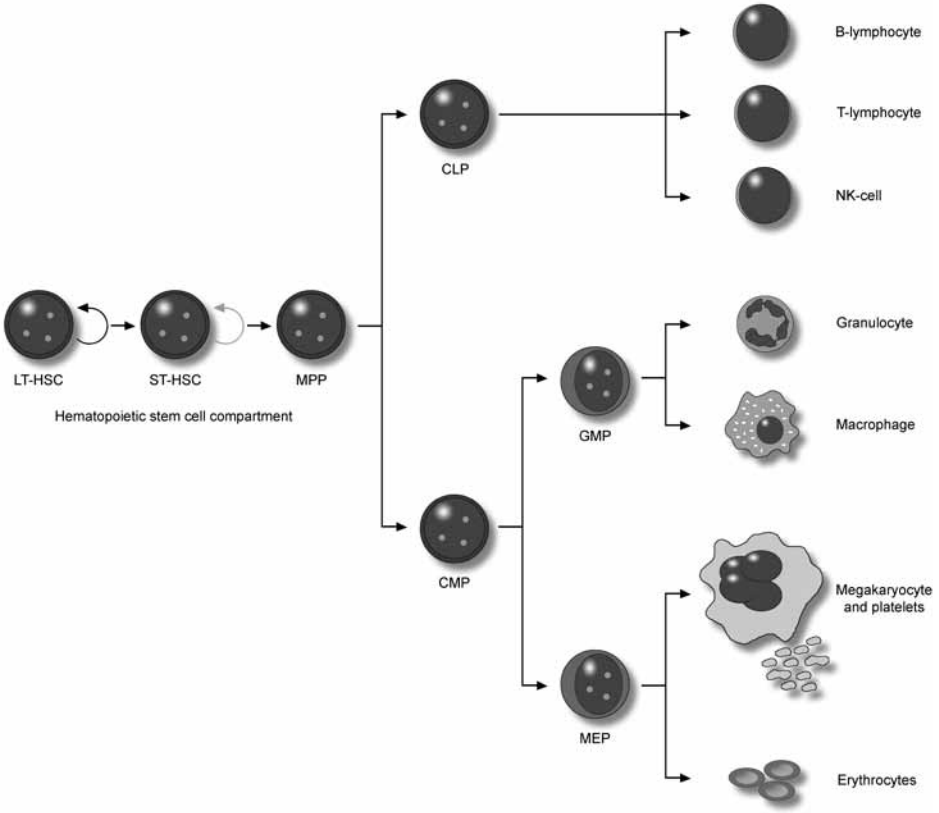


Figure 1. Hematopoiesis. Adapted from Reya et al[1] and Blank et al.[40] Within the hematopoietic stem cell compartment 3 subtypes of hematopoietic progenitors can be distinguished, i.e., long term and short term HSCs (LT-HSCs and ST-HSCs) and multipotent progenitors (MPPs). HSCs differentiate through a number of immature stages, e.g., common lymphoid progenitors (CLPs), common myeloid progenitors (CMPs), granulocyte/macrophage progenitors (GMPs) and megakaryocyte/erythrocyte progenitors (MEPs), into functional blood cells of all different hematopoietic lineages, depicted on the far right. For visualization purposes, not all immature differentiation stages are shown.

to differentiate. Clonal expansion and accumulation of leukemic blasts in the bone marrow and peripheral blood are the result.

2.1. Clinical aspects of AML

The incidence of AML is approximately 3.5 per 100.000 per year with a median age of 67 years at time of diagnosis[9]. AML patients usually present with a classical trias of symptoms, i.e., fatigue, hemorrhage and infections. These symptoms are due to suppression of normal hematopoiesis caused by leukemic blasts within the bone marrow and a subsequent lack of mature hematopoietic cells. Lack of functional blood cells leads to fatigue (lack of erythrocytes), hemorrhage (lack of platelets) and infections (lack of functional leukocytes).

Treatment of AML is divided into 2 phases, a remission-induction phase followed by a consolidation phase[10]. Remission-induction therapy aims at eradication of leukemic cells by treatment with intensive combinatorial chemotherapy. After complete remission is achieved, consolidation therapy aims to eliminate all remaining, undetectable leukemic blasts in order to prevent relapse. This phase consists of either an allogeneic or autologous stem cell transplantation or conventional chemotherapy. The type of consolidation therapy depends on the age of the patient and the presence of a suitable stem cell donor. In addition, a risk score, calculated based on cytogenetic abnormalities, mutations, white blood cell count and time to achieve complete remission, influences the choice of consolidation therapy.

2.2. Molecular mechanisms involved in AML

In the past decade, a high level of heterogeneity in AML became overt with the recognition that many different combinations of genetic and epigenetic alterations can play a role in its pathogenesis[11-15]. These alterations influence pathways involved in proliferation, differentiation, survival, self-renewal and DNA repair. Genetic alterations comprise large cytogenetic defects (translocations, amplifications and deletions) and small genetic aberrations (point mutations and small insertions and deletions). Genomic translocations may disrupt essential pathways in hematopoiesis by engaging expression of oncogenic fusion proteins or by transcriptional deregulation of critical genes that are placed under the influence of an alternative promoter. Moreover, mutations and small deletions and insertions involved in leukemogenesis may result in gain, loss or change of function and/or in transcriptional deregulation of essential genes and the proteins they encode. Examples of recurrent cytogenetic abnormalities and frequently mutated genes in AML as well as their prognostic significance are listed in Table 1 and 2.

In addition to genetic aberrations, epigenetic alterations, comprising DNA methylation and histone modification changes, may be involved in AML by altering expression levels of vital genes. Based on the type of deregulation, these vital genes can be separated into 3 categories, i.e., tumor suppressor genes, haplo-insufficient genes and proto-oncogenes. Reduced expression levels of haplo-insufficient genes and complete loss of tumor suppressor genes on one hand and over expression of proto-oncogenes on the other hand contribute to malignant transformation. Of note, expression levels of many genes are tightly regulated and both their over expression and down regulation may contribute to leukemogenesis. Although many genetic and epigenetic alterations have been described in AML, underlying aberrations remain unidentified in many patients.

2.3. Approaches to study leukemogenesis

Many different processes are deregulated in leukemic cells, conferring uncontrolled growth and maintenance of leukemic blasts. Most of these aberrations are likely down-

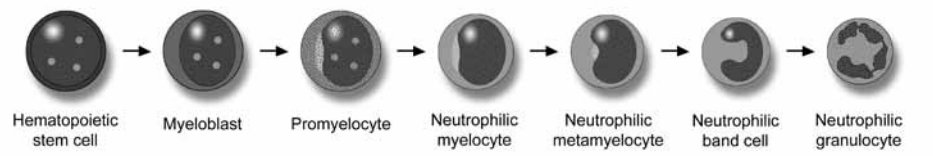


Figure 2. Myelopoiesis. HSCs develop through a number of differentiation stages into neutrophilic granulocytes. In a subset of SCN patients a block in differentiation at the promyelocytic stage of myelopoiesis is seen.

stream effects of a combination of few leukemia initiating events. To better understand these events, it is essential to gain insights into the molecular mechanisms involved in leukemogenesis as well as in the order and timing of events. This knowledge may consequently lead to the detection of new therapeutic targets and parameters predicting malignant transformation.

Leukemogenesis is an evolutionary process, characterized by gain of mutations in hematopoietic stem- and progenitor cells and outgrowth of clones with selective advantage. Consequently, we should distinguish mutations occurring early in disease development, while others arise late(r). Collection and analysis of sequential samples of patients with pre-leukemic conditions that eventually develop leukemia provide the ideal opportunity to investigate leukemogenesis. These samples may comprise bone marrow samples of patients with an increased risk to develop leukemia, like patients with bone marrow failure syndromes or patients treated with chemotherapy[16-18]. Additionally, blood samples of Guthrie cards, used to collect and store blood of newborn babies to test them for a number of hereditary diseases, may be a source of pre-leukemic samples. Usually however, pre-leukemic samples cannot be obtained; in these cases pre-leukemic subclones sorted from leukemia samples may be used.

Broadly, 2 approaches may be adopted to identify leukemia initiating events using pre-leukemic samples. First of all, these samples may be analysed to map the complete landscape of aberrations in subclones that either or not are sustained during leukemogenesis. This information is vital to understand why certain clones are maintained in this process while others disappear. From the other point of the spectrum, one may identify aberrations present in the leukemia followed by tracing these back in earlier samples. This latter approach comes with the inability to identify aberrations in pre-leukemic clones that are not sustained during leukemogenesis. It should be acknowledged that, after identifying potential leukemia initiating events using pre-leukemic and leukemic samples, in vivo models should be employed to study the effect of these aberrations in further detail.

Table 1. Recurrent cytogenetic abnormalities in adult AML.

Cytogenetic abnormality[#]	Frequency (%)[‡]	Genes involved and consequence of abnormality	Prognostic significance[‡]
Complex karyotype (≥ 3 cytogenetic abnormalities)	10.7	Unknown	Unfavorable
+8	9.1	Unknown	Intermediate
-7/7q-	8.4	Unknown	Unfavorable
t(15;17)(q22;q21)	7.6	Fusion of promyelocytic leukemia and retinoic acid receptor, alpha *(PML-RARA)*	Favorable
-5/5q-	7.2	Unknown	Unfavorable
t(8;21)(q22;q22)	5.5	Fusion of acute myeloid leukemia 1* and eight twenty-one (*AML1-ETO*)	Favorable
inv(16)(p13q22) / t(16;16)(p13;q22)	4.7	Fusion of core-binding factor, beta subunit and myosin, heavy chain 11 (*CBFβ-MYH11*)	Favorable
- Y	4.1	Unknown	Intermediate
t/inv(11q23)	3.3	Fusions of myeloid/lymphoid or mixed-lineage leukemia (*MLL*) and different partners	Intermediate / Unfavorable
abn(12p)	2.5	Unknown	Intermediate
+21	2.2	Unknown	Intermediate
abn(17p)	2.2	Unknown	Intermediate / Unfavorable
del(9q)	2.1	Unknown	Intermediate
inv(3)(q21q26) / t(3;3)(q21;q26)	2.0	Overexpression of ecotropic viral integration site 1 (*EVI1*)	Unfavorable
del(11q)	0.9	Unknown	Intermediate
t(9;22)(q34;q11)	0.8	Fusion of breakpoint cluster region and c-abl oncogene 1 (*BCR-ABL*)	Intermediate / Unfavorable
t(6;9)(p23;q34)	0.7	Fusion of DEK oncogene and nucleoporin 214 (*DEK-NUP214*)	Unfavorable

[#]Cytogenetic abnormalities are ordered based on frequency, [‡]based on Mrozek et al.[41], t = translocation, - = loss, + = gain, inv = inversion, abn = abnormality, del = deletion, *a.k.a. runt-related transcription factor 1 (RUNX1).

In the studies described in this thesis, 2 different models were used to identify new molecular events in leukemia and leukemogenesis. These were the human pre-leukemic severe congenital neutropenia model and the murine retroviral integration mutagenesis model. Both models are further introduced in the following sections.

3. SEVERE CONGENITAL NEUTROPENIA

Severe congenital neutropenia (SCN) is a hereditary bone marrow failure syndrome, characterized by a lack of mature neutrophils, severe recurrent infections and a high tendency to develop leukemia. To closely monitor alterations predicting for leukemic transformation bone marrow samples of SCN patients are regularly obtained. These are of great value to study leukemogenesis.

Table 2. Recurrent genetic abnormalities in adult AML.

Gene symbols	Gene names	Frequency‡	Prognostic significance‡
ASXL1	Additional sex combs like 1	10.8	Unfavorable#
CEBPA	CCAAT/enhancer binding protein alpha	4-14	Favorable
DNMT3A	DNA (cytosine-5)-methyltransferase 3 alpha	22	Unfavorable
FLT3-ITD	Fms-related tyrosine kinase 3, internal tandem duplication	20-33	Unfavorable
FLT3-TKD	Fms-related tyrosine kinase 3, tyrosine kinase domain	5-10	Controversial
IDH1	NADP-dependent isocitrate dehydrogenase 1	6.6-9.6	Unfavorable
IDH2	NADP-dependent isocitrate dehydrogenase 2	3.0-8.7	Unfavorable
JAK2	Janus kinase 2	1.5	-
KIT	V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	2-8	Unfavorable in CBF AMLs
KRAS	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	5	-
MLL-PTD	Myeloid/lymphoid or mixed-lineage leukemia, partial tandem duplication	5-11 (CN-AML)	Unfavorable#
NPM1	Nucleophosmin	25-35	Favorable in absence of *FLT3-ITD*
NRAS	Neuroblastoma RAS viral (v-ras) oncogene homolog	10-15	-
PTPN11	Protein tyrosine phosphatase, non-receptor type 11*	2.5-5.1	Unfavorable in absence of NPM1 mutations
RUNX1	Runt-related transcription factor 1	5-13	Unfavorable#
SF3B1	Splicing factor 3b, subunit 1	2.6	-
TET2	Tet oncogene family member 2	23 (CN-AML)	Unfavorable#
TP53	Tumor protein p53	<10	Unfavorable
WT1	Wilms tumor 1	10	Unfavorable#

‡Based on selected references[12-15, 39], *also known as SH2 domain-containing protein tyrosine phosphatase 2 (SHP2), #prognostic significance still under further investigation. CN-AML = cytogenetically normal AML, CBF AMLs = core binding factor AMLs, i.e., with inv(16)/t(16;16) or t(8;21).

3.1. Genetic alterations underlying SCN

SCN patients usually present with recurrent and sometimes even life threatening infections during early childhood. The increased susceptibility to infections in these patients is the result of a developmental defect in myelopoiesis leading to a severe reduction in neutrophilic granulocytes. Mutations in the gene encoding for neutrophil elastase (*ELANE)*, observed in ~50% of these patients, are the most frequent mutations underlying SCN[19, 20]. These mutations are thought to cause misfolding of neutrophilic elastase, a protein initially expressed in the promyelocytic stage of neutrophil development (Figure 2). Its misfolding is suggested to lead to an unfolded protein response (UPR), a stress response within the endoplasmatic reticulum, followed by a differentiation arrest and apoptosis[21]. In some, but not all SCN patients, a characteristic promyelocytic arrest can be observed within the bone marrow. A small minority of patients negative for *ELANE* mutations carry

mutations in other genes, i.e., *CSF3R*, *HAX1, GFI1, WAS* and *G6PC3*; within a large number of SCN patients however, the underlying mutation remains unidentified[19].

The molecular mechanisms underlying leukemogenesis in SCN are introduced in further detail in chapter 2.

4. RETROVIRAL INTEGRATION MUTAGENESIS

Besides using human samples, mouse models can be employed to study leukemic transformation. Many different models have been generated to study the role of specific genetic aberrations as well as their combinatorial effects in leukemogenesis[22, 23]. In addition, murine models are successfully used to discover new genes involved in malignant transformation. For this latter purpose, retroviral integration mutagenesis, in which oncogenic viruses induce murine tumor formation, has been widely used[24, 25]. The source of power of this approach is the capacity of proviral DNA to integrate within the murine genome where it can strongly deregulate transcription of flanking genes. Dependent on the affected genes, transcriptional deregulation may initiate malignant transformation. Clonal expansion of cells harboring specific patterns of proviral integrations will eventually lead to tumor development. Mapping of proviral integrations and their proximal deregulated genes within these malignancies allows identification of new genes involved in malignant transformation[25].

4.1. The Graffi 1.4 Murine Leukemia Virus model

To specifically identify genes involved in leukemogenesis, a retroviral integration mutagenesis model can be employed in which newborn mice are injected with the Graffi 1.4 Murine Leukemia Virus (Gr1.4 MuLV). This leads to infection of hematopoietic progenitors and consequent integration of proviral DNA into the genome of these cells. These integrations may alter expression of genes vital for hematopoiesis, a subsequent selective advantage and leukemic transformation. Ultimately, within 4-6 months after injection, all mice injected with the Gr1.4 MuLV develop leukemia[24, 26]. Mapping of proviral integrations and their flanking genes in these leukemias enabled identification of new genes having a role in leukemogenesis[24].

5. GENOME-WIDE APPROACHES TO STUDY HUMAN DISEASES

Genome-wide approaches are nowadays more and more commonly used to identify new aberrations underlying human diseases. Explanations for this trend can be found in the rapid expansion of possibilities and flexibility in experimental design, the quick progress in devices, experimental tools and reagents to increase the yield of data, and last but not least, the fast decline in experimental costs. In order to use these techniques and to be able to choose the right experimental approach it is essential to be aware of the advantages, disadvantages and caveats that these techniques withhold.

5.1. Array based technologies

A first major breakthrough is represented by array based approaches, which enabled the generation of genome-wide profiles of genetic variations (single nucleotide polymorphism arrays), gene expression levels (microarrays) and epigenetic changes (e.g., by chromatin immunoprecipitation (chIP)-on-chip)[27-29]. These techniques depend on thousands of unique probes spotted on a single chip/array that, based on their specific sequence, represent a large number of RNA transcripts and/or DNA fragments. These probes are used to measure the levels of a large number of RNA or DNA fragments isolated from a single sample. Hybridization signals of the different probes are a read out for the abundance of the specific RNA or DNA fragments in the analyzed sample. However, one has to keep in mind that RNA and DNA fragments that are not covered by the probes cannot be measured.

5.2. Next-generation sequencing

Owing the higher sensitivity and accuracy of datasets generated with next-generation sequencing based methods, e.g., whole genome sequencing (WGS), whole exome sequencing (WES), sequencing subsequent to custom capture, mRNA sequencing (mRNA-seq) and chIP-seq, array based techniques are more and more replaced by this technology[30]. These techniques yield millions of reads, containing RNA or DNA sequences of the sample of interest. Alignment of these reads to the reference genome allows estimation of the quantity of different fragments as well as the identification of genetic alterations. The applicability of the technique does not rely on the availability of probes of interest as is the case for array based technologies; this extends the flexibility and possibilities of next-sequencing based technologies compared to array based approaches.

To analyze genetic aberrations WGS, WES, sequencing subsequent to custom capture and mRNA-seq can be employed. The choice between these different methods depends on the research question, the required coverage and the material available. Independent of the experiment performed, all genetic alterations need to be confirmed with a different method, e.g., by Sanger sequencing, because false positive results may occur as a result of mistakes in sequencing or misalignment. WGS is the method of choice for an unbiased analysis of the complete genetic landscape, without pre-selecting for certain regions of interest. Pre-enrichment of specific genomic locations can be required however to reach a higher coverage necessary to map genetic alterations in minor cell fraction or to enable pooling of many samples in a single experiment. Pre-selection can be achieved by exome capture (total coding region) or custom capture (region of choice), before sequencing is conducted. In essence, mRNA-seq is also a pre-selection method, selecting for expressed transcripts. An advantage of this method is the possibility to simultaneously analyse genetic variants and expression levels. However, only genetic alterations in expressed transcripts can be detected, while unexpressed genetic

alterations, either located in the coding or in the non-coding region of the genome, remain unidentified.

5.3. Germ line variations, single nucleotide polymorphisms and mutations

Germ line variations are variations in the genetic code that determine differences between individuals. Nearly all germ line variations are single nucleotide polymorphisms (SNPs) that are present in "healthy" individuals, while only a few are mutations that underlie single gene disorders like SCN[31]. Frequencies of different SNPs vary within the population and although they are assumed not to cause single gene disorders, few are thought to contribute to multifactorial, polygenic diseases like diabetes mellitus[31].

SNPs can be of use for genome-wide analysis of cancer genomes as heterozygous variations can be useful to identify regions affected by loss-of-heterozygosity (LOH)[32]. However, SNPs are usually encountered a major obstacle in identifying acquired mutations. Especially when no germ line material could be obtained from the patient it is very difficult to distinguish SNPs from acquired mutations. SNP databases[33] are useful to make this discrimination, disadvantages arise however when using these databases. At first, these databases are not saturated; yet, sequencing increasing numbers of "normal" genomes results in identification of new SNPs each day. Secondly, these databases may contain variations which are falsely identified as SNPs. Thirdly, at the position of a known SNP an acquired mutation may have arisen, which may have contributed to disease development.

5.4. Data interpretation

An inevitable consequence of applying genome-wide approaches is that data-analysis is becoming more and more complex and specialized expertise, novel software and new algorithms are crucial to interpret the data. Of note, the possibilities of analyzing large datasets and the ability to answer specific research questions first of all relies on the quality of the dataset, which is on its turn dependent on the quality of the samples, the experimental setup and performance. So far, AML profiles generated using new genome-wide techniques are an invaluable source of information to better understand its complex biology and have led to the understanding that the heterogeneity of genetic and epigenetic alterations in AML is even bigger than initially thought[34-39].

6. SCOPE AND OUTLINE OF THE THESIS

Leukemogenesis is a complex process driven by many different genetic and epigenetic aberrations. Understanding its underlying molecular mechanisms provides an essential basis for identifying new diagnostic tools, prognostic markers and therapeutic targets. The work presented in the thesis focuses on the identification of new alterations in AML by studying 2 different leukemogenesis models; SCN and its derived AMLs in humans

and leukemias in mice induced by retroviral integration mutagenesis. Genetic and epigenetic alterations in these leukemias were mapped by genome-wide approaches. Simultaneously the functional role and prognostic value of some of these aberrations were further investigated.

In **chapter 2,** an outline is given on the molecular mechanisms involved in leukemogenesis in SCN. More specifically, a detailed overview is given on the clinical use of granulocyte-colony stimulating factor (G-CSF), alterations in the G-CSF receptor (colony stimulating factor 3 receptor, CSF3R) and their role in SCN and its progression towards AML. In **chapter 3**, whole exome sequencing was employed to identify (new) genetic alterations in a SCN derived AML, further referred to as the SCN index patient. Additionally, bone marrow samples of the same patient obtained 9 and 15 years before the AML became overt were investigated for the presence of genetic alterations occurring "early" during leukemogenesis. These samples provided the unique opportunity to identify new "early" genetic aberrations and alterations occurring "late" during malignant transformation. Interestingly, in the SCN index patient a new extracellular *CSF3R* mutation was identified as a "late" genetic alteration. In **chapter 4**, we further studied the prevalence and functional characteristics of this newly identified mutation. Furthermore, the prevalence of all "early" and "late" genetic alterations identified in the SCN index patient as well as their close paralogues and interaction partners was investigated in a larger number of SCN and SCN derived leukemia samples using a custom capture approach followed by deep-sequencing. The results of this work are presented in **chapter 5**.

In **chapter 6**, we aimed to identify haplo-insufficient genes contributing to leukemogenesis using murine leukemias obtained from a retroviral integration mutagenesis screen. In this chapter, a promoter array based approach designed to map DNA methylated viral integration sites is described and genes flanking these DNA methylated viral integration sites in leukemias were determined. Next, the transcriptional down regulation of these genes was investigated in these leukemias to identify haplo-insufficient genes. Finally, the prognostic value of 1 identified potential haplo-insufficient gene, i.e., *PTP4A3*, was investigated in a large panel of human AML samples. In **chapter 7**, a computational algorithm is presented that was specifically developed to localize retroviral integrations using an array based approach.

Finally, the results presented in this thesis are summarized and discussed in **chapter 8**.

REFERENCES

1. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature 2001;414(6859):105-11.
2. Lotem J, Sachs L. Cytokine control of developmental programs in normal hematopoiesis and leukemia. Oncogene 2002;21(21):3284-94.

3. Arai F, Hirao A, Suda T. Regulation of hematopoiesis and its interaction with stem cell niches. Int J Hematol 2005;82(5):371-6.
4. Bozzini CE, Alippi RM, Barcelo AC, et al. The biology of stress erythropoiesis and erythropoietin production. Ann N Y Acad Sci 1994;718:83-92; discussion -3.
5. Panopoulos AD, Watowich SS. Granulocyte colony-stimulating factor: molecular mechanisms of action during steady state and 'emergency' hematopoiesis. Cytokine 2008;42(3):277-88.
6. Clark SC, Kamen R. The human hematopoietic colony-stimulating factors. Science 1987;236(4806):1229-37.
7. Kaushansky K. Thrombopoietin: the primary regulator of platelet production. Blood 1995;86(2):419-31.
8. Graber SE, Krantz SB. Erythropoietin and the control of red cell production. Annu Rev Med 1978;29:51-66.
9. SEER Cancer Statistics Review, 1975-2008, National Cancer Institute. 2011. (Accessed at http://seer.cancer.gov/csr/1975_2008/.)
10. Lowenberg B, Downing JR, Burnett A. Acute myeloid leukemia. N Engl J Med 1999;341(14):1051-62.
11. Frohling S, Scholl C, Gilliland DG, Levine RL. Genetics of myeloid malignancies: pathogenetic and clinical implications. J Clin Oncol 2005;23(26):6285-95.
12. Marcucci G, Haferlach T, Dohner H. Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications. J Clin Oncol 2011;29(5):475-86.
13. Dohner K, Dohner H. Molecular characterization of acute myeloid leukemia. Haematologica 2008;93(7):976-82.
14. Renneville A, Roumier C, Biggio V, et al. Cooperating gene mutations in acute myeloid leukemia: a review of the literature. Leukemia 2008;22(5):915-31.
15. Takahashi S. Current findings for recurring mutations in acute myeloid leukemia. Journal of hematology & oncology 2011;4(1):36.
16. Alter BP, Giri N, Savage SA, et al. Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. Br J Haematol 2010;150(2):179-88.
17. Leone G, Pagano L, Ben-Yehuda D, Voso MT. Therapy-related leukemia and myelodysplasia: susceptibility and incidence. Haematologica 2007;92(10):1389-98.
18. Rosenberg PS, Zeidler C, Bolyard AA, et al. Stable long-term risk of leukaemia in patients with severe congenital neutropenia maintained on G-CSF therapy. Br J Haematol 2010;150(2):196-9.
19. Dale DC, Link DC. The many causes of severe congenital neutropenia. N Engl J Med 2009;360(1):3-5.
20. Dale DC, Person RE, Bolyard AA, et al. Mutations in the gene encoding neutrophil elastase in congenital and cyclic neutropenia. Blood 2000;96(7):2317-22.
21. Grenda DS, Murakami M, Ghatak J, et al. Mutations of the ELA2 gene found in patients with severe congenital neutropenia induce the unfolded protein response and cellular apoptosis. Blood 2007;110(13):4179-87.
22. Fortier JM, Graubert TA. Murine models of human acute myeloid leukemia. Cancer Treat Res 2010;145:183-96.
23. McCormack E, Bruserud O, Gjertsen BT. Review: genetic models of acute myeloid leukaemia. Oncogene 2008;27(27):3765-79.

24. Erkeland SJ, Valkhof M, Heijmans-Antonissen C, et al. Large-scale identification of disease genes involved in acute myeloid leukemia. J Virol 2004;78(4):1971-80.
25. Kool J, Uren AG, Martins CP, et al. Insertional mutagenesis in mice deficient for p15Ink4b, p16Ink4a, p21Cip1, and p27Kip1 reveals cancer gene interactions and correlations with tumor phenotypes. Cancer Res 2010;70(2):520-31.
26. Voisin V, Barat C, Hoang T, Rassart E. Novel insights into the pathogenesis of the Graffi murine leukemia retrovirus. J Virol 2006;80(8):4026-37.
27. Forozan F, Karhu R, Kononen J, Kallioniemi A, Kallioniemi OP. Genome screening by comparative genomic hybridization. Trends Genet 1997;13(10):405-9.
28. Kirmizis A, Farnham PJ. Genomic approaches that aid in the identification of transcription factor target genes. Exp Biol Med (Maywood) 2004;229(8):705-21.
29. Watson A, Mazumder A, Stewart M, Balasubramanian S. Technology for microarray analysis of gene expression. Curr Opin Biotechnol 1998;9(6):609-14.
30. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet 2008;24(3):133-41.
31. Nakamura Y. DNA variations in human and medical genetics: 25 years of my experience. Journal of human genetics 2009;54(1):1-8.
32. Mei R, Galipeau PC, Prass C, et al. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. Genome research 2000;10(8):1126-37.
33. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic acids research 2001;29(1):308-11.
34. Figueroa ME, Lugthart S, Li Y, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell 2010;17(1):13-27.
35. Ley TJ, Ding L, Walter MJ, et al. DNMT3A mutations in acute myeloid leukemia. N Engl J Med 2010;363(25):2424-33.
36. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 2008;456(7218):66-72.
37. Mardis ER, Ding L, Dooling DJ, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med 2009;361(11):1058-66.
38. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.
39. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 2011.
40. Blank U, Karlsson G, Karlsson S. Signaling pathways governing stem-cell fate. Blood 2008;111(2):492-503.
41. Mrozek K, Heerema NA, Bloomfield CD. Cytogenetics in acute leukemia. Blood Rev 2004;18(2):115-36.

CHAPTER 2

G-CSF AND ITS RECEPTOR IN MYELOID MALIGNANCY

Renée Beekman[1], Ivo P. Touw[1]

[1]Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

Blood. 2010 Jun 24;115(25):5131-6.

ABSTRACT

Granulocyte colony-stimulating factor (G-CSF) is now used in the clinic for more than two decades to treat congenital and acquired neutropenias and to reduce febrile neutropenia before or during courses of intensive cytoreductive therapy. In addition, healthy stem cell donors receive short term treatment with G-CSF for mobilization of hematopoietic stem cells. G-CSF has also been applied in priming strategies designed to enhance the sensitivity of leukemia stem cells to cytotoxic agents, in protocols aimed to induce their differentiation and accompanying growth arrest and cell death, and in severe aplastic anemia and myelodysplastic syndrome to alleviate anemia. The potential adverse effects of G-CSF administration, particularly the risk of malignant transformation, have fueled ongoing debates, some of which can only be settled in follow-up studies extending over several decades. This specifically applies to children with severe congenital neutropenia who receive life-long treatment with G-CSF and in which the high susceptibility to develop MDS and AML has now become a major clinical concern. Here, we will highlight some of the controversies and challenges regarding the clinical application of G-CSF and discuss a possible role of G-CSF in malignant transformation, particularly in neutropenia patients harboring mutations in the gene encoding the G-CSF receptor.

G-CSF AND ITS RECEPTOR

The growth factor G-CSF, now referred to as colony-stimulating factor 3 (CSF3), is the major regulator of neutrophil production under basal conditions of hematopoiesis, as is evident from the fact that CSF3 or CSF3 receptor-deficient mice are severely neutropenic[1, 2]. CSF3 is also essential for "emergency" granulopoiesis in response to bacterial infections and enhances multiple neutrophil functions[3]. CSF3 exerts its role by inducing proliferation and survival of myeloid progenitor cells, followed by a cell cycle arrest and neutrophilic differentiation[4]. The receptor for CSF3 (CSF3R) belongs to the cytokine receptor type I superfamily, which engage the canonical Janus kinase (Jak)/signal transducer and activator of transcription (STAT), Ras/Raf/MAPkinase and PKB/Akt pathways. When CSF3R mutants were expressed in differentiation competent factor dependent myeloid cell lines, the distal cytoplasmic region of the CSF3R of approximately 100 amino acids was crucial for CSF3-induced neutrophilic differentiation of these cells[5]. While originally being considered as "differentiation domain", later studies demonstrated that this C-terminal region exerts a negative role in STAT5 activation and proliferation signaling in vivo[6, 7]. Negative regulators of CSF3 signaling linked to the distal C-terminus of CSF3R include the protein tyrosine phosphatases SHP-1 and the suppressor of cytokine signaling (SOCS) protein SOCS3. The SOCS protein family is characterized by a so-called SOCS-box, a domain involved in the recruitment of ubiquitin (E3) ligase activity. The negative action of SOCS3 and more specifically of its SOCS-box on CSF3 signaling has

been demonstrated in conditional knockout models[8, 9]. A mechanism for receptor down-regulation has been proposed in which SOCS3 drives ubiquitination of a conserved juxtamembrane lysine residue that is important for lysosomal routing of the CSF3R[4, 10]. A current view is that balanced activation and subsequent attenuation of CSF3R signaling pathways, strongly depending on the kinetics of ligand-induced internalization and intracellular routing of the receptor, is important for neutrophil production, particularly during episodes of emergency granulopoiesis[4, 10].

CSF3 IN THE TREATMENT OF AML

CSF3 as a differentiation inducing agent

Soon after Bradley and Metcalf[11] and Pluznik and Sachs[12] discovered in the mid nineteen-sixties that bone marrow progenitor cells form colonies of differentiated myeloid cells under the influence of external growth factors, it became clear that these crude growth factor preparations also stimulated the proliferation and in part differentiation of leukemic progenitors in acute myeloid leukemia (AML) [13]. Once this was realized, ideas about the potential therapeutic significance of these findings rapidly evolved, which became testable in the mid nineteen-eighties when recombinant technology allowed the large scale production and purification of hematopoietic growth factors, including CSF3[14, 15]. The availability of clinical grade CSF3 and GM-CSF yielded expectations for patients with severe forms of chronic neutropenia, which have proved to be realistic from the outset. Concerning the application of CSF3 in the treatment of myeloid leukemia, one line of thinking was that AML blasts would differentiate upon CSF3 exposure and thereby undergo growth arrest and cell death[16, 17]. These studies provided important insights in the biology of myeloid leukemia and e.g., revealed the hierarchical nature of leukemic cell populations, consisting of leukemic stem cells, progenitors with colony forming potential in vitro (AML-CFU) and partly differentiated nonproliferative end cells[13]. Since then, CSF3 has occasionally been administered to selected AML patients with the objective to induce differentiation of the leukemic cells with variable results and whether the observed therapeutic effects could be ascribed to differentiation induction remained uncertain[18]. Currently the interest in further clinical development of this concept appears to have diminished, arguably because differentiation of the leukemia "bulk" without affecting the leukemic stem cells (LSCs) may not lead to durable therapeutic benefits. Nonetheless, the successful implementation of all-*trans* retinoic acid therapy in the treatment of acute promyelocytic leukemia, serving as the key paradigm that differentiation-inducing agents combined with chemotherapeutic regimens can result in long-lasting remissions[19], leaves the concept of differentiation induction by combinations of agents (including CSF3) open for future application in AML.

CSF3 as a chemo-sensitizer

The use of myeloid growth factors (CSF3, GM-CSF) to activate chemo-resistant dormant LSCs into chemo-sensitive cycling cells has been tested in multiple prospective randomized trials with variable outcome, possibly because of differences in patients groups and study design[20]. For instance, in one study beneficial effects on overall and disease free survival of standard risk AML patients was demonstrated when CSF3 was administered during induction therapy[21], whereas others did not observe favorable responses in a similar study involving elderly AML patients[22]. More recently, the theme of chemo-sensitization of LSCs by growth factor priming has been revisited from another viewpoint, i.e., based on the ability of CSF3 and the CXCR4 antagonist AMD3100 (plerixafor) to push LSCs out of their bone marrow niches that promote self renewal and may be protective against damage by genotoxic compounds. Again, results may be variable and dependent on the subtype of AML as is illustrated by two recent studies in mouse models, one representing acute promyelocytic leukemia (APL), the other AML with high MN1 expression. In the APL model, it was shown that AMD3100 induces the mobilization of leukemic cells from their bone marrow niches into the circulation, thereby increasing their sensitivity to Ara-C or daunorubicin[23]. In contrast, no chemo-sensitizing effects were seen in the AML/MN1 model[24]. Despite the similarities in mobilizing activities of CSF3 and AMD3100, recent studies have shown that CSF3 and AMD3100 synergize in the mobilization of normal stem cells, suggesting that their activities are not entirely overlapping[25]. These observations suggest that combinations of CSF3 and AMD3100 or other agents affecting cell migration and adhesion might be of therapeutic benefit[26].

CSF3 AND MALIGNANT TRANSFORMATION

Leukemia risk in individuals without hematological disorders

The concern that administration of hematopoietic growth factors might accelerate or even cause leukemia has recently received major attention in the context of CSF3 treatment of healthy individuals to mobilize hematopoietic stem cells (HSCs) into the periphery. The adverse effects of CSF3 administration to peripheral stem cell donors have been evaluated in two independent studies involving more than 5000 cases with a follow up of 4-5 years[27, 28]. Both studies reported no statistically significant differences in the incidence of malignancy relative to individuals not exposed to growth factor treatment. On the other hand, in a study from the Research on Adverse Drug Events and Reports (RADAR) project[29] AML was reported in 2 out of 200 HLA-identical siblings donors for AML patients, which significantly exceeds the incidence reported in the other studies. However, irrespective of exposure to CSF3, siblings of AML patients have a 2 to 5-fold increase in the annual incidence of leukemia, which most likely explains this discrepancy[30].

Another context in which a possible leukemogenic effect of CSF3 has been extensively investigated is in adjuvant breast cancer therapy. A retrospective study addressed

the occurrence of AML/MDS in six adjuvant breast cancer trials and showed increased rate of AML/MDS in patients treated with intensified doses of cyclophosphamide requiring CSF3 support[31]. A different study reported a doubling in the risk of AML/MDS in a population of women aged 65 years or older treated with adjuvant chemotherapy and growth factor support for stage I-III breast cancer[32]. Although the absolute risk of secondary leukemia was low in both studies, it was stated that the application of myeloid growth factors and possible leukemia risk should be factored into clinical decisions. However, the benefits of adjuvant chemotherapy in these patients outweighs the risk of secondary MDS or AML and given all of the unknown factors, it remains uncertain whether the weak associations found have a causal relationship to growth factor treatment[33]. Interestingly, a recent study in an as yet small series of patients suggests that the mutational status of *BRCA1* and *BRCA2* genes may contribute to leukemia risk in breast cancer patients, raising the possibility that a relation between CSF3 administration and secondary MDS/AML may specifically apply to these genetically defined subgroups[34]. Although a follow-up of 2000 stem cell donors for at least 10 years might be needed to detect a statistically significant increase in malignant transformation[29], the leukemia incidence associated with CSF3 administration is thus far negligible in stem cell donors and low but not yet conclusively determined in different genetic subtypes in breast cancer patients.

CSF3 treatment and malignant transformation in conditions with increased leukemia risk

CSF3, as a single growth factor or in combination with erythropoietin (EPO), has been used in MDS and severe aplastic anemia (SAA) and MDS but is not generally applied in the treatment of these conditions. In MDS, CSF3 was administered to investigate whether CSF3 would synergize with EPO to alleviate anemia and to reduce transfusion need[35]. A collaborative study that included patients from all risk categories suggested that leukemia risk in MDS patients treated with a combination of CSF3 and EPO was not different from patients not receiving growth factor treatment[36]. However, a complicating factor in this retrospective study is that the EPO+CSF3 treated groups were compared with untreated historical controls from a distinct cohort[36]. In a retrospective survey among 840 SAA patients registered by the EBMT who received immunosuppressive therapy (IST) with or without CSF3, a small but significant increase in hazard (1.9) of AML/MDS was reported in the CSF3-treated group[37]. In contrast, in a meta-analysis of 6 randomized control trials involving a total of 414 patients no statistically different risk of progression to MDS/AML between growth factor treated and control groups was noted[38]. A similar conclusion was reached in an earlier study based in 144 patients[39]. Strikingly, in a Japanese study, CSF3 treatment appeared to be more strongly associated with increased leukemia risk, particularly in cases refractory to IST[40]. Why the leukemia

incidence in this study differed from the European studies[37-39] is unclear but may relate to a more frequent occurrence of chromosome 7 abnormalities (monosomy 7, 7q-) in the Japanese patient group[40]. Supporting this idea, Sloand and colleagues showed that CSF3 preferentially stimulates the clonal expansion of MDS and SAA clones with monosomy 7, which was linked to an increased expression of a CSF3R isoform that lacks a major part of the C-terminal cytoplasmic domain as a result of alternative splicing[41]. On the other hand, IST unresponsive SAA patients not receiving CSF3 therapy may also develop monosomy 7[42]. In summary, although the increase of leukemia risk upon CSF3 treatment of MDS and SAA patients appears to be low, a causal relationship cannot be entirely excluded. Given the limited use of CSF3 in these settings, data from prospective trials further addressing this issue will unlikely become available in the near future.

Severe congenital neutropenia (SCN)

CSF3 therapy alleviates severe neutropenia and related clinical symptoms in more than 90 % of SCN patients and is the preferred choice of treatment of SCN[43]. In the pre-growth factor era, with early mortality due to opportunistic infections being the dominant complication, progression of SCN to acute leukemia was sporadically reported[44-46]. Ever since the introduction of CSF3 therapy, the possibility that CSF3 treatment would increase the risk of MDS/AML development in SCN patients has been an ongoing concern. CSF3 has now been routinely administered to patients with different types of chronic neutropenia for more than two decades. These patients provide an invaluable source for studying the long-term side effects of CSF3 treatment. Since 1994, the Severe Chronic Neutropenia International Registry (SCNIR) has monitored patients with different forms of neutropenia, including SCN, cyclic neutropenia and idiopathic neutropenia[47]. In 2000, the first comprehensive evaluation of the incidence of MDS/AML in SCN patients from the SCNIR was reported[48]. Among 352 SCN patients monitored for an average of 6 yrs (range 0.1-11 yrs) on CSF3 treatment, 31 developed MDS/AML with a cumulative risk of 13% after 8 years of CSF3 treatment. There was no apparent relationship to duration or dose of CSF3 treatment and progression to MDS/AML. A follow-up study published in 2006 involving 374 SCN patients showed that the hazard of MDS/AML increased over time, from 2.9% per year after 6 years to 8.0% per year after 12 years on CSF3[49]. After 10 years, the cumulative incidence for MDS/AML was 21%. This study also specifically addressed the incidence of leukemia in SCN patients relative to CSF3 responsiveness. Patients requiring more than the median dosage of CSF3 (8µg/kg/d) and nonetheless did not reach median absolute neutrophil counts after 6-18 months had a significantly increased MDS/AML incidence (40%) after 12 years compared to patients responding to lower CSF3 doses (11%)[49]. A possible explanation for these associations is that the HSC compartment in SCN patients who respond poorly to CSF3 is more damaged and therefore less susceptible to growth factors. This supports the notion that secondary

leukemia in SCN arises because chronic genotoxic stress in the hematopoietic stem cell compartment leads to the acquisition of oncogenic mutations, with CSF3 possibly playing a role in the clonal expansion of (pre-)leukemic cells. However, whether CSF3 therapy had contributed to MDS/AML development could not be determined in this study[49]. Of note, patients with cyclic or idiopathic neutropenia and neutropenia patients with an underlying metabolic disorder receiving CSF3 treatment regimens comparable to SCN patients treatment do not show an increased propensity to develop MDS or AML[47, 48]. Leukemic progression of neutropenia is thus mainly confined to patients diagnosed with SCN.

CSF3R mutations and malignant transformation in SCN

Direct evidence for a possible role of CSF3 in propagating leukemic expansion comes from SCN/AML cases in which remission of leukemia occurred after termination of CSF3 treatment[50]. However, such patients are exceptional and generally abrogation of CSF3 treatment generally has little or no effect on the leukemic burden in SCN/AML patients. The discovery that patients may harbor nonsense mutations in the *CSF3R* gene, resulting in the expression of truncated CSF3R proteins lacking ~100 amino acids from their C-terminal cytoplasmic domains provided a molecular indication for abnormal CSF3 signaling in SCN[51-53]. Functional studies revealed that these truncated CSF3R were hampered in their ability to transduce signals required for neutrophil differentiation in murine cell line models, a characteristic associated with a possible role of CSF3R dysfunction in leukemic progression of the disease[51-55]. Importantly, a later study showed that the *CSF3R* mutations are usually not constitutive but acquired in hematopoietic stem or progenitor cells during the course of CSF3 treatment[56]. Another major finding of this study was that the time between the first detection of *CSF3R* mutations and the diagnosis of MDS/AML varied greatly. For instance, in one patient a clone with an acquired *CSF3R* mutation appeared just three months before AML became overt, whereas in other patients *CSF3R* mutant clones were already detected four years before the acquisition of monosomy 7 and disease conversion to MDS/AML[56]. In addition, it became clear that patients may harbor multiple distinct acquired *CSF3R* mutations, suggestive of expansion of multiple affected clones[52, 56, 57].

The two major genetically defined subgroups of SCN prone to develop MDS/AML are patients with mutations in *ELA2* and patients with mutations in the *HAX1* gene[58]. More recently, two patients with X-linked neutropenia with mutations in the *WAS* gene were reported in which the disease evolved to MDS/AML[59]. In these three subtypes of SCN, leukemic progression is associated with the acquisition of *CSF3R* mutations and until now no differences in latencies or molecular and cytological features of the arising leukemias have been reported. In an analysis involving 145 SCN cases[57], *CSF3R* mutations were found in approximately one-third of the patients in the neutropenic phase

of the disease. Of 23 patients showing signs of malignant transformation, 18 (78%) harbored *CSF3R* mutations[57, 60], confirming that these mutations are strongly linked to leukemic predisposition [52, 60]. Notably, these mutations have also been detected in lymphoid cells and thus may be acquired in multipotent progenitors [61]. In contrast to SCN, acquisition of *CSF3R* mutations has not been observed in patients with cyclic or idiopathic neutropenia receiving CSF3 therapy[48]. These findings show that long-term CSF3 treatment in neutropenia patients other than SCN is not leukemogenic and further accentuate the correlation between leukemic progression of SCN and the acquisition of *CSF3R* mutations. However, despite all these suggestive correlations the issue whether these mutations are truly "drivers" or just "passengers" in the leukemic process cannot be settled with certainty[62]. For instance, one critical piece of information that is still missing is whether *CSF3R* mutations, once detected in the neutropenic phase, are invariably present in the MDS/AML cells and not "lost" during leukemic progression, as was recently demonstrated for JAKV617F mutations in myeloproliferative disorders[63]. So far, patients harboring clones with *CSF3R* mutations that progress to MDS/SCN without mutations have not been reported, but a systematic analysis is warranted to address this issue.

Molecular mechanisms responsible for leukemic progression of SCN

The critical genetic pathway(s) underlying the leukemic progression of SCN are still largely unknown. Cytogenetic abnormalities that are most frequently found in SCN/AML are chromosome 7 abnormalities (monosomy 7, 7q-) and trisomy 21[48]. Mutations in *Ras* have also been detected in SCN/AML, but their frequency is still controversial[64, 65]. By performing mutational profiling of 14 genes previously implicated in leukemogenesis, Link and colleagues found that mutations of tyrosine kinase genes, *FLT3*, *KIT*, and *JAK*2, were not detected in SCN/AML and neither were other abnormalities, e.g., mutations in *NPM1*, *CEBPA*, *TP53* that are common in de novo AML. As expected, mutations of *CSF3R* were the only regular abnormalities found in SCN/AML, again supporting the hypothesis that the mutant CSF3R may provide an "activated tyrosine kinase signal" important for leukemogenesis[66]. Aberrant signaling from the truncated CSF3R is to a major extent driven by defective ligand-induced receptor internalization owing to the loss of a dileucine-based internalization motif[5] and disturbed lysosomal routing due to the loss of the critical docking site for SOCS3[4, 10]. Prolonged CSF3-induced STAT5 activation and increased reactive oxygen species (ROS) production are two of the major consequences of CSF3R truncations, as demonstrated in vitro and in knock-in mouse models (*Csf3r*-D715) with patient equivalent mutations[6,7, 67]. Both of these mechanisms have been firmly implicated in cancer and may act synergistically in leukemic transformation. For instance, constitutive STAT5 activation by the mutant tyrosine kinase receptor FLT3-ITD has been suggested to drive leukemic cell growth via mechanisms involving direct transcriptional activation and chromatin remodeling[68]. In this respect it is of note that

STAT5 was indeed shown to be crucial for the selective clonal expansion of hematopoietic stem and progenitor cells harboring *Csf3r* mutations[69]. The elevated CSF3-induced ROS levels in bone marrow cells expressing truncated CSF3R may contribute to leukemic transformation by several mechanisms: by causing DNA damage and an increasing mutation rate in the HSC compartment[70] or by inactivation of critical phosphatases such as the lipid phosphatase PTEN and protein tyrosine phosphatases that negatively control growth factor signaling[71, 72].

Despite the proposed leukemogenic role of *CSF3R* mutations, *Csf3r*-D715 mice do not spontaneously develop leukemia[5, 73]. This might be explained by the fact that these mice had not been systematically exposed to CSF3 treatment or that their relatively short lifespan would be prohibitive to unveil the leukemogenic nature of *CSF3R* mutations. Alternatively, a likely hypothesis is that the transforming abilities of *CSF3R* mutations become overt only in the presence of the genetic defects underlying SCN, e.g., mutations in *ELA2*, *HAX1* or *WAS*. Because strains harboring SCN-derived mutations in *Ela2* and mice deficient in *Hax1* expression are available[74, 75] this could be addressed by crossing the *Csf3r*-D715 allele into these mice. However, a complication is that the *Ela2* and *Hax1* mouse models do not copy the neutropenic phenotype found in SCN patients, suggesting that in mice the consequences of these abnormalities for granulopoiesis are less severe or even lacking.

Are CSF3R mutations useful predictors for leukemic progression of SCN?

Because most SCN patients who progress to MDS/AML have a dismal therapy outcome, it is crucial to detect signs of malignant transformation at the earliest possible stage to create the opportunity to timely consider alternative treatments, such as allogeneic stem cell transplantation (allo-SCT)[58, 76]. Regular monitoring of *CSF3R* mutations has been considered to be helpful to screen for the risk of leukemic transformation[58], but when *CSF3R* mutations are present in minor clones, they can easily be missed in direct sequencing protocols. Possibly, next generation sequencing technologies allowing mutation detection in smaller subsets of cells will resolve this problem. Still, the unpredictable time intervals between the first detection of *CSF3R* mutations and the eventual leukemic transformation remains a major dilemma that makes a decision to opt for an allo-SCT in SCN patients who respond favorably to CSF3 treatment difficult. For that reason, the decision to transplant these patients without other additional evidence of leukemic progression (such as acquisition of monosomy 7) remains controversial and "watchful" waiting is being considered the most acceptable option, even though the success rate of treatment at a more advanced stage of malignant transformation will significantly decline[77]. Nonetheless, it must be taken into account that all patients with *CSF3R* mutations will eventually progress to AML[77], with time intervals varying between months, years, or even decades after the initial detection of mutant clones. A striking

example of such a long latency comes from the child in whom a *CSF3R* mutation was first identified[53]. CSF3 treatment of this patient started in 1990 and the *CSF3R* mutation was first detected in a majority of bone marrow cells in 1992[53]. Chronological sampling revealed that the mutant clone persisted and gave rise to RAEBT in 2007, rapidly followed by AML harboring trisomy 21 and a mutation in *RUNX1*.

Irrespective of the possible leukemogenic effects of CSF3 and *CSF3R* mutations in SCN patients, the case reported above stipulates that reliable predictors of leukemic transformation allowing a timely consideration of alternative treatment are urgently needed. Systematic sequential analysis may reveal which (epi-)genetic changes that occur early-on during the neutropenic phase of SCN may be linked to malignant transformation. For instance, SNP-comparative genomic hybridization (CGH) analysis in the above-mentioned SCN patient suggests that copy number neutral loss of heterozygosity (LOH), indicative of acquired uniparental disomy (UPD) in certain chromosomal regions, had already occurred in 1992, i.e., 15 years before malignant transformation (Beekman and Touw, unpublished results). Because UPD is one of the hallmarks of AML, these and other genetic modifications may give new insights in the mechanisms of leukemic progression of SCN and provide valuable indicators of leukemia risk in SCN patients, additional to reduced CSF3 responsiveness and *CSF3R* mutations.

Note added in proof: In a prospective study, Ehlers et al showed a significant correlation between the expression of the CF3R isoform IV and relapse incidence in childhood AML patients receiving CSF3 treatment[78].

ACKNOWLEDGMENTS

This work was supported by grants from the Center for Translational Molecular Medicine (CTMM) and the Dutch Cancer Society for Cancer Research "KWF kankerbestrijding".

REFERENCES

1. Lieschke GJ, Grail D, Hodgson G, et al. Mice lacking granulocyte colony-stimulating factor have chronic neutropenia, granulocyte and macrophage progenitor cell deficiency, and impaired neutrophil mobilization. Blood 1994;84(6):1737-46.
2. Liu F, Wu HY, Wesselschmidt R, Kornaga T, Link DC. Impaired production and increased apoptosis of neutrophils in granulocyte colony-stimulating factor receptor-deficient mice. Immunity 1996;5(5):491-501.
3. Panopoulos AD, Watowich SS. Granulocyte colony-stimulating factor: molecular mechanisms of action during steady state and ‚emergency' hematopoiesis. Cytokine 2008;42(3):277-88.
4. Irandoust MI, Aarts LH, Roovers O, Gits J, Erkeland SJ, Touw IP. Suppressor of cytokine signaling 3 controls lysosomal routing of G-CSF receptor. Embo J 2007;26(7):1782-93.

5. Touw IP, van de Geijn GJ. Granulocyte colony-stimulating factor and its receptor in normal myeloid cell development, leukemia and related blood cell disorders. Front Biosci 2007;12:800-15.
6. Hermans MH, Antonissen C, Ward AC, Mayen AE, Ploemacher RE, Touw IP. Sustained receptor activation and hyperproliferation in response to granulocyte colony-stimulating factor (G-CSF) in mice with a severe congenital neutropenia/acute myeloid leukemia-derived mutation in the G- CSF receptor gene. J Exp Med 1999;189(4):683-92.
7. McLemore ML, Poursine-Laurent J, Link DC. Increased granulocyte colony-stimulating factor responsiveness but normal resting granulopoiesis in mice carrying a targeted granulocyte colony-stimulating factor receptor mutation derived from a patient with severe congenital neutropenia. J Clin Invest 1998;102(3):483-92.
8. Boyle K, Egan P, Rakar S, et al. The SOCS box of suppressor of cytokine signaling-3 contributes to the control of G-CSF responsiveness in vivo. Blood 2007;110(5):1466-74.
9. Croker BA, Metcalf D, Robb L, et al. SOCS3 is a critical physiological negative regulator of G-CSF signaling and emergency granulopoiesis. Immunity 2004;20(2):153-65.
10. Wolfler A, Irandoust M, Meenhuis A, Gits J, Roovers O, Touw IP. Site-specific ubiquitination determines lysosomal sorting and signal attenuation of the granulocyte colony-stimulating factor receptor. Traffic 2009;10(8):1168-79.
11. Bradley TR, Metcalf D. The growth of mouse bone marrow cells in vitro. Aust J Exp Biol Med Sci 1966;44(3):287-99.
12. Pluznik DH, Sachs L. The induction of clones of normal mast cells by a substance from conditioned medium. Exp Cell Res 1966;43(3):553-63.
13. Griffin JD, Lowenberg B. Clonogenic cells in acute myeloblastic leukemia. Blood 1986;68(6):1185-95.
14. Nagata S, Tsuchiya M, Asano S, et al. Molecular cloning and expression of cDNA for human granulocyte colony-stimulating factor. Nature 1986;319(6052):415-8.
15. Souza LM, Boone TC, Gabrilove J, et al. Recombinant human granulocyte colony-stimulating factor: effects on normal and leukemic myeloid cells. Science 1986;232(4746):61-5.
16. Lotem J, Sachs L. Cytokine control of developmental programs in normal hematopoiesis and leukemia. Oncogene 2002;21(21):3284-94.
17. Sachs L. The control of hematopoiesis and leukemia: from basic biology to the clinic. Proc Natl Acad Sci U S A 1996;93(10):4742-9.
18. Piccaluga PP, Martinelli G, Malagola M, et al. Complete remission in acute myeloid leukemia with granulocyte-colony stimulating factor without chemotherapy. Report of cytogenetic remission of a t(9;11)(p22q23) positive AML patient and review of literature. Haematologica 2003;88(8):ECR28.
19. Sanz MA, Grimwade D, Tallman MS, et al. Management of acute promyelocytic leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. Blood 2009;113(9):1875-91.
20. Ravandi F. Role of cytokines in the treatment of acute leukemias: a review. Leukemia 2006;20(4):563-71.
21. Lowenberg B, van Putten W, Theobald M, et al. Effect of priming with granulocyte colony-stimulating factor on the outcome of chemotherapy for acute myeloid leukemia. N Engl J Med 2003;349(8):743-52.

22. Amadori S, Suciu S, Jehn U, et al. Use of glycosylated recombinant human G-CSF (lenograstim) during and/or after induction chemotherapy in patients 61 years of age and older with acute myeloid leukemia: final results of AML-13, a randomized phase-3 study. Blood 2005;106(1):27-34.

23. Nervi B, Ramirez P, Rettig MP, et al. Chemosensitization of acute myeloid leukemia (AML) following mobilization by the CXCR4 antagonist AMD3100. Blood 2009;113(24):6206-14.

24. Heuser M, Kuchenbauer F, Argiropoulos B, et al. Priming reloaded? Blood 2009;114(4):925-6; author reply 6-7.

25. Pitchford SC, Furze RC, Jones CP, Wengner AM, Rankin SM. Differential mobilization of subsets of progenitor cells from the bone marrow. Cell Stem Cell 2009;4(1):62-72.

26. Lane SW, Scadden DT, Gilliland DG. The leukemic stem cell niche: current concepts and therapeutic opportunities. Blood 2009;114(6):1150-7.

27. Holig K, Kramer M, Kroschinsky F, et al. Safety and efficacy of hematopoietic stem cell collection from mobilized peripheral blood in unrelated volunteers: 12 years of single-center experience in 3928 donors. Blood 2009;114(18):3757-63.

28. Pulsipher MA, Chitphakdithai P, Miller JP, et al. Adverse events among 2408 unrelated donors of peripheral blood stem cells: results of a prospective trial from the National Marrow Donor Program. Blood 2009;113(15):3604-11.

29. Bennett CL, Evens AM, Andritsos LA, et al. Haematological malignancies developing in previously healthy individuals who received haematopoietic growth factors: report from the Research on Adverse Drug Events and Reports (RADAR) project. Br J Haematol 2006;135(5):642-50.

30. Confer DL, Miller JP. Long-term safety of filgrastim (rhG-CSF) administration. Br J Haematol 2007;137(1):77-8; author reply 9-80.

31. Citron ML, Berry DA, Cirrincione C, et al. Randomized trial of dose-dense versus conventionally scheduled and sequential versus concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: first report of Intergroup Trial C9741/Cancer and Leukemia Group B Trial 9741. J Clin Oncol 2003;21(8):1431-9.

32. Hershman D, Neugut AI, Jacobson JS, et al. Acute myeloid leukemia or myelodysplastic syndrome following use of granulocyte colony-stimulating factors during breast cancer adjuvant chemotherapy. J Natl Cancer Inst 2007;99(3):196-205.

33. Touw IP, Bontenbal M. Granulocyte colony-stimulating factor: key (f)actor or innocent bystander in the development of secondary myeloid malignancy? J Natl Cancer Inst 2007;99(3):183-6.

34. Cole M, Strair R. Acute myelogenous leukemia and myelodysplasia secondary to breast cancer treatment: case studies and literature review. Am J Med Sci;339(1):36-40.

35. Marsh JC, Ganser A, Stadler M. Hematopoietic growth factors in the treatment of acquired bone marrow failure states. Semin Hematol 2007;44(3):138-47.

36. Jadersten M, Malcovati L, Dybedal I, et al. Erythropoietin and granulocyte-colony stimulating factor treatment associated with improved survival in myelodysplastic syndrome. J Clin Oncol 2008;26(21):3607-13.

37. Socie G, Mary JY, Schrezenmeier H, et al. Granulocyte-stimulating factor and severe aplastic anemia: a survey by the European Group for Blood and Marrow Transplantation (EBMT). Blood 2007;109(7):2794-6.

38. Gurion R, Gafter-Gvili A, Paul M, et al. Hematopoietic growth factors in aplastic anemia patients treated with immunosuppressive therapy-systematic review and meta-analysis. Haematologica 2009;94(5):712-9.

39. Locasciulli A, Arcese W, Locatelli F, Di Bona E, Bacigalupo A, Italian Aplastic Anaemia Study G. Treatment of aplastic anaemia with granulocyte-colony stimulating factor and risk of malignancy. Italian Aplastic Anaemia Study Group. Lancet 2001;357(9249):43-4.

40. Kojima S, Ohara A, Tsuchida M, et al. Risk factors for evolution of acquired aplastic anemia into myelodysplastic syndrome and acute myeloid leukemia after immunosuppressive therapy in children. Blood 2002;100(3):786-90.

41. Sloand EM, Yong AS, Ramkissoon S, et al. Granulocyte colony-stimulating factor preferentially stimulates proliferation of monosomy 7 cells bearing the isoform IV receptor. Proc Natl Acad Sci U S A 2006;103(39):14483-8.

42. Rosenfeld S, Follmann D, Nunez O, Young NS. Antithymocyte globulin and cyclosporine for severe aplastic anemia: association between hematologic response and long-term outcome. Jama 2003;289(9):1130-5.

43. Dale DC, Bonilla MA, Davis MW, et al. A randomized controlled phase III trial of recombinant human granulocyte colony-stimulating factor (filgrastim) for treatment of severe chronic neutropenia. Blood 1993;81(10):2496-502.

44. Gilman PA, Jackson DP, Guild HG. Congenital agranulocytosis: prolonged survival and terminal acute leukemia. Blood 1970;36(5):576-85.

45. Rosen RB, Kang SJ. Congenital agranulocytosis terminating in acute myelomonocytic leukemia. J Pediatr 1979;94(3):406-8.

46. Wong WY, Williams D, Slovak ML, et al. Terminal acute myelogenous leukemia in a patient with congenital agranulocytosis. Am J Hematol 1993;43(2):133-8.

47. Dale DC, Cottle TE, Fier CJ, et al. Severe chronic neutropenia: treatment and follow-up of patients in the Severe Chronic Neutropenia International Registry. Am J Hematol 2003;72(2):82-93.

48. Freedman MH, Bonilla MA, Fier C, et al. Myelodysplasia syndrome and acute myeloid leukemia in patients with congenital neutropenia receiving G-CSF therapy. Blood 2000;96(2):429-36.

49. Rosenberg PS, Alter BP, Bolyard AA, et al. The incidence of leukemia and mortality from sepsis in patients with severe congenital neutropenia receiving long-term G-CSF therapy. Blood 2006;107(12):4628-35.

50. Jeha S, Chan KW, Aprikyan AG, et al. Spontaneous remission of granulocyte colony-stimulating factor-associated leukemia in a child with severe congenital neutropenia. Blood 2000;96(10):3647-9.

51. Dong F, Brynes RK, Tidow N, Welte K, Lowenberg B, Touw IP. Mutations in the gene for the granulocyte colony-stimulating-factor receptor in patients with acute myeloid leukemia preceded by severe congenital neutropenia [see comments]. N Engl J Med 1995;333(8):487-93.

52. Dong F, Dale DC, Bonilla MA, et al. Mutations in the granulocyte colony-stimulating factor receptor gene in patients with severe congenital neutropenia. Leukemia 1997;11(1):120-5.

53. Dong F, Hoefsloot LH, Schelen AM, et al. Identification of a nonsense mutation in the granulocyte-colony-stimulating factor receptor in severe congenital neutropenia. Proc Natl Acad Sci U S A 1994;91(10):4480-4.

54. Dong F, van Buitenen C, Pouwels K, Hoefsloot LH, Lowenberg B, Touw IP. Distinct cytoplasmic regions of the human granulocyte colony-stimulating factor receptor involved in induction of proliferation and maturation. Mol Cell Biol 1993;13(12):7774-81.

55. Fukunaga R, Ishizaka-Ikeda E, Nagata S. Growth and differentiation signals mediated by different regions in the cytoplasmic domain of granulocyte colony-stimulating factor receptor. Cell 1993;74(6):1079-87.

56. Tidow N, Pilz C, Teichmann B, et al. Clinical relevance of point mutations in the cytoplasmic domain of the granulocyte colony-stimulating factor receptor gene in patients with severe congenital neutropenia. Blood 1997;89(7):2369-75.

57. Germeshausen M, Ballmaier M, Welte K. Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis: Results of a long-term survey. Blood 2007;109(1):93-9.

58. Zeidler C, Germeshausen M, Klein C, Welte K. Clinical implications of ELA2-, HAX1-, and G-CSF-receptor (CSF3R) mutations in severe congenital neutropenia. Br J Haematol 2009;144(4):459-67.

59. Beel K, Vandenberghe P. G-CSF receptor (CSF3R) mutations in X-linked neutropenia evolving to acute myeloid leukemia or myelodysplasia. Haematologica 2009;94(10):1449-52.

60. Germeshausen M, Skokowa J, Ballmaier M, Zeidler C, Welte K. G-CSF receptor mutations in patients with congenital neutropenia. Curr Opin Hematol 2008;15(4):332-7.

61. Germeshausen M, Welte K, Ballmaier M. In vivo expansion of cells expressing acquired CSF3R mutations in patients with severe congenital neutropenia. Blood 2009;113(3):668-70.

62. Bernard T, Gale RE, Evans JP, Linch DC. Mutations of the granulocyte-colony stimulating factor receptor in patients with severe congenital neutropenia are not required for transformation to acute myeloid leukaemia and may be a bystander phenomenon. Br J Haematol 1998;101(1):141-9.

63. Levine RL, Gilliland DG. Myeloproliferative disorders. Blood 2008;112(6):2190-8.

64. Germeshausen M, Kratz CP, Ballmaier M, Welte K. RAS and CSF3R mutations in severe congenital neutropenia. Blood 2009;114(16):3504-5.

65. Kalra R, Dale D, Freedman M, et al. Monosomy 7 and activating RAS mutations accompany malignant transformation in patients with congenital neutropenia. Blood 1995;86(12):4579-86.

66. Link DC, Kunter G, Kasai Y, et al. Distinct patterns of mutations occurring in de novo AML versus AML arising in the setting of severe congenital neutropenia. Blood 2007;110(5):1648-55.

67. Hermans MH, Ward AC, Antonissen C, Karis A, Lowenberg B, Touw IP. Perturbed granulopoiesis in mice with a targeted mutation in the granulocyte colony-stimulating factor receptor gene associated with severe chronic neutropenia. Blood 1998;92(1):32-9.

68. Kornfeld JW, Grebien F, Kerenyi MA, et al. The different functions of Stat5 and chromatin alteration through Stat5 proteins. Front Biosci 2008;13:6237-54.

69. Liu F, Kunter G, Krem MM, et al. Csf3r mutations in mice confer a strong clonal HSC advantage via activation of Stat5. J Clin Invest 2008;118(3):946-55.

70. Zhu QS, Xia L, Mills GB, Lowell CA, Touw IP, Corey SJ. G-CSF induced reactive oxygen species involves Lyn-PI3-kinase-Akt and contributes to myeloid cell growth. Blood 2006;107(5):1847-56.

71. Lee SR, Yang KS, Kwon J, Lee C, Jeong W, Rhee SG. Reversible inactivation of the tumor suppressor PTEN by H2O2. J Biol Chem 2002;277(23):20336-42.

72. Ross SH, Lindsay Y, Safrany ST, et al. Differential redox regulation within the PTP superfamily. Cell Signal 2007;19(7):1521-30.

73. van de Geijn GJ, Aarts LH, Erkeland SJ, Prasher JM, Touw IP. Granulocyte colony-stimulating factor and its receptor in normal hematopoietic cell development and myeloid disease. Rev Physiol Biochem Pharmacol 2003;149:53-71.

74. Chao JR, Parganas E, Boyd K, Hong CY, Opferman JT, Ihle JN. Hax1-mediated processing of HtrA2 by Parl allows survival of lymphocytes and neurons. Nature 2008;452(7183):98-102.

75. Grenda DS, Johnson SE, Mayer JR, et al. Mice expressing a neutrophil elastase mutation derived from patients with severe congenital neutropenia have normal granulopoiesis. Blood 2002;100(9):3221-8.

76. Choi SW, Boxer LA, Pulsipher MA, et al. Stem cell transplantation in patients with severe congenital neutropenia with evidence of leukemic transformation. Bone Marrow Transplant 2005;35(5):473-7.

77. Freedman MH, Alter BP. Risk of myelodysplastic syndrome and acute myeloid leukemia in congenital neutropenias. Semin Hematol 2002;39(2):128-33.

78. Ehlers S, Herbst C, Zimmermann M, et al. Granulocyte colony-stimulating factor (G-CSF) treatment of childhood acute myeloid leukemias that overexpress the differentiation-defective G-CSF receptor isoform IV is associated with a higher incidence of relapse. J Clin Oncol 2010;28(15):2591-7.

CHAPTER 3

SEQUENTIAL GAIN OF MUTATIONS IN SEVERE CONGENITAL NEUTROPENIA PROGRESSING TO ACUTE MYELOID LEUKEMIA

Renée Beekman[1], Marijke Valkhof[1], Mathijs A. Sanders[1], Paulette van Strien[1], Jurgen R. Haanstra[1], Lianne Broeders[1], Wendy M. Geertsma-Kleinekoort[1], Anjo J.P. Veerman[2], Peter J.M. Valk[1], Roel G. Verhaak[3], Bob Löwenberg[1], Ivo P. Touw[1]

[1]Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

[2]Department of Pediatric Hematology-Oncology, Vrije Universiteit Medical Center, Amsterdam, The Netherlands;

[3]Department of Bioinformatics and Computational Biology, MD Anderson Cancer Center, Houston TX, USA.

Blood 2012 May 31;119(22):5071-7

ABSTRACT

Severe congenital neutropenia (SCN) is a bone marrow failure syndrome with a high risk to progress towards acute myeloid leukemia (AML). The underlying genetic changes involved in SCN evolution to AML are largely unknown. We obtained serial hematopoietic samples of an SCN patient who developed AML 17 years after initiation of granulocyte-colony stimulating factor (G-CSF) treatment. Next-generation sequencing was done to identify mutations during disease progression. In the AML phase, we found 12 acquired non-synonymous mutations. Three of these, in *CSF3R*, *LLGL2* and *ZC3H18*, co-occurred in a subpopulation of progenitor cells already in the early SCN phase. This population expanded in time, whereas clones solely harboring *CSF3R* mutations disappeared from the bone marrow. The other 9 mutations were only apparent in the AML and affected known AML-associated genes *(RUNX1* and *ASXL1*) and chromatin remodelers (*SUZ12* and *EP300)*. In addition, a novel *CSF3R* mutation was found conferring autonomous proliferation to myeloid progenitors. We conclude that progression from SCN towards AML is a multistep process with distinct mutations arising early during the SCN phase and others later in AML development. Sequential gain of two *CSF3R* mutations implicates abnormal G-CSF signaling as a driver of leukemic transformation in this case of SCN.

INTRODUCTION

Severe congenital neutropenia (SCN) is a bone marrow failure syndrome characterized by strongly reduced neutrophil counts and recurrent, potentially life threatening, opportunistic bacterial infections. Treatment with granulocyte-colony stimulating factor (G-CSF) elevates peripheral neutrophil counts and reduces the risk of infections[1]. Leukemic progression of SCN is a major concern, with an estimated overall cumulative incidence of approximately 20% after 15 years of G-CSF treatment[2].

Constitutional mutations in the gene encoding neutrophil elastase (*ELANE*) are common defects in SCN[3]. In addition, the acquisition of nonsense mutations in the gene encoding the granulocyte-colony stimulating factor receptor (*CSF3R*) is a unique feature in SCN patients[4-7]. These mutations lead to expression of truncated CSF3R proteins, also known as delta forms. In cell line models, truncated CSF3R are hampered in transducing signals required for proper neutrophil differentiation. Additionally, they confer increased proliferative responses to G-CSF treatment but do not cause leukemia in mice[4-6, 8-11]. *CSF3R* delta mutations can be detected in approximately 30% of SCN patients. In some cases, distinct clones with different *CSF3R* delta mutations are present for many years[7, 12]. After evolution of SCN towards AML, *CSF3R* delta mutations are found in approximately 80% of the cases[12]. Until now, all reported SCN/AML cases harboring a *CSF3R* delta mutation in the SCN phase also carry this mutation in the leukemic phase. These observations suggest that leukemic progression in SCN follows a unique pattern, with *CSF3R* delta mutations as an early event, followed by additional genetic and epigenetic events that

are essential for full leukemic transformation. Chromosomal aberrations, such as loss of chromosome 7 and gain of chromosome 21, are apparent in AML arising from SCN and other bone marrow failure syndromes like Fanconi anemia and Shwachman-Diamond syndrome[13]. However, mutations that are quite commonly seen in *de novo* AML have not been reported in AML arising from SCN[14]. Thus, the additional molecular events involved in leukemic progression of SCN remain largely unknown.

To identify the sequential genetic events in leukemic progression of SCN towards AML, we collected serial hematopoietic samples of an SCN patient who developed AML after 17 years of G-CSF therapy. Using whole exome sequencing, we found 12 somatic non-synonymous mutations in the leukemic blasts of this patient. Three of these mutations, the known *CSF3R* mutation and mutations in *LLGL2* and *ZC3H18*, were already present at low frequencies in the early SCN phase, 15 years before AML was diagnosed. Myeloid colony analysis showed that these 3 "early" mutations co-existed in the same hematopoietic progenitors in a small subpopulation of bone marrow cells. Six years later, in the "intermediate" SCN phase, still 9 years before the AML became overt, we observed an expansion of the clone harboring all 3 mutations. The other 9 mutations were only apparent in the AML. The latter "late" appearing mutations comprise a second, novel, *CSF3R* mutation in addition to a series of new and known AML-associated mutations. The novel *CSF3R* mutation is located on the already mutated *CSF3R-d715* allele and causes growth factor independence of myeloid progenitors.

MATERIALS AND METHODS

Case report

A 27-year old male SCN patient was diagnosed with AML 17 years after the start of G-CSF treatment (10µg/kg/day), on which he reached normal neutrophil counts. The patient had a constitutional heterozygous *ELANE* mutation, G174R. At the age of 12, 2 years after G-CSF treatment was initiated, a *CSF3R* delta mutation (*CSF3R-d715*) was discovered in the bone marrow[6]. At the time of AML diagnosis, the peripheral blood contained 24% blasts and dysplasia was observed in the bone marrow. G-CSF treatment was stopped at this point. Six weeks later, a bone marrow analysis revealed 17% blasts. Immunophenotypically, these blasts were of myeloid origin, i.e., positive for CD34, CD117, CD13, CD133, CD33, MPO and CD90. Because no HLA-identical donor was available, the patient received a matched unrelated donor (MUD) allogeneic bone marrow transplantation. Induction therapy was given according to the induction therapy scheme HOVON42A of the Hemato-Oncology Foundation for Adults in the Netherlands[15]. At initiation of induction therapy, the bone marrow contained 15.7% blasts, with 10-50% dysplasia in all lineages. Routine cytogenetic and molecular diagnostics revealed a trisomy 21 (47, XY, +21 [14] /46, XY [4]), with no additional abnormalities (*AML-ETO, CBFB/MYH11, FLT3ITD, FLT3TKD*, mutations in *NPM1, NRAS, KRAS, c-KIT, JAK2* and *CEBPA*). After the second induction cycle trisomy 21, was undetectable in a marrow cytogenetic

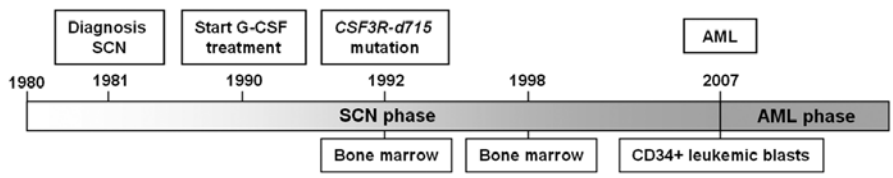


Figure 1. Chronological overview of the clinical course of the SCN/AML patient. Distinct events in the disease course are indicated above the timeline, i.e., the diagnosis of SCN, the initiation of G-CSF therapy, the discovery of the *CSF3R*-d715 mutation and the diagnosis of AML.

analysis. The MUD transplantation was administered after myeloablative conditioning with chemotherapy and total body irradiation. Two months after the transplantation 28% blast were detected in the bone marrow, indicating a recurrence of the AML and the patient died 3.5 months after the transplant. Figure 1 gives a schematic overview of the disease history.

Patient cell samples

Ficoll-gradient separated bone marrow cells from the SCN phases and CD34+ leukemic blasts from the peripheral blood in the leukemic phase were used. Control DNA was isolated from bone marrow-derived fibroblasts. All cell samples were obtained and frozen according to established procedures for viable cell cryopreservation as previously described[16]. The study was performed under the permission of the Institutional Review Board of the Erasmus MC, registration number MEC-2008-387 for biobanking and MEC-2012-030 for the genetic analysis of leukemic progression in SCN patients.

Nucleotide sequencing

Whole Exome sequencing (WES). Sequencing libraries were prepared according to the SureSelect Target Enrichment system for Illumina, protocol version 2.2.1, Nov. 2010. In short, 3 µg genomic DNA was sheared to fragments of approximately 170 base pairs using the Covaris S-series Single Tube Sample Preparation System, Model S2 (Covaris, Woburn, MA, USA). Fragment sizes were checked on the Bioanalyzer (Agilent, Santa Clara, CA). Adapter ligated libraries were prepared according to the manufacturer's protocol using the Paired-End Genomic DNA Sample Prep Kit PE-102-1001 (Illumina, San Diego, CA); 5 cycles of amplification were used. Five hundred ng of prepped library was taken for hybridization with the SureSelect Human All Exon Kit (G3362A, Agilent). A sample concentration of 5.5 picomolar was loaded for sequencing on the Hiseq2000 (Illumina) using 101-bp paired-end reads.

Sequencing reads were processed with the Casava pipeline (version 1.7, Illumina). For alignment the Hg18/NCBI36 assembly (March 2006) was used. Detection of single nucleotide variants, deletions and insertions was performed with otherwise default settings, while snpCovCutoff and indelsCovCutoff were switched off. Variations detected in

the AML sample in 2 independent sequence runs were further analyzed after removal of germ-line variations (present in the fibroblasts) and single nucleotide polymorphisms (SNPs, dbSNP)[17]. Next, non-synonymous variants were determined. Integrative Genome Browser was used for sequence read visualization[18].

Sanger sequencing. WES results were validated by Sanger sequencing, performed according to the manufacturer's protocol (Applied Biosystems, Foster City, CA, USA) using primers indicated in Table S1. Before amplicon generation, genomic DNA or cDNA was first amplified using a Whole Genome Amplification kit (WGA2, Sigma-Aldrich, Zwijndrecht, The Netherlands). DNA was purified with a PCR purification kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol and diluted to 50 ng/µl. Hundred nanograms of amplified DNA was used for amplicon generation; cycling conditions were 30" at 95°C, 30" at the indicated annealing temperature (Table S1) and 45" at 72°C for 35 cycles. In some instances, the unamplified material was used directly for Sanger sequencing (Table S1).

Amplicon-based deep sequencing. Amplicons were generated and purified according to the Amplicon Library Preparation Method Manual (version May 2010, Roche, Basel, Switzerland). Primers and annealing temperatures are indicated in Table S2; 35 cycles were used for amplification. DNA enriched beads, carrying the amplification products, were generated according to the emPCR Amplification Method Manual – Lib-A (version May 2010, Roche); a beads to amplicon ratio of 1:2 was used. Amplicons were analyzed with the GS junior (Roche). Sequence reads were analyzed using the GS Amplicon Variant Analyzer (Roche). For the SCN samples, coverage of at least 1600 was achieved to identify mutations present in minor clones within the bone marrow. For the AML sample coverage of 80 was considered sufficient to validate mutations.

Human myeloid colony assay

Bone marrow was thawed at 37°C, washed twice with IMDM (Gibco Invitrogen, San Diego, CA) with 10% FCS (PAA laboratories, Pasching, Austria). Per 4 ml of culture medium, 2.9 ml MethoCult (H4230, Stem Cell Technologies, Vancouver, Canada), 980 µl IMDM and human GM-CSF (Immunex, Seattle, WA), human G-CSF (Neupogen, Amgen, Thousand Oaks, CA) and human IL-3 (R&D Systems, Minneapolis, USA) in final concentrations of respectively 2 ng/ml, 200 ng/ml and 25 ng/ml were used. Cells were plated at a density of 0.8 x 10^5 /ml. After 2 weeks genomic DNA of single colonies was isolated, followed by amplification using the Whole Genome Amplification kit and Sanger sequencing of *CSF3R-d715*, *LLGL2* and *ZC3H18,* as described above. Results were validated in an independent round of whole genome amplification for i) colonies harboring a mutation, ii) colonies with unclear sequences and iii) a number of randomly chosen non-mutated colonies to rule out amplification artifacts. All colonies harboring mutations in *CSF3R*, *LLGL2* or *ZC3H18* were also analyzed for the presence of the remaining 9 mutations found in the AML sample.

Murine colony assays

Four different CSF3R expression constructs (WT, *d715, T595I, d715/T595I*) were generated and retrovirally transduced into bone marrow cells of *Csf3r* deficient FVB/N mice[19]. Colony assays of these transduced progenitors were performed as previously described[20].

Further details of these procedures are given in the Supplementary Materials and Methods.

RESULTS

Whole exome sequencing reveals acquired mutations in SCN/AML

Whole exome sequencing was done on genomic DNA from the CD34+ leukemic blast fraction and the fibroblast control sample. Acquired non-synonymous mutations were detected by identification of single nucleotide variants and small insertions and deletions, followed by subtraction of variants present in the control fibroblasts and known single nucleotide polymorphisms[17]. Twelve non-synonymous acquired mutations were identified and validated by Sanger sequencing (Table 1, Figure S1). Except for the mutation in *FBXO18*, all mutations occurred in evolutionary conserved amino acids (Figure S2). With the exception of *LAMB1*, all mutant transcripts were detectably expressed in the leukemic blasts (Figure S3). Mutations in *ASXL1* and *RUNX1* are known in myeloid malignancies[21, 22]. Deletions in *EP300*, distinct from the 7-bp deletion found in this patient, have been reported in lymphomas[23, 24]. The ATT insertion in *SUZ12* duplicates an isoleucine at amino acid position 597, located in the conserved VEFS-box. Mutations in this region, which is involved in the interaction between SUZ12 and the histone methyltransferase EZH2 in the polycomb repressor complex 2 (PRC2), have recently also been identified in myelodysplastic/myeloproliferative neoplasms (MDS/MPN) with 17q abnormalities[25]. As expected, the previously identified *CSF3R* delta mutation (*CSF3R-d715)* was present in the leukemic blasts, but remarkably a new *CSF3R* mutation, *T595I*, was now also present. Furthermore, the *CSF3R-T595I* mutation was located on the same allele as the delta mutation, as determined by Sanger sequencing of single amplicons generated from cDNA. Using exome sequencing data from 199 AML cases reported by The Cancer Genome Atlas (TCGA), a similar single *CSF3R-T595I* mutation was detected. Additionally, mutations in *ASXL1* (n=5), *CCDC155* (n=1), *LLGL2* (n=1), *MGA* (n=1), *RUNX1* (n=17), *SUZ12* (n=2) and *ZC3H18* (n=2) were found in the TCGA data set (Table S3; R.G.V. and The Cancer Genome Atlas disease working group, unpublished data, January 2012).

Amplicon-based sequencing reveals an early pre-leukemic clone that expands over time

Amplicon-based deep sequencing was applied to analyze the presence of all 12 somatic mutations in the bone marrow samples obtained at 15 and 9 years before AML was diagnosed (Figure 1). Not only the known *CSF3R-d715* mutation, but also mutations in *LLGL2*

Table 1. Somatic non-synonymous mutations in SCN/AML.

Gene Symbol	RefSeq Reference Transcript	Genomic DNA Change (NCBI36/hg18)	cDNA Change	Mutation Type	Amino Acid Change	Protein Change
ASXL1	NM_015338.5	g.chr20:30485948dupA	c.1772dupA	Indel frameshift	Frameshift and premature stop	p.Y591*
CCDC155	NM_144688.4	g.chr19:54601976C>T	c.820C>T	Missense	Arg>Trp	p.R274W
CSF3R-T595I	NM_000760.3	g.chr1:36706021G>A	c.1853C>T	Missense	Thr>Ile	$p.T595I
CSF3R-d715	NM_000760.3	g.chr1:36704841G>A	c.2215C>T	Nonsense	Gln>*	$p.Q716*
EP300	NM_001429.3	g.chr22:39902447_39902453delTGGAGAC	c.5030_5036delTGGAGAC	Indel frameshift	Frameshift and premature stop	p.V1677Dfs*30
FBXO18	NM_032807.3	g.chr10:6003435C>G	c.2372C>G	Missense	Ala>Gly	p.A791G
LAMB1	NM_002291.2	g.chr7:107387385delG	c.2445delC	Indel frameshift	Frameshift and premature stop	p.P815Pfs*65
LLGL2	NM_004524.2	g.chr17:71070826G>C	c.665G>C	Missense	Arg>Pro	p.R222P
MGA	NM_001164273.1	g.chr15:39787311C>T	c.2282C>T	Missense	Pro>Leu	p.P761L
RUNX1	NM_001754.4	g.chr21:35153662C>T	c.592G>A	Missense	Asp>Asn	p.D198N
SUZ12	NM_015355.2	g.chr17:27346889_27346891dupATT	c.1789_1791dupATT	Indel	Insertion Ile	p.597dupI
ZC3H18	NM_144604.3	g.chr16:87192175delC	c.777delC	Indel frameshift	Frameshift and premature stop	p.P259Pfs*15

All 12 somatic non-synonymous mutations identified in the AML phase are listed. For each mutation, Refseq reference transcripts, the position of the mutation on genomic DNA, cDNA and protein level, the mutation type and the effect on the protein are indicated. See also Figure S1-S3. *Stopcodon, $Amino acid numbers based on earlier publications[4, 6].

and *ZC3H18* were already present in these earlier disease phases (Figure 2A, Table S4). We investigated the clonal hierarchy of these mutations in single myeloid colonies cultured from the earliest bone marrow sample (15 years before AML diagnosis). In the individual colonies (n=88), the mutation status of *CSF3R-d715*, *LLGL2* and *ZC3H18* was determined. Fifteen colonies (17%) harbored both the *CSF3R-d715* and the *LLGL2* mutation, whereas none of the colonies exhibited exclusively either the *LLGL2* or the *CSF3R-d715* mutation (Figure 2B, Table S5). Two of the *CSF3R-d715* and *LLGL2* mutated colonies also carried the *ZC3H18* mutation (Figure 2B, Table S5), indicating that this mutation had emerged later in time. None of the other 9 mutations found in the AML cells was apparent in these colonies (Table S5).

A previous report has shown that multiple *CSF3R* delta mutations can be present in distinct progenitors in the bone marrow of an individual SCN patient[7]. In line with this, we found myeloid colonies with *CSF3R-d717* (n=2) and *CSF3R-d725* (n=1) (Figure 2B, Table S5). Each of these mutations and yet an additional delta mutation (*CSF3R-d730*) were detected in the SCN phase at low frequencies by amplicon-based deep sequencing (Figure 2C, Table S6). None of these variant *CSF3R* mutant clones harbored *LLGL2* or *ZC3H18* mutations, nor were they seen as dominant clones in the AML (Figure 2, Table S5-S6). No viably frozen cells were available from the bone marrow sample obtained 9 years before AML development

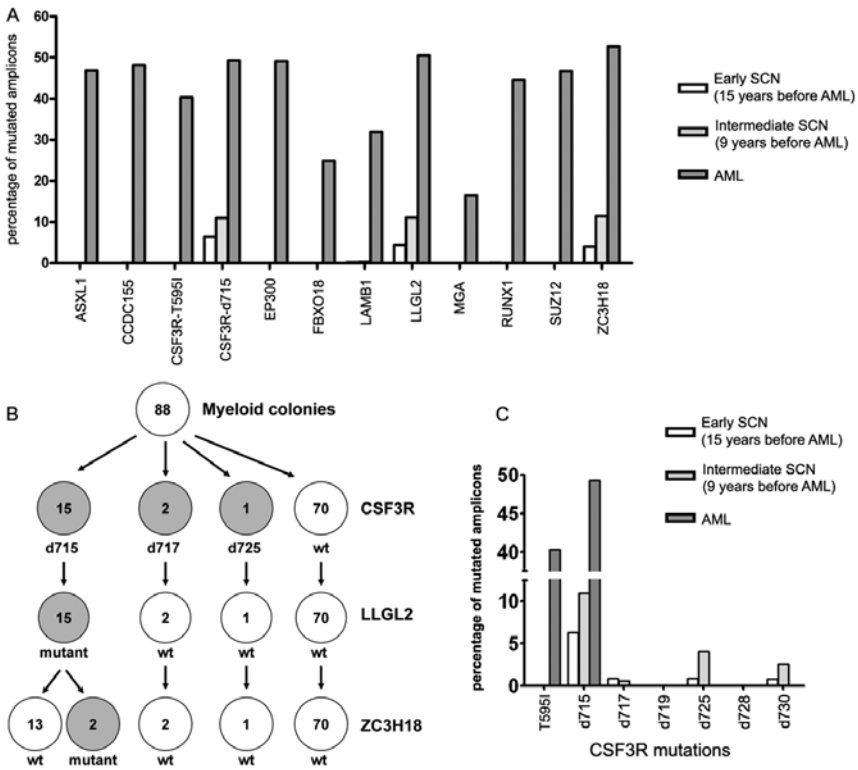


Figure 2. Acquisition of mutations in the evolution of SCN towards AML. (A) The 12 somatic non-synonymous mutations identified in the leukemic blasts were analysed in the SCN phase using amplicon-based deep sequencing. Per mutation, the percentage of mutated amplicons is shown. Based on their frequencies in the AML population, all mutations are considered to be heterozygous, implying that the number of cells carrying the mutations is estimated to be twice the number of mutated amplicons. (B) Single myeloid colonies grown from the bone marrow sample obtained 15 years before leukemia development were analysed for the presence of mutations in *CSF3R*, *LLGL2* and *ZC3H18*. See also Table S5. (C) The presence of different *CSF3R* mutations in the bone marrow obtained 15 and 9 years before leukemia development and in the leukemic phase was investigated by amplicon-based deep sequencing. Per mutation, the percentage of mutated amplicons is shown. *T595I*: *CSF3R* mutation T595I, *d715-d730*: *CSF3R* delta mutations at amino acid position 715 to 730.

and colony analysis could not be performed at this stage. However, by amplicon-based deep sequencing we observed a parallel increase of the *CSF3R-d715*, *LLGL2* and *ZC3H18* mutations from 15 to 9 years before AML development (Figure 2A). Together with the finding that these mutations are present in the same myeloid progenitor cells (Figure 2B), this observation is consistent with a selective outgrowth of clones carrying these 3 mutations.

Sequential gain of a second CSF3R mutation results in G-CSF independence

A new *CSF3R* mutation, which acquired at the *CSF3R-d715* mutant allele, was found exclusively in the AML blasts and changed a polar threonine residue at amino acid position

595 into a nonpolar isoleucine. This residue is located in a highly conserved threonine-rich region in the extracellular domain of the G-CSF receptor (Figure S2). Introduction of human *CSF3R* mutant receptors, carrying this new *T595I* mutation (Figure 3A), into Csf3r deficient primary mouse bone marrow progenitors resulted in the autonomous outgrowth of myeloid colony-forming cells (Figure 3, Table S7). Thus, in the AML phase of disease evolution two different co-existing mutations, i.e., the *T595I* single amino acid substitution and the *CSF3R-d715* mutation had accumulated in the gene encoding the G-CSF receptor. Because expression of the new *CSF3R* mutant without the delta mutation conferred G-CSF independence as did the mutant receptor carrying both the delta and the extracellular mutation, this gain of function can entirely be attributed to the *T595I* mutation. However,

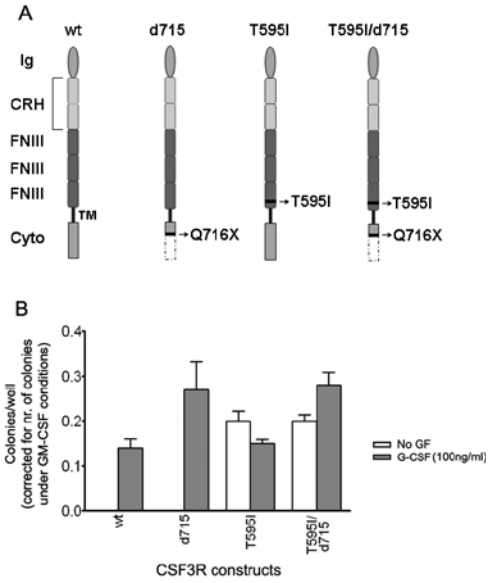


Figure 3. Functional analysis of *CSF3R* mutants in myeloid progenitor cell assays. In vitro colony growth of Csf3r deficient murine hematopoietic progenitor cells expressing different *CSF3R* mutants. (A) Graphical representation of the different *CSF3R* constructs. Wild type (wt), T595I (containing the extracellular mutation at amino acid position 595), d715 (containing the intracellular mutation, Q716X, causing the introduction of a stop codon at amino acid position 716) and T595I/d715, containing both mutations as found in the SCN/AML patient. Ig: Immunoglobulin like domain; CRH: cytokine receptor homology domain; FNIII: fibronectin type III repeats; TM: transmembrane domain; cyto: cytoplasmic domain. Nomenclature has been adopted from Layton et al.[42] (B) Colonies were grown in the presence of puromycin, either without growth factor (no GF) or with 100ng/ml human G-CSF. The induced colony growth is dependent on the transduction efficiency and the type of *CSF3R* construct. The transduction efficiency can be deduced from the number of GM-CSF-induced colonies under puromycin selection as the *CSF3R* constructs confer puromycin resistance, but do not affect GM-CSF-induced colony growth. Hence, by dividing the number of colonies by the number of GM-CSF induced colonies the transduction efficiency was corrected for.

the *T595I/d715* colonies were bigger than the *T595I* colonies (Figure S4), which is suggestive of a higher proliferation capacity by the addition of the *CSF3R-d715* mutant.

DISCUSSION

The results of the present study identified non-synonymous mutations acquired in an SCN patient who progressed to AML.The availability of sequential hematopoietic samples from the childhood SCN phase to overt AML, spanning a period of 17 years, provided the unique opportunity to identify the early and late genetic defects associated with leukemic progression (Figure 4). The *CSF3R-d715* mutation and a mutation in *LLGL2,* encoding the human homologue of the Drosophila lethal giant larvae (*Lgl*) gene, were the first 2 acquired mutations in the early SCN phase. Loss of *Lgl* in Drosophila leads to inadequate distribution of the cell polarity protein Numb, resulting in inappropriate cell fate determinations and tumor formation in epithelial tissues and the brain[26-28]. In man, the NUMB protein has been implicated in controlling the balance between symmetric versus asymmetric hematopoietic stem cell divisions. Interestingly, deregulation of NUMB expression has been associated with blast transformation of chronic myeloid leukemia[29, 30]. How the *LLGL2* mutation found in this study affects hematopoietic stem cell divisions is still unknown; however, the fact that *CSF3R-d715* and *LLGL2* mutations were uniformly present in the same myeloid cells could suggest that they cooperate. Hierarchically, the next genetic abnormality occurring in the early SCN phase in the

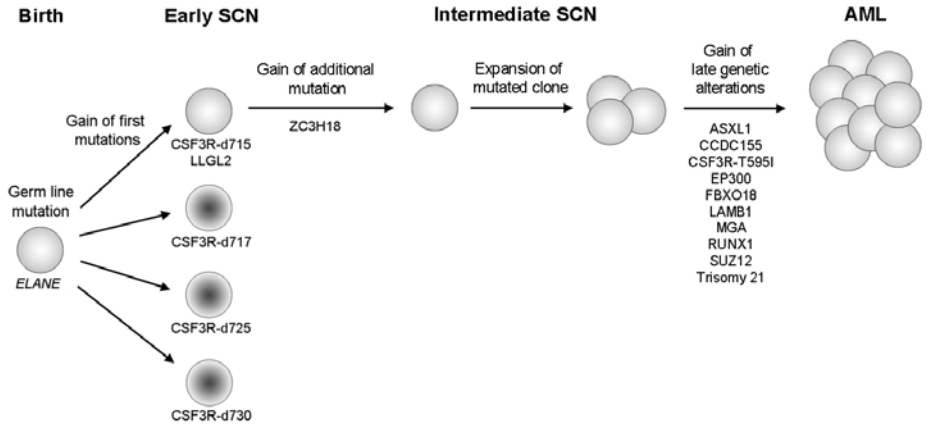


Figure 4. Schematic representation of the clonal evolution of SCN towards AML. The sequential genetic events, starting with the presence of a germ line mutation in the gene encoding neutrophilic elastase (*ELANE)* are indicated. A sequential gain of *CSF3R* delta mutations and an *LLGL2* mutation is observed in the early SCN phase. Only the clone harboring the *CSF3R-d715* and the *LLGL2* mutation gained an additional mutation in *ZC3H18*, followed by its expansion in the intermediate SCN phase. Gain of 9 additional mutations and trisomy 21 in the mutated population preceded complete transformation towards AML. *CSF3R*-d715-d730: *CSF3R* delta mutations at amino acid position 715 to 730.

CSF3R-d715- and *LLGL2*-mutated clone was a mutation in *ZC3H18*. ZC3H18 is a putative mRNA binding protein with a still unknown function, but has been shown to be essential for differentiation in trypanosomes[31].

Additionally, we found small subpopulations harboring distinct *CSF3R* delta mutations in the bone marrow at the early SCN stage. All these clones disappeared during the disease course, except the *CSF3R-d715* clone which evolved towards AML. The different *CSF3R* delta mutations cause expression of distinct truncated G-CSF receptors that all have similar consequences for signaling; they result in a sustained activation of signal transducer and activator of transcription 5 (STAT5)[8]. STAT5 is a transcription factor, implicated in abnormal signaling responses of leukemic cells with mutated forms of the FLT3 receptor (FLT3-ITD) in AML and the BCR-ABL fusion protein in CML[32, 33]. Furthermore, why one of these *CSF3R* delta mutant clones survived *in vivo* and progressed towards a fully transformed AML clone while the other *CSF3R* delta variants extinguished during disease development currently remains unexplained. However, it is conceivable that the additional mutations in *LLGL2* and *ZC3H18*, exclusively present in the *CSF3R-d715* clone, conferred a competitive growth advantage of this particular subclone representative of essential early steps in leukemic progression that cooperate with the aberrant signaling from the truncated G-CSF receptor.

Besides early genetic events, we found 9 mutations that occurred later in the process of leukemic transformation. Of particular interest is the novel *CSF3R* mutation (*T595I*), which appeared exclusively in the AML stage and imposed growth factor independence on an already functionally defective G-CSF receptor. A different mutation in the *CSF3R* transmembrane domain*, CSF3R-T617N,* with a similar downstream effect was previously found as a constitutive mutation in a family with hereditary chronic neutrophilia and as an acquired mutation in 2 AML patients. This mutation is suggested to cause ligand independent homodimerization and induces growth factor independent proliferation and differentiation[34, 35]. The major difference between the *T617N* and the *T595I* mutation in our patient is that the latter one is located on the already affected *CSF3R-d715* allele, which has been shown to cause increased proliferation and impaired differentiation in cell line and animal models[8, 36, 37] and which could explain the increase in colony size between the *T595I* mutant and the *T595I/d715* mutant. The acquisition of autonomous growth abilities by myeloid progenitor cells that already express a hyper-responsive G-CSF receptor mutant strongly suggests that perturbed G-CSF signaling was of vital importance for malignant transformation in this case of SCN. To our knowledge, this is the first example of a gain of 2 different mutations in the same receptor in the process of malignant transformation. An important but still open question is whether the administration of G-CSF to this patient had contributed to the acquisition of this additional mutation. Possibly, the continuous proliferative pressure imposed by G-CSF on clones carrying mutations in *CSF3R-d715* and *LLGL2* and later also in *ZC3H18* may have provided

the context for the selection of a clone harboring this self-activating *CSF3R* mutation, pushing it to become an autonomously proliferating and dominant leukemic clone.

Abnormalities appearing in the AML phase included mutations in *ASXL1*, *SUZ12*, and *EP300*, genes encoding proteins involved in chromatin modification. Mutations in *ASXL1* have been reported previously in AML and are associated with an unfavorable prognosis[38]. SUZ12 is a member of the PRC2 complex that also contains EZH2, the histone methyl transferase responsible for the di- and tri-methylation of lysine 27 in the tail of histone 3 (H3K27), imposing a chromatin mark that represses gene expression. Mutations affecting *EZH2* and less frequently *SUZ12* have been detected in MDS/MPN patients[25, 39, 40]. In contrast, mutations in *EP300* and the highly related *CREBBP*, encoding histone acetyl transferases that act as transcriptional co-activators, have not yet been reported in myeloid malignancies but are the most frequent structural abnormalities in follicular lymphoma and diffuse large B cell lymphoma[23, 24]. Mutations in *CCDC155*, encoding coiled-coil domain containing protein 155 with unknown function; *FBXO18*, encoding a DNA helicase involved in DNA repair and genomic integrity; *LAMB1*, encoding an extracellular matrix protein; and *MGA*, encoding a Max gene associated antagonist of Myc oncoproteins, all represent novel mutations with currently unknown functional significance.

Recurrence is an important criterion to discriminate driver from passenger mutations in the process of malignant transformation. Interestingly, mutations in *CCDC155, LLGL2, MGA* and *ZC3H18* were recently also reported by the TCGA consortium in a panel of AML patients (n=199), albeit at low frequencies. Because frequencies of specific mutations have been shown to vary with the natural history of AML, e.g., *de novo* versus secondary to MDS/MPN or different bone marrow failure syndromes[14, 41], it will be of interest to establish how often the newly identified genes are affected in distinct subtypes of secondary AML. Specifically, it will be important to determine whether *LLGL2*, *ZC3H18* or functionally related genes are more generally affected in bone marrow failure syndromes prone to progress to AML and to establish how these mutations contribute to malignant transformation in conjunction with cooperative gene defects.

ACKNOWLEDGEMENTS

This research was supported by the Center for Translational Molecular Medicine (CTMM), The Dutch Cancer Society "KWF kankerbestrijding" and the E-RARE project ELA2-CN.

REFERENCES

1. Dale DC, Bonilla MA, Davis MW, et al. A randomized controlled phase III trial of recombinant human granulocyte colony-stimulating factor (filgrastim) for treatment of severe chronic neutropenia. Blood 1993;81(10):2496-502.

2. Rosenberg PS, Zeidler C, Bolyard AA, et al. Stable long-term risk of leukaemia in patients with severe congenital neutropenia maintained on G-CSF therapy. Br J Haematol 2010;150(2):196-9.
3. Dale DC, Link DC. The many causes of severe congenital neutropenia. The New England journal of medicine 2009;360(1):3-5.
4. Dong F, Brynes RK, Tidow N, Welte K, Lowenberg B, Touw IP. Mutations in the gene for the granulocyte colony-stimulating-factor receptor in patients with acute myeloid leukemia preceded by severe congenital neutropenia. The New England journal of medicine 1995;333(8):487-93.
5. Dong F, Dale DC, Bonilla MA, et al. Mutations in the granulocyte colony-stimulating factor receptor gene in patients with severe congenital neutropenia. Leukemia 1997;11(1):120-5.
6. Dong F, Hoefsloot LH, Schelen AM, et al. Identification of a nonsense mutation in the granulocyte-colony-stimulating factor receptor in severe congenital neutropenia. Proc Natl Acad Sci U S A 1994;91(10):4480-4.
7. Germeshausen M, Ballmaier M, Welte K. Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis: Results of a long-term survey. Blood 2007;109(1):93-9.
8. Hermans MH, Antonissen C, Ward AC, Mayen AE, Ploemacher RE, Touw IP. Sustained receptor activation and hyperproliferation in response to granulocyte colony-stimulating factor (G-CSF) in mice with a severe congenital neutropenia/acute myeloid leukemia-derived mutation in the G-CSF receptor gene. J Exp Med 1999;189(4):683-92.
9. Touw IP, van de Geijn GJ. Granulocyte colony-stimulating factor and its receptor in normal myeloid cell development, leukemia and related blood cell disorders. Front Biosci 2007;12:800-15.
10. Dong F, van Buitenen C, Pouwels K, Hoefsloot LH, Lowenberg B, Touw IP. Distinct cytoplasmic regions of the human granulocyte colony-stimulating factor receptor involved in induction of proliferation and maturation. Mol Cell Biol 1993;13(12):7774-81.
11. Fukunaga R, Ishizaka-Ikeda E, Nagata S. Growth and differentiation signals mediated by different regions in the cytoplasmic domain of granulocyte colony-stimulating factor receptor. Cell 1993;74(6):1079-87.
12. Germeshausen M, Skokowa J, Ballmaier M, Zeidler C, Welte K. G-CSF receptor mutations in patients with congenital neutropenia. Curr Opin Hematol 2008;15(4):332-7.
13. Freedman MH, Bonilla MA, Fier C, et al. Myelodysplasia syndrome and acute myeloid leukemia in patients with congenital neutropenia receiving G-CSF therapy. Blood 2000;96(2):429-36.
14. Link DC, Kunter G, Kasai Y, et al. Distinct patterns of mutations occurring in de novo AML versus AML arising in the setting of severe congenital neutropenia. Blood 2007;110(5):1648-55.
15. Clinical picture: AML (Acute Myeloid Leukemia). Trial: HOVON 42 A AML / SAKK. (Accessed at http://www.hovon.nl/studies/studies-per-ziektebeeld/aml.html?action=showstudie&studie_id=2&categorie_id=4.)
16. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. The New England journal of medicine 2004;350(16):1617-28.
17. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29(1):308-11.
18. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011;29(1):24-6.

19. Hermans MH, van de Geijn GJ, Antonissen C, et al. Signaling mechanisms coupled to tyrosines in the granulocyte colony-stimulating factor receptor orchestrate G-CSF-induced expansion of myeloid progenitor cells. Blood 2003;101(7):2584-90.

20. Palande K, Roovers O, Gits J, et al. Peroxiredoxin-controlled G-CSF signalling at the endoplasmic reticulum-early endosome interface. Journal of cell science 2011;124(Pt 21):3695-705.

21. Carbuccia N, Trouplin V, Gelsi-Boyer V, et al. Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias. Leukemia 2010;24(2):469-73.

22. Taketani T, Taki T, Takita J, et al. AML1/RUNX1 mutations are infrequent, but related to AML-M0, acquired trisomy 21, and leukemic transformation in pediatric hematologic malignancies. Genes Chromosomes Cancer 2003;38(1):1-7.

23. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011.

24. Pasqualucci L, Dominguez-Sola D, Chiarenza A, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature 2011;471(7337):189-95.

25. Score J, Hidalgo-Curtis C, Jones AV, et al. Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. Blood 2011.

26. Ohshiro T, Yagami T, Zhang C, Matsuzaki F. Role of cortical tumour-suppressor proteins in asymmetric division of Drosophila neuroblast. Nature 2000;408(6812):593-6.

27. Gateff E. Malignant neoplasms of genetic origin in Drosophila melanogaster. Science (New York, NY 1978;200(4349):1448-59.

28. Peng CY, Manning L, Albertson R, Doe CQ. The tumour-suppressor genes lgl and dlg regulate basal protein targeting in Drosophila neuroblasts. Nature 2000;408(6812):596-600.

29. Ito T, Kwon HY, Zimdahl B, et al. Regulation of myeloid leukaemia by the cell-fate determinant Musashi. Nature 2010;466(7307):765-8.

30. Kharas MG, Lengner CJ, Al-Shahrour F, et al. Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. Nature medicine 2010;16(8):903-8.

31. Benz C, Mulindwa J, Ouna B, Clayton C. The Trypanosoma brucei zinc finger protein ZC3H18 is involved in differentiation. Molecular and biochemical parasitology 2011;177(2):148-51.

32. Mizuki M, Fenski R, Halfter H, et al. Flt3 mutations from patients with acute myeloid leukemia induce transformation of 32D cells mediated by the Ras and STAT5 pathways. Blood 2000;96(12):3907-14.

33. Shuai K, Halpern J, ten Hoeve J, Rao X, Sawyers CL. Constitutive activation of STAT5 by the BCR-ABL oncogene in chronic myelogenous leukemia. Oncogene 1996;13(2):247-54.

34. Forbes LV, Gale RE, Pizzey A, Pouwels K, Nathwani A, Linch DC. An activating mutation in the transmembrane domain of the granulocyte colony-stimulating factor receptor in patients with acute myeloid leukemia. Oncogene 2002;21(39):5981-9.

35. Plo I, Zhang Y, Le Couedic JP, et al. An activating mutation in the CSF3R gene induces a hereditary chronic neutrophilia. J Exp Med 2009;206(8):1701-7.

36. McLemore ML, Poursine-Laurent J, Link DC. Increased granulocyte colony-stimulating factor responsiveness but normal resting granulopoiesis in mice carrying a targeted granulocyte colony-stimulating factor receptor mutation derived from a patient with severe congenital neutropenia. J Clin Invest 1998;102(3):483-92.

37. Ward AC, Smith L, de Koning JP, van Aesch Y, Touw IP. Multiple signals mediate proliferation, differentiation, and survival from the granulocyte colony-stimulating factor receptor in myeloid 32D cells. J Biol Chem 1999;274(21):14956-62.

38. Pratcorona M, Abbas S, Sanders M, et al. Acquired mutations in ASXL1 in acute myeloid leukemia: prevalence and prognostic value. Haematologica 2011.

39. Ernst T, Chase AJ, Score J, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nature genetics 2010;42(8):722-6.

40. Nikoloski G, Langemeijer SM, Kuiper RP, et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nature genetics 2010;42(8):665-7.

41. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 2011;478(7367):64-9.

42. Layton JE, Hall NE, Connell F, Venhorst J, Treutlein HR. Identification of ligand-binding site III on the immunoglobulin-like domain of the granulocyte colony-stimulating factor receptor. J Biol Chem 2001;276(39):36779-87.

SUPPLEMENTARY MATERIALS AND METHODS

Sorting CD34+ leukemic blasts

Peripheral blood from the leukemic phase was thawed at 37°C and washed once with PBS (Gibco Invitrogen, San Diego, CA) with 2% FCS (PAA laboratories, Pasching, Austria), followed by isolation of CD34+ leukemic blasts.

FACS sorting: Peripheral blood cells were diluted to 5x10^6 cells per 50 µl (=100x10^9 cells/liter) and incubated for 1 hour at 4°C with CD34-Pe-Cy7 (BD, Franklin Lakes, NJ) in a final dilution of 1/50. Cells were washed, spun down (5', 1600 rpm = 550g) and resuspended in PBS (Gibco Invitrogen) with 2% FCS (PAA laboratories) and 7-AAD (Invitrogen, San Diego, CA) in a final dilution of 1/1000. Cells were filtered, followed by FACS sorting on the BD FACSAria Cell Sorter (BD); 7-AAD-/CD34+ cells were collected. Next, genomic DNA was isolated.

MACS: CD34+ cells for RNA isolation were isolated using the CD34 MicroBead Kit (Miltenyi, Bergisch Gladbach, Germany) according to the manufacturer's protocol. After this procedure, CD34+ purity was 97% of total 7-AAD- cells as determined using FACS. Next cDNA was isolated.

cDNA preparation

cDNA preparation was performed as previously described[1].

Genomic DNA isolation

Cells were resuspended in 300 µl SE buffer (31.25 mM EDTA, 93.75 mM NaCl). Next, 2.5 µl proteinase K (20µg/µL, Roche, Basel, Switzerland) and 15 µl 20% SDS (Biorad, Hercules, CA) were added followed by incubation o/n at 37°C. One-hundred µl saturated NaCl (Merck, Whitehouse Station, NJ) was added, samples were vortexed for 15", followed by 15' centrifugation at 10.000 rpm (9600g). This step was repeated until transparency of the supernatant was reached. DNA was precipitated in 800 µl 100% EtOH (Sigma-Aldrich, Zwijndrecht, The Netherlands), centrifuged at 10.000 rpm (9600g) for 15' and washed in 70% EtOH (Sigma-Aldrich). Pellets were air dried for 15' at RT, resuspended in TE-4 buffer (0.01 M Tris, 1mM EDTA) and incubated for 1 hour at 60°C.

Because of high RNA contamination in the fibroblasts DNA, this sample was purified with the ZR genomic DNA II kit (Zymo research, Orange, CA, USA) before whole exome sequencing according to the manufacturer's protocol. To increase the concentration, DNA was precipitated with 0.1M NaAc and 65% EtOH at -80°C for 30' followed by centrifugation (15', 4°C, 10.000 rpm = 9600 g). Pellets were washed with 70% EtOH, centrifuged (15', 4°C, 10.000 rpm = 9600 g), air dried for 15' at RT and resuspended in TE-4 buffer followed by incubation for 1 hour at 60°C.

Fibroblast culture

Bone marrow cells of the SCN phase, obtained 15 years before the leukemia, were thawed at 37°C, washed twice with and cultured in DMEM (Gibco Invitrogen) containing 10% FCS (PAA laboratories) and 10µg/ml penicillin/streptomycin (Gibco Invitrogen). Fibroblasts grew out after two weeks.

ELANE mutation analysis

PCR primers are listed in Table S8. Three amplicons, comprising the complete *ELANE* coding sequence, were generated using 50 ng of genomic DNA. Cycling conditions were 30" at 95° C, 30" at 62°C and 90" at 72°C for 35 cycles. Exon 1-2 and 3 were amplified using Q-solution (Qiagen, Hilden, Germany). Amplicons were purified using the Qiaquick PCR purification kit (Qiagen), followed by Sanger sequencing according to the manufacturer's protocol (Applied Biosystems, Foster City, CA, USA).

Allele specific analysis *CSF3R* mutations

For PCR amplification, cDNA from the CD34+ leukemic blasts was used. A cDNA fragment of the *CSF3R*, containing both mutation sites, was amplified using 5'-CACATCCACCTCATGGCTGC-3' and 5'-CAGAGGTTCTCATAGGACTTG-3'. Cycling conditions were: 30" at 95°C, 30" at 60°C and 1' at 72°C for 35 cycles. The amplified fragment was ligated into the pCR2.1 vector (TA-Cloning Kit, Invitrogen) according to the manufacturer's protocol and transformed into DH5α competent cells (Invitrogen). Plasmids from single bacterial colonies were sequenced with M13 primers (forward: 5'-GTAAAACGACGGCCAG-3' and reverse: 5'-CAGGAAACAGCTATGAC-3') for the presence of mutations *CSF3R*-d715 and *CSF3R*-T595I according to the manufacturer's protocol (Applied Biosystems). The presence of both mutations in the same sequence indicates that they were located on the same allele.

CSF3R expression constructs

A schematic overview of the different *CSF3R* constructs is given in Figure 3A. For amplification, cDNA from the CD34+ leukemic blasts was used. The *CSF3R* coding sequence was amplified in 2 parts which were ligated into the pCR2.1 vector (TA-Cloning Kit, Invitrogen) according to the manufacturer's protocol. The 5' part of the *CSF3R* was amplified using 5'-CTATGGCAAGGCTGGGAAACT-3' and 5'-CAGGTTAACAACAGCAGGAGG-3', cycling condition were: 30" at 95°C, 30" at 56°C (first 10 cycles) and 60°C (next 25 cycles) and 2,5' at 72°C for 35 cycles. The *CSF3R*-T595I mutation was introduced with the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent, Santa Clara, CA), according to the manufacturer's protocol, using the 5'-CCTCACCCTGATGATCTTGACCCCAGAGG-3' (sense) and 5'-CCTCTGGGGTCAAGATCATCAGGGTGAGG-3' (antisense) primers. The 3' part of the *CSF3R*, with and without the d715 mutation, was amplified with 5'-TTGTTAACCTGCCTCT-

GTGGAACTG-3' x 5'-GTTAACCCCTAGAAGCTCCCCAGCG-3'. Cycling condition were: 30" at 95°C, 30" at 60°C and 1' at 72°C for 35 cycles. Both parts were combined in pCR2.1 using the *Hpa*I-restriction sites (GTTAAC) introduced by PCR. Next, all 4 constructs were ligated into pBABE-puro and sequences were checked by restriction analysis and Sanger sequencing according to the manufacturer's protocol (Applied Biosystems).

Murine colony assays

Production of retroviruses: HEK293FT cells were cultured in DMEM (Gibco Invitrogen) containing 10% FCS (PAA laboratories) and 10µg/ml penicillin/streptomycin (Invitrogen). Transfection of HEK293FT with the 4 different *CSF3R* expression constructs in pBABE-puro was carried out using Fugene6 (Roche) according to the manufacturer's protocol.

Bone marrow of Csf3r deficient FVB/N mice[2] was isolated, followed by lysis of erythrocytes using lysing solution (Beckham Coulter, Brea, CA) and lineage depletion using the Mouse Hematopoietic Progenitor Cell Enrichment Set (BD) both according to the manufacturer's protocol. Cells were pre-stimulated for 2 days with murine SCF, TPO, FLT3 (All from Preprotech, Rocky Hill, NJ) and COS cell-derived IL-3. Next, these cells were retrovirally transduced as previously described[3], with a double round of transduction, using virus containing supernatant from the 293FT cells. After 48 hours, human CSF3R expression levels were analyzed using FACS (PE-conjugated mouse-α-human CD114, BD, Figure S5). Based on the FACS staining, similar amounts of CSF3R positive cells were plated in methocult (Stem Cell Technologies) without growth factors, or with different concentrations human G-CSF (Neupogen, Amgen, Thousand Oaks, CA) or murine GM-CSF (100 ng/ml, Preprotech). All colonies were grown in the presence of puromycin (1µg/ml, Sigma-Aldrich). After 7 days, colonies were counted. G-CSF-induced colony numbers were adjusted on the basis of the number of puromycin resistant GM-CSF-induced colonies to correct for variations in transduction efficiencies of the different CSF3R expression constructs.

Mutations in 199 AML genomes from The Cancer Genome Atlas

To determine if genes with mutations identified in the AML stage of the case studied here are recurrently mutated in AML, mutations found in a series of 199 AML samples whose exomes were sequenced by the TCGA AML project were downloaded from TCGA[4, 5] (and TCGA AML working group, unpublished). (Table S3).

SUPPLEMENTARY FIGURES

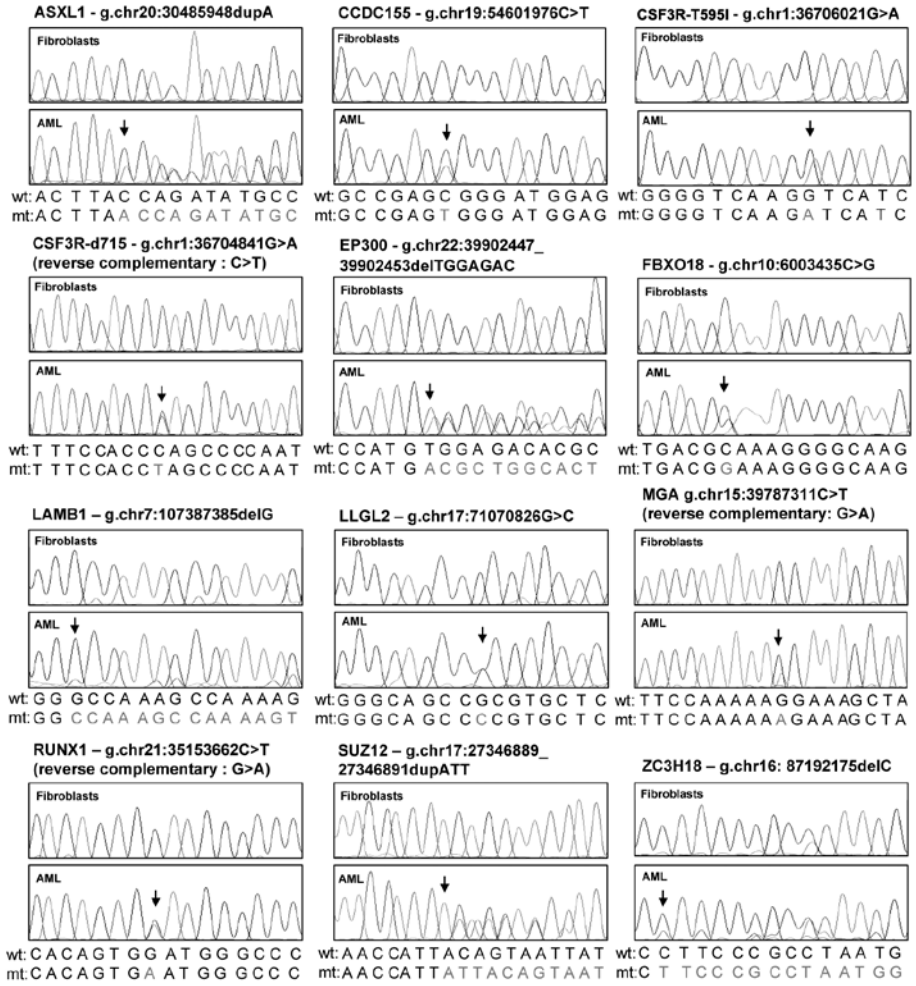


Figure S1. Sanger sequencing of mutations (gDNA). All 12 non-synonymous somatic mutations identified with whole exome sequencing were validated with Sanger sequencing, both in CD34+ leukemic blasts and fibroblasts (germ line control). For each mutation (indicated by arrows), the affected gene, its genomic position (hg18), the effect of the mutation on genomic DNA as well as the wild type (wt) and mutant (mt) sequence are depicted. See also Table 1.

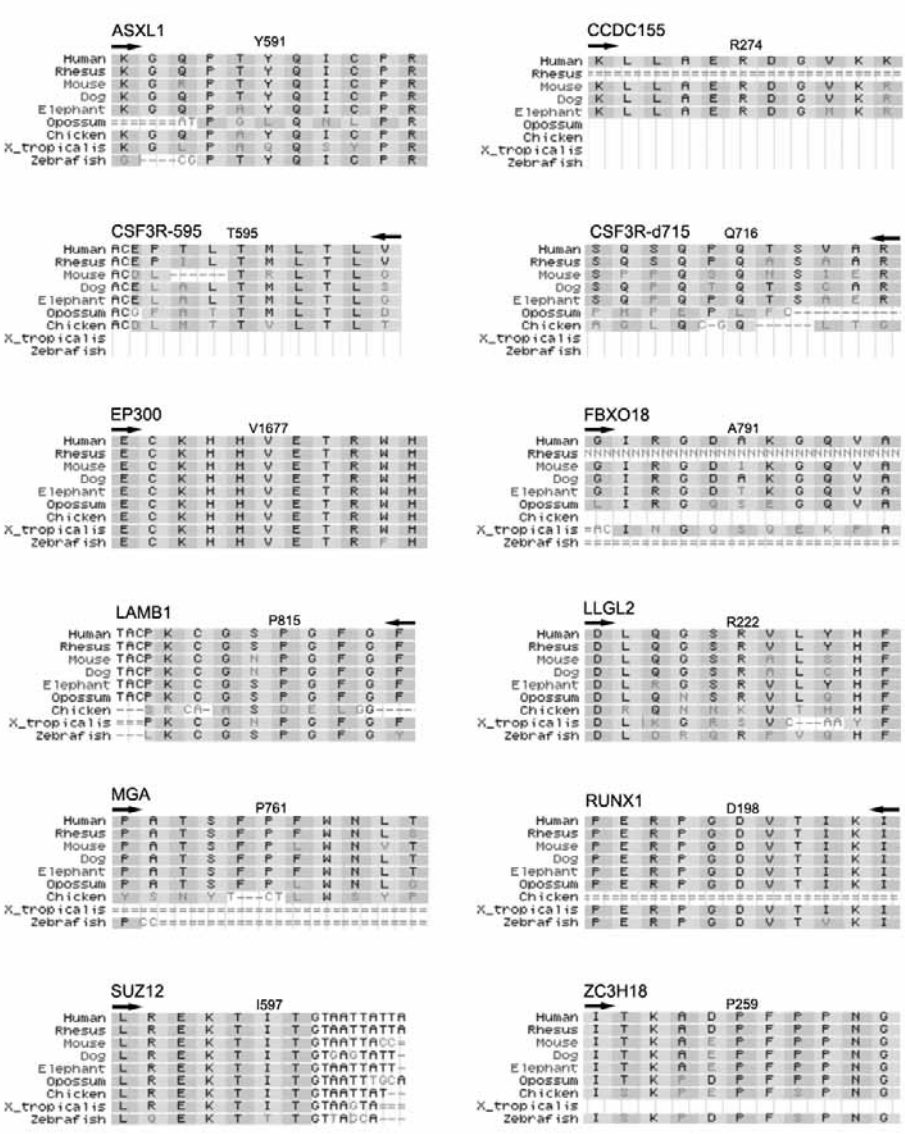


Figure S2. Evolutionary conservation of mutated amino acids. All 12 identified somatic mutations identified in the AML phase cause an amino acid change. The amino acids at these positions in other species are indicated. Arrows indicate the direction of translation. See also Table 1.

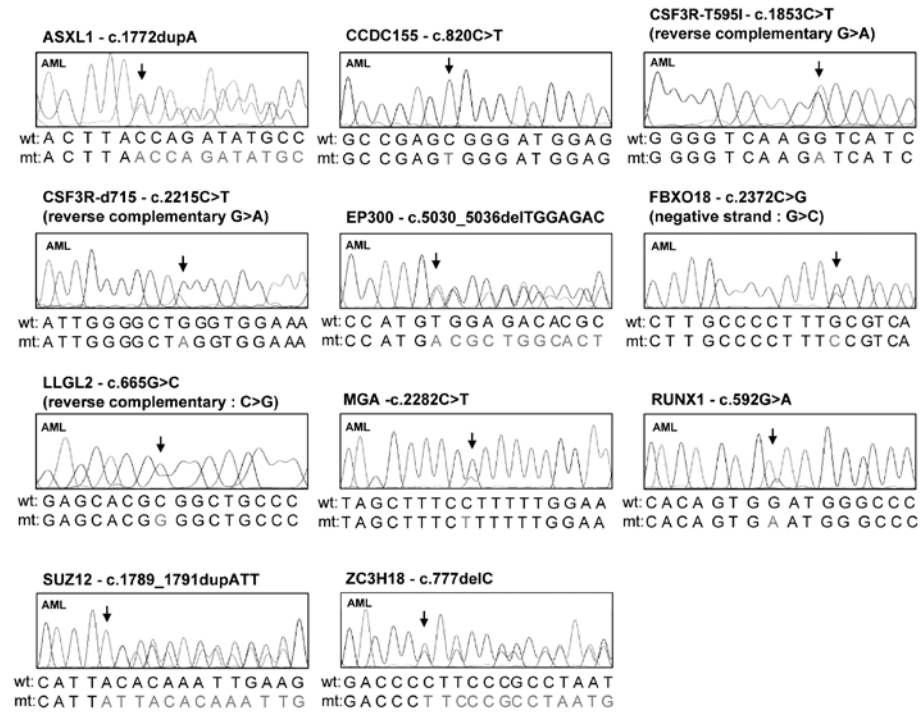


Figure S3. Sanger sequencing mutations (cDNA). All 12 identified non-synonymous somatic mutations were analysed by Sanger sequencing on cDNA of the CD34+ leukemic blasts. LAMB1 was not expressed within the leukemic blasts, for all other genes, except CCDC155 that only has mutant allele expression, both wild type and mutant transcripts were expressed. For each mutation, the affected gene (indicated by arrows), its position on cDNA, the effect of the mutation on cDNA as well as the wild type (wt) and mutant (mt) sequence are depicted. See also Table 1.

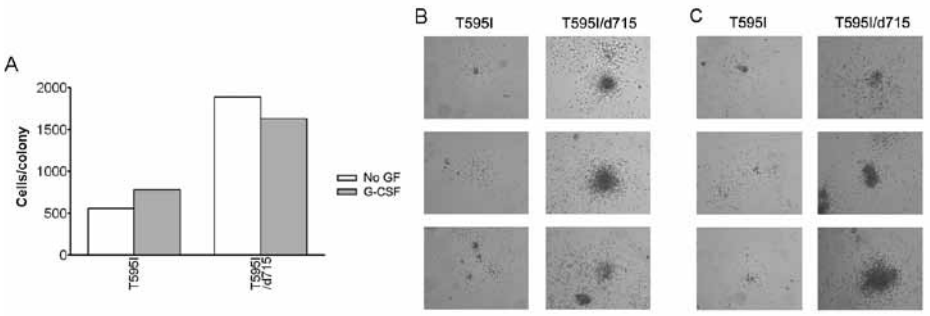


Figure S4. Colony size *CSF3R* mutants. In vitro colony growth of Csf3r deficient murine hematopoietic progenitor cells expressing different *CSF3R* mutants. See also Figure 3. (A) Cells per colony are indicated for colonies grown without growth factor (no GF) and with G-CSF. (B-C) Representative photomicrographs of G-CSF-induced myeloid colonies (B) and colonies grown without growth factors (C).

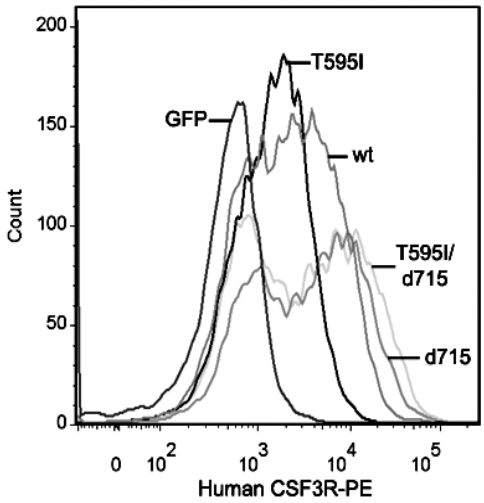


Figure S5. Membrane expression of *CSF3R* mutants. Murine lineage depleted Csf3r deficient bone marrow cells were retrovirally transduced with different *CSF3R* constructs (Figure 3A) or GFP (negative control). Histograms show membrane expression of the different expression constructs analysed by FACS.

SUPPLEMENTARY TABLES

Table S1. Primers for Sanger sequencing to validate somatic mutations.

Gene	gDNA/ cDNA	Forward/ Reverse/ Sequencing Primer	Sequence	Annealing temperature (°C)	Amplified or unamplified gDNA/cDNA
ASXL1	gDNA	Forward	**TGATTCTGTATGCCATGACC**	60	Amplified
		Reverse	CTCTCTATGGCAGTGGTGAC		
	cDNA	Forward	**CGTCAGTCCTTTCGTAACAC**	60	Amplified
		Reverse	CTCTCTATGGCAGTGGTGAC		
CCDC155	gDNA	Forward	**TCCCACTGGAATGGTTGGAG**	60	Amplified
		Reverse	AGCCTGACAGCTTGATGGAAG		
	cDNA (1st PCR)	Forward	CAGCTTCACAGCACCCAGCAGGC	60	Unamplified
		Reverse	GTCGTCACTCTGTATTCTTC		
	cDNA (nested PCR)	Forward	**GAGGAGCTGGAGGACCTGAAG**	60	Unamplified
		Reverse	CACACTCAAAGAGCTGCCGC		
CSF3R-T595I	gDNA	Forward	CACCAACAGTACAGTCCTCAC	58	Amplified
		Reverse	**CTCTCAAAGGGACTCACTTG**		
	cDNA	Forward	CACCAACAGTACAGTCCTCAC	60	Amplified
		Reverse	CAGAGGTTCTCATAGGACTTG		
		Sequencing	**GGAGTCACAGCGGAGATAGT**		
CSF3R-d715	gDNA	Forward	**CCATCACCAAGCTCACAGTG**	60	Amplified
		Reverse	CAGAGGTTCTCATAGGACTTG		
	cDNA	Forward	CACCAACAGTACAGTCCTCAC	60	Amplified
		Reverse	**CAGAGGTTCTCATAGGACTTG**		
EP300	gDNA	Forward	**CGTTTCTCACGCTGGCAAGG**	60	Amplified
		Reverse	CTCAGAACAACCCTGCGGAG		
	cDNA	Forward	**TGCCTCCCATTGTTGATCCT**	60	Amplified
		Reverse	GGTGATACACAAGTCATAATC		
FBXO18	gDNA	Forward	**CCATCAGAGTAAGGCTTTTAG**	60	Amplified
		Reverse	GGGTATGTTCGTCTACAACG		
	cDNA	Forward	GTCAACGCCCTGTTCACAGTG	60	Amplified
		Reverse	CTCCGTTCTTCCTCTGGCTG		
		Sequencing	**CAAATATCAATGATTCTGTC**		
LAMB1	gDNA	Forward	GAGAGCCCGATGATGTGCTGC	58	Amplified
		Reverse	**GGAGGGAAGCATCGGCTCTG**		
LLGL2	gDNA	Forward	**GTGTGTTCGAGATGGTGGAG**	60	Amplified
		Reverse	TAGCTGCCGTCAGAGTGACA		
	cDNA	Forward	**GTGTGTTCGAGATGGTGGAG**	60	Amplified
		Reverse	CATCGTGGATCACTGAGATG		

Table S1. Primers for Sanger sequencing to validate somatic mutations. (continued)

Gene	gDNA/ cDNA	Forward/ Reverse/ Sequencing Primer	Sequence	Annealing temperature (°C)	Amplified or unamplified gDNA/cDNA
MGA	gDNA	Forward	CAACAATGAGCATTGATCTT	56	Amplified
		Reverse	**GTTAAATGTCACCAATAAAGG**		
	cDNA	Forward	**TACATCCTGGTCTTCAAGAAG**	60	Amplified
		Reverse	GGTACACCACAGGAGATGTG		
RUNX1	gDNA	Forward	**AACTGGTAACTTGTGCTGAAG**	60	Amplified
		Reverse	CTGAGACATGGTCCCTGAGT		
	cDNA	Forward	**GAAATGCTACCGCAGCCATG**	60	Amplified
		Reverse	ATCTAGTTTCTGCCGATGTC		
SUZ12	gDNA	Forward	**AATCGTCTGTATTTCCATAGTGA**	60	Amplified
		Reverse	AACCTGTACATTTCCGGATTC		
	cDNA	Forward	**GAACAGCAAAGAACATATAGTAG**	60	Amplified
		Reverse	AAACAGCATACAGGCATGATTC		
ZC3H18	gDNA	Forward	**CTTGTGGACCCTTTCTGATC**	60	Amplified
		Reverse	ATTCAGACACAGCTCAGTTG		
	cDNA	Forward	**GTGACACCATTATCCACTCT**	60	Amplified
		Reverse	ACTCTCTGTTGGGGGCTCTG		

Primers in bold were used for sequencing. Due to scarcity of the material, amplified DNA was used; in case the amplicon generation was suboptimal on amplified material, unamplified material was used.

Table S2. Primers for amplicon-based deep sequencing.

Gene	Forward/ Reverse Primer	Sequence	MIDs	Annealing temperature (°C)
ASXL1	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-TGATTCTGTATGCCATGACC	1-4	60
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-CTCTCTATGGCAGTGGTGAC	5	
CCDC155	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-TCCCACTGGAATGGTTGGAG	1-4	60
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-AGCCTGACAGCTTGATGGAAG	5	
CSF3R-T595I	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-CACCAACAGTACAGTCCTCAC	1-4	58
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-CTCTCAAAGGGACTCACTTG	5	
CSF3R-d715	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-CCATCACCAAGCTCACAGTG	1-4	60
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-CAGAGGTTCTCATAGGACTTG	5	

Table S2. Primers for amplicon-based deep sequencing. (continued)

Gene	Forward/ Reverse Primer	Sequence	MIDs	Annealing temperature (°C)
EP300	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-GAGTTCTCTTCACTCCGAAG	1-4	60
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-ACTAGCCTCTTCAGCTCTGC	5	
FBXO18	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-CCATCAGAGTAAGGCTTTTAG	1-4	60
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-GGGTATGTTCGTCTACAACG	5	
LAMB1	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-GAGAGCCCGATGATGTGCTGC	1-4	58
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-GGAGGGAAGCATCGGCTCTG	5	
LLGL2	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-GTGTGTTCGAGATGGTGGAG	1-4	60
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-TAGCTGCCGTCAGAGTGACA	5	
MGA	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-CAACAATGAGCATTGATCTT	1-4	56
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-GTTAAATGTCACCAATAAAGG	5	
RUNX1	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-AACTGGTAACTTGTGCTGAAG	1-4	60
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-CTGAGACATGGTCCCTGAGT	5	
SUZ12	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-AATCGTCTGTATTTCCATAGTGA	1-4	60
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-AACCTGTACATTTCCGGATTC	5	
ZC3H18	**Forward**	CGTATCGCCTCCCTCGCGCCATCAG-MID-CTTGTGGACCCTTTCTGATC	1-4	60
	Reverse	CTATGCGCCTTGCCAGCCCGCTCAG-MID-CATCAGCGGGTGAGGTCCGAG	5	

All primers contain different multiplex identifiers (MIDs)[6] to distinguish different samples.

Table S3. Somatic mutations in The Cancer Genome Atlas (TCGA) AML dataset.

Gene	protein change (SCN/AML patient)	Nr. of patients with somatic mutation in the TCGA AML dataset	Patient number	Validation Status	Mutation type	Protein change
ASXL1	p.Y591*	5	LAML-AB-2807	Unknown	Frame_Shift_Del	p.G738fs
			LAML-AB-2821	Unknown	Nonsense_Mutation	p.Q733*
			LAML-AB-2864	Unknown	Frame_Shift_Ins	p.S921fs
			LAML-AB-2907	Valid	Nonsense_Mutation	p.R1415*
			LAML-AB-2927	Unknown	Splice_Site	e13-1
CCDC155	p.R274W	1	LAML-AB-2833	Unknown	Missense_Mutation	p.M286L
CSF3R	§p.T595I and §p.Q716*	2	LAML-AB-2808	Unknown	Missense_Mutation	***§p.T595I***
			LAML-AB-2854	Unknown	Frame_Shift_Ins	§p.G659fs
EP300	p.V1677Dfs*30	0	xx	xx	xx	xx
FBXO18	p.A791G	0	xx	xx	xx	xx
LAMB1	p.P815Pfs*65	0	xx	xx	xx	xx
LLGL2	p.R222P	1	LAML-AB-2891	Unknown	Frame_Shift_Del	p.S924fs
MGA	p.P761L	1	LAML-AB-2950	Unknown	Frame_Shift_Del	p.V2389fs
RUNX1	p.D198N	17	LAML-AB-2805	Unknown	Frame_Shift_Ins	p.A142fs
			LAML-AB-2805	Unknown	Nonsense_Mutation	p.R201*
			LAML-AB-2807	Unknown	In_Frame_Del	p.SG167in_frame_del
			LAML-AB-2807	Unknown	Missense_Mutation	p.G168R-p.S167R
			LAML-AB-2865	Unknown	Missense_Mutation	***p.D198N***
			LAML-AB-2890	Unknown	Nonsense_Mutation	p.R201*
			LAML-AB-2899	Unknown	Missense_Mutation	p.R162G
			LAML-AB-2907	Valid	Splice_Site	e5-2
			LAML-AB-2912	Unknown	Missense_Mutation	p.R162K
			LAML-AB-2916	Unknown	Frame_Shift_Ins	p.N96fs
			LAML-AB-2927	Unknown	Missense_Mutation	p.R162S
			LAML-AB-2927	Unknown	Nonsense_Mutation	p.R201*
			LAML-AB-2933	Unknown	Missense_Mutation	p.R162G
			LAML-AB-2936	Unknown	Nonsense_Mutation	p.R320*
			LAML-AB-2949	Unknown	Frame_Shift_Del	p.S314fs
			LAML-AB-2959	Unknown	Missense_Mutation	p.R201Q
			LAML-AB-2970	Valid	Missense_Mutation	p.P113L
			LAML-AB-2978	Valid	Nonsense_Mutation	p.R201*
			LAML-AB-2983	Valid	Missense_Mutation	p.A149P
			LAML-AB-3009	Valid	Frame_Shift_Ins	p.G164fs
SUZ12	p.597dupI	3	LAML-AB-2820	Unknown	Missense_Mutation	p.R103Q
			LAML-AB-2949	Unknown	Frame_Shift_Del	p.L308fs
			LAML-AB-3009	Valid	Missense_Mutation	p.G163D

Table S3. Somatic mutations in The Cancer Genome Atlas (TCGA) AML dataset. (continued)

Gene	protein change (SCN/AML patient)	Nr. of patients with somatic mutation in the TCGA AML dataset	Patient number	Validation Status	Mutation type	Protein change
ZC3H18	p.P259Pfs*15	2	LAML-AB-2868	Unknown	Missense_Mutation	p.R360G
			LAML-AB-2972	Valid	Nonsense_Mutation	p.R521*

TCGA harbours next-generation sequencing data of 199 AML patients. Listed are the genes found in the SCN/AML index case (See also Table 1) followed by the mutations in these genes found in the TCGA AML dataset. Marked in bold, italic are the exact mutations also found in the SCN/AML case. [$]Amino acid numbers based on earlier publications[7, 8]

Table S4. Amplicon-based deep sequencing of somatic non-synonymous mutations.

	Disease phase		
Gene	**SCN 15 years before AML**	**SCN 9 years before AML**	**AML**
ASXL1	0/1446 (0%)	0/1629 (0%)	166/354 (46,9%)
CCDC155	0/1527 (0%)	1/1441 (0,1%)	163/339(48,1%)
CSF3R-T595I	0/1397 (0%)	0/1500 (0%)	127/315 (40,3%)
CSF3R-d715	87/1375 (6,3%)	171/1563(10,9%)	173/351 (49,3%)
EP300	0/1769 (0%)	0/1918 (0%)	161/328(49,1%)
FBXO18	0/1604 (0%)	0/1648 (0%)	95/382 (24,9%)
LAMB1	3/1377 (0,2%)	4/1487 (0,3%)	111/348 (31,9%)
LLGL2	61/1396 (4,4%)	187/1690 (11,1%)	146/289 (50,5%)
MGA	0/1790 (0%)	0/1751 (0%)	54/328 (16,5%)
RUNX1	1/1862 (0,1%)	0/1809 (0%)	204/458 (44,5%)
SUZ12	0/833 (0%)	0/906 (0%)	63/135 (46,7%)
ZC3H18	64/1616 (4,0%)	204/1787 (11,4%)	189/359 (52,7%)

All 12 somatic non-synonymous mutations indentified in the leukemic blasts were analysed in SCN samples of the same patient using amplicon-based sequencing. SCN samples were obtained 9 and 15 years before the AML was diagnosed. For each mutation the number of mutated amplicons is presented as a fraction of total amplicons. Because all mutations are heterozygous, the number of cells carrying the mutations is estimated to be twice the number of mutated sequence reads.

Table S5. Mutation analysis of single myeloid colonies.

Colony	*CSF3R-truncating mutation*	*LLGL2*	*ZC3H18*	*ASXL1*	*CCDC155*	*CSF3R-T595I*	*EP300*	*FBXO18*	*LAMB1*	*MGA*	*SUZ12*	*RUNX1*
1	N	N	N									
2	ND	ND	N									
3	N*	N	N*									
4	N	N	N									
5	N	N	N									
6	N	N	N									
7	N*	N	N*									
8	N	N	N									
9	N	N	N									
10[$]	d715*	Y	N*	N	N	N	N	N	N	N	N	N
11	N	N	N									
12	N	N	N									
13	N	N	N									
14	N	N	N									
15	N*	N*	N*									
16	N*	N*	N*									
17	N	N	N									
18	N	N	N									
19	N*	N*	N*									
20	N	N	N									
21[#]	d725	N	N*	N	N	N	N	N	N	N	N	N
22[$]	d715	Y	N*	N	N	N	N	N	N	N	N	N
23	N	N	N									
24	N*	N*	N*									
25	N	N	N									
26[$]	d715*	Y*	N*	N	N	N	N	N	N	N	N	N
27	NA*	N*	N*									
28	N	N	N									
29[#]	d717	N	N	N	N	N	N	N	N	N	N	N
30	N	N	N									
31[#]	d717	N	N*	N	N	N	N	N	N	N	N	N
32[$]	d715	Y*	Y*	N	N	N	N	N	N	N	N	N
33	N	N	N									
34	N*	N*	N*									
35	N	N	N									
36	N	N	N									
37	N*	N	N*									

Table S5. Mutation analysis of single myeloid colonies. (continued)

Colony	*CSF3R-truncating mutation*	*LLGL2*	*ZC3H18*	*ASXL1*	*CCDC155*	*CSF3R-T595I*	*EP300*	*FBXO18*	*LAMB1*	*MGA*	*SUZ12*	*RUNX1*
38	N	N	N									
39	N*	N*	N*									
40	N*	N	N*									
41	N	N	N									
42	N	N	N									
43	N*	N*	N*									
44	N	N	N									
45	N*	N*	N*									
46	N*	N*	N*									
47$	d715*	Y	N*	N	N	N	N	N	N	N	N	N
48$	d715	Y	N	N	N	N	N	N	N	N	ND	N
49	N	N	N									
50	N	N	N									
51	N	N	N									
52$	d715	Y	N	N	N	N	N	N	N	N	N	N
53$	d715*	Y	N*	N	N	N	N	N	N	N	N	N
54	N*	N*	N*									
55$	d715	Y	N	N	N	N	N	N	N	N	N	N
56	N	N	N									
57	N	N	N									
58	N*	N	N*									
59	N	N	N									
60$	d715	Y	N	N	N	N	N	N	N	N	N	N
61	ND	ND	ND									
62$	d715*	Y	N*	N	N	N	N	N	N	N	N	N
63	d715*	NA*	NA*									
64	N	N	N									
65	N*	N*	N*									
66	N	N	N									
67	N	N	N									
68$	d715*	Y *	N*	ND	N	ND	N	N	N	ND	N	N
69	N*	NA	N*									
70	N*	NA	N									
71	N*	N	N*									
72	N	N	N									
73	N*	N	N*									
74	N*	N*	N*									
75	N*	N*	N*									

Table S5. Mutation analysis of single myeloid colonies. (continued)

Colony	*CSF3R-truncating mutation*	*LLGL2*	*ZC3H18*	*ASXL1*	*CCDC155*	*CSF3R-T595I*	*EP300*	*FBXO18*	*LAMB1*	*MGA*	*SUZ12*	*RUNX1*
76	N*	N*	N*									
77	N*	ND	N*									
78	N*	N	N*									
79	N	N	N									
80$	d715*	Y*	Y *	N	N	N	N	N	N	N	N	N
81	N	N	N									
82	N	N	N									
83	N*	N*	N*									
84$	d715*	Y	N*	N	N	N	N	N	N	N	N	N
85	N	N	N									
86	N*	N	N*									
87	N	N	N									
88	N	N	N									
89	N*	N*	N*									
90	N*	N	N*									
91	N	N	N									
92$	d715*	Y*	N*	N	N	N	N	N	N	N	N	N
93	d715*	NA	N*									
94	N*	N	N*									
95	N*	N	N*									
96	N	N	N									

Single myeloid colonies were grown from the SCN sample obtained 15 years before AML development. These were analysed for the presence of the 3 mutations observed in the SCN phase by amplicon based deep-sequencing (Table S4 and S6), i.e., *CSF3R* truncating mutations and mutations in *LLGL2* and *ZC3H18*. All colonies positive for either of these were subsequently analysed for the presence of the other somatic mutations. *d715* = *CSF3R* truncation at position 715, *d717* = *CSF3R* truncation at position 717, *d725* = *CSF3R* truncation at position 725, Y = positive for mutation, N = negative for mutation, ND = not determined because amplification or sequence reaction did not work, NA = non-assigned due to unclear sequence, NA* = non-assigned due to discrepancy between 2 independent amplification and sequence runs, * = status could be determined in 2 independent amplification and sequence runs. $colonies positive for *CSF3R-d715* and *LLGL2* mutation, #colonies positive for other *CSF3R* truncating mutations than d715, grey = colonies for which the status of *ZC3H18*, *LLGL2* and/or *CSF3R-d715* mutations could not be assigned.

Table S6. Amplicon based deep-sequencing to quantify *CSF3R* truncating mutations.

	Disease phase		
Gene	**SCN 15 years before AML**	**SCN 9 years before AML**	**AML**
CSF3R-d715	87/1375 (6,3%)	171/1563 (10,9%)	173/351 (49,3%)
CSF3R-d717	12/1375 (0,9%)	9/1563 (0,6%)	0/351 (0%)
CSF3R-d719	1/1375 (0,1%)	0/1563 (0%)	0/351 (0%)
CSF3R-d725	12/1375 (0,9%)	64/1563 (4,1%)	0/351 (0%)
CSF3R-d728	0/1375 (0%)	0/1563 (0%)	0/351 (0%)
CSF3R-d730	11/1375 (0,8%)	40/1563 (2,6%)	0/351 (0%)

Amplicon based deep-sequencing was employed to analyse the frequency of the most frequent *CSF3R* truncating mutations found in SCN[9] in the SCN and leukemic phase. The different mutations introduce a stopcodon at amino acid position 716 (d715), 718 (d717), 720 (d719), 726 (d725), 729 (d728), 731 (d730). For each mutation the number of mutated amplicons is presented as a fraction of total amplicons. As all mutations are considered to be heterozygous, the number of cells carrying the mutations is estimated to be twice the number of mutated amplicons.

Table S7. Colony numbers murine colony assay *CSF3R* mutants.

		Number of colonies/well				Number of colonies/well (corrected for number of GM-CSF colonies)			
culture condition	**value**	**wt**	**d715**	**T595I**	**T595I/d715**	**wt**	**d715**	**T595I**	**T595I/d715**
no growth factor	mean	0	0	227	217	0,00	0,00	0,20	0,20
	sd	0	0	22,5	15,3	0,000	0,000	0,023	0,014
G-CSF (5ng/ml)	mean	90	230	238	262	0,11	0,31	0,21	0,24
	sd	17,3	50,7	20,2	41,6	0,021	0,092	0,021	0,038
G-CSF (10ng/ml)	mean	95	255	205	263	0,11	0,34	0,18	0,24
	sd	18	21,8	13,2	28,4	0,022	0,075	0,016	0,026
G-CSF (50ng/ml)	mean	88	195	250	287	0,11	0,26	0,22	0,26
	sd	12,6	39,1	21,8	47,5	0,016	0,074	0,023	0,043
G-CSF (100ng/ml)	mean	118	205	170	313	0,14	0,27	0,15	0,28
	sd	16,1	21,8	5,0	31,8	0,020	0,063	0,010	0,029
GM-CSF (100ng/ml)	mean	830	750	1140	1105	xx	xx	xx	xx
	sd	37,7	152,2	65,4	22,9	xx	xx	xx	xx

Murine lineage depleted Csf3r deficient bone marrow cells were retrovirally transduced with different *CSF3R* constructs (Figure 3A). Myeloid colonies were grown under different growth factor conditions. All data are average values of 3 wells. GM-CSF colony growth was performed in the presence of puromycin, and was used to correct for the transduction efficiencies of the different *CSF3R* constructs. The delta method was used to calculate the standard deviation of the ratios. See also Figure 3B. sd = standard deviation.

Table S8. Primers to identify *ELANE* mutation by Sanger sequencing.

Exon	Forward/Reverse/ Sequencing Primer	Sequence	Annealing temperature (°C)
Exon 1-2	Forward	GAGCCAATCCAGCGTCTTGTC	62 + Q-solution
	Reverse	CCTGAGGGCGAAGGTGCTCG	
Exon 3	Forward	TGGCGAATGTGTGAGTAGCCG	62 + Q-solution
	Reverse	CCACGATGCCACCCCAGG	
Exon 4-5	Forward	TCCAACGCCCTGAGCCTTG	62
	Reverse	CTGGTCTTGAACTCCTGAGC	
Exon 1	Sequencing	GAGCCAATCCAGCGTCTTGTC	xx
Exon 2	Sequencing	TGCCTCTCCGTGCCTCAGT	xx
Exon 3	Sequencing	CCACGATGCCACCCCAGG	xx
Exon 4	Sequencing	TCCAACGCCCTGAGCCTTG	xx
Exon 5	Sequencing	GCAGCAACAGGCACCGTGGCT	xx

REFERENCES SUPPLEMENTARY DATA

1. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. The New England journal of medicine 2004;350(16):1617-28.
2. Hermans MH, van de Geijn GJ, Antonissen C, et al. Signaling mechanisms coupled to tyrosines in the granulocyte colony-stimulating factor receptor orchestrate G-CSF-induced expansion of myeloid progenitor cells. Blood 2003;101(7):2584-90.
3. Aarts LH, Roovers O, Ward AC, Touw IP. Receptor activation and 2 distinct COOH-terminal motifs control G-CSF receptor distribution and internalization kinetics. Blood 2004;103(2):571-9.
4. http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp.
5. Ding L, Ley TJ, Larson DE, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature 2012;481(7382):506-10.
6. http://www.eurofinsdna.com/fileadmin/PDFs/product_services/TB-0905_Technical_Bulletin_Roche_MID_Adaptors.pdf.
7. Dong F, Brynes RK, Tidow N, Welte K, Lowenberg B, Touw IP. Mutations in the gene for the granulocyte colony-stimulating-factor receptor in patients with acute myeloid leukemia preceded by severe congenital neutropenia. The New England journal of medicine 1995;333(8):487-93.
8. Dong F, Hoefsloot LH, Schelen AM, et al. Identification of a nonsense mutation in the granulocyte-colony-stimulating factor receptor in severe congenital neutropenia. Proc Natl Acad Sci U S A 1994;91(10):4480-4.
9. Germeshausen M, Ballmaier M, Welte K. Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis: Results of a long-term survey. Blood 2007;109(1):93-9.

CHAPTER 4

PREVALENCE AND FUNCTIONAL CHARACTERISATION OF THE AUTO-ACTIVATING COLONY STIMULATING FACTOR 3 MUTATION *CSF3R-T595I* IN AML

Renée Beekman[1], Marijke Valkhof[1], Onno Roovers[1], Tanja Jevdjovic[1], Paulette van Strien[1], Peter.J.M. Valk[1], Ivo P. Touw[1]

[1]Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

Manuscript in preparation

ABSTRACT

Both granulocyte-colony stimulating factor (G-CSF) and its receptor (colony stimulating 3 receptor, CSF3R) are essential for neutrophil production. Upon ligand binding, CSF3R activates different signal transduction pathways to exert its function. Neutrophil production is disturbed in severe congenital neutropenia (SCN) patients that furthermore have a high tendency to develop acute myeloid leukemia (AML). The neutropenia in SCN patients can be successfully treated with G-CSF administration. In time however, these patients may acquire truncating mutations in *CSF3R*, with an incidence of up to 80% in SCN patients who have progressed to AML. In a previous study we have reported the accumulation of 2 *CSF3R* mutations on the same allele during leukemogenesis in a SCN patient. The first mutation, present already 15 years before the AML became overt, was a truncating mutation that has been shown previously to induce proliferation, while hampering differentiation. The second mutation, *CSF3R-T595I*, located in the extracellular domain of the receptor, was only present in the AML phase and induced G-CSF independent myeloid colony growth. Here, we show that the prevalence of the *CSF3R-T595I* mutation in *de novo* AML is low (5/1446 = 0.3%). In addition, we found that 2 of the *de novo* AML patients carried a mutation substituting a threonine at amino acid position 617 to either an isoleucine or an asparagine. None of these *CSF3R* changes (*T595I, T617I* or *T617N*) coincide with the presence of *CSF3R* truncating mutations. Furthermore, we show that the CSF3R-T595I mutant induces G-CSF independent nuclear translocation of STATs (signal transducer and activator of transcription), indicating G-CSF independent stimulation of the receptor which can explain the observed G-CSF independent myeloid colony growth.

INTRODUCTION

G-CSF, also known as colony stimulating factor 3 (CSF3), is the main growth factor for neutrophil production both under physiological conditions and during stress myelopoieisis[1-3]. G-CSF and its receptor, CSF3R, are essential for myelopoiesis, as mice lacking either of these are severely neutropenic[1, 2]. CSF3R is a member of the cytokine receptor I superfamily and exerts its function via the Janus kinase (JAK)/ signal transducer and activator of transcription (STAT), RAS/MAF/MAPkinase and PKB/AKT pathways[4]. CSF3R activation is transient and its inactivation is regulated by various negative regulators, while receptor internalisation and degradation also play an important role in this process[4]. A tight balance between activation and signal attenuation of CSF3R is essential for neutrophil production[5, 6].

G-CSF is successfully used to treat neutropenia in SCN, a rare bone marrow failure syndrome characterized by highly reduced neutrophil counts[7]. Constitutive mutations in *CSF3R* are infrequently reported to underlie SCN refractory to G-CSF treatment[8, 9]. In contrast, acquisition of truncating *CSF3R* mutations in a small intracellular glutamine

stretch of the receptor is a common phenomenon in SCN patients[10-13]. Extensive analyses in different models has shown that expression of these truncated receptors leads to an increased proliferative response, while they are hampered in transducing required differentiation signals[4, 10-12, 14-16]. *CSF3R* truncating mutations are observed in one third of SCN patients, with an increasing incidence to approximately 80% in SCN patients that have progressed to AML, suggesting a role for *CSF3R* mutations in leukemic progression[11, 13, 17]. In other AML patients however, *CSF3R* mutations are rare[18-20].

We previously have identified a new extracellular *CSF3R* mutation (*T595I*) in the leukemic blasts of a SCN patient that progressed to AML[21]. Initial studies have shown that this mutation, located on the *CSF3R* allele that already carried an intracellular *CSF3R-d715* truncating mutation, confers growth factor independent myeloid colony growth[21]. Here, we have investigated the occurrence of the *CSF3R-T595I* mutation in a large cohort of 1446 *de novo* AML patients and we show that its prevalence is 0.3%. In addition, 2 *de novo* AML patients carried a mutation in *CSF3R,* substituting the threonine at amino acid position 617 for an isoleucine or an asparagine. The latter one has been reported to induce G-CSF independent proliferation and differentiation[19, 22]. For all but one of the reported *CSF3R* mutations the coincidence with *CSF3R* truncating mutations could be investigated, revealing that they do not coincide in this *de novo* AML cohort. Additionally, we show that expression of mutant CSF3R-T595I induces nuclear translocation of STATs independent of ligand binding, which indicates that the mutant receptor is activated in the absence of G-CSF.

MATERIALS AND METHODS

Patient cell samples

Ficoll-gradient separated bone marrow cells from leukemic blasts of AML samples were used. All cell samples were obtained, frozen and anonymously stored according to established procedures for viable cell cryopreservation as previously described[23]. The study was performed under the permission of the Institutional Review Board of the Erasmus MC, registration number MEC-2008-387.

cDNA preparation

cDNA preparation was performed as previously described[23].

CSF3R-T595I mutation analysis

Amplicons were generated using the 5′-GCTCAGAACCAGTCCTTCTC-3′ and 5′-CTGCTGTGAGCTGGGTCTG-3′ primer. Cycling conditions were 30″ at 95°C, 30″ at 60°C and 30″ at 72°C for 35 cycles. Amplicons were analysed on a denaturing high-performance liquid chromatography (dHPLC) using a WAVE device (Transgenomics, Omaha, NE, USA) at a temperature of 63.1°C. Amplicons showing an aberrant dHPLC pattern compared to

the wildtype control were analysed by Sanger sequencing using the 5'-GCTCAGAAC-CAGTCCTTCTC-3' primer according to the manufacturer's protocol (Applied Biosystems, Foster City, CA, USA). Patients carrying the *T595I, T617I and T617N* mutations were subsequently analysed for the presence of *CSF3R* truncating mutations (at amino acid position 715-730) by Sanger sequencing as previously described[21].

Expression constructs

CSF3R constructs. *CSF3R* expressing constructs in pBABE-puro were generated as previously described[21]. From pBABE-puro these constructs were ligated into pLNCX2. Four different construct were used: wildtype, d715 (containing a truncating mutation at amino acid position 715), T595I (containing a missense mutation at amino acid position 595, converting a threonine into an isoleucine) and d715/T595I (containing both mutations described above). The T595V mutation was introduced in pBABE-puro at the position of the T595I mutation using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent, Santa Clara, CA) according to the manufacturer's protocol, using the 5'-GTCCTCACCCT-GATGGTCTTGACCCCAGAGG-3' (sense) and 5'-CCTCTGGGGTCAAGACCATCAGGGTGAG-GAC-3' (antisense) primers.

STAT constructs. The STAT3-YFP construct in pCDNA5 was a kind gift from Prof. Dr. G. Müller-Newen[24]. GFP was amplified from a GFP containing vector (Clontech Laboratories, Mountain View, CA) and YFP was exchanged for GFP by restriction and ligation in pCDNA5. The STAT5-GFP construct was amplified from the pSG5 vector and ligated into pCDNA5.

Generation stably expressing CSF3R mutant HeLa cell lines

Production of retroviruses: Phoenix A cells were cultured in DMEM (Gibco, Life Technologies Corporation, Carlsbad, CA) containing 10% FCS (PAA laboratories, Pasching, Austria) and 10µg/ml penicillin/streptomycin (Gibco, Life Technologies Corporation, Carlsbad, CA). Transfection of Phoenix A cells with the 4 different *CSF3R* expression constructs in pLNCX2 was carried out using TransIT-LT1 Transfection Reagent (Mirus Bio Corporation, Madison, WI) according to the manufacturer's protocol.

Parental HeLa cells were grown in DMEM (Gibco, Life Technologies Corporation, Carlsbad, CA) containing 10% FCS (PAA laboratories, Pasching, Austria) and 10µg/ml penicillin/streptomycin (Gibco, Life Technologies Corporation, Carlsbad, CA). HeLa cells were retrovirally transduced as previously described[25], with a double round of transduction, using virus containing supernatant from the Phoenix A cells. Fourty-eight hours post transduction cells were put on neomycin selection for 4 weeks (G418, 1 mg/ml, Invitrogen, Life Technologies Corporation, Carlsbad, CA). Transduction efficiency was determined by FACS analysis measuring human CSF3R expression (PE-conjugated mouse-α-human CD114, BD, Franklin Lakes, NJ).

Murine colony assays

Murine colony assays with the *CSF3R-T595V* mutant were performed as previously described[21].

Imaging

HeLa cells stably expressing the different CSF3R mutants were transiently transfected with STAT3-GFP or STAT5-GFP expressing constructs using TransIT-LT1 Transfection Reagent (Mirus Bio Corporation, Madison, WI) according to the manufacturer's protocol. The HeLa cells were grown in DMEM (Gibco, Life Technologies Corporation, Carlsbad, CA) containing 10% FCS (PAA laboratories, Pasching, Austria) and 10µg/ml penicillin/streptomycin (Gibco, Life Technologies Corporation, Carlsbad, CA). Prior to fixation, cells were starved by culturing in plain DMEM medium (Gibco, Life Technologies Corporation, Carlsbad, CA) for 4 hours, either or not followed by 30 minutes stimulation with 100ng/ml human G-CSF (Neupogen, Amgen, Thousand Oaks, CA). Cells where fixed with 3% paraformaldehyde (Polysciences, Warrington, PA) on ice, and subsequently stained with mouse anti-human CSF3R CD114 (Becton Dickinson, Franklin Lakes, NJ) and donkey anti-mouse Cy5 (Jackson ImmunoResearch, Newmarket, UK) in 0.05% Saponin (Sigma-Aldrich, Zwijndrecht, The Netherlands) in PBS (Gibco Invitrogen, San Diego, CA). Cells where imaged with a Leica SP5 CLSM equipped with Argon/HeNe lasers with a 63x Planapochromat oil immersion objective (Leica Microsystems, Wetzlar, Germany). Between 50 and 60 stacks where scanned per cell, deconvolved and visualized in Huygens Professional 4.1 (Scientific Volume Imaging, Hilversum, the Netherlands). Next, all scans were background subtracted, and the nuclear/cytoplasmic GFP intensity ratios were determined in ImageJ[26] by dividing the mean grey scale value of a circular ROI in the nucleus, with the mean grey scale value of the same sized circular ROI in the cytoplasm.

RESULTS

The *CSF3R-T595I* mutation has a low prevalence in de novo AML and does not coincide with *CSF3R* truncating mutations

We previously have identified a new extracellular *CSF3R* mutation, *CSF3R-T595I,* in a SCN patient who progressed to AML[21]. We have investigated the prevalence of this mutation in a larger cohort of 1446 *de novo* AML patients. The molecular and cytogenetic characteristics of this cohort are listed in Table 1. Five of the investigated *de novo* AML patients (0.3%) carried the same *CSF3R-T595I* mutation. Clinical, molecular and cytogenetic characteristics of these patients are listed in Table 2. Furthermore we identified 2 patients with a *CSF3R* mutation substituting a threonine at amino acid position 617 to an isoleucine (*T617I*) or an asparagine (*T617N*), of which the latter substitution has been shown to induce G-CSF independent proliferation and differentiation[19, 22]. In the previously investigated SCN patient, we have shown that the *CSF3R-T595I* mutation coincided

Table 1. Cytogenetic and molecular characteristics of the 1446 *de novo* AML patients.

Cytogenetic abnormality*	Number of patients
inv(16)/t(16;16)	54
t(8;21)	61
t(15;17)	22
+8	89
abn5/5q	72
abn7/7q	98
abn3q	40
t(6;9)	6
t(9;22)	3
t(11q23)	58
CN**	459
Complex (>3 abn.)	136
Other	116
ND	419
Molecular abnormalities*	**Number of patients**
NPM1	293
FLT3ITD	261
FLT3TKD	59
N-RAS	81
K-RAS	5
CEBPA	69

Abn = abnormality, CN = cytogenetic normal, ND = non determined, ITD = internal tandem duplication, TKD = tyrosine kinase domain. *All patients with the specific alteration were taken into account, irrespective of the presence of other abnormalities, **contains patients with –X/-Y.

Table 2. Clinical and cytogenetic characteristics *de novo* AML patients carrying *CSF3R* mutations.

Case	WBC (x10^9/l)	Age at diagnosis (yrs)	FAB classification	Blast percentage in BM	Cytogenetic abnormalities	Molecular abnormalities*	Identified *CSF3R* mutation	*CSF3R* truncating mutations§
2187	48.9	50	1	93	CN**	negative	*T595I*	No
3491	38	72	1	95	46,XY [22]	negative	*T595I*	No
14331	32.1	57	2	37	46,XY[20]	negative‡	*T595I*	UD
16252	NA	NA	NA	NA	NA	NA	*T595I*	No
19206	NA	NA	NA	NA	NA	NA	*T617I*	No
21997	NA	NA	NA	NA	NA	NA	*T595I*	No
21999	20.7	39	2	58	46, XX [23]	FLT3ITD$^{\#,\ddagger}$	*T617N*	No

WBC = white blood cell count, BM = bone marrow, CN = cytogenetic normal, NA = non-assigned. UD = undetermined due to lack of material. *Considering mutations in *NPM1, FLT3ITD, FLT3TKD, N-RAS, K-RAS* and *CEBPA*. **may harbour –X/-Y. $^{\#}$*FLT3TKD* and *N-RAS* mutations were not determined. ‡*K-RAS* mutations were not determined. §Investigated mutations are truncating mutations at amino acid positions 715, 717, 719, 725, 728 and 730.

with a *CSF3R* truncating mutation on the same allele[21]. The coincidence of the newly identified *CSF3R* mutations with *CSF3R* truncating mutations could be investigated in all but one patient; the combination of *CSF3R* mutations with either the *T595I, the T617I* or the *T617N* mutations was not observed in our study (Table 2).

The change in amino acid polarity of the CSF3R-T595I mutant induces G-CSF independent colony growth

Previously, we have shown that the CSF3R-T595I mutation induces G-CSF independent growth in *in vitro* murine colony assays[21]. The substitution of a threonine to an isoleucine at amino acid position 595 results in a structural and polarity change in the extracellular membrane proximal region of the receptor. Expression of CSF3R-T595V, a different mutant which causes a similar polar but no structural amino acid change, similarly induces growth factor independent growth (Figure 1). Hence, the change in polarity of the CSF3R-T595I mutant rather than the structural variation causes the observed phenotype. Polarity changes are associated with spontaneous conformation changes and dimerisation properties of receptors, rather than with altering docking sites for other molecules. This highly suggests that the G-CSF independent activation of the T595I and the T595V mutant can be explained by their spontaneous dimerisation.

Expression of the CSF3R-T595I mutant induces nuclear translocation of STATs

An important signal transduction pathway activated by CSF3R is the JAK/STAT pathway; upon activation of CSF3R, STATs are phosphorylated and subsequently translocated to the nucleus where they regulate gene expression[27, 28]. To study the effect of the CSF3R-

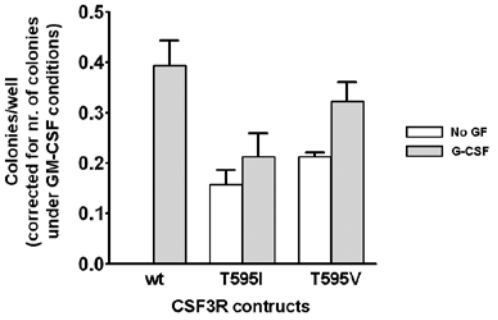


Figure 1. Functional analysis of mutant CSF3R-T595V in myeloid progenitor cell assay. In vitro colony growth of Csf3r deficient murine hematopoietic progenitor cells expressing the wild type human CSF3R receptor (wt), the T595I and the T595V mutant, substituting a threonine at amino acid position 595 for an isoleucine or a valine respectively. Colonies were grown in the presence of puromycin, either without growth factor (no GF) or with G-CSF. The transduction efficiency was corrected for by dividing the number of colonies by the number of GM-CSF induced colonies under puromycin selection as the CSF3R constructs confer puromycin resistance, but do not affect GM-CSF-induced colony growth.

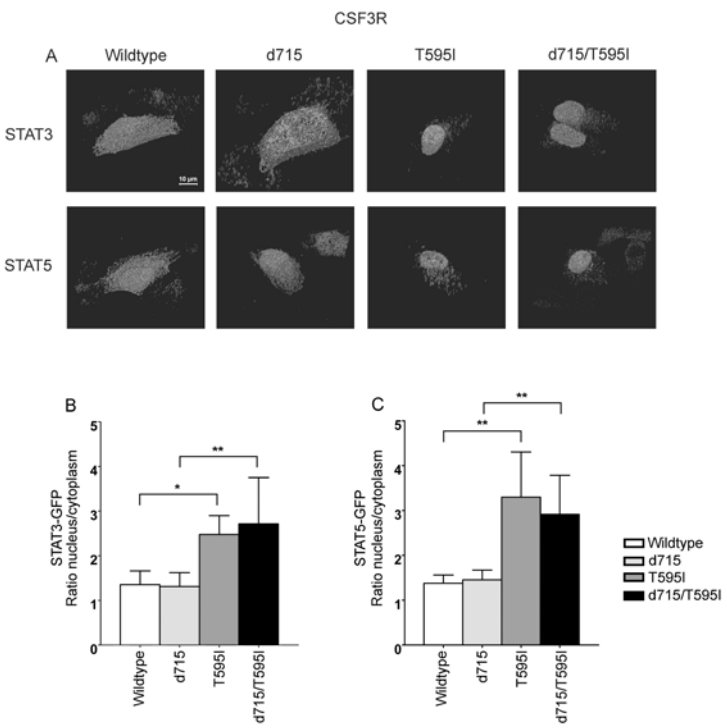


Figure 2. Expression of mutant receptor CSF3R-T595I translocates STATs to the nucleus. A. STAT3/STAT5 (GFP-tagged, light grey) distribution in HeLa cells expressing different CSF3R mutants (dark grey speckles in the cytoplasm) in a non-stimulated state. B and C. Nuclear/cytoplasmic ratio of GFP intensity for STAT3 (B) and STAT5 (C). Number of cells analysed for STAT3-GFP: Wildtype and T595I (n=11), d715 (n=9), d715/T595I (n=13) and for STAT5-GFP: Wildtype and d715 (both n=8), T595I and d715/T595I (both n=9). Error bars indicate standard deviations. Significant differences were calculated using a student's t-test, *p-value <0.0001, assuming equal variances, **p-value <0.001, assuming unequal variances.

T595I mutant on JAK/STAT activation, HeLa cells that stably expressed different CSF3R mutants, were transiently transfected with GFP-tagged STAT3 or STAT5, followed by monitoring their nuclear translocation. In cells expressing the CSF3R-T595I mutants an increase in nuclear localisation of both STAT3 and STAT5 was observed in a non-stimulated state, while this was not seen in cells expressing the CSF3R constructs without this mutation (Figure 2). In these latter cells we observed nuclear translocation of STAT3 and STAT5 upon ligand binding (data not shown), indicating that the expressed CSF3R constructs were functional. The observations described above were independent of the presence of the *CSF3R-d715* truncating mutation in the same construct (Figure 2).

DISCUSSION

In this study, we have investigated the prevalence of a previously identified[21] auto-activating mutation, *CSF3R-T595I,* in the G-CSF receptor in a large cohort of *de novo* AML patients. Simultaneously, we further analysed the functional characteristics of this *CSF3R* mutant. Previously, a different auto-activating *CSF3R* mutation (*T617N*), located in its transmembrane domain, was shown to be present in 2 out of 555 AML patients (0.4%)[19]. We observed a similar prevalence for the *CSF3R-T595I* mutation in the investigated AML cohort, while we also identified a patient with a similar *T617N* mutation and one with a *T617I* mutation. These results indicate that auto-activating mutations in *CSF3R* in *de novo* AML are rare. Additionally, we show that the *CSF3R-T595I, -T617N* or *-T617I* mutations do not coincide with *CSF3R* truncating mutations, in contrast with the previously investigated SCN patient[21]. The dual *CSF3R* mutation found in the previous patient induces both G-CSF independent proliferation, due to the *CSF3R-T595I* mutation, and a block in differentiation, which can be attributed to the *CSF3R* truncating mutation. In the *de novo* AML patients carrying the *CSF3R-T595I, -T617N* or *-T617I* mutations as identified in this study, the differentiation block is likely caused by mutations in different genes. As far as we have a clinical history of the investigated patients, we do not have indications that the patients carrying the different *CSF3R* mutations were treated with G-CSF in the past. This indicates that the *CSF3R-T595I, -T617N* or *-T617I* mutations likely occur without selective pressure of G-CSF, which could similarly explain why these patients do not carry *CSF3R* truncating mutations.

Besides determining the prevalence in *de novo* AML, we have shown that the *CSF3R-T595I* mutant activates the JAK/STAT pathway independent of G-CSF stimulation. This explains the observation that this mutant induces growth factor independent myeloid colony formation as previously described[21]. As this mutant is ligand independent, it does not have to reach the plasma membrane to be activated by G-CSF. Hence, it would of interest to investigate from which cellular compartment the CSF3R-T595I mutant induces signal transduction. Furthermore, if activation at the plasma membrane does not occur, because the mutant either might internalise immediately or does not even reach the plasma membrane, this may have implications for activation of other signal transduction pathways, like the PKB/AKT pathway, which is exclusively activated at the plasma membrane[29].

Finally, it would be of interest to determine the effect of the CSF3R-T595I mutant, either or not in combination with a truncating mutation, on proliferation and differentiation *in vivo.* Therefore, transplantation studies of csf3r -/- bone marrow retrovirally transduced with the different CSF3R expressing constructs should be conducted, followed by monitoring the hematopoietic phenotype.

REFERENCES

1. Lieschke GJ, Grail D, Hodgson G, et al. Mice lacking granulocyte colony-stimulating factor have chronic neutropenia, granulocyte and macrophage progenitor cell deficiency, and impaired neutrophil mobilization. Blood 1994;84(6):1737-46.
2. Liu F, Wu HY, Wesselschmidt R, Kornaga T, Link DC. Impaired production and increased apoptosis of neutrophils in granulocyte colony-stimulating factor receptor-deficient mice. Immunity 1996;5(5):491-501.
3. Panopoulos AD, Watowich SS. Granulocyte colony-stimulating factor: molecular mechanisms of action during steady state and ‚emergency' hematopoiesis. Cytokine 2008;42(3):277-88.
4. Touw IP, van de Geijn GJ. Granulocyte colony-stimulating factor and its receptor in normal myeloid cell development, leukemia and related blood cell disorders. Front Biosci 2007;12:800-15.
5. Irandoust MI, Aarts LH, Roovers O, Gits J, Erkeland SJ, Touw IP. Suppressor of cytokine signaling 3 controls lysosomal routing of G-CSF receptor. Embo J 2007;26(7):1782-93.
6. Wolfler A, Irandoust M, Meenhuis A, Gits J, Roovers O, Touw IP. Site-specific ubiquitination determines lysosomal sorting and signal attenuation of the granulocyte colony-stimulating factor receptor. Traffic 2009;10(8):1168-79.
7. Dale DC, Bonilla MA, Davis MW, et al. A randomized controlled phase III trial of recombinant human granulocyte colony-stimulating factor (filgrastim) for treatment of severe chronic neutropenia. Blood 1993;81(10):2496-502.
8. Sinha S, Zhu QS, Romero G, Corey SJ. Deletional mutation of the external domain of the human granulocyte colony-stimulating factor receptor in a patient with severe chronic neutropenia refractory to granulocyte colony-stimulating factor. J Pediatr Hematol Oncol 2003;25(10):791-6.
9. Ward AC, van Aesch YM, Gits J, et al. Novel point mutation in the extracellular domain of the granulocyte colony-stimulating factor (G-CSF) receptor in a case of severe congenital neutropenia hyporesponsive to G-CSF treatment. J Exp Med 1999;190(4):497-507.
10. Dong F, Brynes RK, Tidow N, Welte K, Lowenberg B, Touw IP. Mutations in the gene for the granulocyte colony-stimulating-factor receptor in patients with acute myeloid leukemia preceded by severe congenital neutropenia. N Engl J Med 1995;333(8):487-93.
11. Dong F, Dale DC, Bonilla MA, et al. Mutations in the granulocyte colony-stimulating factor receptor gene in patients with severe congenital neutropenia. Leukemia 1997;11(1):120-5.
12. Dong F, Hoefsloot LH, Schelen AM, et al. Identification of a nonsense mutation in the granulocyte-colony-stimulating factor receptor in severe congenital neutropenia. Proc Natl Acad Sci U S A 1994;91(10):4480-4.
13. Germeshausen M, Ballmaier M, Welte K. Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis: Results of a long-term survey. Blood 2007;109(1):93-9.
14. Hermans MH, Antonissen C, Ward AC, Mayen AE, Ploemacher RE, Touw IP. Sustained receptor activation and hyperproliferation in response to granulocyte colony-stimulating factor (G-CSF) in mice with a severe congenital neutropenia/acute myeloid leukemia-derived mutation in the G-CSF receptor gene. J Exp Med 1999;189(4):683-92.

15. Dong F, van Buitenen C, Pouwels K, Hoefsloot LH, Lowenberg B, Touw IP. Distinct cytoplasmic regions of the human granulocyte colony-stimulating factor receptor involved in induction of proliferation and maturation. Mol Cell Biol 1993;13(12):7774-81.

16. Fukunaga R, Ishizaka-Ikeda E, Nagata S. Growth and differentiation signals mediated by different regions in the cytoplasmic domain of granulocyte colony-stimulating factor receptor. Cell 1993;74(6):1079-87.

17. Germeshausen M, Skokowa J, Ballmaier M, Zeidler C, Welte K. G-CSF receptor mutations in patients with congenital neutropenia. Curr Opin Hematol 2008;15(4):332-7.

18. Link DC, Kunter G, Kasai Y, et al. Distinct patterns of mutations occurring in de novo AML versus AML arising in the setting of severe congenital neutropenia. Blood 2007;110(5):1648-55.

19. Forbes LV, Gale RE, Pizzey A, Pouwels K, Nathwani A, Linch DC. An activating mutation in the transmembrane domain of the granulocyte colony-stimulating factor receptor in patients with acute myeloid leukemia. Oncogene 2002;21(39):5981-9.

20. Dong F, van Paassen M, van Buitenen C, Hoefsloot LH, Lowenberg B, Touw IP. A point mutation in the granulocyte colony-stimulating factor receptor (G-CSF-R) gene in a case of acute myeloid leukemia results in the overexpression of a novel G-CSF-R isoform. Blood 1995;85(4):902-11.

21. Beekman R, Valkhof MG, Sanders MA, et al. Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. Blood 2012;119(22):5071-7.

22. Plo I, Zhang Y, Le Couedic JP, et al. An activating mutation in the CSF3R gene induces a hereditary chronic neutrophilia. J Exp Med 2009;206(8):1701-7.

23. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

24. Herrmann A, Vogt M, Monnigmann M, et al. Nucleocytoplasmic shuttling of persistently activated STAT3. J Cell Sci 2007;120(Pt 18):3249-61.

25. Aarts LH, Roovers O, Ward AC, Touw IP. Receptor activation and 2 distinct COOH-terminal motifs control G-CSF receptor distribution and internalization kinetics. Blood 2004;103(2):571-9.

26. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA. 1997-2011. (Accessed at http://imagej.nih.gov/ij/.)

27. de Koning JP, Dong F, Smith L, et al. The membrane-distal cytoplasmic region of human granulocyte colony-stimulating factor receptor is required for STAT3 but not STAT1 homodimer formation. Blood 1996;87(4):1335-42.

28. Tian SS, Tapley P, Sincich C, Stein RB, Rosen J, Lamb P. Multiple signaling pathways induced by granulocyte colony-stimulating factor involving activation of JAKs, STAT5, and/or STAT3 are required for regulation of three distinct classes of immediate early genes. Blood 1996;88(12):4435-44.

29. Blume-Jensen P, Hunter T. Oncogenic kinase signalling. Nature 2001;411(6835):355-65.

CHAPTER 5

TARGETED IDENTIFICATION OF ACQUIRED MUTATIONS IN SEVERE CONGENITAL NEUTROPENIA AND ITS PROGRESSION TOWARDS LEUKEMIA

Renée Beekman[1], Marijke Valkhof[1], Remco Hoogenboezem[1], Paulette van Strien[1], Nico G. Hartwig[2], Taco W. Kuijpers[3], Marrie C. Bruin[4], Göran Carlsson[5], Peter Vandenberghe[6], Jean Donadieu[7], Ivo P. Touw[1]

[1]Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

[2]Department of Pediatrics, Erasmus University Medical Center - Sophia Children's Hospital, Rotterdam, The Netherlands

[3]Department of Pediatrics, Academic Medical Center - Emma Children's Hospital, Amsterdam, The Netherlands

[4] Department of Hematology, University Medical Center - The Wilhelmina Childrens Hospital, Utrecht, The Netherlands

[5]Childhood Cancer Research Unit, Department of Women's and Children's Health, Karolinska Institutet, Karolinska University Hospital Solna, Stockholm, Sweden

[6]Center for Human Genetics, Universitaire Ziekenhuizen Leuven, Leuven, Belgium

[7]AP-HP Registre Français des Neutropénies Congénitales, Hôpital Trousseau, Service d'Hémato-oncologie Pédiatrique, APHP, Paris, France

Manuscript in preparation

ABSTRACT

Severe congenital neutropenia (SCN) is a bone marrow failure syndrome characterized by severely reduced neutrophil counts and a high risk to progress towards acute myeloid leukemia (AML). SCN is frequently associated with constitutional mutations in the gene encoding neutrophilic elastase *(ELANE)* and acquired mutations in the gene encoding colony stimulating factor 3 receptor *(CSF3R)*. So far, *CSF3R* mutations are the only commonly acquired alterations found in SCN patients. Furthermore, mutations frequently observed in *de novo* AML are rare in SCN patients who have progressed to leukemia. Hence, the molecular mechanisms involved in SCN and its leukemogenesis remain largely unknown. Previously we have identified 12 mutations in sequential hematopoietic samples of a SCN patient who progressed to AML. We have investigated these mutated genes, interaction partners and close paralogues, in a larger cohort of 26 SCN patients, of which 5 had progressed to leukemia. Therefore we used a custom capture approach to enrich for genes of interest followed by deep-sequencing of the captured fragments. Alterations identified in the SCN phase were mutations in *CSF3R* and *LAMB2*. Leukemia related alterations were mutations in *CSF3R* (3 patients), *RUNX1* (2 patients) and *CREBBP* (1 patient). The combination of *RUNX1* and *CSF3R* mutations are frequently found in SCN patients who have progressed to leukemia suggesting that they might have an important, possibly synergistic, role in leukemic transformation in SCN patients.

INTRODUCTION

SCN is a bone marrow failure syndrome characterized by severe neutropenia, which can be successfully treated with colony-stimulating factor 3 (CSF3)[1]. In addition however, these patients have a high risk to progress towards (mainly myeloid) leukemia[2]. SCN is frequently associated with constitutional mutations in *ELANE*, genes encoding HCLS1 associated protein X-1 (*HAX1*), growth factor independent 1 (*GFI1*), *CSF3R*, glucose 6 phosphatase, catalytic subunit 3 (*G6PC3*) and Wiskott-Aldrich syndrome protein (*WAS*)[3-6]. The incidence of leukemic transformation varies between different SCN subgroups; mainly patients with mutations in *ELANE, HAX1* and *WAS* have been reported to progress towards leukemia[7,8]. A unique phenomenon in SCN, which is strongly linked to leukemic progression, is the acquisition of truncating mutations in *CSF3R*. The prevalence of these mutations in SCN patients is around 30% and increases to approximately 80% in SCN patients who have developed AML[9-13].

So far, high incidences of acquired mutations in pathways other than the CSF3 signal transduction pathway have not been reported in SCN patients. Furthermore, a previous study has shown that genes commonly mutated in *de novo* AML are not altered in SCN/AML patients[14], indicating that different molecular mechanisms underlie malignant transformation in the latter group of patients. In contrast however, we have recently

reported a SCN/AML patient carrying acquired mutations in *ASXL1, RUNX1* and *SUZ12*[15], which are known to be mutated in myeloid neoplasms including *de novo* AML[16-18]. In the same SCN/AML patient we have found 8 mutations which have not been described previously in AML, among which were mutations in *LLGL2* and *ZC3H18* that were present in pre-leukemic clones in the SCN phase already 15 years before the AML became overt[15].

By a custom capture approach followed by deep-sequencing we have extended the mutation analysis of the genes involved in the affected pathways in the SCN/AML case described above to a larger cohort of 26 SCN patients of which 5 had progressed towards leukemia. In this way we observed, besides mutations in *CSF3R,* a mutation in *LAMB2* in the neutropenic phase. Furthermore, we show that *RUNX1* mutations are recurrent in the leukemic phase and coincide with *CSF3R* mutations, indicating that RUNX1 may have an important role in leukemic transformation in SCN patients.

MATERIALS AND METHODS

Samples

Ficoll-gradient separated bone marrow cells and bone marrow-derived fibroblasts were obtained and stored as previously described[15, 19]. Amplified genomic DNA of 4 leukemia samples were obtained via the French neutropenia registry[20]. Patient characteristics are listed in Table 1; additional information is given in Table S1. The study was performed under the permission of the Institutional Review Board of the Erasmus MC, registration number MEC-2012-030.

Genomic DNA isolation

Genomic DNA was isolated as previously described[15]. Genomic DNA of bone marrow-derived fibroblasts was, if contaminated with RNA as determined by gel electrophoresis, purified as previously described[15].

Whole genome amplification

Whole genome amplification of samples with less than 1 µg DNA (as indicated in Table S1) was performed using the REPLI-g Mini Kit according the manufacturer's protocol (Qiagen, Hilden, Germany).

Design capture library

A SeqCap EZ library (Roche Nimblegen, Madison, WI) was designed covering protein coding regions of (i) 11 mutated genes in a previously reported SCN/AML patient[15], (ii) their close paralogues and (iii) their main interaction partners. Simultaneously, 6 SCN related genes were covered on the capture library. All investigated genes are indicated in Table 2.

Table 1. Characteristics of SCN patient cohort.

Patient number	Gender	Sample Number	Disease stage	Age at obtaining sample	Clinical symptoms (SCN phase)	Constitutional SCN related mutations
1	F	9820	SCN	1 yr and 11 mos	Recurrent gingivitis	No mutation
2	M	10701	SCN	2 yrs	NA	No mutation
3	F	10819	SCN	7 yrs and 1 mo	Recurrent infections, splenomegaly	No mutation
4	M	10892	SCN	4 yrs	Recurrent furuncles	ELANE exon 4 C158X†
5	F	12386	SCN	1 yr and 1 mo	Recurrent infections	No mutation
6	F	12690	SCN	3 mos	NA	ELANE exon 2 I31F†
		13885	SCN	1 yr and 5 mos		
7	F	13502B	SCN	35 yrs	NA	No mutation
		24269	SCN	46 yrs		
		30359	SCN	51 yrs		
8	M	16799	SCN	1 yr and 7 mos	Recurrent gingivitis	HAX1 exon 2 W44X
9	F	10784	SCN	10 yrs	NA	No mutation
		18929	SCN	19 yrs		
10	M	19470	SCN	10 yrs	Recurrent infections	No mutation
		30535	SCN	20 yrs		
11	F	19805	SCN	1 yr and 6 mos	NA	ELANE exon 3 Q68P†
12	M	19917	SCN	22 yrs	NA	ELANE intron 4, 1st basepair$
13	F	20315	SCN	27 yrs	NA	ELANE exon 3 L92H†
14	M	20858	SCN	11 mos	Recurrent infections	ELANE exon 3 R74L†
15	M	22084	SCN	1 yr and 3 mos	Skin abscesses	ELANE exon 3 L55P†
16	M	25928	SCN	3 yrs and 1 mo	Recurrent gingivitis	No mutation
17	M	27366	SCN	4 yrs and 6 mos	NA	ELANE exon 4 W127R†
18	F	29578	SCN	17 yrs	NA	HAX1 exon 2 W44X
19	M	26921	SCN	1 yrs and 10 mos	NA	No mutation
		29784	SCN	3 yr and 10 mos		
20	M	30339	SCN	2 yrs and 1 mo	NA	ELANE exon 2 L18P†
21	M	29569	SCN	11 mos	NA	HAX1 exon 2 W44X
22	M	459139	SCN	36 yrs	Recurrent ear infections, mastoiditis, bronchitis	WAS L270P‡
		566005	AML	38 yrs		
23	M	S18480	AML	NA	NA	ELANE exon 3 L92P†°
24	NA	S18481	AML	NA	NA	ELANE exon 2 H24L†
25	NA	S18482	AML	NA	NA	ELANE exon 5 G185R†
26	F	S18483	ALL	NA	NA	ELANE exon 5 G185R†°

For each patient, gender, sample number(s), disease stage at obtaining the sample, age at obtaining the sample, clinical symptoms and the presence of constitutional SCN related mutations (in *CSF3R, ELANE, GFI1, G6PC3, HAX1* or *WAS*) are indicated. Multiple sample numbers per patient indicate sequential sampling. M = male, F = female, SCN = severe congenital neutropenia, AML = acute myeloid leukemia, ALL = acute lymphoid leukemia, yr = year, mo = month. NA = non-assigned. Additional information on capture procedures, fibroblast control samples and constitutional mutations can be found in Table S1. †Amino acid number based on previous publication[29], $indicated by Dale et al.[3] to induce a splicing defect, °Previously reported by Link et al.[14], ‡previously reported by van Beel et al.[7]

Table 2. Genes covered by the custom capture library.

Genes mutated in previously reported SCN/AML patient
ASXL1
CCDC155
CSF3R
EP300
FBXO18
LAMB1
LLGL2
MGA
RUNX1
SUZ12
ZC3H18
Close paralogues and main interaction partners
BAP1
CBFB
CREB1
CREBBP
CUL1
EED
EZH2
GPSM2/LGN
HIF1A
LAMB2
LAMB4
LLGL1
MAX
PARD6B/Par-6
PRKCI/aPKC
RBX1
RUNX2
RUNX3
SKP1
ZC3H10
ZC3H13
SCN related genes
CSF3R
ELANE
GFI1
G6PC3
HAX1
WAS

Eleven genes were previously reported to be mutated in an SCN/AML patient[15]. These genes and their close interaction partners and paralogues were investigated in a larger SCN cohort. Simultaneously, 6 SCN related genes were covered.

Nucleotide sequencing

Targeted sequencing. Sequencing libraries of genomic DNA were prepared according to the TruSeq™ DNA Sample Preparation v2 Guide (August 2011, Illumina, San Diego, CA), the NimbleGen SeqCap EZ Library SR User's Guide version 3.0 (November 2011, Roche Nimblegen) and the NimbleGen SeqCap EZ Dual-Capture Library SR protocol (Roche Nimblegen). Primers were ordered at Sigma-Aldrich (Zwijndrecht, The Netherlands). In short, 1 µg genomic DNA was sheared to fragments of approximately 250 base pairs using the Covaris S-series High Performance, Single Tube Sample Preparation System, Model S220 (Covaris, Woburn, MA, USA) using the following settings: duty cycle 10%, cycles per burst 200, duration 80 seconds, peak incident power 140. Fragment sizes were checked on the Bioanalyzer (Agilent, Santa Clara, CA). Multiplex identifier (MID) containing, adapter ligated libraries were prepared according to the manufacturer's protocol using the Paired-End Genomic DNA Sample Prep Kit FC-121-2001 (Illumina). A pool of 1 µg of prepped library (containing 8 samples on average) was taken for hybridization with the custom designed SeqCap EZ library (Roche Nimblegen). A sample concentration of 8 picomolar was loaded for sequencing on the Hiseq2000 (Illumina) using 101-bp paired-end reads.

Whole exome sequencing (WES). WES of 3 genomic DNA of bone marrow-derived fibroblasts (patient 7, 10 and 21, Table S1) was performed as described previously[15].

Amplicon-based deep sequencing and Sanger sequencing. Amplicon-based deep sequencing and Sanger Sequencing were performed as previously described[15]. Amplification products for sequencing were generated with primers listed in Table S2.

Data-analysis

Alignment to reference genome. Paired end alignment (Hg19) was done using the Burrows-Wheeler Aligner version 0.5.9-r16 [21]. Further analysis was performed using samtools 0.1.18[22] and varscan v2.2.8[23, 24]. First an mpileupfile was generated using samtools with adjusted settings to allow a maximum depth of 1000000 (-d1000000, -L1000000), while only reads with a minimum mapping quality of 40 and base calls with a minimum phred quality of 30 were taken into account (-q40 -Q30) and the BAQ calculation was skipped (-B).

Identification of acquired mutations. Single nucleotide variants (SNVs) with a variant frequency of at least 2.5% in the SCN samples were detected using the mpileup2snp function of varscan (settings: --min-coverage 100 --min-reads2 40 --min-var-freq 0.025 --p-value 99e-02). In the same samples, small insertions and deletions (indels) were detected by determining all indels with an allele frequency of >2.5% using the output of the readcounts function of varscan. For both the SNVs and the indels, non-synonymous variations, not present in dbSNP135[25] were selected using annovar[26]. For these variants, the output of the readcounts function of varscan and the Integrative Genome Viewer

(IGV[27]) were used to determine the allele frequency in the matching fibroblast control. All variants with both forward and reverse supporting reads and a fold change of at least 2 fold compared to the fibroblast control were further analysed by Sanger sequencing or amplicon-based deep sequencing. In the 5 leukemia samples, an allele frequency threshold of 10% was used. As for these samples no matching fibroblast controls were present, other fibroblast controls were used in IGV to estimate the sequencing error rate at the detected positions.

Identification of constitutional SCN related mutations. SNVs and small indels in *CSF3R, ELANE, GFI1, G6PC3, HAX1* and *WAS* with a minimum allele frequency of 10% in the SCN and AML samples were detected using the mpileup2snp and mpileup2indel function in varscan (settings: --min-coverage 100 --min-reads2 40 --min-var-freq 0.10 --p-value 99e-02). Non-synonymous variations, not present in dbSNP135 were selected using annovar. Presence of these mutations in the matching fibroblast control (if analysed) was confirmed in IGV. Confirmation of *ELANE* mutations by Sanger sequencing was performed as previously described[15].

RESULTS

A cohort of 26 SCN patients (Table 1 and S1) was analysed for the presence of acquired mutations in genes previously described to be mutated in a SCN patient who developed AML[15]. Simultaneously, close paralogues and interaction partners were investigated as well as SCN related genes (Table 2). Thirteen of the analysed patients (50%) carried a constitutional *ELANE* mutation, 3 (11.5%) a *HAX1* mutation and 1 (3.8%) a *WAS* mutation (Table 1 and S1). We designed a custom capture strategy to enrich for genes of interest and to reach a high coverage to enable identification of mutations in minor cell populations. Captured DNA of SCN samples and matching fibroblast controls were sequenced. To determine the capability of the adopted approach to identify mutations in small cell fractions, samples of the previously described SCN/AML patient[15] were analysed. All mutations, including the ones present in minor subclones (allele frequency as low as 3%), could be detected using this strategy (data not shown).

The acquired mutations identified in this study are indicated in Table 3. Four patients (15.4%) carried acquired truncating mutations in *CSF3R*; 1 of these (patient 13) was a SCN patient who received an allogeneic bone marrow transplantation, the other 3 were patients that progressed towards leukemia. The latter 3 were previously reported[7, 14, 20]. Another mutation detected during the SCN phase was a mutation in *LAMB2* (patient 22), which however could not be detected in the subsequent AML phase of the same patient, indicating outgrowth of a different clone in the leukemic phase. Identified leukemia related mutations comprise mutations in *RUNX1* found in 2 patients and a mutation in *CREBBP* found in 1 patient. These mutations coincide with mutations in *CSF3R*.

Table 3. Acquired mutations.

Patient number	Sample number	Disease stage	Gene	Genomic alteration (Hg19)	Effect on protein	Frequency (% of mutated reads)	Validated and/or previously described
13	20315	SCN	CSF3R	chr1:36932224G>A	Q726X[$]	8.8%	Amplicon-based deep sequencing
			CSF3R	chr1:36932248G>A	Q718X[$]	5.6%	Amplicon-based deep sequencing
22	459139	SCN	LAMB2	chr3:49159403G>A	R1633W	2.8%	Amplicon-based deep sequencing
	566005	AML	CSF3R	chr1:36932213A>C	Y729X[$]	33.7%	Amplicon-based deep sequencing and Sanger sequencing
			RUNX1	chr21:36259172G>T	R107S, R80S[‡]	32.8%	Sanger sequencing
23	S18480	AML	CREBBP	chr16:3777947T>C	I2367M, I2329M[‡]	49.1%	Sanger sequencing
			CSF3R	chr1:36932248G>A	Q718X[$]	36.8%	Sanger sequencing, Link et al.[†] and Cassinat et al.[º]
26	S18483	ALL	CSF3R	chr1:36932296G>A	Q703X[$]	31.0%	Sanger sequencing, Link et al.[†] and Cassinat et al.[º]
			RUNX1	chr21:36231825C>T	A187T, A160T[‡]	51.6%	Link et al.[†]
			RUNX1	chr21:36252940G>T	S141X, S114X[‡]	34.4%	Sanger sequencing

Indicated are the identified acquired mutations in a selected group of genes (see Table 2) in a cohort of 26 SCN patients. [$]Amino acid numbers based on previous publications[9, 11], [‡]existence of multiple transcripts, [†] Link et al.[14], [º]Cassinat et al.[20]

DISCUSSION

We investigated a cohort of 26 SCN patients for the presence of mutations in a selected group of genes. This selection was based on a previous study in which we mapped the sequential acquisition of mutations in an SCN patient who progressed to AML[15]. In our study, the prevalence of *CSF3R* mutations in the 5 leukemia samples was 60% which is in line with earlier publications; in contrast, the incidence of *CSF3R* mutations at the SCN phase was only 4.5% (1 out of 22 patients), which is small compared to previous reports[9-13]. A likely explanation for this observation is that many SCN samples were obtained at an early age, while acquisition of *CSF3R* mutations may occur later in life. The same may be true for other mutations that are acquired in the SCN phase.

Besides mutations in *CSF3R,* leukemia specific alterations found in this study were mutations in *RUNX1* and *CREBBP.* However, for 2 *RUNX1* mutations (patient 26) and the *CREBBP* mutation (patient 23) we cannot rule out that these are germline variations, as we did not have a matched germline control. *RUNX1* mutations are known to coincide

with trisomy 21, a phenomenon regularly seen in SCN patients who progress to AML[18, 28]. In line with this, one of the leukemia patients carrying a *RUNX1* mutation (patient 26) simultaneously harbored a trisomy 21[20], the other (patient 22) however did not. Both the observation that trisomy 21 is a common phenomenon in SCN progression towards AML and that *RUNX1* mutations are recurrently found in these patients suggest an essential role for *RUNX1* alterations in leukemic transformation in SCN patients.

To obtain more insights into the molecular mechanism of SCN and its progression towards AML, further research should be performed in a larger cohort of SCN patients, preferably on samples obtained at later stages of disease development. This will furthermore reveal whether malignant transformation in these patients follows a general pattern of mutation acquisition or whether many different alterations can be involved in its leukemogenesis.

REFERENCES

1. Dale DC, Bonilla MA, Davis MW, et al. A randomized controlled phase III trial of recombinant human granulocyte colony-stimulating factor (filgrastim) for treatment of severe chronic neutropenia. Blood 1993;81(10):2496-502.
2. Rosenberg PS, Zeidler C, Bolyard AA, et al. Stable long-term risk of leukaemia in patients with severe congenital neutropenia maintained on G-CSF therapy. British journal of haematology 2010;150(2):196-9.
3. Dale DC, Person RE, Bolyard AA, et al. Mutations in the gene encoding neutrophil elastase in congenital and cyclic neutropenia. Blood 2000;96(7):2317-22.
4. Donadieu J, Fenneteau O, Beaupain B, Mahlaoui N, Chantelot CB. Congenital neutropenia: diagnosis, molecular bases and patient management. Orphanet J Rare Dis 2011;6:26.
5. Klein C. Genetic defects in severe congenital neutropenia: emerging insights into life and death of human neutrophil granulocytes. Annu Rev Immunol 2011;29:399-413.
6. Dale DC, Link DC. The many causes of severe congenital neutropenia. N Engl J Med 2009;360(1):3-5.
7. Beel K, Vandenberghe P. G-CSF receptor (CSF3R) mutations in X-linked neutropenia evolving to acute myeloid leukemia or myelodysplasia. Haematologica 2009;94(10):1449-52.
8. Zeidler C, Germeshausen M, Klein C, Welte K. Clinical implications of ELA2-, HAX1-, and G-CSF-receptor (CSF3R) mutations in severe congenital neutropenia. British journal of haematology 2009;144(4):459-67.
9. Dong F, Brynes RK, Tidow N, Welte K, Lowenberg B, Touw IP. Mutations in the gene for the granulocyte colony-stimulating-factor receptor in patients with acute myeloid leukemia preceded by severe congenital neutropenia. N Engl J Med 1995;333(8):487-93.
10. Dong F, Dale DC, Bonilla MA, et al. Mutations in the granulocyte colony-stimulating factor receptor gene in patients with severe congenital neutropenia. Leukemia 1997;11(1):120-5.

11. Dong F, Hoefsloot LH, Schelen AM, et al. Identification of a nonsense mutation in the granulocyte-colony-stimulating factor receptor in severe congenital neutropenia. Proc Natl Acad Sci U S A 1994;91(10):4480-4.
12. Germeshausen M, Ballmaier M, Welte K. Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis: Results of a long-term survey. Blood 2007;109(1):93-9.
13. Germeshausen M, Skokowa J, Ballmaier M, Zeidler C, Welte K. G-CSF receptor mutations in patients with congenital neutropenia. Curr Opin Hematol 2008;15(4):332-7.
14. Link DC, Kunter G, Kasai Y, et al. Distinct patterns of mutations occurring in de novo AML versus AML arising in the setting of severe congenital neutropenia. Blood 2007;110(5):1648-55.
15. Beekman R, Valkhof MG, Sanders MA, et al. Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. Blood 2012;119(22):5071-7.
16. Carbuccia N, Trouplin V, Gelsi-Boyer V, et al. Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias. Leukemia 2010;24(2):469-73.
17. Score J, Hidalgo-Curtis C, Jones AV, et al. Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. Blood 2011.
18. Taketani T, Taki T, Takita J, et al. AML1/RUNX1 mutations are infrequent, but related to AML-M0, acquired trisomy 21, and leukemic transformation in pediatric hematologic malignancies. Genes Chromosomes Cancer 2003;38(1):1-7.
19. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. The New England journal of medicine 2004;350(16):1617-28.
20. Cassinat B, Bellanne-Chantelot C, Notz-Carrere A, et al. Screening for G-CSF receptor mutations in patients with secondary myeloid or lymphoid transformation of severe congenital neutropenia. A report from the French neutropenia register. Leukemia 2004;18(9):1553-5.
21. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 2010;26(5):589-95.
22. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078-9.
23. Koboldt DC, Chen K, Wylie T, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 2009;25(17):2283-5.
24. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22(3):568-76.
25. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29(1):308-11.
26. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38(16):e164.
27. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011;29(1):24-6.
28. Freedman MH, Bonilla MA, Fier C, et al. Myelodysplasia syndrome and acute myeloid leukemia in patients with congenital neutropenia receiving G-CSF therapy. Blood 2000;96(2):429-36.
29. Xia J, Bolyard AA, Rodger E, et al. Prevalence of mutations in ELANE, GFI1, HAX1, SBDS, WAS and G6PC3 in patients with severe congenital neutropenia. British journal of haematology 2009;147(4):535-42.

Table S1. Additional information SCN patient cohort.

Patient	Sample Number	Disease stage/ sample	Whole genome amplification of DNA	Capture method	Constitutional mutations in CSF3R, ELANE, GFI1, G6PC3, HAX1 and WAS	Alteration genomic DNA (Hg19)	heterozygous/ homozygous/ hemizygous	Validated or described in earlier publication
1	9820	SCN	no	Custom capture	No mutation	x	x	x
		Fibroblasts	yes	Custom capture				
2	10701	SCN	no	Custom capture	No mutation	x	x	x
		Fibroblasts	no	Custom capture				
3	10819	SCN	no	Custom capture	No mutation	x	x	x
		Fibroblasts	no	Custom capture				
4	10892	SCN	no	Custom capture	ELANE exon 4 C158X†	chr19:855758C>A	heterozygous	validated by Sanger sequencing
		Fibroblasts	yes	Custom capture				
5	12386	SCN	no	Custom capture	No mutation	x	x	x
		Fibroblasts	yes	Custom capture				
6	12690	SCN	no	Custom capture	ELANE exon 2 I31F†	chr19:852986A>T	heterozygous	validated by Sanger sequencing
	13885	SCN	yes	Custom capture				
	12690	Fibroblasts	yes	Custom capture				
7	13502B	SCN	no	Custom capture	No mutation	x	x	x
	24269	SCN	yes	Custom capture				
	30359	SCN	no	Custom capture				
	24269	Fibroblasts	no	Whole exome				
8	16799	SCN	no	Custom capture	HAX1 exon 2 W44X	chr1:154245888insA	homozygous	no
		Fibroblasts	no	Custom capture				
9	18929	SCN	no	Custom capture	No mutation	x	x	x
	10784	SCN	yes	Custom capture				
	18929	Fibroblasts	no	Custom capture				

Table S1. Additional information SCN patient cohort. (continued)

Patient	Sample Number	Disease stage/ sample	Whole genome amplification of DNA	Capture method	Constitutional mutations in CSF3R, ELANE, GFI1, G6PC3, HAX1 and WAS	Alteration genomic DNA (Hg19)	heterozygous/ homozygous/ hemizygous	Validated or described in earlier publication
10	19470	SCN	no	Custom capture	No mutation	x	x	x
	30535	SCN	no	Custom capture				
	19470	Fibroblasts	no	Whole exome				
11	19805	SCN	no	Custom capture	ELANE exon 3 Q68P†	chr19:853327A>C	heterozygous	validated by Sanger sequencing
		Fibroblasts	yes	Custom capture				
12	19917	SCN	no	Custom capture	ELANE intron 4, 1st basepair§	chr19:855795G>A	heterozygous	validated by Sanger sequencing
		Fibroblasts	yes	Custom capture				
13	20315	SCN	no	Custom capture	ELANE exon 3 L92H†	chr19:853399T>A	heterozygous	validated by Sanger sequencing
		Fibroblasts	yes	Custom capture				
14	20858	SCN	no	Custom capture	ELANE exon 3 R74L†	chr19:853345G>T	heterozygous	validated by Sanger sequencing
		Fibroblasts	no	Custom capture				
15	22084	SCN	no	Custom capture	ELANE exon 3 L55P†	chr19:853288T>C	heterozygous	validated by Sanger sequencing
		Fibroblasts	no	Custom capture				
16	25928	SCN	no	Custom capture	No mutation	x	x	x
		Fibroblasts	yes	Custom capture				
17	27366	SCN	no	Custom capture	ELANE exon 3 W127R†	chr19:855663T>C	heterozygous	validated by Sanger sequencing
		Fibroblasts	yes	Custom capture				
18	29578	SCN	no	Custom capture	HAX1 exon 2 W44X	chr1:154245888insA	homozygous	no
		Fibroblasts	no	Custom capture				
19	29784	SCN	no	Custom capture	No mutation	x	x	x
	26921	SCN	no	Custom capture				
	29784	Fibroblasts	no	Custom capture				

Table S1. Additional information SCN patient cohort. (continued)

Patient	Sample Number	Disease stage/ sample	Whole genome amplification of DNA	Capture method	Constitutional mutations in CSF3R, ELANE, GFI1, G6PC3, HAX1 and WAS	Alteration genomic DNA (Hg19)	heterozygous/ homozygous/ hemizygous	Validated or described in earlier publication
20	30339	SCN	no	Custom capture	ELANE exon 2 L18P[†]	chr19:852948T>C	heterozygous	validated by Sanger sequencing
		Fibroblasts	no	Custom capture				
21	29569	SCN	no	Custom capture	HAX1 exon 2 W44X	chr1:154245888insA	homozygous	no
		Fibroblasts	no	Whole exome				
22	459139	SCN	yes	Custom capture	WAS L270P	chrX:48546720T>C	hemizygous	van Beel et al.
	566005	AML	no	Custom capture				
23	S18480	AML	yes	Custom capture	ELANE exon 3 L92P[†]	chr19:853399T>C	heterozygous	Link et al.
24	S18481	AML	yes	Custom capture	ELANE exon 2 H24L[†]	chr19:852966A>T	heterozygous	no
25	S18482	AML	yes	Custom capture	ELANE exon 5 G185R[†]	chr19:856000G>A	heterozygous	no
26	S18483	ALL	yes	Custom capture	ELANE exon 5 G185R[†]	chr19:856000G>A	heterozygous	Link et al.

For each patient, sample number(s) and disease stage/sample are indicated. Fibroblasts were grown from the bone marrow of the corresponding SCN sample. Additionally it is indicated whether the DNA was whole genome amplified prior to sequencing and which capture method was used; custom capture refers to the designed capture library, containing protein coding regions of the genes listed in Table 2. SCN = severe congenital neutropenia, AML = acute myeloid leukemia, ALL = acute lymphoid leukemia. For the constitutional SCN related mutations, additional information is given on chromosome position, zygosity and confirmation of these mutations by Sanger sequencing or description in earlier publications (van Beel et al.[1], Link et al.[2]). [†]Amino acid number based on previous publication[3], [$]indicated by Dale et al[4] to induce a splicing defect.

Table S2. Primers used for validation.

Gene	Alteration	Forward primer	Reverse primer	Annealing temperature (°C)
CREBBP	chr16:3777947T>C	TCAATCCACCCTTCCATGGC	CAACATCCAGCAAGCCCTG	60
CSF3R	All alterations	CCATCACCAAGCTCACAGTG	CAGAGGTTCTCATAGGACTTG	60
LAMB2	chr3:49159403G>A	CTGCCTGCCGTTTCTTCTGC	CCTCCAGGTTCTGATTCCCC	60
RUNX1	chr21:36259172G>T	CCAGTACCTTGAAAGCGATG	GCAAGATGAGCGAGGCGTTG	60
RUNX1	chr21:36252940G>T	ACGTACCTCTTCCACTTCG	TGCTATTCCTCTGCAACCT	60

The identified alterations were validated by amplicon-based deep sequencing or Sanger sequencing. Amplicon generation was performed with indicated primers.

REFERENCES SUPPLEMENTARY DATA

1. Beel K, Vandenberghe P. G-CSF receptor (CSF3R) mutations in X-linked neutropenia evolving to acute myeloid leukemia or myelodysplasia. Haematologica 2009;94(10):1449-52.
2. Link DC, Kunter G, Kasai Y, et al. Distinct patterns of mutations occurring in de novo AML versus AML arising in the setting of severe congenital neutropenia. Blood 2007;110(5):1648-55.
3. Xia J, Bolyard AA, Rodger E, et al. Prevalence of mutations in ELANE, GFI1, HAX1, SBDS, WAS and G6PC3 in patients with severe congenital neutropenia. British journal of haematology 2009;147(4):535-42.
4. Dale DC, Person RE, Bolyard AA, et al. Mutations in the gene encoding neutrophil elastase in congenital and cyclic neutropenia. Blood 2000;96(7):2317-22.

CHAPTER 6

RETROVIRAL INTEGRATION MUTAGENESIS IN MICE AND COMPARATIVE ANALYSIS IN HUMAN AML IDENTIFY REDUCED *PTP4A3* EXPRESSION AS A PROGNOSTIC INDICATOR

Renée Beekman[1], Marijke Valkhof[1], Stefan J. Erkeland[1], Erdoğan Taşkesen[1], Veronika Rockova[1,2], Justine K. Peeters[1], Peter J. M. Valk[1], Bob Löwenberg[1], Ivo P. Touw[1]

Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

[2]Department of Biostatistics, Erasmus University Medical Center, Rotterdam, The Netherlands

PLoS One. 2011 Oct 20;6(10):e26537.

ABSTRACT

Acute myeloid leukemia (AML) results from multiple genetic and epigenetic aberrations, many of which remain unidentified. Frequent loss of large chromosomal regions marks haplo-insufficiency as one of the major mechanisms contributing to leukemogenesis. However, which haplo-insufficient genes (HIGs) are involved in leukemogenesis is largely unknown and powerful experimental strategies aimed at their identification are currently lacking. Here, we present a new approach to discover HIGs, using retroviral integration mutagenesis in mice in which methylated viral integration sites and neighbouring genes were identified. In total we mapped 6 genes which are flanked by methylated viral integration sites (mVIS). Three of these, i.e., *Lrmp*, *Hcls1* and *Prkrir*, were up regulated and one, i.e., *Ptp4a3,* was down regulated in the affected tumor. Next, we investigated the role of *PTP4A3* in human AML and we show that *PTP4A3* expression is a negative prognostic indicator, independent of other prognostic parameters. In conclusion, our novel strategy has identified *PTP4A3* to potentially have a role in AML, on one hand as a candidate HIG contributing to leukemogenesis in mice and on the other hand as a prognostic indicator in human AML.

INTRODUCTION

Acute myeloid leukemia (AML) is a complex disease driven by multiple cytogenetic abnormalities, such as inv(16), t(8;21), t(15;17), 3q abnormalities, deletions of (the q-arms) of chromosome 5 and 7 and by aberrant expression and/or mutations of genes e.g., *EVI1, FLT3, RAS, RUNX1, CKIT, WT1, CEBPA* and *NPM1*[1, 2]. The frequent occurrence of chromosomal deletions suggests that haplo-insufficiencies contribute to the pathogenesis of AML. However, because deleted regions often harbor numerous genes, it remains difficult to pin point critical haplo-insufficient genes (HIGs) involved in the pathogenesis of AML. Gene expression profiling (GEP) focusing on down regulated genes could be informative, however differences in expression levels may relate to differentiation status of the AML blasts, rather than to mechanisms underlying leukemogenesis[3]. In addition, mapping of minimal affected regions in combination with GEP to identify HIGs often is cumbersome because these regions may still contain numerous genes and differences in their expression level may be subtle. Even in chromosomal regions frequently lost upon leukemic progression, e.g., the q-arm of chromosome 7, identification of critical HIGs remains difficult.

Retroviral insertion mutagenesis in mouse models has been used to discover novel genes involved in the development of different types of cancer[4-6]. Most of these genes have been classified as proto-oncogenes, owing to the fact that proviral integrations preferentially occur in 5' promoter regions, supposedly leading to increased or sustained expression of flanking genes. Only a small minority of identified genes have been classified as tumor suppressor genes or HIGs, based on disruption of coding sequences by the

proviral integration[7, 8]. Gene therapy studies using murine leukemia virus (MLV)-based vectors have shown that epigenetic changes of long terminal repeats (LTRs) of integrated proviruses often result in silencing of therapeutic genes[9, 10], and that preventing methylation of the CpG islands within LTRs overcomes this problem[11]. Based on these observations, we hypothesized that methylation of viral sequences not only results in silencing of retroviral genes themselves but may also affect host genes located proximal to proviral integrations. Methylated LTRs located in proximity of promoter regions may thus identify genes that are deregulated leading to haplo-insufficiency.

To discover potential HIGs relevant for human AML, we used murine leukemia samples induced by Graffi 1.4 Murine Leukemia Virus (Gr1.4 MLV), classified as mixed lineage or myeloid leukemias by immunophenotyping[6, 12]. By methylation specific PCR (MSP) and methylated DNA immunoprecipitation (MeDIP)[13] we observed an extensive variation in the level of DNA methylated proviral integrations in these tumors. We designed a strategy to map methylated proviral integrations by combining MeDIP, inverse PCR (iPCR) and promoter array hybridization. We identified 6 genes to be flanked by methylated viral integration sites (mVIS), of which *Lrmp*, *Hcls1* and *Prkrir* were transcriptionally up regulated and *Ptp4a3* was transcriptionally down regulated. Further studies in human AML samples revealed a negative prognostic value of *PTP4A3* expression levels, independent of other prognostic indicators. In conclusion, by mapping DNA methylated viral integration sites in murine leukemias induced by retroviral integration mutagenesis followed by comparative analysis in human AML, we identified *PTP4A3* not only as a candidate HIG contributing to leukemogenesis in mice but also as an independent prognostic indicator in human AML.

RESULTS

Viral integrations sites of the Graffi1.4 MuLV are subject to DNA methylation

In this study murine leukemia samples induced by Gr1.4 MLV were analysed[6]. First, a methylation specific PCR (MSP) was performed to determine the level of DNA methylation of the Gr1.4 MLV LTRs. To this end, amplification products from methylated LTRs were quantified with quantitative PCR (qPCR) and corrected for total LTRs in these samples (Figure 1A). A considerable variation in LTR methylation was seen between different tumors (data not shown). Based on these methylation levels, leukemia samples were divided into 4 methylation categories of equal sample size (1 = highest LTR methylation level, 4 = lowest LTR methylation level).

Subsequently, MeDIP was used on a subset of samples to enrich for methylated LTRs and flanking genomic regions. As a control, genomic DNA of normal bone marrow, spleen and liver was used. MeDIP enrichment relative to input levels was determined for the LTR, the non-methylated actin B locus (*ActB*) and the hemi-methylated imprinting control region 1 (ICR1) of *H19*. As expected, *H19* enrichment scores were high and *ActB*

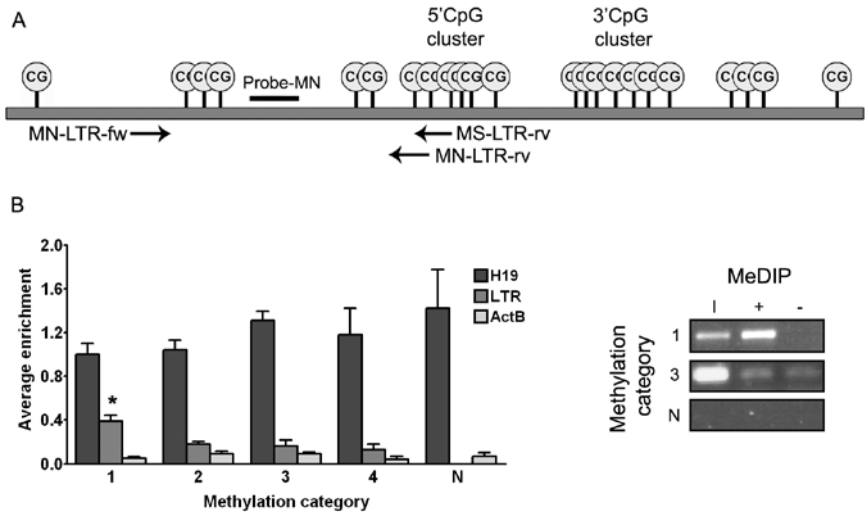


Figure 1. LTR methylation analysis. (A) Overview of the methylation specific PCR (MSP) approach. Depicted is a schematic representation of the Gr1.4 MuLV LTR, containing 23 CpGs. The MSP was performed, after bisulphite treatment, with a methylation neutral forward primer (MN-LTR-fw), and a methylation specific (MS-LTR-rv) or neutral (MN-LTR-rv) reverse primer. Amplification products were quantified using methylation neutral probe-MN. Tumor samples were divided into 4 equal groups based on the methylation status of their LTRs (1 = highest LTR methylation level, 4 = lowest LTR methylation level). (B) *Left panel.* For *H19*, the LTR and *ActB*, average enrichment after MeDIP compared to input levels were calculated for each MSP-defined methylation category as well as for normal bone marrow, spleen and liver (N). In category 1 to 4 respectively 18, 15, 4 and 3 samples were analysed; error bars indicate standard deviations. P-values were calculated using a Wilcoxon test; *significantly higher than other categories, p-value <0.001. *Right panel*. Example of LTR enrichment after MeDIP (I = input, + = IP with anti-5-methylcytidine, - = IP with pre-immune serum IgG, 1 and 3 = methylation categories, N = normal spleen).

enrichment scores were low in all categories (Figure 1B). Additionally, samples in the highest methylation category showed a significantly higher LTR enrichment after MeDIP compared to the samples in other categories (p-value <0.001), confirming the specificity of the MSP (Figure 1B).

Ptp4a3 is flanked by a methylated viral integration site and is transcriptionally down regulated

Genes located near methylated viral integration sites (mVIS) may be down regulated due to the proximity of a methylated regulatory sequence, and, their transcriptional down regulation may contribute to murine leukemogenesis. Therefore, after showing that a proportion of viral integration sites are subject to DNA methylation, we set out to identify genes flanking these viral integration sites. To this end, iPCR, to amplify regions flanking viral integration sites, and MeDIP, to enrich for DNA methylated fragments,

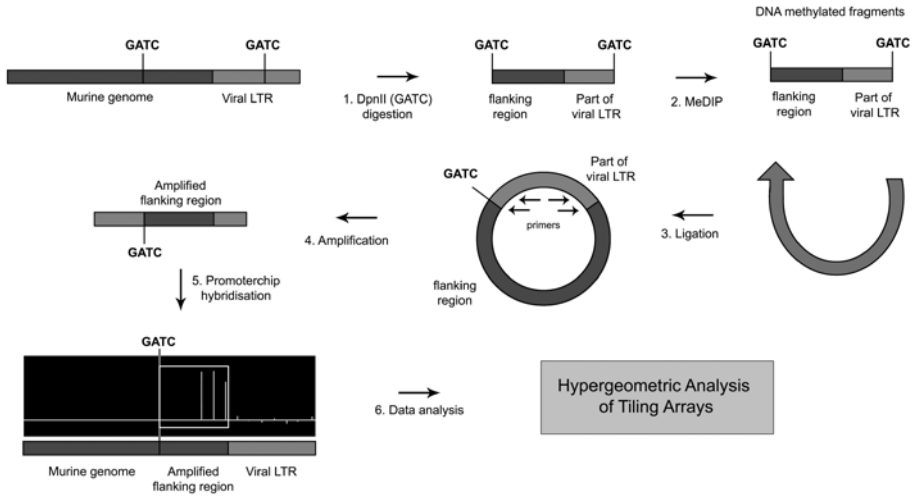
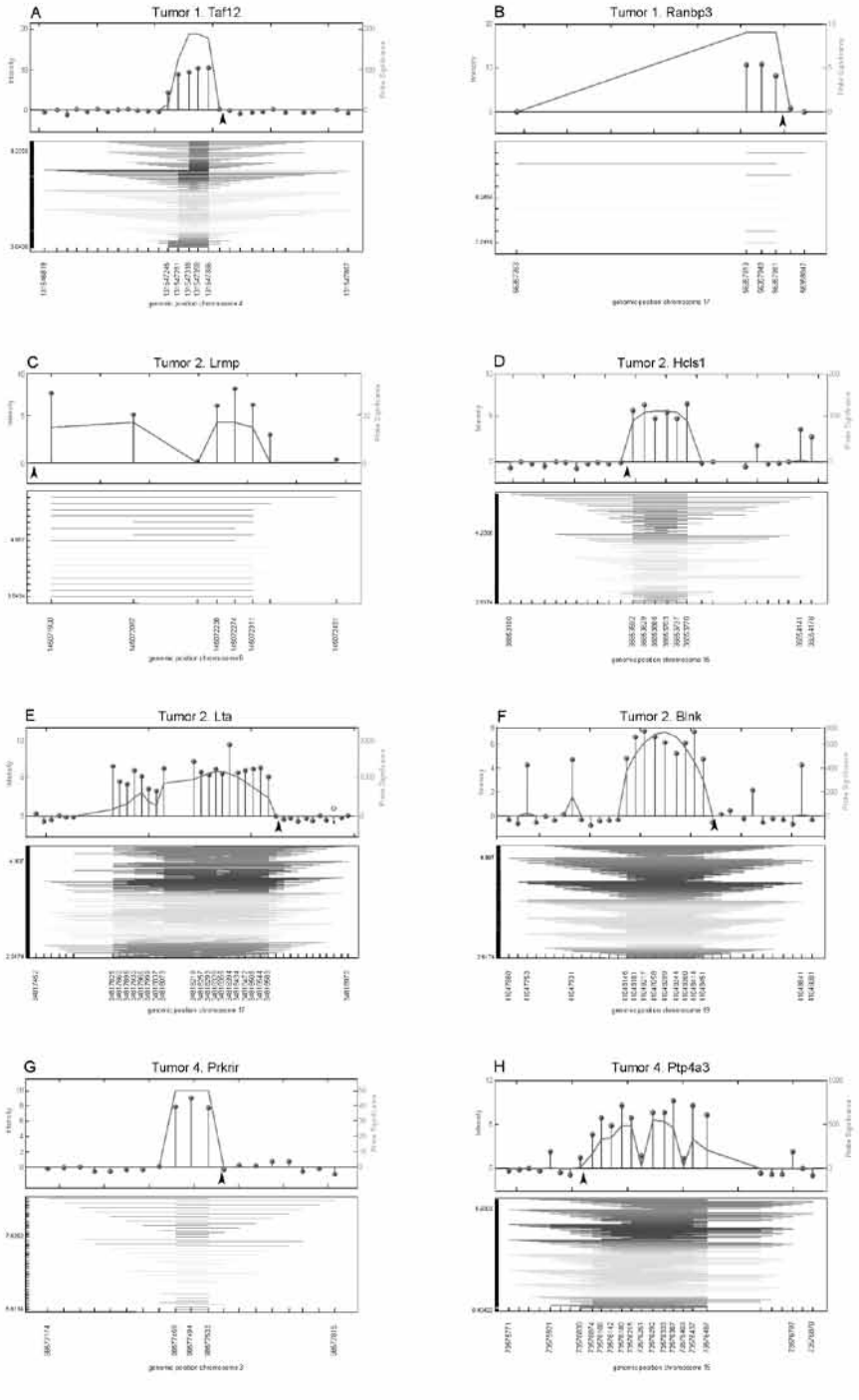


Figure 2. Identification of mVIS. Strategy outline for identification of regions flanking DNA methylated viral integration sites (mVIS) within murine leukemias. Genomic DNA was digested with DpnII (step 1), followed by methylated DNA immunoprecipitation (MeDIP, step 2). MeDIP enriched fragments were ligated (step 3) and amplified using primers within the LTR (step 4). These fragments were hybridized on a DNA promoter array (step 5). Hypergeometric Analysis of Tiling Arrays (HAT) was used to identify regions flanking mVIS (step 6).

were combined to amplify regions flanking mVIS (Figure 2). Amplified fragments of 6 tumor samples were hybridized to Murine 1.0 R promoter arrays and, using hypergeometric analysis of tiling arrays (HAT)[14], 15 amplified regions were mapped in these tumors (Table S1). Eight of these integrations were validated by directed PCR followed by Sanger sequencing (Figure 3, Table S1). Because MLVs tend to integrate within 10 kb around the transcriptional start site[15], the nearest genes within 10 kb downstream of these 8 mVIS were determined (Figure 3, Table S1).

To support that regions identified in this way were indeed flanked by methylated LTRs, we performed a methylation sensitive digestion followed by directed PCR. Using this

Right: **Figure 3. Identified viral integration sites.** Eight viral integrations identified with HAT could be confirmed with directed PCR and Sanger sequencing (see Table S1 for further details). The graphical output of HAT is represented in graph A-H. Above each graph, the tumor in which the integration was identified as well as the nearby located gene are indicated. The upper panel of each graph shows normalized intensities of the different probes (lollipops) on the mouse promoter 1.0R arrays and their significance (indicated by the continuous grey line) as calculated with HAT. The black arrowhead indicates the exact position of the proviral integration, as determined by directed PCR followed by Sanger sequencing. In the lower panel the lowest and highest probe intensity threshold with a significant outcome are given on the left. The stripes indicate significantly enriched regions at different probe intensity thresholds, calculated with HAT, which are merged into the final viral integration site. Below each graph, the genomic position is indicated (assembly mm8, February 2006).


A
Tumor 1. Taf12
B
Tumor 1. Ranbp3
C
Tumor 2. Lrmp
D
Tumor 2. Hcls1
E
Tumor 2. Lta
F
Tumor 2. Blnk
G
Tumor 4. Prkrir
H
Tumor 4. Ptp4a3

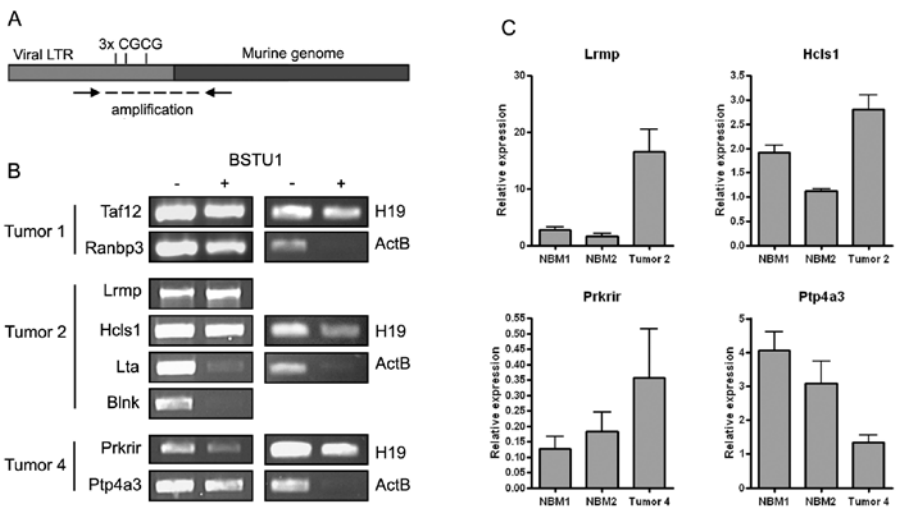


Figure 4. Methylation sensitive restriction analysis of viral integration sites and expression of nearby located genes. (A) Schematic overview of the methylation specific restriction approach. Genomic DNA was digested with BstU1 (CGCG, blocked by DNA methylation), followed by mVIS amplification with primers as indicated by arrows. If the flanking LTR is methylated, mVIS amplification is unaffected upon BstU1 digestion. (B) All 8 identified viral integration sites, identified in tumor 1, 2 and 4, were amplified before (-) and after (+) BstU1 digestion. As controls, *H19* (hemi-methylated) and *ActB* (unmethylated), both containing 2 BstU1 digestion sites, were analysed in each tumor. (C) Expression levels of 4 genes flanked by methylated viral integration sites were determined by qPCR in the respective tumors. Expression levels relative to housekeeping gene *Tbp* are shown; error bars indicate standard deviations. NBM = normal bone marrow.

approach, only viral integration sites flanked by methylated LTRs could be amplified (Figure 4A), as was the case for 6 out of 8 identified integrations (Figure 4B, Table S1). Subsequently, expression levels of genes flanking these mVIS were quantified by qPCR and compared to normal bone marrow expression levels. Unfortunately, RNA of tumor 1 was lacking, therefore this analysis could not be performed for *Taf12* and *Ranbp3*. Of the other 4 genes, *Ptp4a3* expression was 2-3 fold reduced in the respective tumor (Figure 4C, Table S1).

PTP4A3 is an independent prognostic factor in human AML

The human orthologue of murine *Ptp4a3*, i.e., *PTP4A3*, was further studied in human AML. Transcript levels of *PTP4A3* were assessed in 454 AML samples, diagnosed under the age of 60, profiled using the HGU133 2.0 plus gene expression arrays[16]. *PTP4A3* expression values are represented by 2 probesets with a high correlation (Pearson correlation coefficient = 0.90). Survival analysis with these probesets gave similar results; all results shown are based on expression levels of probeset 206574_s_at. *PTP4A3* expression levels were negatively correlated with prognostic outcome both for overall survival

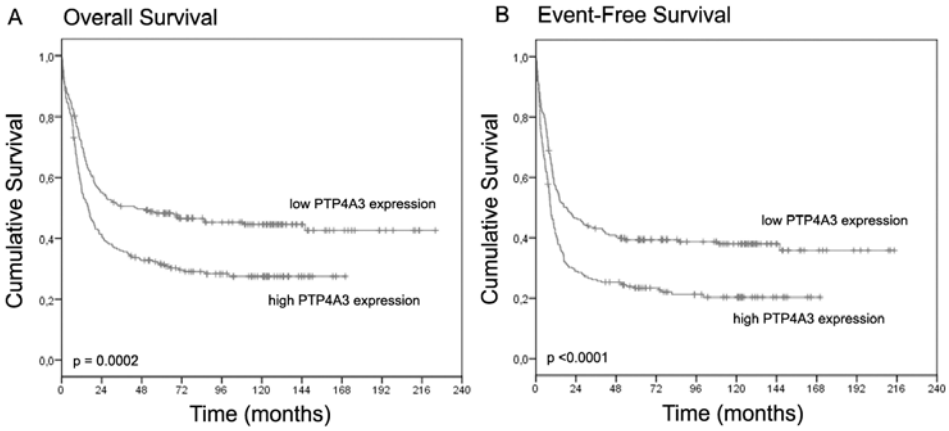


Figure 5. Survival analysis. A cohort of 454 de novo AML cases diagnosed under the age of 60 was divided into 2 groups of equal size based on MAS5 normalised expression of *PTP4A3* (probe 206574_s_at). Overall survival (A) and event-free survival (B) were analysed. P-values were calculated with a log rank test.

Table 1. Multivariate survival analysis.

	Overall Survival		Event-Free Survival	
Risk factor	HR (95% CI)	P-value	HR (95% CI)	P-value
PTP4A3 expression	1.112 (0.995-1.243)	0.061	1.131 (1.019-1.255)	0.021*
Age (decades)	1.134 (1.024-1.256)	0.016*	1.068 (0.969-1.177)	0.186
WBC∞	1.373 (1.063-1.773)	0.015*	1.296 (1.020-1.648)	0.034*
Favorable cytogenetic risk†	0.376 (0.257-0.548)	<0.0001*	0.469 (0.335-0.658)	<0.0001*
Unfavorable cytogenetic risk†	1.432 (1.059-1.935)	0.020*	1.507 (1.124-2.020)	0.006*
NPM1+FLT3ITD-‡	0.473 (0.317-0.705)	0.0002*	0.578 (0.398-0.839)	0.004*
CEBPA double mutant$	0.591 (0.418-0.836)	0.003*	0.560 (0.384-0.815)	0.002*

Multivariate analysis in 454 de novo AML patients under the age of 60. ∞WBC higher than $20x10^9$/L versus lower than $20x10^9$/L, †compared to intermediate cytogenetic risk, ‡compared to no *NPM1*$^+$*FLT3ITD*$^-$, $compared to no *CEBPA* double mutation. *Statistically significant. HR = hazard ratio, CI = confidence interval, WBC = white blood cell count, *FLT3ITD* = internal tandem duplication of *FLT3*.

(OS, p-value <0.0001, hazard ratio = 1.269) and event-free survival (EFS, p-value <0.0001, hazard ratio = 1.261). Kaplan-Meier curves are shown in Figure 5. A permutation test predicted a probability of 0.0036 for a random gene locus to be a significant prognostic indicator with a p-value <0.0001 for both OS and EFS. Multivariate analysis showed that the negative correlation of *PTP4A3* expression with event-free survival was independent of other prognostic parameters, i.e., age, white blood cell count, cytogenetic risk, *CEBPA* mutation status and *NPM1*$^+$*FLT3ITD*$^-$ status (Table 1).

DISCUSSION

We designed a strategy to identify candidate HIGs in AML using retroviral integration mutagenesis, by mapping DNA methylated proviral integrations. By using HAT[14], we deliberately aimed at detecting integrations present in the majority of the leukemic cells, which are most likely involved in the early phase of leukemogenesis. At the same time, integrations present in subclones that contribute to later stages of leukemic progression will be missed using this approach. We identified 6 genes that are flanked by methylated viral integrations. Expression analysis showed that *Lrmp* (lymphoid-restricted membrane protein), *Hcls1* (hematopoietic cell specific Lyn substrate 1) and *Prkrir* (protein-kinase, interferon-inducible double stranded RNA dependent inhibitor, repressor of (P58 repressor)) were up regulated and *Ptp4a3* (protein tyrosine phosphatase type IVA), a phosphatase also known as *Prl3* (phosphatase of regenerating liver 3) was down regulated in the respective murine tumor. These results indicate that a flanking methylated viral integration site does not necessarily lead to transcriptional repression. As 1 out of 4 genes flanked by a mVIS was transcriptionally down regulated and expression of the 2 other genes could not be investigated, the efficiency to detect potential HIGs by identifying mVIS would approximately be 17-25%. However, the number of analysed tumors is too small to allow an accurate estimation of the efficiency.

Ptp4a3 expression is controlled by p53 induced after DNA damage in mouse embryonic fibroblasts (MEFs) and its activity is involved in inducing a G1 cell cycle arrest in these cells [17]. Surprisingly however, the same study also demonstrated a cell cycle arrest upon reduction of *PTP4A3* expression[17]. Apparently, depending on expression level dosage, *PTP4A3* may have both positive and negative effects on cell cycle regulation. Hence, *PTP4A3* haplo-insufficiency, but not its complete loss, may lead to an impairment of cell cycle arrest after DNA damage. Dosage effects of *PTP4A3* expression in relation to cellular responses may be more complex, particularly in cancer cells. For example, in carcinoma cell lines *PTP4A3* expression may lead to down regulation of p53[18] and it is variably induced by γ-irradiation[19]. Finally, high *PTP4A3* expression has been linked to increased tumor aggressiveness in different types of solid tumors, e.g., melanoma, gastric cancer, colon cancer, hepatocellular carcinoma and breast cancer[20-24], possibly because high *PTP4A3* expression leads to increased epithelial-mesenchymal transition[25].

The role of *PTP4A3* in hematopoietic malignancies has not been studied as extensively as in carcinoma. Only a few studies report differences in expression levels of *PTP4A3* in ALL and myeloma subgroups, based on gene expression profiling[26-28]. Interestingly however, in a recent study, *PTP4A3* has been proposed to have a role in drug-resistance in AMLs with internal tandem duplication of *FLT3 (FLT3ITD)*[29]. This finding, together with the observation that high *PTP4A3* expression negatively correlates with prognostic outcome, indicates that PTP4A3 might be a potential therapeutic target in AML.

In conclusion, using a retroviral mutagenesis screen in which we enriched for DNA methylated viral integration sites we identified *PTP4A3* as a potential haplo-insufficient gene with an independent prognostic value in human de novo AML. Challenges for the future are to determine the dose-effect of *PTP4A3* expression in myeloid development and to extend the screens to additional myeloid neoplasms, e.g., myelodysplasia, therapy-related AML, AML secondary to bone marrow failure and myeloproliferative disorders.

MATERIALS AND METHODS

Ethics statement

For this study no novel murine leukemias were generated, all experiments described were performed on material generated in a previous study[6]. All animal procedures for the use of control bone marrow fractions were approved by the animal care and use committee of the Erasmus MC (approval # 119-10-05).

All human cell samples were obtained after written informed consent and stored anonymously in a biobank. The study was performed under the permission of the Institutional Review Board of the Erasmus MC, registration number MEC-2008-387.

Mouse leukemia and normal cell samples

DNA and RNA samples from a previously generated panel of Gr1.4-induced leukemias[6], and control samples (bone marrow, spleen, liver) from normal FVB/N mice were used.

Methylation specific PCR

Primer and probe sequences are shown in Table S2. Two µg of genomic DNA was treated with bisulphite using the EZ DNA Methylation kit according to the manufacturer's protocol (Zymo research, Orange, CA, USA). LTRs were amplified with bsLTRfw and bsLTRrv using 1 µL out of 10 µL of bisulphite-treated DNA. Cycling conditions were 30" at 94° C, 30" at 50°C and 1' at 72°C for 10 cycles in a total volume of 50 µL. Two µL was used in a nested qPCR (Figure 1A) using MN-LTR-fw x MS-LTR-rv/MN-LTR-rv (MN = methylation neutral, MS = methylation specific). Cycling conditions were 15" at 94°C, 30" at 57°C and 30" at 60°C for 45 cycles. Amplified LTRs, methylated and unmethylated, were quantified using a methylation neutral probe (probe-MN, Sigma-Aldrich, Zwijndrecht, The Netherlands). Delta cycle threshold-values (dCt), representing the number of methylated LTRs as a fraction of total LTRs, were calculated as follows: dCt = Ct(Methylated LTRs)-Ct(All LTRs) = Ct(MN-LTR-fw x MS-LTR-rv) – Ct(MN-LTR-fw x MN-LTR-rv). PCRs were performed in duplicate and mean dCt values were calculated.

MeDIP

Ten µg genomic DNA was digested overnight with 100U of DpnII (New England Biolabs, Ipswich, MA, USA). Four µg digested DNA was denatured for 10' at 95°C and incubated with either 2.5 µg anti-5-methylcytidine (BI-MECY-1000, Eurogentec, Liège, Belgium) or mouse pre-immune IgG (Sigma-Aldrich, Zwijndrecht, The Netherlands) in 500 µL IP-buffer (PBS with 0.05% Triton X-100) for 2 hrs at 4°C, followed by incubation with 30 µL of washed beads (M-280 sheep-anti-mouse IgG, Invitrogen, San Diego, CA, USA) for 2 hrs at 4°C. Beads were washed 3 times with 700 µL IP-buffer. As a 10% input reference, 400 ng digested DNA not subjected to MeDIP was used. Beads and the 10% input reference DNA were resuspended in 100 µL IP-buffer and incubated for 3 hrs at 50°C after adding 20 µg proteinase K (Roche, Basel, Switzerland). Supernatants, containing immunoprecipitated DNA, and the input DNA were purified using the MinElute Reaction Cleanup Kit (Qiagen, Hilden, Germany) and were eluted in 40 µL elution buffer. Two µL immunoprecipitated DNA was used to amplify the imprinting control region 1 (ICR1) of *H19* with *H19*ICR1fw x *H19*ICR1rv, *ActB* with *ActB*fw x *ActB*rv and the LTR with LTRfw x LTRrv using (q)PCR. Primer sequences are shown in Table S2. Cycling conditions were 30" at 95° C, 30" at 58°C and 45" at 72°C for 30 cycles (PCR) or 15" at 94° C, 30" at 59°C and 30" at 60°C for 45 cycles (qPCR). Amplification products were analysed using gel electrophoresis (PCR) or quantified (qPCR) using SYBRgreen Master mix (Applied Biosystems, Foster City, CA, USA).

Inverse PCR

Primer sequences are shown in Table S2. Six murine leukemias with high LTR enrichment (more than 10% of input) and low *ActB* enrichment (less than 10% of input) were selected for inverse PCR. Eight µL MeDIP-DNA was denatured for 3' at 95°C, renatured by a temperature decrease of 0.1°C/sec to 20°C, and ligated for 45' at room temperature using a rapid DNA ligation kit (Roche, Basel, Schwitzerland). Two µL out of 20 µl ligated product was amplified with primers mL1 and mL2, followed by a nested PCR with primers mL1N and mL2N using 2 µL of the first PCR product. Cycling conditions were 30" at 95°C, 30" at 60°C (first PCR) or 56°C (nested PCR) and 3' at 72°C for 30 cycles. In the nested PCR 10 mM dCTP, dATP, dGTP, 8mM dTTP and 2mM dUTPs were used.

Promoter array hybridization

PCR products of 10 nested PCR reactions were purified with a PCR purification kit (Qiagen, Hilden, Germany) and pooled. A total of 7.5 µg of these amplified fragments was fragmented and labeled using the GeneChip WT Double-stranded DNA terminal labeling kit (Affymetrix, Santa Clara, CA, USA). Fragmentation to 66 bp was checked on a Bioanalyser (Agilent, Santa Clara, CA). Labeled DNA was hybridized to mouse promoter 1.0R arrays (Affymetrix, Santa Clara, CA, USA) for 16 hrs at 45°C. Arrays were washed with the FS_450_0001 protocol using the Fluidics Station 450 (Affymetrix, Santa Clara, CA,

USA), followed by scanning. Probe values were normalized with model-based analysis of tiling-arrays (MAT)[30] and mVIS were determined using hypergeometric analysis of tiling arrays (HAT)[14], both for HAT and MAT default settings were used. Genes located nearby amplified regions were identified using UCSC (assembly mm8, Feb. 2006).

Directed PCR and Sanger sequencing

Primers are shown in Table S2; amplification of the integration site was performed with VIS(*corresponding gene*) x LTRfw2, for *Lrmp* a nested PCR was performed with VIS(*Lrmp_nested*) x LTRfw. As input, 200 ng of the corresponding tumor DNA was used; cycling conditions were 30" at 95°C, 30" at 58°C and 45" at 72°C for 30 cycles. Products were purified using the Multiscreen HTS 66-well filtration system (Millipore, Billerica, MA, USA). Sanger sequencing was performed with primer LTRfw according to the manufacturer's protocol (Applied Biosystems, Foster City, CA, USA).

Methylation sensitive restriction analysis

Primers are shown in Table S2. Two and a half µg of tumor DNA was digested with 25U of BstU1 (New England Biolabs, Ipswich, MA, USA) o/n at 60°C, purified using the Multiscreen HTS 66-well filtration system (Millipore, Billerica, MA, USA), eluted in 30 µl and diluted to 50 ng/µl. Amplification of the integration site was performed as described under directed PCR and Sanger sequencing, with 100 instead of 200 ng input of DNA. As controls *H19* ICR1 (*H19*ICR1fw x *H19*ICR1rv) and *ActB* (*ActB*fw x *ActB*rv) were amplified. Cycling conditions were 30" at 95°C, 30" at 58°C and 45" at 72°C for 30 cycles. Amplification products were analysed using gel electrophoresis.

RNA isolation, cDNA preparation and qPCR

RNA of murine samples was isolated using Trizol (Invitrogen, San Diego, CA) according to the manufacturer's protocol. One µg of RNA was used for cDNA preparation, using SuperScript II Reverse Transcriptase (Invitrogen, San Diego, CA) according to the manufacturer's protocol. One µl cDNA was used as input for the qPCR. Genes of interest were amplified with their respective forward and reverse primers (Table S2), as an input control, TATA box binding protein (*Tbp)* was analysed. Cycling conditions were 3" at 95°C and 30" at 60°C for 45 cycles. Amplification products were quantified using Fast SYBRgreen Master mix (Applied Biosystems, Foster City, CA, USA). Expression levels relative to *Tbp* were calculated.

Survival analysis human AML samples

Purified AML blasts were obtained following informed consent as described[31]. Gene-expression profiles of 454 de novo AML patients under the age of 60 were used for this

analysis[16]. Expression levels were MAS5 normalised (Scaling factor 100), values <30 were set at 30, followed by log2 transformation.

For *Ptp4a3*, univariate and multivariate survival analyses were performed using expression levels of probesets 206574_s_at or 209695_at in a Cox regression model. In the multivariate analysis age, white blood cell count, cytogenetic risk group, *NPM1*$^{+}$*FLT3ITD*-status and *CEBPA* mutation status were used as additional prognostic parameters. We recognised the following cytogenetic risk groups: favorable = t(15;17), inv(16) and t(8;21), unfavorable = t(3;3), inv(3), -7/7q-, -5/5q-, complex karyotype, t(11q23) except t(9;11), t(9;22) and t(6;9), intermediate = all other cases with known cytogenetics. Kaplan-meier graphs were generated by dividing the AML cohort in 2 groups of equal sample size based on *PTP4A*3 expression of probe 206574_s_at. Analyses were performed in SPSS (version 17, SPPS Inc, Chicago, IL).

For the permutation test, all probesets with an annotated gene symbol (based on HG-U133_Plus_2.na32.annot.csv, Affymetrix, Santa Clara, CA, USA) were selected. Next probesets with expression levels <30 in all 454 patients were discarded, leaving a total of 40720 probesets. The permutation test was performed by randomly selecting 6 probesets (representing 6 mVIS), followed by randomly selecting 1 out of these 6 probesets (representing 1 down regulated gene). For this probeset a univariate Cox regression analysis was performed for overall survival (OS) and event-free survival (EFS). A p-value of <0.0001 (as observed for *PTP4A3)* was considered significant. This analysis was repeated 100.000 times, followed by calculating the frequency, i.e., probability, of observing a significant p-value for both OS and EFS. Analyses were performed in Matlab (version 2008b, Mathworks, Natick, MA).

ACKNOWLEDGEMENTS

The authors thank M. Sanders for advice concerning the permutation test. This research is supported by the Center for Translational Molecular Medicine (www.ctmm.nl, CTMM), the Netherlands Genomics Initiative (NGI) and the Dutch Cancer Society (www.kwfkankerbestrijding.nl, KWF Kankerbestrijding).

REFERENCES

1. Lowenberg B. Acute myeloid leukemia: the challenge of capturing disease variety. Hematology Am Soc Hematol Educ Program 2008:1-11.
2. Swerdlow S.H. CE, Harris N.L. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, IARC, Lyon; 2008.
3. Wouters BJ, Lowenberg B, Delwel R. A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects. Blood 2009;113(2):291-8.
4. Suzuki T, Shen H, Akagi K, et al. New genes involved in cancer identified by retroviral tagging. Nat Genet 2002;32(1):166-74.

5. Theodorou V, Kimm MA, Boer M, et al. MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. Nat Genet 2007;39(6):759-69.
6. Erkeland SJ, Valkhof M, Heijmans-Antonissen C, et al. Large-scale identification of disease genes involved in acute myeloid leukemia. J Virol 2004;78(4):1971-80.
7. Kool J, Berns A. High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. Nat Rev Cancer 2009;9(6):389-99.
8. Suzuki T, Minehata K, Akagi K, Jenkins NA, Copeland NG. Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice. Embo J 2006;25(14):3422-31.
9. Lorincz MC, Schubeler D, Goeke SC, Walters M, Groudine M, Martin DI. Dynamic analysis of proviral induction and De Novo methylation: implications for a histone deacetylase-independent, methylation density-dependent mechanism of transcriptional repression. Mol Cell Biol 2000;20(3):842-50.
10. Yao S, Sukonnik T, Kean T, Bharadwaj RR, Pasceri P, Ellis J. Retrovirus silencing, variegation, extinction, and memory are controlled by a dynamic interplay of multiple epigenetic modifications. Mol Ther 2004;10(1):27-36.
11. Swindle CS, Kim HG, Klug CA. Mutation of CpGs in the murine stem cell virus retroviral vector long terminal repeat represses silencing in embryonic stem cells. J Biol Chem 2004;279(1):34-41.
12. Voisin V, Barat C, Hoang T, Rassart E. Novel insights into the pathogenesis of the Graffi murine leukemia retrovirus. J Virol 2006;80(8):4026-37.
13. Weber M, Davies JJ, Wittig D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 2005;37(8):853-62.
14. Taskesen E, Beekman R, de Ridder J, et al. HAT: hypergeometric analysis of tiling-arrays with application to promoter-GeneChip data. BMC Bioinformatics 2010;11:275.
15. Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. Science 2003;300(5626):1749-51.
16. Verhaak RG, Wouters BJ, Erpelinck CA, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. Haematologica 2009;94(1):131-4.
17. Basak S, Jacobs SB, Krieg AJ, et al. The metastasis-associated gene Prl-3 is a p53 target involved in cell-cycle regulation. Mol Cell 2008;30(3):303-14.
18. Min SH, Kim DM, Heo YS, Kim HM, Kim IC, Yoo OJ. Downregulation of p53 by phosphatase of regenerating liver 3 is mediated by MDM2 and PIRH2. Life Sci 2009;86(1-2):66-72.
19. Pfeifer D, Wallin A, Holmlund B, Sun XF. Protein expression following gamma-irradiation relevant to growth arrest and apoptosis in colon cancer cells. J Cancer Res Clin Oncol 2009;135(11):1583-92.
20. Hao RT, Zhang XH, Pan YF, et al. Prognostic and metastatic value of phosphatase of regenerating liver-3 in invasive breast cancer. J Cancer Res Clin Oncol 2010;136(9):1349-57.
21. Laurent C, Valet F, Planque N, et al. High PTP4A3 phosphatase expression correlates with metastatic risk in uveal melanoma patients. Cancer Res 2010;71(3):666-74.
22. Mollevi DG, Aytes A, Padulles L, et al. PRL-3 is essentially overexpressed in primary colorectal tumours and associates with tumour aggressiveness. Br J Cancer 2008;99(10):1718-25.
23. Wang Z, Cai SR, He YL, et al. Elevated PRL-3 expression was more frequently detected in the large primary gastric cancer and exhibits a poor prognostic impact on the patients. J Cancer Res Clin Oncol 2009;135(8):1041-6.

24. Zhao WB, Li Y, Liu X, Zhang LY, Wang X. Evaluation of PRL-3 expression, and its correlation with angiogenesis and invasion in hepatocellular carcinoma. Int J Mol Med 2008;22(2):187-92.
25. Wang H, Quah SY, Dong JM, Manser E, Tang JP, Zeng Q. PRL-3 down-regulates PTEN expression and signals through PI3K to promote epithelial-mesenchymal transition. Cancer Res 2007;67(7):2922-6.
26. Juric D, Lacayo NJ, Ramsey MC, et al. Differential gene expression patterns and interaction networks in BCR-ABL-positive and -negative adult acute lymphoblastic leukemias. J Clin Oncol 2007;25(11):1341-9.
27. Broyl A, Hose D, Lokhorst H, et al. Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients. Blood 2010;116(14):2543-53.
28. Fagerli UM, Holt RU, Holien T, et al. Overexpression and involvement in migration by the metastasis-associated phosphatase PRL-3 in human myeloma cells. Blood 2008;111(2):806-15.
29. Zhou J, Bi C, Chng WJ, et al. PRL-3, a metastasis associated tyrosine phosphatase, is involved in FLT3-ITD signaling and implicated in anti-AML therapy. PLoS One 2011;6(5):e19798.
30. Johnson WE, Li W, Meyer CA, et al. Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci U S A 2006;103(33):12457-62.
31. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

Table S1. Retroviral integrations.

Tumor name	Chr.	Region start (mm8)	Region stop (mm8)	Region Size	Confirmed by directed PCR and Sanger sequencing	Exact position integration determined with Sanger sequencing (mm8)	Nearest gene	Distance from gene	Flanked by methylated LTR as confirmed by methylated restriction analysis	Down regulated in tumor
Tumor1	chr4	131546819	131547867	1048	Yes	131547433	*Taf12*	1st intron	Yes	Not determined due to lack of material
Tumor1	chr8	75652987	75653413	426	No	xx	xx	xx	xx	xx
Tumor1	chr17	56357383	56358047	664	Yes	56357998	*Ranbp3*	379 bp upstream	Yes	Not determined due to lack of material
Tumor2	chr1	133792492	133792605	113	No	xx	xx	1st intron	xx	xx
Tumor2	chr2	164274288	164274798	510	No	xx	xx	xx	xx	xx
Tumor2	chr6	145071900	145072481	581	Yes	145071861	*Lrmp*	7037 bp upstream	Yes	No
Tumor2	chr16	36853190	36854178	988	Yes	36853575	*Hcls1*	647 bp upstream	Yes	No
Tumor2	chr17	34817452	34818970	1518	Yes	34818633	*Lta*	5230 bp upstream	No	xx
Tumor2	chr19	41047680	41048881	1201	Yes	41048491	*Blnk*	645 bp upstream	No	xx
Tumor4	chr2	85563399	85563837	438	No	xx	xx	xx	xx	xx
Tumor4	chr7	98577174	98577815	641	Yes	98577566	*Prkrir*	989 bp upstream	Yes	No
Tumor4	chr15	73575771	73576870	1099	Yes	73576038	*Ptp4a3*	1st intron	Yes	Yes
Tumor5	chr2	164334437	164334992	555	No	xx	xx	xx	xx	xx
Tumor5	chr3	20436615	20436752	137	No	xx	xx	xx	xx	xx
Tumor6	chr15	73575994	73576487	493	No	xx	xx	xx	xx	xx

Retroviral integrations identified with HAT are listed. For each integration the murine tumor and the genomic position are indicated as well as whether the integration could be confirmed with directed PCR and Sanger sequencing. For all integrations that could be confirmed, nearby located genes are given, their distance to the retroviral integration and whether the flanking viral integration was DNA methylated as analysed by methylation sensitive restriction analysis. Finally, for the 6 genes with a flanking DNA methylated viral integration site is indicated if they were down regulated in the respective tumor.

Table S2. Primers and probes.

Name	Sequence
bsLTRfw	GAGAAATAGGGAAGTTTAGATTAA
bsLTRrv	CCCAAAATAAACAATCAATCAATC
MN-LTR-fw	GGTTAAATAGGATATTTGTGGTGAGTAG
MN-LTR-rv	AACGAACTAATTAATTAATTCAAATAAAAC
MS-LTR-rv	CGAACAAAAACGAAAAACGAA
Probe-MN	FAM-AAACCATATCTAAAAACCATCTATTCTTACCCCC-TAMRA
*H19*ICR1fw	ACATTCACACGAGCATCCAGG
*H19*ICR1rv	GCTCTTTAGGTTTGGCGCAAT
*ActB*fw	AGCCAACTTTACGCCTAGCGT
*ActB*rv	TCTCAAGATGGACCTAATACG
LTRfw	AAAGACCTGAAACGACCTTGC
LTRrv	AAGGACCAGCGAGACCACG
mL1	CAACCTGGAAACATCTGATGG
mL2	CCCAAGAACCCTTACTCGGC
mL1N	CTTGAAACTGCTGAGGGTTA
mL2N	AGTCCTCCGATAGACTGTGTC
LTRfw2	CCAGGTTGCCCCAAAGACCTG
VIS(*Taf12*)	CAAGATCCGGGCTTTCAGAC
VIS(*Ranbp3*)	GACCAGGCTGCTCTCAAACG
VIS(*Lrmp*)	GGACACTACACTCATATTTG
VIS(*Lrmp_nested*)	GTGTGCTATGGGTAATTCAG
VIS(*Hcls1*)	TTCTCCTCCTTGCTTTCTGC
VIS(*Lta*)	CTAGGAGTCTTGTGCATCGTC
VIS(*Blnk*)	GAGGACAAGCCTAGTGATTTC
VIS(*Prkrir*)	CTGCTTGTTCACACAAAGTC
VIS(*Ptp4a3*)	CAGCCTCCTCTAGCAGTATC
Tbp fw	GCTGACCCACCAGCAGTTCAGTA
Tbp rv	AAGGAGAACAATTCTGGGTTTGA
Lrmp fw	CACAAGGCGAAGAGGCAGTG
Lrmp rv	GTGCTCTGTTGGCTCTTCTG
Hcls1 fw	CCCTTCTCTGTCCTACCAAG
Hcls1 rv	CCTTCATCCACCATCTCAAT
Prkrir fw	CTTACCAGTCATTTGAACAAC
Prkrir rv	CTTCAAGGGTTAAAGGCAGC
Ptp4a3 fw	CCATCCAGTTCATCCGACAG
Ptp4a3 rv	GACACAGATGTAATGAGGTAC

CHAPTER 7

HAT: HYPERGEOMETRIC ANALYSIS OF TILING-ARRAYS WITH APPLICATION TO PROMOTER-GENECHIP DATA

Erdoğan Taşkesen[1, 2], Renée Beekman[1], Jeroen de Ridder[2, 3, 4], Bas J. Wouters[1], Justine K. Peeters[1], Ivo P. Touw[1], Marcel J.T. Reinders[2, 3], Ruud Delwel[1]

[1]Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

[2]Delft Bioinformatics Lab (DBL), Delft University of Technology, Delft, The Netherlands

[3]Netherlands Bioinformatics Centre (NBIC), The Netherlands

[4]Bioinformatics and Statistics, Department Molecular Biology, Netherlands Cancer Institute, Amsterdam, The Netherlands

BMC Bioinformatics. 2010 May 21;11:275.

ABSTRACT

Background: Tiling-arrays are applicable to multiple types of biological research questions. Due to its advantages (high sensitivity, resolution, unbiased), the technology is often employed in genome-wide investigations. A major challenge in the analysis of tiling-array data is to define regions-of-interest, i.e., contiguous probes with increased signal intensity (as a result of hybridization of labelled DNA) in a region. Currently, there is no standard criteria to define these regions-of-interest as there is no single probe intensity cut-off level, different regions-of-interest can contain various numbers of probes, and can vary in genomic width. Furthermore, the chromosomal distance between neighboring probes can vary across the genome among different arrays.

Results: We have developed Hypergeometric Analysis of Tiling-arrays (HAT), and evaluated the performance for two different tiling-array datasets, a Chromatin Immunoprecipitation study on chip (ChIP) for the identification of genome-wide DNA binding profiles of transcription factor *cebpa* (used for method comparison). Using this assay, we can refine the detection of regions-of-interest by illustrating that regions detected by HAT are more highly enriched for expected motifs in comparison with an alternative detection method (MAT). In addition, retroviral insertional mutagenesis on chip data was used to examine the performance of HAT among different applications of tiling-array datasets. In both studies, detected regions-of-interest have been validated with (q)PCR.

Conclusions: We demonstrate that HAT has increased specificity for analysis of tiling-array data in comparison with the alternative method, and that it accurately detected regions-of-interest in two different applications of tiling-arrays. HAT has several advantages over previous methods: (i) as there is no single cut-off level for probe-intensity, HAT can detect regions-of-interest at various thresholds, (ii) it can detect regions-of-interest of any size, (iii) it is independent of probe-resolution across the genome, and across tiling-array platforms and (iv) it employs a single user defined parameter: the significance level. Regions-of-interest are detected by computing the hypergeometric-probability, while controlling the Family Wise Error. Furthermore, the method does not require experimental replicates, common regions-of-interest are indicated, a sequence-of-interest can be examined for every detected region-of-interest, and flanking genes can be reported.

BACKGROUND

Tiling-arrays are used for the identification of specific genomic DNA regions that can be enriched using various procedures to study certain molecular biological features. For example, DNA fragments that are bound by a protein of interest, e.g., a transcription factor can be enriched by using chromatin Immunoprecipitation (ChIP). When these

enriched fragments are hybridized to an array, a genome-wide protein binding profile can be obtained that is associated with this particular protein of interest in the cell type that was studied (ChIP-on-chip[1]). Other applications of tiling-arrays[2] are: Methylated-DNA immunoprecipitation (MeDIP-on-chip[3]), transcriptome mapping[4], recognition of hypersensitive sites such as segments of open chromatin that are cleaved more readily by DNaseI (DNase-chip[5]), or identification of copy number variations or breakpoints (Array CGH[6]). The use of tiling-arrays to detect enriched DNA regions has several advantages such as: (i) high sensitivity, which allows the detection of small DNA fragments associating with rare molecules, and (ii) high probe-resolution, which results in accurate acquisition of unbiased data.

A tiling-array is an array of short DNA fragments, which represent 'probes' that cover the entire genome, or contigs of the genome. The hybridization of labelled DNA to an array (for example DNA enriched using ChIP), will produce a quantitative signal intensity for each probe. Multiple contiguous probes with increased signal intensity across a particular genomic region, is a putative region-of-interest, and suggests the presence of a protein binding site.

As there are no standard criteria to accurately define a region-of-interest, a major challenge in the analysis of tiling-array data is to define such a region, and discriminate a positive signal from non-specific signals[7]. Defining regions-of-interest requires intensity thresholds on continuous probe intensity levels. Following this, the decision of the number of consecutive probes above the threshold needs to be made before a region-of-interest is called. This threshold, and the number of probes above the threshold directly influence the size of the region-of-interest that can be detected. As biologically relevant regions may vary in intensity, employing a single threshold is therefore insufficient. Additionally, as the probe-resolution varies across the genome, and across different tiling-array platforms, choosing a fixed number of consecutive probes as a region-of-interest is also inadequate. Various methods have been developed to detect regions-of-interest in ChIP-on-chip data such as Welch t-test, HMM, TileMap, MAT, Mixture model approach, CMARRT, Starr and Ringo[8-15]. MAT (Model-based analysis of tiling-arrays for ChIP-chip)[9] has been one of the most cited methods for analysing ChIP-on-chip data and it has been shown to outperform the following methods: Welch t-test, HMM and TileMap[8, 10, 12]. MAT uses various user-defined parameters to model a region-of-interest; such as maximum bandwidth, maximum gap size between probes, the minimum number of probes in a region and the use of a fixed threshold. A major limitation of this method is that it assumes a uniform probe-resolution across the genome, and depends on many user-defined parameters.

Here we propose a statistical framework (HAT: Hypergeometric Analysis of Tiling-arrays) to identify regions-of-interest in tiling-array data. Our method has several advantages over previous methods (such as MAT): (i) as there is no single cut-off level for probe-

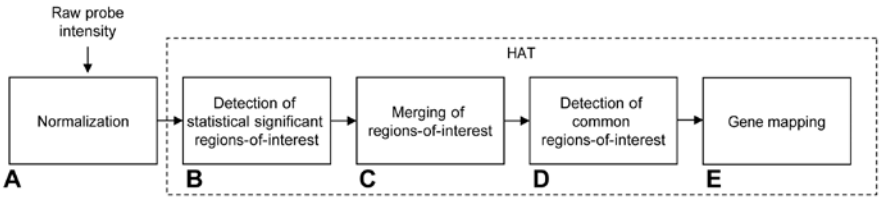


Figure 1. Illustration of the method. The different steps of the method, illustrated as blocks (A, B, C, D and E), are needed to process raw probe-intensity data, detection of unique candidate regions and mapping of the detected regions-of-interest to the 5′ transcriptional start site of nearby located genes. HAT is indicated with the blocks B, C, D and E. These are representative for the detection of unique candidate regions-of-interest in single, as well as multiple samples.

intensity; our method can detect regions-of-interest for a large number of thresholds, (ii) it can detect regions-of-interest of any size, (iii) it is independent of probe-resolution across the genome and across tiling-array platforms and (iv) it employs only a single user defined parameter: the significance level. The method can be seen as a generalization of the transcript discovery approach used in Bertone *et al*[4].

A detailed description of our framework (Figure 1) can be found in the method section. Briefly, instead of a single probe-intensity cut-off level, our method evaluates a large number of thresholds. Each threshold transforms the continuous signal intensity levels into discrete calls for each probe; referred to as positive probes where the probe intensity exceeds the threshold, and negative probes where it does not. In order to define regions-of-interest, all probes within the window of each positive probe are evaluated and the *p*-value is defined based on the ratio of both positive and negative probes using the hypergeometric distribution. To detect regions-of-interest of any size, the width of the window is also varied across all relevant window widths, where a relevant window is defined by the expected fragment size in the experimental procedure (e.g., due to sonication). The resulting regions-of-interest for each setting of the threshold and each window width are combined by taking the union of the significant window positions. The Family Wise Error (FWE) is controlled by employing a Bonferroni correction.

We have used two datasets using promoter tiling-arrays to evaluate our method. In the first assay, tiling-array data was employed to identify genome-wide DNA binding profiles of the transcription factor *cebpa*, in a cell line model. Using these data, we have shown that although our method detected less regions-of-interest than MAT, the detected regions are more highly enriched for *CEBP* binding motifs, and include known *cebpa* target genes. In the second experiment; a retroviral insertional mutagenesis assay, HAT identified novel putative transforming loci that may play a role in tumor development. Two of these loci were subsequently validated using PCR.

HAT can also detect and compare regions-of-interest across multiple samples. Each sample is analysed independently, but when multiple samples within one experiment

are used, detected regions-of-interest at the same genomic location among different samples are combined into 'common regions-of-interest', thereby increasing the confidence. In addition, our method can incorporate sequence information for the detection of pre-defined sequences (e.g., binding location within or near the region). These are highlighted in the graphical output for every detected region-of-interest and indicated in the output file.

RESULTS AND DISCUSSION

Data

Two distinct experimental datasets were used in this study: ChIP-on-chip data derived from an inducible *cebpa* expressing myeloid cell line model and data obtained from retrovirally integrated genomic DNA from retroviral induced murine leukemias. Data was generated using the Affymetrix GeneChip Mouse Promoter 1.0 Array. This chip generates 4.6 million perfect match probes over 28000 mouse promoter regions. Promoter regions cover 6Kb upstream to 2.5Kb downstream of 5' transcription start sites. Each probe has a size of 25nt.

Detection of regions-of-interest for C/ebpα chromatin immunoprecipitation by applying HAT

To compare different methods and to analyse the promoter array data, we made use of a dataset that was obtained from a ChIP of beta-estradiol induced *cebpa* in a myeloid cell line, 32D, followed by promoter array hybridizations. The data was used to examine the validity of detected regions-of-interest in two ways: (i) at the molecular sequence level; *cebpa* interacts with the nucleotide sequence *'CCAAT'* within the promoter regions represented on the chip, therefore *CEBP* binding motifs are expected to be enriched, and (ii) on the gene-level; examination of the presence of known *cebpa*-interacting genes, by taking the genes flanking the detected region-of-interest into account. Furthermore, one selected region-of-interest was validated by Real Time Quantitative PCR (qPCR).

The experimental setup was as follows: clones were derived from a myeloid cell line model (32D) that expresses either beta-estradiol inducible *cebpa*-ER (3 clones) or control ER (2 clones). Chromatin immunoprecipitations were carried out using an antibody directed against ER in the beta-estradiol treated cells and the DNA obtained from these cells, after immunoprecipitation, was hybridized to Affymetrix promoter chips.

For method comparison we used Model-based analysis of tiling-arrays for ChIP-chip (MAT), with the default parameters for the detection of regions-of-interest (bandwidth of 300bp; resulting in 2*bandwidth probe positions, 300bp of maximum gap size between positive probes, minimum of 8 probes for MAT-score, and enriched fragments at the $1x10^{-5}$ significance level). The default settings agree with the average sonicated fragment sizes, being 600bp, and the distance between two consecutive probes being

approximately 35bp. Using the default criteria in MAT, 4784 unique regions-of-interest were detected in at least one of the 32D-*cebpa*-ER clones (n=3) and absent in control samples 32D-ER (n=2). Using HAT, the same significance level and maximum fragment size ($1x10^{-5}$ and 600bp respectively) were chosen to detect statistically significant regions-of-interest. Applying these parameters, 1679 statistically significant regions-of-interest were detected in any of the 32D-*cebpa*-ER clones; 80% (1318) of these regions were detected in two or more clones (common regions-of-interest). This corresponds to 856 unique chromosomal regions-of-interest. HAT detected approximately one fifth of the regions-of-interest in comparison with MAT for the same significance level, and 99.9% (855) of these unique detected regions in HAT overlapped with the regions detected by MAT (Figure 2).

To investigate the validity of these detected regions-of-interest (for both HAT and MAT) on the sequence level, a motif enrichment analysis was performed. This was carried out using the Cis-regulatory Element Annotation System (CEAS)[16], where a *p*-value is computed for each known motif, and the motifs that are significantly enriched in the regions-of-interest are reported. The top 10 enriched motifs are indicated in Table 1 for both methods. These data showed that HAT detects regions that are highly enriched for

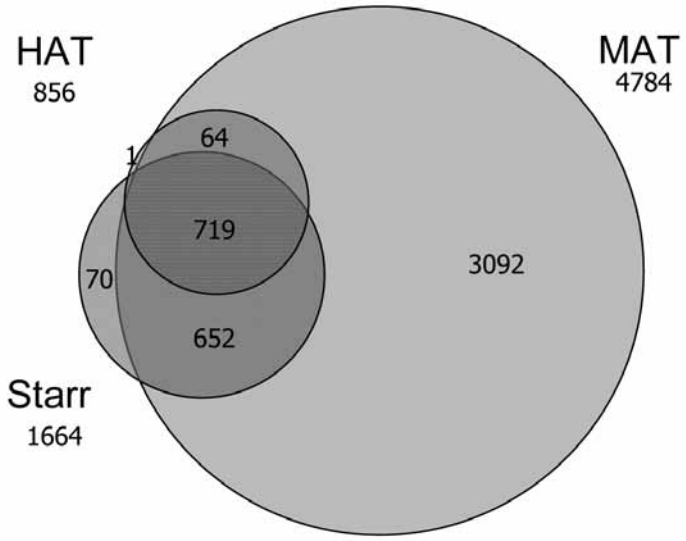


Figure 2. Venn-digram depiction the overlapping regions-of-interest between HAT, Starr and MAT. Detected regions-of-interest by HAT (856), Starr (1664) and MAT (4784) are indicated with the number of overlapping regions between the methods, circle sizes represent number of detected regions-of-interest. The overlap of regions detected by all three methods (719) showed high enrichment for *CEBP* binding motifs. Overlapping regions between HAT and MAT (64) and Starr and MAT (652) also showed high enrichment for *CEBP* binding motifs. Uniquely detected regions by Starr (70) showed no significantly enriched motifs, and MAT (3092) showed limited motifs enriched for *CEBP*. Note that the number of overlapping regions can contain multiple regions-of-interest detected by a single method.

Table 1. Motif enrichment analysis.

	MAT				HAT			
Nr	Motif	Hits	Fold-change	*p*-value	Motif	Hits	Fold-change	*p*-value
1	AP2alpha	9735	1.606	0.0	M00117.CEBPbeta	1532	2.325	2.84E-185
2	Elk-1	5380	1.707	9.23E-286	M00770.CEBP	3076	1.766	2.23E-183
3	M00470.AP-2gamma	5938	1.641	1.82E-274	M00912.C-EBP	3036	1.715	1.31E-164
4	M00109.CEBPbeta	6170	1.617	3.52E-269	cEBP	1928	1.965	1.89E-157
5	M00695.ETF	3449	1.885	3.05E-250	M00116.CEBPalpha	2689	1.722	4.30E-148
6	M00025.Elk-1	2979	1.949	1.76E-237	M00109.CEBPbeta	1278	2.161	2.35E-132
7	M00446.Spz1	4863	1.665	3.90E-237	M00190.CEBP	2402	1.719	9.67E-132
8	M00008.Sp1	5135	1.625	1.04E-228	M00098.Pax-2	1799	1.578	8.57E-73
9	E74A	3635	1.691	7.08E-188	M00496.STAT1	1909	1.545	8.88E-71
10	M00771.ETS	3756	1.674	2.37E-187	M00971.Ets	1917	1.508	4.42E-64

The top 10 motifs enriched in the detected regions-of-interest ($\alpha = 1x10^{-5}$) by HAT and MAT for the *cebpa*-study (ChIP-on-chip). Among the top 10 motifs enriched in the regions-of-interest detected with HAT, seven contained the *CEBP* binding motif whereas for MAT, only one contained the *CEBP* binding motif. For each reported motif, the number of hits within the regions-of-interest are counted, their fold change computed, and the *p*-value derived using the binomial test.

the *CEBP* motif binding sites, whereas MAT does not show a clear enrichment for these sites. Note that the detected regions-of-interest by HAT, are a subset of MAT.

To investigate detected regions-of-interest based on their flanking genes, regions-of-interest were mapped to the closest 5' transcriptional start site of a gene. Mapping is applied on the forward and reverse DNA strands, with a maximum distance of 300kb up- and down-stream (NCBI murine genome build 36). This resulted in 2174 unique genes for the 856 unique detected regions-of-interest using HAT (10.7% out of the total set of unique genes present in mouse). These mouse genes were subsequently overlayed with 169 known homologous human *cebpa* interacting genes (derived from Ingenuity Pathway Analysis, IPA), demonstrating that 40 *cebpa* interacting genes being detected by HAT ($p \leq 4x10^{-7}$) and 86 by MAT ($p \leq 3x10^{-5}$). Note; MAT has approximately five times more regions-of-interest resulting in 7238 unique genes (35.8% out of the total set of unique genes present in mouse). Some of the detected *cebpa* interacting genes have previously been described, such as: *myc, hp, mpo and il6ra*[17-20]. Enrichment of the *il-6 receptor alpha* (*il6ra*) transcriptional start site (Figure 3) was subsequently validated by qPCR.

An alternative comparison can be performed using the number of regions-of-interest, instead of the significance level. For HAT; 856 unique regions-of-interest were detected with a significance level $\alpha=1x10^{-5}$. To gain approximately the same number of regions-of-interest using MAT, we would need to set the α level at $1x10^{-19}$, resulting in 893 regions-of-interest. The overlapping regions-of-interest between HAT and MAT is 84% (718 out

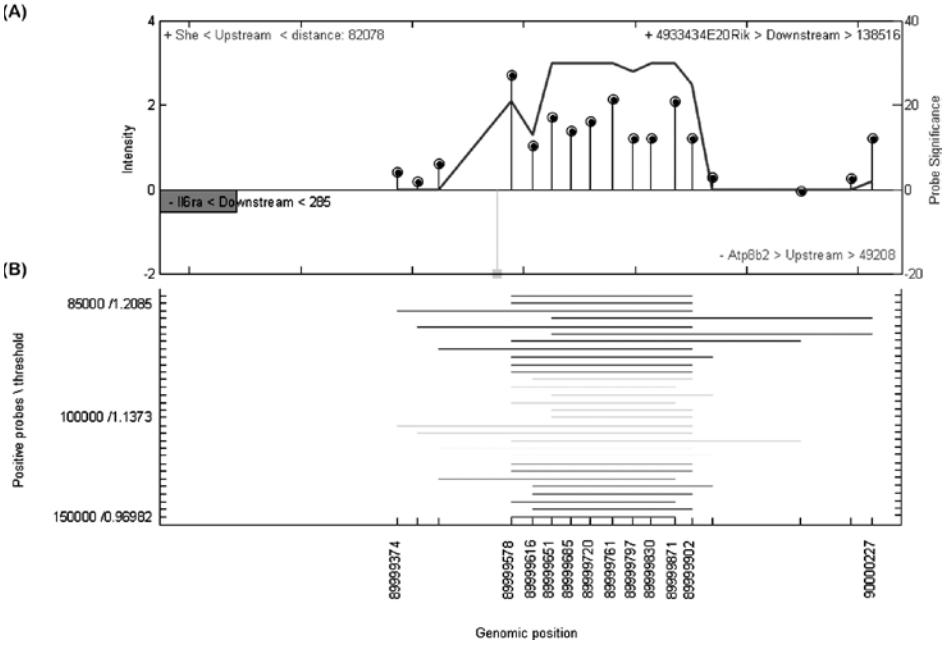


Figure 3. Graphical output of a detected region-of-interest from the *cebpa*-study. It was confirmed with qPCR that the *cebpa* protein targets and regulates the proximal promoter region of the *il-6 receptor alpha* gene, which lies downstream of the region-of-interest (negative DNA strand). The top panel (A), indicates the probes, represented as vertical lollipops, the left y-axis the probe-intensities, and the right y-axis illustrates the contribution of each probe separately to the region (probe-significance, indicated by the continuous grey line). The x-axis indicates the genomic probe positions, and illustrates with a downwards facing light grey bar, ending in a square; the sequence-of-interest. The sequence, 'CCAAT', was found on the negative DNA strand. Furthermore, flanking genes to this detected region are indicated with distances in base pairs to the 5' transcriptional start site. In the bottom panel (B), the detected regions-of-interest for various windows and probes are shown. They represent the detection of regions-of-interest, for a number of different top probes and window sizes. The merged region-of-interest has a fragment width of 853bp, and lies in the proximal promoter region of *il6ra* on the negative DNA strand.

of 856) and 83% (742 out of 893) respectively. Both methods show a high enrichment for the *CEBP* binding motifs. Comparing the detected regions-of-interest with respect to MAT (4827 with $\alpha=1x10^{-5}$), we need to set the α level higher than 0.05 in HAT, but this may compromise the reliability of detected regions-of-interest. For this reason, we have set the α level at 0.05 and hereby detected 1910 unique regions-of-interest. These were highly enriched for *CEBP* binding motifs based on the motif enrichment analysis (Table 2), whereas the detected regions-of-interest by MAT were not highly enriched for *CEBP* binding motifs (Table 1). The region-of-interest detected by HAT showed 98% (1879 out of 1910) overlap with MAT, whereas the overlap of detected regions of MAT with HAT was 39% (1874 out of 4784).

Table 2. HAT: Motif enrichment analysis using α = 0.05.

Nr	Motif	Hits	Fold-change	*p*-value
1	**M00117.CEBPbeta**	3236	2.082	6.19E-304
2	**M00770.CEBP**	6688	1.628	8.95E-299
3	**M00912.C-EBP**	6609	1.583	6.70E-265
4	**M00116.CEBPalpha**	5858	1.591	1.12E-239
5	**cEBP**	4068	1.758	1.88E-238
6	**M00190.CEBP**	5245	1.592	2.81E-215
7	**M00716.ZF5**	3927	1.706	2.12E-208
8	**M00109.CEBPbeta**	2645	1.896	3.06E-195
9	**M00098.Pax-2**	4355	1.619	1.76E-191
10	**M00428.E2F-1**	4374	1.572	4.67E-171

The top 10 motifs enriched in the 1910 detected regions-of-interest using HAT (α =0.05) in the *cebpa*-study. There is a high enrichment for binding motif *CEBP*. For each reported motif, the number of hits within the regions-of-interest are counted, their fold change computed, and the *p*-value derived using the binomial test.

In addition, the HAT and MAT results were also compared with the detected regions of Starr[15]. Starr implements the CMARRT algorithm[11] and thereby incorporates the correlation structure for the identification of regions-of-interest in tiling-array data. For the detection of regions-of-interest, we have utilized similar parameter settings (fragment size = 600bp, minimum number of probes in a region = 8 and $\alpha=1x10^{-5}$) as used in HAT and MAT. Using these parameter settings, Starr detected 1664 regions-of-interest and showed high enrichment for *CEBP* binding motifs (Table S1). Following this, we have examined the overlap of regions-of-interest detected by all methods as depicted in Figure 2. All regions-of-interest detected by HAT (except one) were also detected by MAT alone or together with Starr (64 and 791 respectively). Note that the number of overlapping regions can contain multiple regions-of-interest detected by a single method. To assess the validity of the detected regions-of-interest by HAT, Starr and MAT, we have examined the enrichment for *CEBP* binding motifs for the different parts in the venn-diagram, depicted in Figure 2. High enrichment for *CEBP* motifs are found for; (i) the overlap of HAT with the other two methods (n=719 regions-of-interest), (ii) the overlap of HAT with MAT (n=64 regions-of-interest) and, (iii) the overlap between Starr and MAT (n=652 regions-of-interest). No significant enriched motifs are found in the regions detected only by Starr (n=70 regions-of-interest) and limited motifs are enriched for *CEBP* in the regions detected only by MAT (n=3092 regions-of-interest). Therefore we can conclude that HAT showed to have the highest specificity as it was able to detect regions-of-interest highly enriched for *CEBP* binding motifs.

Detection of retroviral insertion sites by HAT

Retroviral Integration Mutagenesis (RIM) in mice is a powerful tool to identify new genes playing an important role in oncogenesis. Mice are injected with retroviruses that potentially integrate into the murine genome upon infection. Viral integration can lead to gene deregulation, and depending on the genes affected, tumors may develop. Genes located proximal to viral integration sites are potentially oncogenic, leading to tumor development. Genomic regions that have been targeted by proviral DNA in multiple tumors are called common viral integration sites (VIS), and are likely driving tumor development. Using retroviral insertional mutagenesis, many oncogenes have been identified using large sequencing screens in multiple tumors[21-24]. We hypothesise that within tumors, genes may be silenced as a result of proviral integration caused by hypermethylation of the CpGs in the viral long terminal repeat, and subsequently in the promoters of their target genes. The identification of methylated genes by means of retroviral insertional mutagenesis may be studied by Methyl-DNA immunopreciptitation (MeDIP-on-chip), followed by inverse PCR, using long terminal repeat (LTR) specific primers. After combining these two technologies, we hybridized samples to Affymetrix

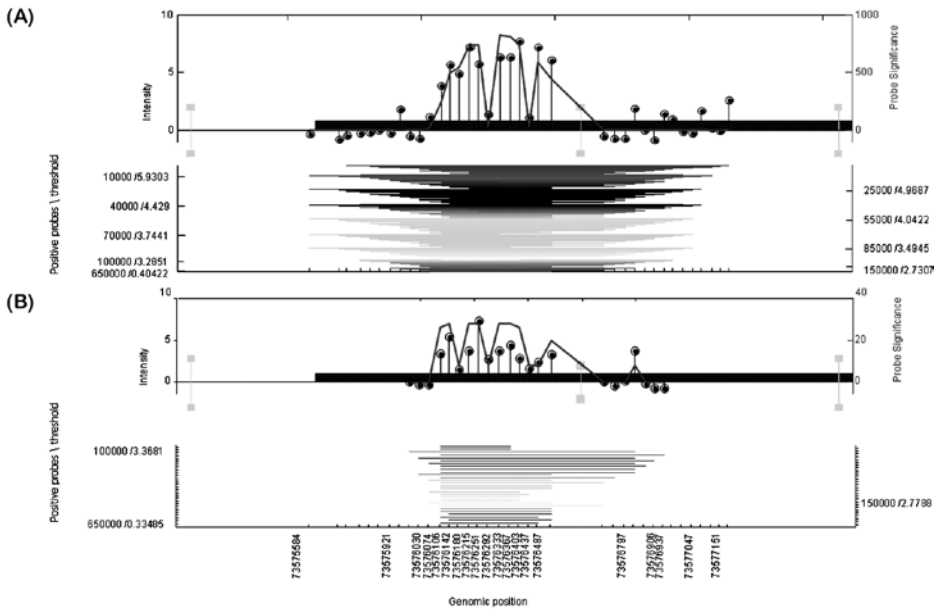


Figure 4. Graphical output of a detected cmVIS in the MeDIP-study. A region-of-interest detected in two samples, is illustrated in Panels A and B. Panel A shows 840 sub-regions that are merged with a total length of 1567bp. The restriction sites, indicated as light grey bars ending in a square, are located in and around the detected region, and are present on both DNA strands due to the palindrome sequence; *'GATC'*. The region-of-interest detected in the second tumor (Panel B), exists of 28 subregions, with a fragment width of 949bp.

promoter chips to identify genomic locations involved in viral integration that potentially harbour new tumor suppressor genes (TSG).

Regions-of-interest within this dataset differ from the *cebpa*-study as they have; (i) a higher variability in fragment sizes and, (ii) contain specific sequences within the identified regions. Therefore these data are used to examine the performance and broad applicability of HAT among different applications of tiling-array data. Using HAT, we have identified candidate TSGs in mouse tumors by considering regions with a maximum fragment size of 1000bp and a significance level $\alpha = 0.05$. With these parameters, we detected 15 methylated Viral Integration Sites (mVIS); of which one appeared to be a common methylated VIS (cmVIS) among two samples (Figure 4).

Besides the detection of candidate regions based on a statistical framework, we have attached additional mouse genomic sequence information (mm8) to the model,

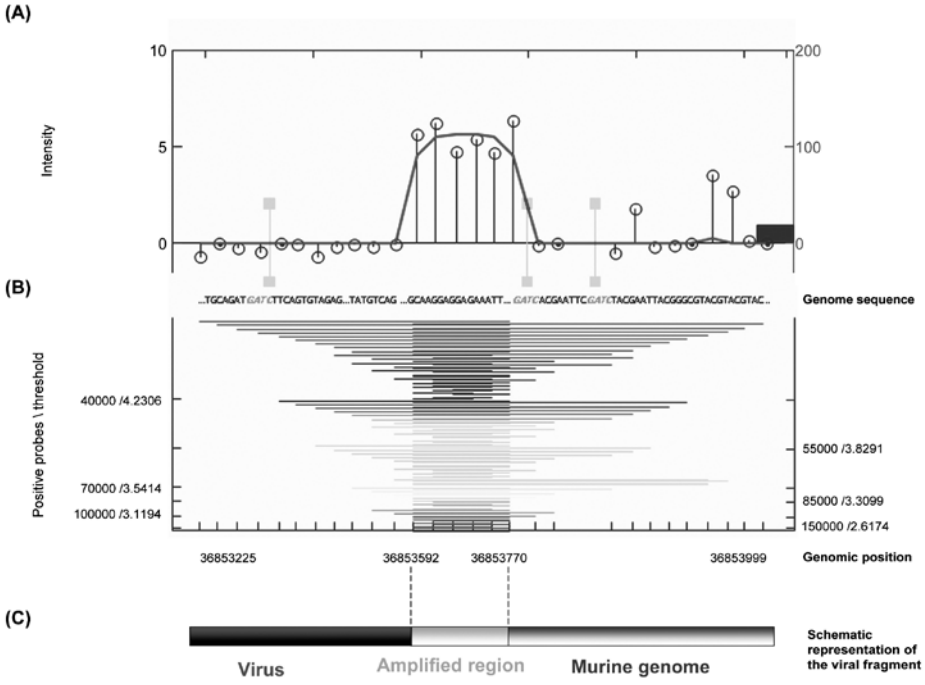


Figure 5. Graphical output of a detected and validated mVIS in the MeDIP-study. Panel A illustrates the detected mVIS which are subject to DNA methylation. Only a section of the detected region-of-interest has an increased probe-intensity; the probe-significance signifying this subregion. Directly beside the increased probe-significance, a restriction cleavage site is indicated by means of a light grey bar, ending in a square. Due to the palindrome sequence, these sites are indicated at the same genomic position on both DNA strands. Panel B shows the detected statistically significant regions among the different thresholds, and window sizes. A schematic representation of the amplified genomic region, with the virus- and the murine contribution, is shown in Panel C.

in order to determine the sequence-of-interest based on the restriction enzyme used in the inverse PCR. Within this assay, a restriction enzyme (DPNII) will cleave DNA at sequence 'GATC', within the integrated viral sequence and the flanking genome. Note, that because of this property, it is expected that every detected region must contain a nearby restriction site, which can be easily verified with HAT. HAT showed that all detected mVISs contain a nearby restriction site, confirming specificity of the identified region as being a viral insertion site. For PCR validation of the method, two mVISs were selected based on their location to a nearby 5' transcriptional start site, and confirmed. One of the validated regions is illustrated in Figure 5.

Extended applications of HAT

The scope of this method is not limited to the presented studies (i.e., detecting transcription factor binding sites and DNA methylated regions). Moreover, we have successfully applied HAT for the detection of regions enriched for histone modifications such as, trimethylation of histone 3 at lysine 4 or lysine 27 (H3K4me3 and H3K27me3) (data not shown). Some of the detected regions-of-interest were selected for further validation and confirmed by qPCR. Regarding tiling-array data spanning the entire genome[25] (e.g., RNA transcript mapping data[4], we do not expect changes in algorithm performance (detection of regions-of-interest) due to an increased variability in hybridization consistency since the applied normalization method[9, 26] corrects for two major causes of differences in hybridization consistency, i.e., probe sequence and presence of repeats within the genome. Furthermore, in addition to one-colour arrays (e.g., Affymetrix tiling-arrays) we envision that HAT can also be applied on data stemming from two-colour arrays (e.g., Nimblegen tiling-arrays) because data structure remains similar. We stress however that the normalization procedure is an important step and strongly depends on the type of tiling-array dataset.

CONCLUSIONS

Here we propose a statistical framework; HAT (Hypergeometric Analysis of Tiling-arrays) to analyse tiling-array data. We show that the method is robust and has increased specificity in the detection of regions-of-interest in comparison with two alternative methods. This is achieved by computing the hypergeometric-probability for every detected region-of-interest, among different threshold levels of probe-intensities, and window sizes while keeping control of the Family Wise Error (FWE) by employing Bonferroni correction. Besides the detection of regions-of-interest, HAT also determines sequences-of-interest, flanking genes and the distances to 5' transcriptional start sites among both DNA strands. We describe the performance of HAT, when applied to different experimental tiling-array datasets. For each experimental dataset, the selected downstream genes flanking the detected regions-of-interest are successfully confirmed

by (q)PCR. We compared the detected regions-of-interest of HAT with two other methods (MAT[9] and Starr[15]), and showed that HAT resulted in a reduced number of detected regions-of-interest using the same significance for both MAT and Starr. However, using motif enrichment analysis we showed that the regions-of-interest detected by HAT were more enriched for the expected binding motifs, i.e., *CEBP*, compared to MAT and showed similar enrichment for Starr, illustrating increased specificity using HAT.

Besides analysing ChIP-on-chip data, HAT is also suitable for the analysis of other types of tiling-array data. Applying HAT to the data from the MeDIP inverse-PCR and promoter-GeneChip hybridization experiment, we discovered mVIS and cmVIS that are subject to DNA methylation and identified the genes (unpublished data) that flank these methylated viral integration sites (Figure 4 and 5).

HAT is applicable to detect regions-of-interest among the different applications of tiling-arrays, and has the advantage of being independent for thresholds, number of probes in a region and probe-resolution. It does not depend on setting various user defined parameters, except for the significance level and an optional maximum fragment size.

METHODS

Extracting candidate gene-regions based on high throughput data using tiling-arrays is a multi-step process (Figure 1). The first step is to normalize the probe-intensity data from the chip (Figure 1A). For this purpose, we utilize the normalization from Model-based analysis of tilling-arrays for ChIP-chip (MAT)[9, 26], but other normalization procedures can also be applied. The normalization procedure prevents systematic variation between experimental conditions, which are unrelated to biological differences. As a result of this normalization, the probe-intensity values follow a normal distribution with a negative mean; hence the majority of probes have values below zero, and are ignored in all subsequent analyses. Probe-intensities that may be the result of hybridization of labelled DNA on the chip (e.g., were present in the immunopreciptitated chromatin sample), have values greater then zero and are used to determine candidate regions-of-interest.

After normalization, probe-intensities are discretized using a varying threshold and the significance of the probes within a varying window is determined. Significant window positions are then merged into the final regions-of-interest. We illustrate this approach in the simplified schematic representation shown in Figure 6. In Figure 6A, eight probes are shown at an arbitrary genomic location. Their intensities are represented by vertical lollipops. The positive probes (six in this example) are assumed to be part of a possible candidate region. Probes with higher intensity levels are more likely to be the results of hybridization on chip, but the exact level of intensity for which this is the case is unknown. Therefore, multiple probe intensity levels are taken into account by varying the discretization threshold *t*. The number of probes that exceed this threshold (called

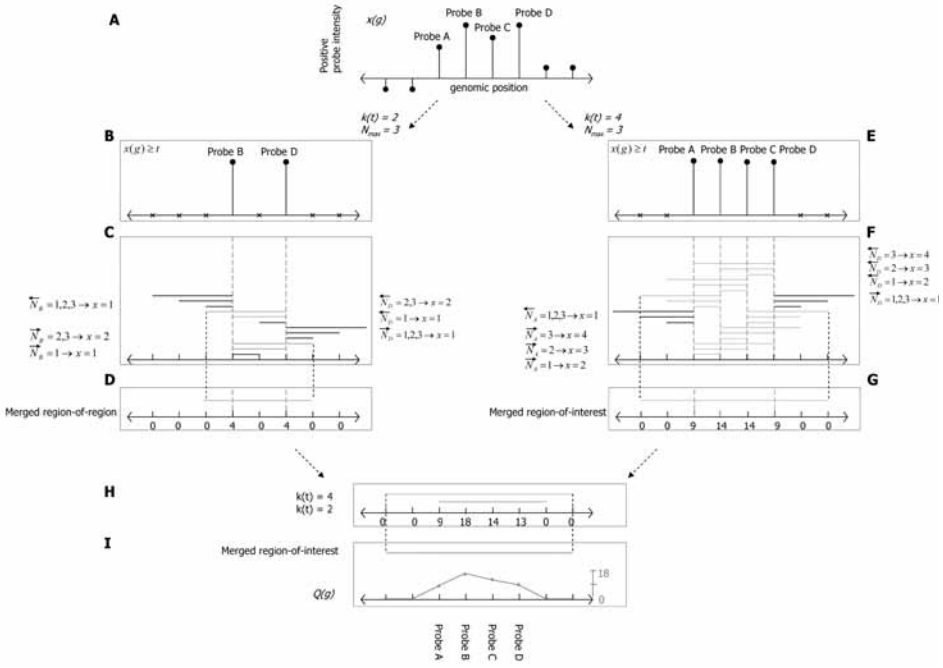


Figure 6. Schematic depiction for the detection of regions-of-interest. Schematic depiction for the detection of regions-of-interest, based on probe-intensities. Eight probes, with their genomic location, are shown in Panel A. Four of these have positive probe-intensities. The use of multiple thresholds, transforms continuous data into discrete data; as shown in Panel B and E. Various window scales *N*, are used to examine neighboring probes for their probe-intensities in Panel C and F. These windows will contain different number of positive probes. The hypergeometric probability is computed for every region-of-interest, and excludes a region-of-interest when the region is not statistically significant (black lines, C and F) after correcting for a single positive probe in a region-of-interest and multiple testing. The remaining statistically significant regions (light grey, C and F) are merged for each *k(t)* (illustrated in Panel D, G, H) and then among all *k(t)* to a single region-of-interest (Panel I). To determine how often probes were detected in statistically significant regions, the probe-significance is computed (Panel D and E), and indicated with continuous line that signifies the statistically significant probes in the detected region-of-interest.

positive probes) is denoted by *k(t)*. Figure 6B and 6E, illustrates the thresholds *k*(*t*)=2 and *k(t)*=4, respectively. All probes exceeding *t* are set to one, and those not exceeding the threshold *t* are set to zero.

To define a region-of-interest, we determine the significance of all possible window positions *g*, for which the window contains at least one positive probe. To account for the fact that the exact number of probes in a region-of-interest is undefined, and may differ greatly between different regions-of-interest due to differences in local probe-resolution; the window width *N* is varied. To prevent evaluating many highly similar windows, thereby incurring a high multiple testing penalty, only those window widths for which the number of probes in the window varies are evaluated. Therefore, *N* is de-

fined in terms of the number of probes contained in the window. The number of positive probes in a window of width *N*, at genomic position *g*, for threshold value *t*, is denoted by *x(g,t,N)*. In the example presented in Figure 6, we varied *N* from 1 through to 3. For the case *k(t)*=2 (Panel B and C), *x(g,t,N)* ranges from 1 through to 2, and in case *k(t)*=4 (Panel E and F), *x(g,t,N)* ranges from 1 through to 4.

For each window, a *p*-value is determined; defined as the probability of observing at least *x* positive probes in the window. For any window position *g*, threshold level *t* and window width *N*, *p(g,t,N)* is computed as:

$$p(g,t,N) = P(X \geq x | g,t,N, x \geq 1) = \frac{P(X \geq x | g,t,N)}{P(X \geq 1 | g,t,N)} \tag{1}$$

Note that since we restrict each window to contain at least one positive probe to prevent evaluating useless window positions, this probability is conditioned on $X \geq 1$. All probabilities are computed using the hypergeometric distribution:

$$P(X \geq x | g,t,N) = 1 - \sum_{0}^{x-1} \frac{\binom{k}{x}\binom{K-k}{N-x}}{\binom{K}{N}}, \tag{2}$$

where *K* is a fixed parameter and represents the total number of probes present on the (e.g., promoter) chip. To correct for the number of tests performed, we apply Bonferroni correction, controlling the Family Wise Error per value of the threshold level as follows:

$$p^*(g,t,N) = p(g,t,N) \cdot k(t) \cdot N \tag{3}$$

Based on this *p*-value, it is possible to exclude regions that do not reach a predefined significance level (α):

$$S(g,t,N) = \begin{cases} 1 & \text{if} \quad P^*(g,t,N) \leq \alpha \\ 0 & \text{else} \end{cases} \tag{4}$$

Due to the use of various values for *t* and *N*, similar or partly overlapping regions are found. In order to find a single region-of-interest at the same genomic location, these overlapping regions are merged by joining regions with one or more overlapping probes. In our example, we assume for simplicity, that windows with *x(g,t,N)* ≥ 2 are statistically significant. These statistically significant regions are indicated in light grey in Figure 6C and Figure 6F. The merging procedure is illustrated in Figure 6D, where four regions are merged into a single region, and in Figure 6G where 18 regions are merged.

Finally, regions found for different threshold levels *t* are also merged (Figure 6H) into the final region-of-interest (Figure 6I). Regions-of-interest tend to be larger than the regions detected at a single setting of the threshold level, or single window width due to the merging of all these individual regions. To determine the most important parts of the region-of-interest, we introduce a probe-significance score *Q(g)*, which reports how often probes were part of the statistically significant region. This score is illustrated by the curve in Figure 6I, and computed as follows:

$$\begin{aligned} \mathrm{Q}(g) &= \textstyle\sum_{\forall t}\sum_{\forall N} S(g,t,N)\cdot I(x(g,t,N),t) \\ \text{where} & \\ I\left(x(g,t,N)\right) &= \begin{cases} 1 & \text{if} \quad x(g)\geq t \\ 0 & \text{else} \end{cases} \end{aligned} \tag{5}$$

In our example so far, regions are detected within a single sample. When multiple samples are available (for the same experiment), array-wise detection of regions-of-interest is examined in order to detect common regions-of-interest (Figure 1D). A radius, defined in base pairs, can be defined to set the maximum distance between regions over multiple samples (default is zero).

Additional properties of HAT

The HAT method includes two additional properties beside the detection of regions-of-interest; (i) The determination of sequences-of-interest surrounding and within the detected regions-of-interest, e.g., the enhancer binding protein *cebpa* is known to interact with *'CCAAT'* sequences, and it is therefore expected that detected regions-of-interest contain the *CEBP* sequence in a chromatin IP experiment. The presence, and positions of the sequences-of-interest can be indicated in the (graphical) output of HAT. In this graphical output, sequences are indicated with an upward facing grey bar, indicating that the sequence is detected on the positive strand, or a downward facing grey bar representing a sequence on the negative strand. (ii) The determination of genes flanking the detected regions-of-interest. For every detected region-of-interest (for both up- and down-stream and forward and reverse DNA strands), the genes with the closest distance to the transcriptional start site are determined, and indicated in the (graphical) output.

To include these regions-of-interest and genes into the HAT method, the public genome-sequence (available for different model systems) can be utilized from the UCSC genome browser.

AVAILABILITY AND REQUIREMENTS

HAT is implemented in Matlab R2009b and is tested on Unix and MS-Windows. It requires an installation of Matlab and the scripts used for implementation of the method are available on *http://www.erasmusmc.nl/hematologie/*. The runtime depends on the number of used threshold cut-off's as the computation complexity increases linear with the used number of probes for the detection of regions-of-interest. In addition, runtime also depends on the different steps in the method (Figure 1 B-F). On average, for the *cebpa*-study, 28 minutes were needed per sample for the detection of regions-of-interest, while MAT required on average a runtime of 23 minutes per sample. Note, however, that in our algorithm the data was analysed using a multitude of window sizes and thresholds. A more detailed overview of the runtime for each step in the method can be found in the supplementary material (Figure S1).

ACKNOWLEDGEMENTS

The authors thank Erik van den Akker, Martin van Vliet and Mathijs Sanders for the discussions. This research is supported by the Center for Translational Molecular Medicine (CTMM), the Netherlands Genomics Initiative (NGI) and the Dutch Cancer Society (KWF Kankerbestrijding).

REFERENCES

1. Aparicio O, Geisberg JV, Struhl K. Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. Current protocols in cell biology / editorial board, Juan S Bonifacino [et al 2004;Chapter 17:Unit 17 7.
2. Liu XS. Getting started in tiling microarray analysis. PLoS computational biology 2007;3(10):1842-4.
3. Weber M, Davies JJ, Wittig D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature genetics 2005;37(8):853-62.
4. Bertone P, Stolc V, Royce TE, et al. Global identification of human transcribed sequences with genome tiling arrays. Science (New York, NY 2004;306(5705):2242-6.
5. Crawford GE, Davis S, Scacheri PC, et al. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nature methods 2006;3(7):503-9.
6. Heidenblad M, Lindgren D, Jonson T, et al. Tiling resolution array CGH and high density expression profiling of urothelial carcinomas delineate genomic amplicons and candidate target genes specific for advanced tumors. BMC medical genomics 2008;1:3.
7. Royce TE, Rozowsky JS, Bertone P, et al. Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping. Trends Genet 2005;21(8):466-75.
8. Ji H, Wong WH. TileMap: create chromosomal map of tiling array hybridizations. Bioinformatics (Oxford, England) 2005;21(18):3629-36.

9. Johnson WE, Li W, Meyer CA, et al. Model-based analysis of tiling-arrays for ChIP-chip. Proceedings of the National Academy of Sciences of the United States of America 2006;103(33):12457-62.
10. Keles S, van der Laan MJ, Dudoit S, Cawley SE. Multiple testing methods for ChIP-Chip high density oligonucleotide array data. J Comput Biol 2006;13(3):579-613.
11. Kuan PF, Chun H, Keles S. CMARRT: a tool for the analysis of ChIP-chip data from tiling arrays by incorporating the correlation structure. Pacific Symposium on Biocomputing 2008:515-26.
12. Li W, Meyer CA, Liu XS. A hidden Markov model for analyzing ChIP-chip experiments on genome tiling arrays and its application to p53 binding sequences. Bioinformatics (Oxford, England) 2005;21 Suppl 1:i274-82.
13. Sun W, Buck MJ, Patel M, Davis IJ. Improved ChIP-chip analysis by a mixture model approach. BMC bioinformatics 2009;10:173.
14. Toedling J, Skylar O, Krueger T, Fischer JJ, Sperling S, Huber W. Ringo--an R/Bioconductor package for analyzing ChIP-chip readouts. BMC bioinformatics 2007;8:221.
15. Zacher B, Kuan PF, Tresch A. Starr: Simple Tiling ARRay analysis of Affymetrix ChIP-chip data. BMC bioinformatics 2010;11:194.
16. Ji X, Li W, Song J, Wei L, Liu XS. CEAS: cis-regulatory element annotation system. Nucleic acids research 2006;34(Web Server issue):W551-4.
17. Ramji DP, Foka P. CCAAT/enhancer-binding proteins: structure, function and regulation. The Biochemical journal 2002;365(Pt 3):561-75.
18. Tinel M, Berson A, Elkahwaji J, Cresteil T, Beaune P, Pessayre D. Downregulation of cytochromes P450 in growth-stimulated rat hepatocytes: role of c-Myc induction and impaired C/EBP binding to DNA. Journal of hepatology 2003;39(2):171-8.
19. Wang W, Wang X, Ward AC, Touw IP, Friedman AD. C/EBPalpha and G-CSF receptor signals cooperate to induce the myeloperoxidase and neutrophil elastase genes. Leukemia 2001;15(5):779-86.
20. Zhang P, Iwama A, Datta MW, Darlington GJ, Link DC, Tenen DG. Upregulation of interleukin 6 and granulocyte colony-stimulating factor receptors by transcription factor CCAAT enhancer binding protein alpha (C/EBP alpha) is critical for granulopoiesis. The Journal of experimental medicine 1998;188(6):1173-84.
21. Erkeland SJ, Valkhof M, Heijmans-Antonissen C, et al. Large-scale identification of disease genes involved in acute myeloid leukemia. Journal of virology 2004;78(4):1971-80.
22. Suzuki T, Shen H, Akagi K, et al. New genes involved in cancer identified by retroviral tagging. Nature genetics 2002;32(1):166-74.
23. Theodorou V, Kimm MA, Boer M, et al. MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. Nature genetics 2007;39(6):759-69.
24. Touw IP, Erkeland SJ. Retroviral insertion mutagenesis in mice as a comparative oncogenomics tool to identify disease genes in human leukemia. Mol Ther 2007;15(1):13-9.
25. Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. Genomics 2005;85(1):1-15.
26. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics (Oxford, England) 2003;19(2):185-93.

SUPPLEMENTARY FIGURE

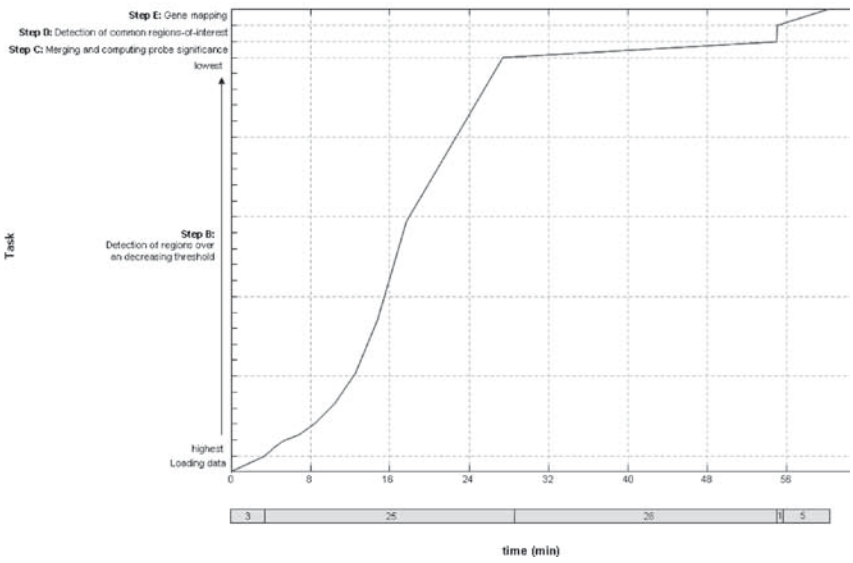


Figure S1. HAT Computation performance. Runtime of the various steps in the method.The *cebpa*-study is used to analyse the runtime for the different steps in the method; Step B: loading data and detection of regions-of-interest, Step C: Merging of regions-of-interest and computation of the probe-significance, Step D: detection of common-regions-of-interest and Step E: gene mapping. Per sample, 62 minutes were needed on average to process all the steps in the method.

SUPPLEMENTARY TABLE

Table S1. Starr: Motif enrichment analysis.

Nr	Motif	Hits	Fold-change	*p*-value
1	**M00117.CEBPbeta**	1992	2.153	2.74E-203
2	**M00912.C-EBP**	4084	1.643	9.46E-190
3	**cEBP**	2570	1.866	2.94E-181
4	**M00770.CEBP**	3980	1.627	1.90E-178
5	**M00116.CEBPalpha**	3523	1.607	7.15E-151
6	**M00109.CEBPbeta**	1670	2.011	5.45E-145
7	**M00190.CEBP**	3138	1.599	3.13E-132
8	**HLF**	666	2.023	3.80E-60
9	**M00260.HLF**	608	2.017	9.94E-55
10	**M00771.ETS**	845	1.731	6.80E-49

The top 10 motifs enriched in the 1664 detected regions-of-interest using Starr (fragment size = 600bp, minimum number of probes in a region = 8, $\alpha = 1x10^{-5}$) in the *cebpa*-study. There is a high enrichment for binding motif *CEBP*. For each reported motif, the number of hits within the regions-of-interest are counted, their fold change computed, and the *p*-value derived using the binomial test.

CHAPTER 8

SUMMARY AND GENERAL DISCUSSION

1. SUMMARY

The work presented in this thesis focuses on the identification of new genetic and epigenetic abnormalities involved in acute myeloid leukemia (AML) and their potential role in leukemogenesis. To this end, two different leukemogenesis models were used; severe congenital neutropenia (SCN) in humans, a pre-leukemic condition which may progress towards AML, and retroviral integration mutagenesis in mice, a model in which murine leukemias are induced by oncogenic viruses.

In **chapter 1**, a general introduction to SCN and AML is given, including an overview of the known molecular mechanisms underlying these diseases. SCN is a bone marrow failure syndrome characterized by severely reduced neutrophils counts, recurrent infections and a high risk to develop AML. The neutropenia in SCN patients is successfully treated with administration of granulocyte-colony stimulating factor (G-CSF), however leukemic progression remains a major concern. The role of G-CSF and its receptor in myeloid malignancies and related to this, the molecular mechanisms behind leukemic transformation in SCN patients are discussed in **chapter 2**. Mutations in the gene encoding the G-CSF receptor (colony stimulating factor 3 receptor, *CSF3R*), leading to the expression of truncated receptors that are hyper responsive to G-CSF, as well as trisomy 21 and monosomy 7 are so far the only common aberrations linked to leukemic progression in these patients. While underlying aberrations in SCN/AML remain largely unknown, much more is known about *de novo* AML. Current deep-sequencing analyses show that *de novo* AML is a highly heterogeneous disease; many different genetic and epigenetic alterations are detected in AML, few of which are common while the large majority has a low prevalence. In the last part of **chapter 1**, genome-wide techniques are introduced with a main focus on deep-sequencing; their advantages, disadvantages and caveats are discussed.

Chapters 3 to 5 focus on the identification and functional characterization of new mutations involved in progression of SCN towards AML. In **chapter 3**, whole exome sequencing was employed to identify mutations in longitudinal samples of an SCN patient that developed AML after 17 years of G-CSF treatment, further referred to as the SCN index patient. Twelve acquired non-synonymous mutations were detected in the leukemic phase. Interestingly, 3 of these mutations were already present in low frequencies 15 years before the AML became overt. One of these mutations was a *CSF3R-d715* truncating mutation which was previously identified in this patient, while the other 2 were new mutations in *LLGL2* and *ZC3H18*. Furthermore, it is shown that these "early" mutations co-existed in the same pre-leukemic clone and that the *ZC3H18* mutation was acquired later in time than the mutations in *LLGL2* and *CSF3R*. For the first 2 mutations, no hierarchical pattern of acquisition could be determined. The presence of these 3 mutations in the early SCN phase suggests that they play an important role in the early steps of leukemic progression. Nine other mutations were only detected in the leukemic

blasts of the SCN index patient. Strikingly, 1 of these was a yet additional mutation in the extracellular domain of CSF3R, further referred to as the *CSF3R-T595I* mutation. This mutation was located on the already affected *CSF3R-d715* allele and conferred G-CSF independent growth. In **chapter 4**, it is shown that the *CSF3R-T595I* mutation is present in a low frequency in *de novo* AML patients, however not in combination with *CSF3R* truncating mutations. Furthermore, we show that the *CSF3R-T595I* mutant confers G-CSF independent nuclear translocation of transcription factors STAT3 and STAT5. Of the other "late" mutations present in the leukemic blasts of the SCN index patient, *ASXL1, EP300, RUNX1* and *SUZ12* were previously reported to be mutated in hematopoietic malignancies, while mutations located in *CCDC155*, *FBXO18, LAMB1* and *MGA* were not. Although, unpublished results of The Cancer Genome Atlas consortium, containing deep-sequencing data of 199 AML patients, show that mutations in *MGA* and *CCDC155* are rare but present. In **chapter 5**, the presence of genetic alterations in the mutated genes identified in the SCN index patient as well as in their paralogues and interactions partners was determined in a larger cohort of 26 SCN patients, of which 5 progressed towards leukemia. This analysis demonstrated that mutations in the identified genes, except in *CSF3R*, are rare in the SCN phase; of the additional genes only mutations in *LAMB2* was detected in 1 patient. The only common mutations found during the leukemic phase were *RUNX1* mutations in combination with mutations in *CSF3R*.

In **chapter 6**, a murine retroviral integration mutagenesis model was adopted to identify new haplo-insufficient genes involved in AML. This model is briefly introduced in **chapter 1**. Haplo-insufficiency occurs when normal expression levels or the functional activity of a gene/protein is reduced. Decreased expression and/or reduced functional activity of haplo-insufficient genes/proteins contribute to malignant transformation. Retroviral integration mutagenesis is a powerful tool to discover new genes involved in oncogenesis. To specifically screen for haplo-insufficient genes we aimed to detect DNA methylated viral integration sites and their flanking genes as DNA methylation is a known mechanism of gene silencing. In order to do this, inverse PCR, to amplify retroviral integration sites, methylated DNA immunoprecipitation, to enrich for DNA methylated fragments and promoter array hybridization were combined to locate the immunoprecipitated fragments. To adequately analyze these data, a new algorithm named "hypergeometric analysis of tiling arrays" (HAT) was developed (**chapter 7**). This method was specifically designed to detect enriched regions of different lengths and enrichment values. The analysis led to the identification of the potential haplo-insufficient gene *Ptp4a3*, which was flanked by a DNA methylated viral integration site while its expression was reduced compared to expression levels in normal murine bone marrow. Further analysis showed that *PTP4A3* expression was of prognostic significance in a cohort of AML patients diagnosed under the age of 60, predicting increased tumor aggressiveness upon high *PTP4A3* expression.

2. GENERAL DISCUSSION

In this final part, the different findings described in the thesis will be placed in perspective of the current knowledge and their possible implications will be addressed. Additionally, a perspective will be given on leukemia research in the future, with a focus on using and interpreting whole exome and whole genome sequencing data.

2.1. Leukemogenesis in severe congenital neutropenia

We had the unique opportunity to study sequential hematopoietic samples from a SCN patient from the childhood SCN phase to overt AML, spanning a period of 15 years. This allowed us to identify the "early" and "late" genetic defects associated with leukemic progression in this SCN patient.

2.1.1. Hierarchical acquisition of mutations and dynamics of pre-leukemic clones

In the SCN/AML case described in this thesis, further referred to as the SCN index case, we used acquired mutations detected in the leukemic blasts as a starting point to find "early" mutations present in the SCN phase of the disease. A drawback of this approach is that it does not allow monitoring the dynamics, i.e., appearance, expansion and disappearance, of clones that harbor other mutations. Despite this limitation, we detected distinct clones with different *CSF3R* delta mutations in the SCN phase that were not found in the leukemic blasts, a phenomenon which has been described previously[1]. Should these clones be considered pre-leukemic? The existence of clones with different *CSF3R* mutations highly suggests that leukemic transformation is an evolutionary "Darwinian" process characterized by acquisition of mutations in different hematopoietic progenitors followed by a selective outgrowth of clones with specific combinations of mutations, as was addressed previously by M. Greaves[2]. In line with this reasoning, we should consider clones harboring mutations to be pre-leukemic. However, their selective advantage and potential to transform into a full blown leukemia varies extensively and depends on (i) the aberrations they harbor, (ii) the acquisition of additional alterations that may or may not arise as a consequence of preceding aberrations and (iii) extrinsic selective pressure. Hence, identifying and monitoring these pre-leukemic clones will give insights into the dynamic process of leukemic evolution and specifically into which aberrations, alone or in combination, are potentially vital in leukemogenesis.

Initial mutations found in the SCN index case in progression towards AML were the *CSF3R-d715* and the *LLGL2* mutation, followed by a mutation in *ZC3H18*. The *CSFR-d715* mutation is known to cause hyperproliferation of myeloid progenitors in cell line and animal models[3, 4]. Important to note is that *CSF3R* delta mutations always arise in the context of a SCN underlying mutation, e.g., in *ELANE*, *HAX1* or *WAS*, while studies performed to investigate the functional consequences of *CSF3R* delta mutations were so far not performed in the context of these mutations. Hence, we cannot state with

certainty that expansion of the pre-leukemic clone during the SCN phase is caused by the *CSF3R-d715* mutation alone. It is more likely that the "early" mutations in *LLGL2* and *ZC3H18* contribute to expansion and leukemic progression of the clone carrying the *CSF3R-d715*, as clones carrying other *CSF3R* delta mutations disappear in the leukemic phase. This hypothesis is supported by the finding that the *ZC3H18* mutation is found in a frequency which is approximately 6-7 times lower than the mutation frequency of *LLGL2* and *CSF3R* in the early SCN sample (single colony assay, Chapter 3, Figure 2 and custom capture approach, mutated reads *ZC3H18* = 1%, *LLGL2* = 6%, *CSF3R* = 7%), while the frequency in the intermediate SCN sample is equal (custom capture data, mutated reads *ZC3H18*, *LLGL2* and *CSF3R* = 11%). These data suggest that clones carrying the *ZC3H18* mutation have a proliferation and/or survival advantage compared to clones that do not. Interestingly however, the frequency of 2 other delta *CSF3R* clones (*d725* and *d730*) increase as well during the SCN phase, while the frequency of the *CSF3R-d717* clone does not (Chapter 3, Figure 2). This observation might represent normal fluctuations in hematopoietic stem cell proliferation but a different explanation might be that the *CSF3R-d725* and *-d730* clones harbor mutations in other genes, which may influence their proliferation and/or survival capacity. This hypothesis could be addressed by whole exome or whole genome sequencing (WES, WGS) of single bone marrow cells in the SCN phase of the index patient. With the same approach, it would be possible and of interest to investigate whether other pre-leukemic clones without *CSF3R* mutations exist.

The 9 year gap between the final investigated SCN sample and the leukemic blasts is a weakness in this study as the 9 "late" mutations may have occurred years before the AML became overt and could still represent "early" mutations. Hence, for these mutations it will be difficult to elucidate the hierarchical pattern of acquisition. From previous studies in ALL it is known that genetic alterations may arise in an hierarchical way, either in a linear or a branching pattern[5]. Furthermore, the existence of subclones in some but not all AML samples, based on mutation frequencies determined by deep-sequencing, suggests similar evolutionary patterns in human AML[6]. Hence, frequencies of mutations in the leukemic blasts of the SCN index patient, as determined by methods reaching a high coverage, e.g., amplicon-based sequencing or custom capture followed by deep-sequencing, may be informative. "Early" mutations are expected to be present in all leukemic cells as the leukemia arose from pre-leukemic cells carrying these mutations. In line with this hypothesis, the 3 "early" mutations are present in approximately 50% of the reads, indicative of heterozygous mutations in 100% of the leukemic blasts. Furthermore, we should consider that the mutations *in CSF3R-T595I, FBXO18, LAMB1* and *MGA*, respectively present in 40,3%, 24,9%, 31,9% and 16,5% of the reads as determined by amplicon-based sequencing, were acquired late in the process of leukemogenesis and may represent subclones. Custom capture followed by deep-sequencing similarly revealed a frequency of less than 50% for these 4 mutations (data not shown). WES or

WGS of single leukemic blasts could be employed to investigate the combinatorial pattern of these mutations in these cells in order to understand the hierarchical acquisition of these mutations in leukemic progression in further detail.

2.1.2. Biological significance of "early" mutations

In addition to studying their dynamics, we should ask what the functional relevance of "early" mutations, solely or in combination with other aberrations, might be. They may cause (subtle) changes in proliferation and differentiation capacity, like has been shown for the *CSF3R-d715* mutation[3, 4, 7]. Furthermore, they may influence stem cell potential, survival, cellular stress responses and/or genomic stability. Important to note is that early mutations may initially be important for disease progression but their effect may be overruled by other aberrations occurring later in time. Simultaneously, early mutations may be passengers rather than drivers in leukemogenesis. Passenger mutations coincide by chance with other aberrations in a single progenitor cell but do not contribute to leukemic transformation. A selective advantage caused by any of the other aberrations present in the same progenitor however may result in an increased frequency of these cells, which also carry the passenger mutation. In this way, passenger mutations may mistakenly be considered relevant for disease development. Until now, the functional consequences of both the *LLGL2* and the *ZC3H18* mutation remain unknown. *LLGL2* encodes the human orthologue of the Drosophila lethal giant larvae (*Lgl*) gene. Loss of Lgl in Drosophila leads to altered patterning of apoptosis in eye development as well as to a shift of asymmetric towards symmetric cell divisions of neuroblasts and tumor formation in the brain, caused by aberrant distribution of the cell polarity protein Numb[8-11]. Furthermore, epidermal loss of *Lgl* in Zebrafish results in epithelial-mesenchymal-transition (EMT)[12]. In mammals, *LLGL2* is implicated in the formation of protein complexes that regulate cell-polarity in epithelial tissues, like the PAR-3-αPKC-PAR-6 and the αPKC-PAR-6-LGN complex[13-15]. In line with this finding, loss of LLGL2 via suppression of the transcriptional repressor ZEB1 is implicated in loss of cell polarity, increased EMT and metastasis formation in human colorectal cancers[16, 17]. Additionally, LLGL2 is aberrantly distributed or lost in gastric dysplasia and adenocarcinoma[18, 19]. So far, indications for a role of LLGL2 in hematopoiesis and/or hematopoietic malignancies are weak. *LLGL2* mRNA expression levels are low to absent in CD34+ bone marrow cells, normal myeloid differentiation and AML (Verhaak et al.[20], S. Sun and M. Jongen-Lavrencic, data not shown). Deregulation of NUMB protein levels however has been implicated in blast transformation in chronic myeloid leukemia[21, 22]. Furthermore, over expression of Gpsm2 (a.k.a Lgn), a direct binding partner of LLGL2 in a complex together with αPKC and PAR-6, leads to increased hematopoietic reconstitution after transplantation in sublethally irradiated mice[15, 23]. In contrast, αPKC (aPKCζ and aPKCλ)

is dispensable for hematopoietic stem cell (HSC) self-renewal and reconstitution during serial competitive transplantation studies[24].

The *LLGL2* mutation found in the SCN index case is located in one of its WD40 domains and may disrupt direct interactions with PAR-6, leading to a shift in composition of polarity regulating protein complexes[14, 15]. To address this hypothesis, protein-protein interactions of wild type and mutant LLGL2, as well as the effect of their expression on cell polarity in epithelial *in vitro* models[13] could be studied. Based on the fact that loss of the *Lgl* gene affects asymmetric cell division in Drosophila[9] we may hypothesize that the *LLGL2* mutation affect the balance between symmetric and asymmetric cell divisions and proliferation potential of HSCs. This could be addressed by studying colony growth and replating capacity of early hematopoietic progenitors expressing mutant LLGL2, however, effects may be subtle and only competitive repopulation or serial transplantation studies may reveal the effect. Experiments conducted so far show that over expression of either wild type or mutant LLGL2 seems to have an unfavorable effect on colony formation of murine hematopoietic progenitors (data not shown). It is important to keep in mind that the *LLGL2* and *CSF3R-d715* mutations coincide in hematopoietic progenitors of the SCN index patient suggesting a co-operative or additive effect of these mutations. Possibly, the combination of the *LLGL2* mutation and the *CSF3R-d715* mutation may affect the balance between proliferation and differentiation respectively at the hematopoietic stem cell and the myeloid progenitor level. The combinatorial effects of these 2 "early" mutations might be investigated by studying the functional consequences of the *LLGL2* mutation both in a *CSF3R* wild type and a *CSF3R-d715* background. This could for example be done by crossing the *Llgl2* knockout mouse[25] with the *Csf3r-d715* knock in mouse[26], followed by studying the hematopoietic phenotype.

ZC3H18 has been reported to act as a putative RNA binding protein involved in differentiation in trypanosoma[27]. In mammals, it is predicted to be a nuclear protein, containing a ZnF_C3H1 domain. A similar domain has been found in the human splicing factor U2AF35, which is frequently mutated in MDS and AML[28]. Furthermore, ZC3H18 recently has been postulated to activate the NF-κB pathway by facilitating IκBα turnover[29]. Studying the interactions of wild type and mutant ZC3H18 with other proteins, RNA and chromatin as well as its effects on IκBα turnover and NF-κB signaling may result in further understanding of its role in normal and malignant hematopoiesis. One possibility is that mutated ZC3H18 induces a pro-survival effect by enhancing NF-κB signaling, which would complement the putative effects of mutated LLGL2 on stem cell renewal and increased proliferative signaling from mutated CSF3R.

2.1.3. Functional relevance of "late" mutations

Although the "early" mutations in the SCN index patient may play an initiating role in the early phase of SCN progression towards AML, they do not lead to full leukemic trans-

formation, as clones carrying these mutations existed already 15 years before the AML became overt. Hence, additional aberrations are essential for leukemic transformation. Strikingly, one of the "late" mutations was the *CSF3R-T595I* mutation that confers growth factor independence to myeloid progenitors; obtaining growth factor independence is an important evolutionary step in leukemogenesis. For the other "late" mutations we should consider what could be their functional significance. Because the frequency of some mutations suggests that these are present in subclones, we should assume that not all leukemic blasts contain all 12 mutations; it is likely that there is one major leukemic clone, containing the majority of mutations and additional subclones, carrying one or few extra mutations. Hence, the combination of 12 mutations is not essential for leukemic transformation and mutations in the major clone are likely to be more relevant for leukemogenesis than mutations in minor subclones.

Mutations in chromatin remodelers and transcription factors ASXL1, EP300, RUNX1 and SUZ12 are known in AML and/or hematopoietic malignancies and are therefore likely relevant for disease development[30-34]. SUZ12 interacts with EZH2 in the polycomb repressor complex 2 (PRC2) which is essential for methylation of histone 3 lysine 27 (H3K27). The mutation in SUZ12 leads to a duplication of an isoleucine in the VEFS-box, the domain involved in interaction with EZH2[35]. It would be of interest to investigate if the SUZ12 mutation causes an increase, a decrease or a loss of interaction with EZH2 and/or whether it may lead to a change in H3K27 methylation. Before proceeding to study the functional relevance of the other mutations in *CCDC155, FBXO18, LAMB1* and *MGA* it would be important to investigate if and how frequently they are mutated in additional AML patients. Both *MGA* and *CCDC155* are mutated in 1 out of 199 AML patients investigated by "The Cancer Genome Atlas (TCGA)" consortium[36] (Chapter 3, Table S3). This may still be regarded as rather weak evidence to claim that they are likely relevant for leukemic development and extensive analyses in larger AML cohort are essential to determine their exact frequency.

2.1.4. Recurrence of the newly identified mutations in SCN, leukemia and other disorders

It is still too early to predict if and in which patients we will find recurrence of the mutations identified in our SCN index case. It seems an obvious choice to start with SCN and SCN/AML samples, as they are closely related to our index case. Furthermore, *CSF3R* delta mutations are highly recurrent in these patients, indicative of commonly affected pathways[1]. Strikingly, in the cohort of 26 SCN patients that we have screened, the recurrence of the mutations identified in the SCN index patients is low. This may however be explained by the fact that many investigated SCN patients were young at the time bone marrow samples were obtained, which would furthermore be an explanation why the incidence of *CSF3R* mutations was low. In the leukemic cells of 5 SCN patients that we had available for this study recurrent mutations were found in *RUNX1* (2 patients). These

mutations coincided with the presence of *CSF3R* mutations as was seen for our SCN index case, suggesting a mechanism of cooperation or synergism between mutated CSF3R and mutated RUNX1. Obviously, the numbers of patients analyzed are still too small to draw firm conclusions about the mutation frequencies in early and late stages of leukemic progression of SCN.

Besides studying the frequency and relevance of the newly identified mutations in SCN, we should consider what could be the relevance of these mutations in other (hematopoietic) neoplasms. Two striking examples of mutations recurrently found in different disorders are (i) *BRAF* mutations, originally discovered in melanoma and found in 100% of the hairy cell leukemias[37, 38] and (ii) *SF3B1* mutations observed in approximately 65-75% of patients with refractory anemia with ring sideroblasts (RARS) and later also described in chronic lymphoid leukemia (CLL)[28, 39-42]. Hence, it seems rather unpredictable in which spectrum of diseases mutations in a single gene may play an important role. Therefore it seems best to select, in an unbiased way, large cohorts of patients with different types of neoplasms to screen for these mutations. In the current deep-sequencing era however, it would be even easier to obtain information regarding the frequencies of the complete spectrum of somatic genetic alterations in different types of neoplasms. In this way we do not only obtain information on the genes of interest but we get information on the complete spectrum of mutations in these disorders. Next, the neoplasms in which mutations in our genes of interest are recurrent can be identified, followed by studying the functional significance of these mutations in these diseases.

2.2. Retroviral integration mutagenesis

Retroviral integration mutagenesis (RIM) has been used as a powerful screening strategy to discover genes involved in malignant transformation in different types of cancer[43-45]. In its "classical" form, this technology is geared towards the discovery of proto-oncogenes, due to the fact that proviral integrations preferentially occur in activate promoter regions.

2.2.1. Identification of haplo-insufficient genes using retroviral integration mutagenesis

In this thesis, we have attempted to adapt RIM with the objective to identify possible tumor suppressor or haplo-insufficient genes (HIGs) in leukemia. Based on experience from gene therapy studies, it is well established that CpG islands in proviral long terminal repeats (LTRs) can be methylated, leading to silencing of the transduced therapeutic gene[46, 47]. Hence, we assumed that silencing by DNA methylated LTRs could also extend to nearby located genes. Although a new candidate haploinsufficient gene (*Ptp4a3)* was identified using this method, this new strategy also had major limitations. First of all, the polyclonality of the murine tumors limits the detection of retroviral integrations. Only the ones present in a high number of cells will be identified, while other integra-

tions, present in subclones remain undetected. Secondly, we observed that, although the number of investigated DNA methylated integrations was small, DNA methylation of viral integration sites not necessarily leads to transcriptional repression of flanking genes. Further evidence that *Ptp4a3* is a true haplo-insufficient gene in hematopoietic malignancies should come from additional studies, in which the effect of changes in its expression levels will be monitored. Furthermore, to prove the prognostic value of *PTP4A3* in human AML, the prognostic significance of its expression levels should be analysed in an independent AML cohort.

2.2.2. Future applications of retroviral integration mutagenesis

RIM has primarily been used to identify new aberrations in human cancer. A general drawback of RIM however is that genes involved in murine oncogenesis may be irrelevant for human malignancies. Furthermore, the heterogeneity of retroviral integrations in murine leukemias makes identification of the ones critical for disease development complex. Putting it into perspective of the leukemia research field as it stands today, one might even wonder what yet the applicability of RIM could be as the use of upcoming genome-wide techniques are nowadays easily employed to directly identify aberration in human leukemias. The first application we may consider is to identify co-operating genes in predisposed mouse models carrying a single oncogenic aberration. However, an alternative and possibly better approach in this case seems to be direct identification of alterations in the specific subgroup of AML patients carrying this aberration. A more potential application of RIM will be to study oncogenesis. The evolutionary process of leukemogenesis in these mice, e.g., the sequential accumulation of different retroviral integrations in time, may reveal vital information on "early" and "late" events in leukemogenesis. New techniques that will be designed to study single cells or small cell fractions may aid in this analysis as they will enable monitoring of pre-leukemic (sub)clones.

2.3. Data-analysis: new tools and algorithms

The rapid increase in using genome-wide approaches to study human diseases comes with the necessity to develop software tools and algorithms to interpret the data. The drive to design new tools usually originates from the lack of applications that match the specific needs of the user to analyze the generated data. This was similarly true for the RIM data described in this thesis. Therefore we designed a new software tool named HAT.

2.3.1. Applicability of HAT

HAT was developed to calculate significant enrichment of genomic regions in tiling array data and can be used to analyze protein-DNA and DNA-DNA interactions. The main advantage of this method compared to other methods, like model-based analysis of

tiling arrays for ChIP-chip (MAT)[48] and CMARRT[49] is that both region size and intensity threshold values do not have to be defined by the user. Other methods use the given region size to calculate mean probe intensity values by implementing neighboring probes. This may be useful when sheared DNA is used as starting material, like for chromatin immunoprecipitation, but not when fragments in the starting material have more strict boundaries and differ in length, like after digestion with restriction enzymes or in case of insertions and deletions. Because the other methods listed above consequently use the defined threshold to determine enriched regions, large regions containing many probes with enrichment scores just below the threshold value will be missed. As HAT varies threshold values and region sizes in a single analysis, both small regions (with few, but highly enriched probes) and large regions (with many, but lowly enriched probes) can be simultaneously determined to be significantly enriched. A disadvantage of HAT on the other hand is that it is mainly useful to identify regions which are highly enriched compared to the complete genome, because the probe intensity on the array is used as an internal reference. In other words, HAT is not the method of choice when analyzing interactions with a more global binding pattern. In these cases one should detect significantly enriched regions in comparison with a negative control. Of further note, in the near future enriched genomic regions will be mainly determined on next-generation sequencing platforms rather than by using tiling arrays, which may decrease the usefulness of HAT. The developed algorithm can however easily be changed in such a way that it will be applicable to analyze next-generation sequencing data, namely by using the coverage per base pair as intensity value.

2.3.2. Detection of somatic mutations in next-generation sequencing data

Once analyzing next-generation sequencing data, one realizes quickly that the existing data-analysis tools, like Varscan[50, 51], the Genome Atlas Toolkit (GATK)[52, 53] and SomaticSniper[54] do not exactly analyze the data as preferred by the user and/or do not clearly indicate the used algorithm such that the general user is able to understand which parameters, e.g., coverage, mapping quality, phred scores and allele frequency, are considered most important. Furthermore, we have observed that the statistical algorithms used to calculate reliability scores tend to overestimate the importance of coverage, while underestimating the significance of the allele frequency, especially when comparing tumor samples with germ line controls. This may be beneficial when identifying mutations in minor (sub)clones, but to identify mutations in major clones, this does not seem to be the right approach. In the latter case, if the coverage is acceptable (>10x), the allele frequency should be considered one of the most important parameters.

It is generally known that all sequencing platforms have a certain sequencing error rate and that alignment algorithms may cause misalignment artefacts, leading to false positive variant calling. This problem differs per experimental platform and is usually in-

herent to the genomic region, which for example may contain many repeats or stretches of similar base pairs. Hence, sequencing of germ line controls may show a similar result at the same genomic position and the variant will be considered a germ line variation. However, this is not always the case. Therefore, it is highly recommendable always to validate next-generation sequencing data on a different sequencing platform, as it will reduce the number of falsely determined genetic alterations.

2.4. Perspective on leukemia research and clinical practice

The number of studies in which next-generation sequencing approaches are employed to map genetic alterations in leukemias are increasing enormously. Inevitably they warrant for specific strategies to be able to understand the biological significance of the many genetic aberrations identified in these studies.

2.4.1. Will more sequencing lead to more simplicity or more complexity?

At the time of writing, we are passing a phase in which whole genome sequencing techniques are widely used to identify genetic aberrations in leukemia samples that up to now were stored in biobanks. We may expect that in a few years all genetic alterations in these samples are determined. Furthermore, it is likely that in the near future all newly diagnosed leukemias will be analysed by WGS. Consequently, databases, like the "Catalogue Of Somatic Mutations In Cancer (COSMIC)" database[55] and the dataset provided by the TCGA consortium[36], will emerge in which the genetic landscape of leukemias will be publically available. However, one should interpret these data with care as many reported mutations in hematopoietic malignancies are either not validated or are validated using the same, rather than a different sequencing platform, which will introduce the same sequencing bias[6, 56].

In what way will the extensive next-generation sequencing data influence our understandings of leukemia and leukemogenesis? Will it lead to more simplicity or to more complexity, to divergence or to convergence? Based on the spectrum of mutations so far observed in leukemias and on recent studies that provide new insights in the complete genetic landscape of MDS/AML by next-generation sequencing[28, 36, 38, 39, 55, 57-59] we should anticipate that a convergence will be seen for only few genetic alterations, while a divergence in the complete genetic landscape will be observed. In other words, it is becoming increasingly clear that even a higher heterogeneity should be expected than could be anticipated based on methods adopted so far to sub classify leukemias, i.e., morphology, immunophenotyping, karyotyping and gene expression profiling.

How can we explain the observed heterogeneity in genetic aberrations? One explanation could be that the variety in mutations congregates in commonly affected pathways that are or still have to be elucidated. Secondly, it may be that only common mutations play a role in disease development and that other, less frequent mutations solely have

occurred coincidently in leukemic clones but should be considered passenger mutations. A more likely explanation however will be a combination of the 2 above. The presence of one or few common mutations are likely vital for leukemogenesis and affect essential, common pathways, while a combination of less frequent mutations cause effects that by themselves are not sufficient for disease development, however combinations of few are necessary for malignant transformation.

2.4.2. Understanding the heterogeneity in leukemia

When considering the analysis of genetic variants in cancer, the challenge in the long run will not be to identify acquired mutations but to determine the ones relevant for disease development. What could be the following steps, once the complete genetic landscape of AML is known? At first the focus should lie on understanding the role of non-synonymous mutations as by altering protein functions they are likely relevant for disease development. However, even by reducing the number of mutations by primarily focusing on non-synonymous alterations it may be an illusion to expect that the full functional significance of each of these different mutations will be captured in the near future. An important and feasible step however should be to study how common genetic alterations contribute to disease. In order to investigate this, functional studies in different experimental *in vitro* and *in vivo* models can be conducted. The focus of these studies could lie on the elucidation if and how mutually exclusive common mutations affect similar pathways. Recent examples of mutually exclusive mutations disrupting common pathways are TET2 and IDH1/IDH2 mutations both influencing conversion of 5-methylcytidine into 5-hydroxymethylcytidine[60] and mutations in different components of the splicing machinery[28]. Studying commonly affected pathways could consequently be useful to define new therapeutic targets. Besides studying their functional role, it is of interest to gain insights into sequential gain of these genetic aberrations in the process of leukemogenesis as well as to determine their prognostic significance.

When considering the thousands of low frequent mutations it is much more difficult to come up with a strategy to study if and how these may be relevant for leukemia and we need to make a few assumptions before being able to proceed. First of all, we should assume that the number of patients in which a specific mutation is found, reflects its relevance for disease development. We may determine a predicted relevance for each mutation, based on its frequency in AML. Secondly, we should suppose that mutations exerting similar effects by being located (i) in the same gene, (ii) in genes with similar functions or (iii) in genes present in similar pathways, are mutually exclusive. Following the same line of reasoning, mutations that have a synergistic effect should significantly coincide. Keeping these assumptions in mind, we should combine all reported (validated) mutations and determine which mutations significantly coincide and which are mutually exclusive. Hereby we have to take into account the predicted relevance of

these different mutations based on their frequencies in AML, as both mutual exclusiveness as well as coincidence of mutations with a high recurrence is more suggestive of a respectively similar or synergistic effect. Important to note is that the power of the proposed strategy is dependent on the number of patients. After determining groups of mutations which likely have a synergistic or similar effect, literature studies followed by *in vitro* and *in vivo* experimental approaches can be adopted to further elucidate the role of these mutations in leukemia.

Determining the relevance of synonymous mutations as well as mutations in intronic and intergenic regions will prove to be even more difficult. Here, we may expect answers from integrating different datasets like WGS, mRNA-seq, miRNA-seq, methyl-seq and chIP-seq. This may allow us to understand the possible consequences of genetic variations on epigenetic regulation and gene expression. However, many epigenetic modifications as well as gene expression, microRNA and protein levels may be perturbed in leukemia independent of genetic aberrations. Hence, by solely focusing on genetic alterations we would ignore other important cellular regulatory mechanisms that have a major role leukemic transformation. Large scale profiling of microRNA and gene expression levels and DNA methylation patterns in AML[61-63] show that a similar level of heterogeneity is observed at these levels of regulation as is seen for genetic alterations. The approach to identify alterations in these datasets that play a key role in leukemia will be as proposed above, by integrating different datasets. In contrast to genetic alterations however, alterations at these regulatory levels are dynamic. Furthermore, identification of these alterations in minor clones is difficult, as the effect is diluted out by "normal" cells while the sensitivity of the adopted experimental methods might not be sufficient to monitor these changes. Hence, identification of "early" epigenetic changes and aberrations in expression levels will prove to be a bigger challenge than for genetic alterations. Unfortunately, a limitation for all the proposed studies is that sufficient samples are needed to reach enough statistical power. Hence, consequences of alterations present in single or few patients will remain difficult to interpret.

2.4.3. How to deal with heterogeneity in clinical practice?

Nowadays, with the upcoming use of genome-wide techniques, we are mapping genetic and epigenetic aberrations in leukemia more rapidly than that we are able to understand their biological relevance. The new findings may be used in clinical practice as predictive markers, diagnostic tools and prognostic parameters. However, with the high heterogeneity that is unraveling at the time of writing, defining candidate genes to perform extensive functional studies in order to elucidate their functional role in leukemia will be more difficult. This nevertheless is essential to understand the molecular mechanisms behind leukemia and to find new therapeutic targets. The first step in this process, i.e., selecting candidate genes, should be based on the assumption that genes altered in a

large number of patients are likely driving disease development. Hence, the selection of candidate genes should first of all be dependent on the frequency of specific alterations in AML. Secondly, it should be explored whether proliferation and/or survival of the leukemic cell is dependent on the presence of the alteration in the candidate gene, as only then, intervening with the alteration will have a cytostatic or cytotoxic effect. Thirdly, the mechanism of action of the alteration should be investigated to specifically target the affected molecular pathway. Fourthly, different chemical compounds can be generated and/or tested to monitor their effect on proliferation and survival of cells with the specific alteration, preferably in comparison to cells not carrying the alteration. In the end, this will lead to the identification of specific compounds that can be tested in mouse models, followed by further analysis in clinical trials. In this way, we should be able to assign combinations of few therapeutic compounds to leukemia patients based on the presence and absence of specific aberrations. This is an elegant treatment strategy, as we tackle different essential pathways in the leukemic blasts in order to eradicate the complete leukemic clone, without leaving undetectable leukemic blasts that may cause relapse.

In conclusion, with the accumulating results of studies using genome-wide techniques, a general view is unraveling that leukemia is a highly heterogeneous disease and that the combination of molecular events leading to leukemic transformation are likely unique in each single case. The challenge in the future will not be to add more heterogeneity by performing additional large scale screens but to understand the molecular mechanism behind the disease and to assign the right therapy to each single patient.

REFERENCES

1. Germeshausen M, Ballmaier M, Welte K. Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis: Results of a long-term survey. Blood 2007;109(1):93-9.
2. Greaves M. Cancer stem cells: back to Darwin? Seminars in cancer biology 2010;20(2):65-70.
3. Hermans MH, Antonissen C, Ward AC, Mayen AE, Ploemacher RE, Touw IP. Sustained receptor activation and hyperproliferation in response to granulocyte colony-stimulating factor (G-CSF) in mice with a severe congenital neutropenia/acute myeloid leukemia-derived mutation in the G- CSF receptor gene. J Exp Med 1999;189(4):683-92.
4. McLemore ML, Poursine-Laurent J, Link DC. Increased granulocyte colony-stimulating factor responsiveness but normal resting granulopoiesis in mice carrying a targeted granulocyte colony-stimulating factor receptor mutation derived from a patient with severe congenital neutropenia. J Clin Invest 1998;102(3):483-92.
5. Anderson K, Lutz C, van Delft FW, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature 2011;469(7330):356-61.
6. Ding L, Ley TJ, Larson DE, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature 2011;481(7382):506-10.

7. Touw IP, van de Geijn GJ. Granulocyte colony-stimulating factor and its receptor in normal myeloid cell development, leukemia and related blood cell disorders. Front Biosci 2007;12:800-15.
8. Gateff E. Malignant neoplasms of genetic origin in Drosophila melanogaster. Science (New York, NY 1978;200(4349):1448-59.
9. Ohshiro T, Yagami T, Zhang C, Matsuzaki F. Role of cortical tumour-suppressor proteins in asymmetric division of Drosophila neuroblast. Nature 2000;408(6812):593-6.
10. Peng CY, Manning L, Albertson R, Doe CQ. The tumour-suppressor genes lgl and dlg regulate basal protein targeting in Drosophila neuroblasts. Nature 2000;408(6812):596-600.
11. Grzeschik NA, Amin N, Secombe J, Brumby AM, Richardson HE. Abnormalities in cell proliferation and apico-basal cell polarity are separable in Drosophila lgl mutant clones in the developing eye. Developmental biology 2007;311(1):106-23.
12. Reischauer S, Levesque MP, Nusslein-Volhard C, Sonawane M. Lgl2 executes its function as a tumor suppressor by regulating ErbB signaling in the zebrafish epidermis. PLoS genetics 2009;5(11):e1000720.
13. Yamanaka T, Horikoshi Y, Izumi N, Suzuki A, Mizuno K, Ohno S. Lgl mediates apical domain disassembly by suppressing the PAR-3-aPKC-PAR-6 complex to orient apical membrane polarity. Journal of cell science 2006;119(Pt 10):2107-18.
14. Yamanaka T, Horikoshi Y, Sugiyama Y, et al. Mammalian Lgl forms a protein complex with PAR-6 and aPKC independently of PAR-3 to regulate epithelial cell polarity. Curr Biol 2003;13(9):734-43.
15. Yasumi M, Sakisaka T, Hoshino T, et al. Direct binding of Lgl2 to LGN during mitosis and its requirement for normal cell division. J Biol Chem 2005;280(8):6761-5.
16. Spaderna S, Schmalhofer O, Wahlbuhl M, et al. The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. Cancer research 2008;68(2):537-44.
17. Aigner K, Dampier B, Descovich L, et al. The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. Oncogene 2007;26(49):6979-88.
18. Lisovsky M, Dresser K, Baker S, et al. Cell polarity protein Lgl2 is lost or aberrantly localized in gastric dysplasia and adenocarcinoma: an immunohistochemical study. Mod Pathol 2009;22(7):977-84.
19. Lisovsky M, Ogawa F, Dresser K, Woda B, Lauwers GY. Loss of cell polarity protein Lgl2 in foveolar-type gastric dysplasia: correlation with expression of the apical marker aPKC-zeta. Virchows Arch 2010;457(6):635-42.
20. Verhaak RG, Wouters BJ, Erpelinck CA, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. Haematologica 2009;94(1):131-4.
21. Ito T, Kwon HY, Zimdahl B, et al. Regulation of myeloid leukaemia by the cell-fate determinant Musashi. Nature 2010;466(7307):765-8.
22. Kharas MG, Lengner CJ, Al-Shahrour F, et al. Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. Nature medicine 2010;16(8):903-8.
23. Ting SB, Deneault E, Hope K, et al. Asymmetrical segregation and self-renewal of hematopoietic stem and progenitor cells with endocytic Ap2a2. Blood 2011.

24. Sengupta A, Duran A, Ishikawa E, et al. Atypical protein kinase C (aPKCzeta and aPKClambda) is dispensable for mammalian hematopoietic stem cell activity and blood formation. Proceedings of the National Academy of Sciences of the United States of America 2011;108(24):9957-62.

25. Sripathy S, Lee M, Vasioukhin V. Mammalian Llgl2 is necessary for proper branching morphogenesis during placental development. Molecular and cellular biology 2011;31(14):2920-33.

26. Hermans MH, Ward AC, Antonissen C, Karis A, Lowenberg B, Touw IP. Perturbed granulopoiesis in mice with a targeted mutation in the granulocyte colony-stimulating factor receptor gene associated with severe chronic neutropenia. Blood 1998;92(1):32-9.

27. Benz C, Mulindwa J, Ouna B, Clayton C. The Trypanosoma brucei zinc finger protein ZC3H18 is involved in differentiation. Molecular and biochemical parasitology 2011;177(2):148-51.

28. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 2011;478(7367):64-9.

29. Gewurz BE, Towfic F, Mar JC, et al. Genome-wide siRNA screen for mediators of NF-kappaB activation. Proceedings of the National Academy of Sciences of the United States of America 2012;109(7):2467-72.

30. Carbuccia N, Trouplin V, Gelsi-Boyer V, et al. Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias. Leukemia 2010;24(2):469-73.

31. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011.

32. Pasqualucci L, Dominguez-Sola D, Chiarenza A, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature 2011;471(7337):189-95.

33. Score J, Hidalgo-Curtis C, Jones AV, et al. Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. Blood 2011.

34. Taketani T, Taki T, Takita J, et al. AML1/RUNX1 mutations are infrequent, but related to AML-M0, acquired trisomy 21, and leukemic transformation in pediatric hematologic malignancies. Genes, chromosomes & cancer 2003;38(1):1-7.

35. Yamamoto K, Sonoda M, Inokuchi J, Shirasawa S, Sasazuki T. Polycomb group suppressor of zeste 12 links heterochromatin protein 1alpha and enhancer of zeste 2. J Biol Chem 2004;279(1):401-6.

36. The Cancer Genome Atlas. 2011. (Accessed 2011, at http://cancergenome.nih.gov/.)

37. Davies H, Bignell GR, Cox C, et al. Mutations of the BRAF gene in human cancer. Nature 2002;417(6892):949-54.

38. Tiacci E, Trifonov V, Schiavoni G, et al. BRAF mutations in hairy-cell leukemia. The New England journal of medicine 2011;364(24):2305-15.

39. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. The New England journal of medicine 2011;365(15):1384-95.

40. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nature genetics 2011;44(1):47-52.

41. Rossi D, Bruscaggin A, Spina V, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. Blood 2011;118(26):6904-8.

42. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. The New England journal of medicine 2011;365(26):2497-506.

43. Erkeland SJ, Valkhof M, Heijmans-Antonissen C, et al. Large-scale identification of disease genes involved in acute myeloid leukemia. J Virol 2004;78(4):1971-80.
44. Suzuki T, Shen H, Akagi K, et al. New genes involved in cancer identified by retroviral tagging. Nat Genet 2002;32(1):166-74.
45. Theodorou V, Kimm MA, Boer M, et al. MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. Nat Genet 2007;39(6):759-69.
46. Lorincz MC, Schubeler D, Goeke SC, Walters M, Groudine M, Martin DI. Dynamic analysis of proviral induction and De Novo methylation: implications for a histone deacetylase-independent, methylation density-dependent mechanism of transcriptional repression. Molecular and cellular biology 2000;20(3):842-50.
47. Yao S, Sukonnik T, Kean T, Bharadwaj RR, Pasceri P, Ellis J. Retrovirus silencing, variegation, extinction, and memory are controlled by a dynamic interplay of multiple epigenetic modifications. Mol Ther 2004;10(1):27-36.
48. Johnson WE, Li W, Meyer CA, et al. Model-based analysis of tiling-arrays for ChIP-chip. Proceedings of the National Academy of Sciences of the United States of America 2006;103(33):12457-62.
49. Kuan PF, Chun H, Keles S. CMARRT: a tool for the analysis of ChIP-chip data from tiling arrays by incorporating the correlation structure. Pacific Symposium on Biocomputing 2008:515-26.
50. Koboldt DC, Chen K, Wylie T, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics (Oxford, England) 2009;25(17):2283-5.
51. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research 2012;22(3):568-76.
52. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics 2011;43(5):491-8.
53. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 2010;20(9):1297-303.
54. Larson DE, Harris CC, Chen K, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics (Oxford, England) 2012;28(3):311-7.
55. Bamford S, Dawson E, Forbes S, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. British journal of cancer 2004;91(2):355-8.
56. Walter MJ, Shen D, Ding L, et al. Clonal Architecture of Secondary Acute Myeloid Leukemia. The New England journal of medicine 2012.
57. Ley TJ, Ding L, Walter MJ, et al. DNMT3A mutations in acute myeloid leukemia. The New England journal of medicine 2010;363(25):2424-33.
58. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 2008;456(7218):66-72.
59. Mardis ER, Ding L, Dooling DJ, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. The New England journal of medicine 2009;361(11):1058-66.
60. Figueroa ME, Abdel-Wahab O, Lu C, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer cell 2011;18(6):553-67.
61. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. The New England journal of medicine 2004;350(16):1617-28.

62. Figueroa ME, Lugthart S, Li Y, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer cell 2010;17(1):13-27.

63. Jongen-Lavrencic M, Sun SM, Dijkstra MK, Valk PJ, Lowenberg B. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. Blood 2008;111(10):5078-85.

LIST OF ABBREVIATIONS

NEDERLANDSE SAMENVATTING

DANKWOORD

CURRICULUM VITAE, AWARDS AND LIST OF PUBLICATIONS

PHD PORTFOLIO

LIST OF ABBREVIATIONS

AML	Acute myeloid leukemia
CEBPA	CCAAT/enhancer binding protein alpha
ChIP-on-chip	Chromatin immunoprecipitation followed by analysis on array
ChIP-seq	Chromatin immunoprecipitation followed by deep-sequencing
CSF3	Colony-stimulating factor 3
CSF3R	Colony-stimulating factor 3 receptor
CSF3R-d715 to d730	Mutated CSF3R, truncated receptors at amino acid position 715 to 730 (intracellular)
CSF3R-T595I	Mutated CSF3R, substitution of a threonine to an isoleucine at amino acid position 595 (extracellular)
CSF3R-T595V	Mutated CSF3R, substitution of a threonine to a valine at amino acid position 595 (extracellular)
ELANE	Neutrophil elastase
G6PC3	Glucose-6-phosphatase 3
G-CSF	Granulocyte colony-stimulating factor (see also CSF3)
GFI1	Growth factor independent protein 1
Gr1.4 MuLV	Graffi 1.4 murine leukemia virus
HAT	Hypergeometric analysis of tiling-arrays
HAX1	HCLS1 associated protein X-1
HIG	Haplo-insufficient gene
HSC	Hematopoietic stem cell
Indel	Insertion or deletion
iPCR	Inverse PCR
LLGL2	Lethal giant larvae homolog 2 (Drosophila)
LTR	Long terminal repeat
MDS	Myelodysplastic syndrome
MeDIP	Methylated DNA immunoprecipitation
MeDIP-on-chip	Methylated DNA immunoprecipitation followed by analysis on array
methyl-seq	DNA methylation profiling by deep-sequencing
miRNA-seq	microRNA profiling by deep-sequencing
mRNA-seq	mRNA profiling by deep-sequencing
MSP	Methylation specific PCR
mVIS	Methylated viral integration site
PTP4A3	Protein tyrosine phosphatase type IVA, member 3
RIM	Retroviral integration mutagenesis
SCN	Severe congenital neutropenia
SCN/AML	SCN derived AML

SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
STAT3	Signal transducer and activator of transcription 3
STAT5	Signal transducer and activator of transcription 5
WAS	Wiskott-Aldrich syndrome protein
WES	Whole exome sequencing
WGS	Whole genome sequencing
ZC3H18	Zinc finger CCCH-type containing 18

NEDERLANDSE SAMENVATTING

De aanmaak van bloed is een continu proces waarbij dagelijks miljoenen nieuwe rode en witte bloedcellen en bloedplaatjes worden aangemaakt. Als dit proces verstoord wordt, kan bloedkanker ofwel leukemie ontstaan. Afwijkingen die bij het ontstaan van leukemie een rol spelen zijn verworven veranderingen in de genetische code ofwel het DNA van de leukemische cellen, deze worden mutaties genoemd. Ook veranderingen in hoe de genetische code wordt afgelezen in deze cellen, de zogenaamde epigenetische veranderingen, kunnen hierin een rol spelen. Acute myeloïde leukemie (AML) is een vorm van leukemie die in het beenmerg ontstaat, de plaats waar de normale bloedcelvorming plaatsvindt. Het werk beschreven in dit proefschrift richt zich op de identificatie van nieuwe genetische en epigenetische afwijkingen die mogelijk betrokken zijn bij het ontstaan van AML. Daartoe werden twee verschillende modellen voor het ontstaan van leukemie gebruikt. Ten eerste werden analyses gedaan in patiënten met ernstige aangeboren neutropenie (SCN), een aandoening met een verhoogde kans op het ontwikkelen van AML. Daarnaast werd gebruik gemaakt van virale insertie mutagenese in muizen, waarbij leukemie wordt geïnduceerd door infectie met leukemie veroorzakende virussen.

In **hoofdstuk 1** wordt een algemene introductie gegeven over SCN en AML, met daarbij een overzicht van de huidige inzichten op het gebied van de onderliggende genetische en epigenetische veranderingen in deze ziektebeelden. SCN is een aandoening waarbij patienten een sterk verlaagd aantal witte bloedcellen (neutrofiele granulocyten) in het bloed hebben, een conditie die neutropenie wordt genoemd. Omdat neutrofiele granulocyten essentieel zijn voor de afweer tegen micro-organismen, met name bacteriën, hebben SCN patiënten een sterk verhoogde kans op terugkerende en soms zelfs levensbedreigende infecties. De neutropenie kan in veel gevallen succesvol worden behandeld met G-CSF, de groeifactor voor neutrofiele granulocyten. Echter, de sterk verhoogde kans op het ontwikkelen van AML in deze patiënten blijft een belangrijk punt van zorg. De rol van G-CSF in het ontstaan van leukemie in deze en andere patiënten wordt bediscussieerd in **hoofdstuk 2**. G-CSF bindt aan de G-CSF receptor, een eiwit dat zich presenteert op de buitenkant van neutrofiele granulocyten. Hierop volgend worden er in de cel mechanismen geactiveerd die de groei en uitrijping van deze cellen stimuleren. In SCN patiënten worden frequent mutaties gevonden in de G-CSF receptor. Deze leiden tot expressie (i.e., het presenteren) van verkorte vormen van de receptor, die versterkt gevoelig zijn voor G-CSF. Mutaties in de G-CSF receptor, evenals bepaalde afwijkingen in chromosoomaantallen, te weten het verwerven van een extra kopie van chromosoom 21 (trisomie 21) en het verlies van een kopie van chromosoom 7 (monosomie 7) of een gedeelte daarvan (7q-) zijn tot dus ver de enige frequent voorkomende afwijkingen die gerelateerd worden aan leukemie ontwikkeling in SCN patiënten. Derhalve zijn de onderliggende afwijkingen in SCN en de leukemie die hieruit voortkomt voor het grootste deel nog onbegrepen. Er is voornamelijk meer bekend over

de groep patiënten met *de novo* AML. Deze AML patiënten hebben geen voorgeschiedenis van beenmerg falen, myelodysplastisch syndroom (een aandoening waarbij afwijkingen in het beenmerg worden gezien die ook bij leukemie voorkomen) of behandeling met chemotherapie, en vormen de grootste groep van AML patiënten. De huidige genoombrede technieken waarmee de gehele genetische code snel in kaart kan worden gebracht hebben laten zien dat de onderliggende veranderingen in *de novo* AML patiënten zeer variabel kunnen zijn; veel verschillende genetische en epigenetische afwijkingen worden gevonden in deze patiëntengroep, waarvan slechts enkele frequent voorkomen. In het laatste deel van **hoofdstuk 1** worden de voor- en nadelen van genoombrede technieken besproken.

In **hoofdstuk 3 tot 5** van dit proefschrift ligt de focus op de identificatie van nieuwe mutaties die betrokken zijn bij de progressie van SCN naar AML, evenals op de mogelijke rol die deze mutaties kunnen hebben in dit proces. In **hoofdstuk 3** worden de genetische afwijkingen in sequentiële bloed- en beenmergmonsters van een SCN patiënt in kaart gebracht die uiteindelijk na 17 jaar G-CSF behandeling leukemie ontwikkelde. Deze patiënt wordt hierna de index patiënt genoemd. In deze analyse lag de nadruk op het identificeren van verworven mutaties in de bloed- en/of beenmergcellen die dus niet in andere lichaamscellen van de patiënt voorkomen. Verworven mutaties kunnen in verschillende regio's in het DNA voorkomen; als een mutatie in een regio voorkomt die codeert voor een eiwit dan kan dit een verandering teweeg brengen in de structuur en functie van dit eiwit. In de index patiënt werden 12 van dit soort mutaties gevonden. Een interessante bevinding is dat drie van deze mutaties al aanwezig waren in een kleine hoeveelheid cellen in het beenmerg 15 jaar voordat de AML zich manifesteerde. Eén van deze mutaties was een mutatie in de G-CSF receptor die al eerder was geïdentificeerd in deze patiënt, terwijl de andere twee nieuwe mutaties waren in de genen coderend voor *LLGL2* en *ZC3H18*. De drie "vroege" mutaties zijn samen aanwezig in dezelfde (pre-leukemische) cel, waarbij de mutatie in *ZC3H18* later is ontstaan dan de mutaties in de G-CSF receptor en *LLGL2*. Voor deze twee eerdere mutaties kon geen hiërarchie in het ontstaan in de tijd worden aangetoond. De aanwezigheid van drie mutaties in de vroege fase van de ziekte suggereert dat ze een belangrijke rol spelen in de vroege stappen van leukemie ontwikkeling. Negen andere mutaties werden alleen gevonden in de leukemische fase in de SCN index patiënt. Eén daarvan was een tweede mutatie in de G-CSF receptor die voorkomt in combinatie met de eerder gevonden G-CSF receptor mutatie. Cellen die G-CSF receptoren met deze tweede mutatie tot expressie brengen groeien in de afwezigheid van G-CSF en andere groeifactoren. Het verwerven van groeifactor onafhankelijke groei kan gezien worden als een belangrijke stap in leukemie ontwikkeling. In **hoofdstuk 4** wordt beschreven dat deze tweede G-CSF receptor mutatie ook laag frequent voorkomt in *de novo* AML patiënten. Bovendien wordt er beschreven dat expressie van receptoren met deze mutatie G-CSF onafhankelijke stimulatie van de transcriptiefactoren STAT3 en STAT5 veroorzaakt in de cel, iets wat normaliter alleen wordt gezien na G-CSF stimulatie. Wat betreft de andere

mutaties die gevonden werden in de leukemie cellen van de index patiënt; mutaties in de genen *ASXL1, EP300, RUNX1* en *SUZ12* werden eerder gerapporteerd in andere leukemieën en lymfklierkanker, terwijl mutaties in de genen *CCDC155, FBXO18, LAMB1* en *MGA* niet eerder werden gevonden. Echter, ongepubliceerde data van The Genome Atlas Consortium die de volledige genetische informatie van 199 AML patiënten bevat, laat zien dat mutaties in *MGA* en *CCDC155* in een zeer lage frequentie aanwezig zijn. In **hoofdstuk 5** wordt de aanwezigheid van mutaties in de genen die gemuteerd zijn in de SCN index patiënt in een grotere groep van 26 SCN patiënten onderzocht, evenals mutaties in genen die er sterk op lijken en genen die coderen voor eiwitten waarmee ze een interactie aan kunnen gaan. Vijf van de onderzochte SCN patiënten ontwikkelden leukemie. Deze analyse laat zien dat mutaties in de geïdentificeerde genen in de index patiënt, behalve in de G-CSF receptor, zeldzaam zijn in andere SCN patiënten in de vroege fase van de ziekte; in 1 patiënt werd een mutatie in *LAMB2* gevonden. Daarnaast zijn *RUNX1* mutaties in combinatie met mutaties in de G-CSF receptor de enige frequent gevonden afwijkingen in SCN patiënten die leukemie hebben ontwikkeld.

In **hoofdstuk 6** werd retrovirale insertie mutagenese gebruikt als model om nieuwe haplo-insufficiënte genen op te sporen die betrokken zijn bij AML. De gebruikte retrovirussen in dit model kunnen beenmergcellen van muizen infecteren en vervolgens integreren in het DNA van deze cellen. Hier kunnen ze de regulatie van dichtbij liggende genen verstoren, waardoor ze leukemie kunnen veroorzaken. Hierbij kan er sprake zijn van een verhoogde of verlaagde expressie van de aangedane genen. We spreken van haplo-insufficiëntie als de normale expressie niveaus en/of functionele activiteit van een gen/eiwit gereduceerd is en deze verlaging in expressie en/of functionele activiteit bijdraagt tot leukemie ontwikkeling. Om haplo-insufficiënte genen te detecteren hebben we virale integraties opgespoord die op DNA niveau gemethyleerd zijn, omdat DNA methylatie een bekend mechanisme is dat expressie van genen kan reduceren. Daartoe hebben we een methode ontwikkeld waarbij verrijkt wordt voor DNA fragmenten van gemethyleerde retrovirale inserties, die vervolgens op een chip werden geladen. De data van de chip hebben we gebruikt om de flankerende genen van deze gemethyleerde inserties op te sporen. Om de data van de chip goed te kunnen analyseren hebben we een nieuwe data-analyse methode ontwikkeld genaamd "hypergeometric anaylsis of tiling arrays" (HAT), welke wordt beschreven in **hoofdstuk 7**. Deze methode is speciaal gericht op het detecteren van verrijkte genomische regio's van verschillende grootte en intensiteit. Middels deze methode werd het gen *Ptp4a3* geïdentificeerd dat werd geflankeerd door een DNA gemethyleerde retrovirale insertie en een verlaagde expressie had in de leukemiecellen ten opzichte van normale expressie niveaus in het beenmerg in muizen, een teken van haplo-insufficiëntie. *Ptp4a3* is dus een potentieel haplo-insufficiënt gen. Verder bleken *PTP4A3* expressie niveaus van prognostisch voorspellende waarde in een cohort van AML patiënten gediagnosticeerd onder de leeftijd van 60 jaar.

DANKWOORD

Lieve allemaal, het is zover. Met trots kan ik zeggen: "Mijn boekje is klaar" en dat voelt geweldig! Uiteraard had dit boekje niet voor jullie gelegen zonder de hulp van velen. Ik maak graag gebruik van dit laatste, doch meest gelezen deel van mijn proefschrift om een aantal mensen persoonlijk te bedanken.

Mijn eerste woorden zijn uiteraard voor jou, **Ivo**. Ik weet nog goed dat ik inmiddels bijna 6 jaar geleden naar je toe stapte om te vragen of je een promotieplek voor me had. Ik was overtuigd van jouw vertrouwen in mij, ik wist dat ik de vrijheid zou krijgen om mijn eigen keuzes te kunnen maken binnen mijn onderzoeksproject en ik wist dat ik me in jouw groep op mijn plek zou voelen. Deze verwachtingen heb je helemaal waargemaakt. Het is niet altijd makkelijk geweest, maar de vrijheid die ik gekregen heb heeft de belangrijkste basis gevormd voor het wetenschappelijk inzicht dat ik de afgelopen jaren heb ontwikkeld. Dankjewel voor je enthousiasme, de discussies, je altijd oprechte en duidelijke mening en de inzichten die en het vertrouwen dat je mij de afgelopen jaren hebt gegeven. Ik waardeer je grote inzet voor het behouden van wetenschappelijke kwaliteit en je belangrijke bijdrage aan de ontwikkeling van de generatie wetenschappers van de toekomst enorm, zoals je onder andere laat zien door het opzetten van het TRTH programma, je betrokkenheid bij de ontwikkeling van het onderwijs en je hulp bij het aanvragen van projecten door jonge wetenschappers. Ik hoop ons contact in de toekomst te mogen voortzetten.

Beste **Ruud, Riccardo** en **Joop,** jullie wil ik bedanken voor jullie bereidheid plaats te nemen in de beoordelingscommissie van mijn proefschrift. Beste **Ruud**, we kennen elkaar alweer 10 jaar. Jouw enthousiasme voor je vak heeft mij tijdens mijn research master uiteindelijk doen kiezen voor een stage op de afdeling hematologie. Dat was in 2004 en is een hele goede beslissing gebleken. Ik weet nog goed dat ik, voordat ik die keus maakte, je vertelde dat ik heel veel onderwerpen leuk vond en dat jij hierop zei: "dan maak je dus nooit een verkeerde keus". Dat advies heb ik niet alleen toen, maar ook later nog vaak meegenomen en is heel waardevol gebleken. Inmiddels zijn we 10 jaar en een proefschrift verder! Dankjewel voor je enthousiasme en voor je luisterend oor op wetenschappelijk en persoonlijk gebied. Beste **Riccardo**, we kennen elkaar alweer een aantal jaren via het master programma molecular medicine. Dank je wel voor je kritische commentaar op mijn proefschrift, voor je enthousiasme en je aanmoedigingen om door te gaan in de wetenschap.

Beste **Bob**, niet alleen als afdelingshoofd maar nu nog steeds beteken je veel voor de afdeling. Jouw inzicht voor alle verschillende aspecten van de hematologie op klinisch en wetenschappelijk gebied en voor het belang van medewerkers met verschillende

deskundigheden, zoals je zelf hebt geschreven toen je afscheid nam als afdelingshoofd in 2011, zijn van uitzonderlijke waarde voor de afdeling. Daarnaast wil ik je bij deze bedanken voor je wetenschappelijke en persoonlijke interesse in mijn toekomstplannen. Beste **Peter V.**, dank je voor alle momenten van overleg. Ik waardeer je kritische blik, je interesse en je begrip en inzicht in veel uiteenlopende wetenschappelijke en persoonlijke situaties. Deze kwaliteiten maken het zeer prettig om met je samen te werken. Beste **Marieke v L.**, het is alweer 2 jaar geleden dat je naar Sanguin in Amsterdam bent vertrokken. Sinds die tijd mis ik je enthousiasme en positivisme. Overleg met jou moedigde me altijd aan "moeilijke" keuzes te maken en te kiezen voor de dingen die goed voelden.

Beste **Marijke**, dit boekje had er absoluut niet gelegen zonder jouw enorme inzet. Het is een bekroning op ons werk. Ik hoop dat je er net zo trots op bent als ik. Ik heb ervan genoten de afgelopen jaren met je samen te werken. Een zeer goed team, dat waren we! Het was geweldig dat er aan een half woord genoeg was voor het uitvoeren van experimenten. Ik heb bewondering voor je enorme werklust (er werd nooit geklaagd als er teveel werk was, maar wel als er te weinig te doen was), je vaktechnisch inzicht en je doorzettingsvermogen. Ik waardeer je eigenwijsheid en je altijd duidelijke mening. Zonder deze eigenschappen van jou waren een aantal experimenten in dit proefschrift waarschijnlijk nooit uitgevoerd. Daarnaast wil ik je meegeven dat ik heel trots op je ben dat je moeilijke keuzes niet uit de weg gaat, maar dat je ze hebt aangegrepen als een kans om jezelf beter te leren kennen en een nieuwe weg in te slaan. Petje af! Beste **Paulette**, de afgelopen jaren heb je ons team versterkt en daar ben ik heel blij mee. Je bent een ontzettend harde werker, je hebt veel inzicht en je enthousiasme werkt aanstekelijk. Daarnaast ben je ontzettend lief en een steun en toeverlaat voor velen, dat is een enorm mooie eigenschap waar velen een voorbeeld aan zouden kunnen nemen! Lieve meiden, ik ben er enorm trots op wat wij samen de afgelopen jaren hebben bereikt en ik ben heel blij dat jullie op 18 januari naast mij staan. Ik mis jullie, jullie moeten maar snel langskomen in Barcelona.

Beste **Annemarie** en **Karishma**, mijn mede AIO's in de Touw-groep. Wat was het leuk om met jullie samen te werken! Dank jullie wel voor de momenten waarop we lief en leed konden delen. **Karishma**, thank you for the great trip to India! It was a great experience and who could have thought that it would change my life completely. I wish you all the best together with Girish. **Annemarie**, dank je wel dat ik je paranimf mocht zijn tijdens jouw promotie, dat was een hele eer! Ik wens je veel succes met het afronden van je opleiding tot klinisch chemicus en veel geluk samen met Joost, die een geweldige acteur is gebleken. Beste **Jurgen**, wat gaat de tijd snel, je bent alweer ruim een jaar weg op de afdeling. We hebben veel meegemaakt samen de afgelopen jaren, zowel op het

werk als privé. Dank je wel voor je scherpe visie, je inzicht en je luisterend oor. Dit laatste is voor mij zeer belangrijk geweest om mijn gevoelens en frustraties te kunnen delen wanneer ik dit nodig had. Het was geweldig om met je samen te werken en ik hoop ons contact in de toekomst voort te kunnen zetten. Dear **Tanja**, it was very nice working with you. I enjoyed the personal talks that we had and I keep very good memories about shooting the movie for Annemarie. You were a great director and camera woman! I wish you all the best in Serbia. Beste **Onno**, je bent een zeer behulpzame en geweldige collega. Dank je wel voor je bijdrage aan dit boekje. Ik wens je veel geluk samen met Bia en ik hoop dat jullie snel een knoop kunnen doorhakken, zodat jullie in de toekomst samen kunnen zijn. Beste **Stefan E.**, dank je wel voor al je input de afgelopen jaren en voor de supervakantie in Boston, nu alweer heel wat jaartjes geleden. In jouw proefschrift in 2005 schreef je speciaal voor mij: "Het is moeilijk, maar het gaat zeker lukken……., denk ik." Nu, 7 jaar verder mogen we concluderen dat je "denk ik" weg mag halen, daar ben ik trots op! Ik heb bewondering voor je harde werklust, voor je doorzettingsvermogen en voor hoe je je eigen groep leidt. Het is moeilijk, maar het gaat zeker lukken! Dear **Albert**, you were an inspiration for me. You have a great personality, I admire your calmness, your willingness to help and your dedication to your work. These characteristics made you a great colleague. Beste **Lotte**, het is alweer een tijdje geleden dat je als student op de afdeling rondliep. Ik heb er bewondering voor hoe je een moeilijk project tot een goed einde hebt gebracht, dat uiteindelijk ook nog bekroond is met een publicatie. Dank je wel voor al je inzet. Veel succes in Groningen met je promotie en je specialisatie.

Beste **Su Ming** en **Simone**, jullie waren mijn (kop)kamergenootjes van het eerste uur en zijn dit gebleven tot het eind. Dank jullie wel voor al het lief en (AIO-)leed dat we samen gedeeld hebben. De gesprekken met jullie waren inspirerend, inzichtelijk en gezellig, soms iets te gezellig. **Su Ming**, je mag trots zijn op jezelf, op wat je de afgelopen jaren hebt bereikt tijdens je promotie en op je postdoc positie in Leiden. Ik wil je 1 ding meegeven voor de toekomst: probeer het leven van de positieve kant te benaderen. Ik wens je veel geluk samen met Xiwen. **Simone**, dank je wel voor de vele discussies die we gevoerd hebben, waarin "wie is de mol" toch redelijk vaak aan bod kwam. Ik waardeer je altijd duidelijke mening. Ik wens je veel succes met het kiezen van je toekomstige postdoc positie en met de keuzes die daarna gaan komen. Keuzes maken is niet altijd even makkelijk, maar ik heb er alle vertrouwen in dat je ergens uitkomt waar je je op je plek voelt. Beste **Erdogan**, 5 jaar geleden begonnen we samen op de kopkamer. Het was altijd leuk en gezellig om met je te praten over wetenschappelijke en niet-wetenschappelijke zaken. Het was heel waardevol om via jou een kijk op de bio-informatica te krijgen en ik ben heel blij dat je me ooit hebt geleerd hoe ik matlab moet gebruiken! En jij ook denk ik, want dat scheelde toch heel wat vragen van mijn kant. En wie had gedacht dat we er zelfs samen nog een publicatie uit zouden slepen.

Ik wens je veel succes met het afronden van je promotie en natuurlijk alle geluk samen met Ankie en Senna. Dear **Rasti**, I have appreciated your always critical view the past few years. This made you a very good and well appreciated colleague. You can be proud of your postdoc position at EMBL. I wish you good luck with finishing your thesis. Beste overige kopkamergenootjes van het eerste uur, **Kerim, Lucila, Stein** en **Eric V.**, het is alweer lang geleden dat we met elkaar op de kopkamer zaten. Dank jullie wel voor de gezellige momenten en veel succes in de toekomst. Dear **Roberto**, I regret that the time has been too short to get to know each other very well. I have enjoyed the inspiring talks we had the past few months. I am sure that you will find your way in science as long as you keep believing in yourself! Beste **Shirley**, dank je wel dat ik mijn gevoelens en onzekerheden over belangrijke levenskeuzes met je kon delen. Wat dat betreft hebben meer gemeen dan we op het eerste gezicht misschien dachten. Ik wens je veel geluk toe in de toekomst en ik weet zeker dat je de juiste keuzes zal maken om dit te bereiken. Beste **Michelle**, we hebben maar kort bij elkaar op de kamer gezeten, desalniettemin was het erg gezellig. Succes de komende jaren met je promotie!

Beste **Lianne**, je bent sinds jaar en dag betrokken bij het SCN project, al jaren voordat ik op de afdeling als AIO begon. Jouw inzet is zeer belangrijk geweest voor de studies die in dit proefschrift beschreven staan. Dank je wel voor je grote inzet, dat je altijd bereid was om te helpen, de geweldige data-opslag waarin altijd alles terug te vinden was (daar kan ik nog heel wat van leren!) en voor je interesse in het project. En daarnaast natuurlijk voor je persoonlijke interesse. Beste collegae van de BMT groep, beste **Eric B., Lianne, Peter v. G., Arie, Hans H., Hannie, Pia, Anita** en **Larissa**, ik ben jullie zeer dankbaar voor de goede opslag van alle samples die binnenkomen op de afdeling. Dit is niet alleen belangrijk geweest voor de totstandkoming van dit proefschrift, maar is van zeer grote waarde voor de afdeling in het algemeen.

Beste **Antoinette**, het is alweer lang geleden dat ik als student begon in de Delwel groep. Het was heel fijn dat ik zoveel van jou kon leren in die beginperiode, dank je wel voor de goede begeleiding, de goede gesprekken en je gezelligheid. Beste **Claudia**, je bent een zeer ervaren analiste en je stond altijd klaar om te helpen. Ik heb veel van je geleerd. Dank je wel voor alle hulp gedurende de afgelopen jaren. Beste collegae van de moleculaire diagnostiek, beste **Peter V., Antoinette, Pauline, Sonja, Wendy, Chantal, Isabel** en **Marloes**, dank jullie wel voor alle hulp en alle samples die ik de afgelopen jaren van jullie gekregen heb en voor alle persoonlijke interesse! Jullie werk is van zeer belangrijke waarde voor de afdeling.

Dear **Mathijs, Remco, Erdogan** and **Veronika**, this thesis could not have been generated without you. The past years I have learned that the collaboration between biologists,

bio-informaticians and statisticians is of great importance in science; that everyone of us has his/her own expertise, but that you can reach a lot when joining forces! Collaborating with you was a great experience and I learned a lot about the different aspects of bio-informatics and statistics. This knowledge will be of great value in my future career and I hope that it will be possible to keep in contact every now and then to share some questions with you.

Beste **Ans,** dank je wel voor al je hulp de afgelopen jaren. Voor alle brieven en andere dingen die je voor me geregeld hebt. Beste LDC-genootjes, beste **Mark, Merel** en **Isabel**. Ik heb de labdagcommissie als een hele leuke tijd ervaren. Dank jullie wel voor jullie inzet, creativiteit en gezelligheid gedurende deze periode! Beste **Jurgen, Sophie** en **Marleen**, ik heb goede herinneringen aan onze spelletjes avonden, dank jullie wel hiervoor!

Dear **Meri, Roel V., Godfrey, Gert-Jan, Dominik, Marieke B., Bart, Fokke, Eric vd. A., Mahban, Elwin, Jan, Leenke, Tomasia, Sevilay, Natasja, Diana, Hannah, Saman, Justine, Bas, Sanne, Jurjen, Jo, Annemiek, Goran, Marianne, Janine, Eric B., Martijn, Yvonne, Joyce, Reinilde, Nicole, Hans de L., Irene, Marije, Carola, Arturo, Rogier, Natalie, Ferry, Jasper, Francois, Annelieke, Mojca, Menno, Yvette, Athina, Stefan G.** and **Helen**, thank you for all your help, you were great colleagues! I wish you all the best for the future. Dear **Roel P., Davine, Farshid, Elnaz, Andrzej, Kasia, Rowan, Marshall, Patricia, Aysegül, Lucia, Amiet** and **Noemi**, I wish you all the luck with finishing your PhD-projects and writing your theses. Dear **Julia**, I wish you a great time at the department!

Beste **Judith** en **Anne**, ik vind het altijd erg leuk om met jullie af te spreken. Ook al is het niet heel frequent, het is altijd gezellig en vertrouwd. Dank jullie wel voor hiervoor! Ik hoop dat jullie snel een keer naar Barcelona komen.

Beste **(oud-)collegae van de afdeling hematologie**, ik wil jullie allemaal bedanken voor jullie wetenschappelijke input, voor de inspirerende en leuke discussies op wetenschappelijk en persoonlijk gebied, voor alle hulp die ik van jullie heb gekregen de afgelopen jaren en voor alle gezelligheid!

Dear **Nico Hartwig, Taco Kuijpers, Marrie Bruin, Göran Carlsson, Peter Vandenberghe, Jean Donadieu** and **Anjo Veerman**. Thank you for sending in SCN patient samples during the past years as well as for your clinical input in the different projects described in this thesis.

Dear mentors of the research master Molecular Medicine, dear **Claire, Anton, Elaine, Riccardo, Dick** and **Ruud**, I would like to thank you for starting the initiative to allow medical students to simultaneously get trained in different aspects of fundamental research. This initiative recognizes that it is of great importance to train students both in medicine and fundamental research to allow them to implement research findings in the clinic and vice versa. This is of great importance in translational research. For me, it was a great opportunity to follow the research master and it has been of great value for my scientific career!

Dear mentors and fellow-students of TRTH group 1, dear **Eva, Hal, Radek, Frank, Wim, Felicitas, Thomas, Philip** and **Rafael**, it was great to discuss our research projects together and to get a broader view on different aspects of fundamental and translational hematology. Your input in my project has been of great importance to finish this thesis! Dear **Donna**, thank you for all your help and your personal interest! The TRTH has been a great opportunity for me to set up new collaborations and to develop as a fundamental researcher in hematology. To all the mentors and fellow-students of the TRTH program of 2011, it was great meeting you, we will keep in contact!

Lieve **Sandra, Christine** en **Daniël**, dit proefschrift had er niet gelegen zonder jullie hulp in moeilijke tijden. Ik ben ontzettend dankbaar voor onze goede vriendschap, het is geweldig om samen de mooie momenten van het leven te kunnen delen en het voelt heel geruststellend om te weten dat ik altijd op jullie terug kan vallen. Tijden veranderen en we zullen elkaar minder vaak zien nu ik in Barcelona woon, maar dat betekent niet dat onze vriendschap minder hecht zal worden. Lieve **Sandra**, jij bent een geweldige vriendin. Ik mis onze hardlooprondjes, onze ontbijtjes op zaterdag, onze kopjes thee en je gezelligheid. Gelukkig hebben we skype, dat maakt het iets makkelijker. Ik ben er trots op hoe je je door moeilijke tijden in het leven slaat en dat je ze aangrijpt om jezelf beter te leren kennen. Heb vertrouwen in jezelf, dat is wat ik je graag mee wil geven! Dank je wel dat ik bij je in huis mocht wonen, dat voelde heel fijn en vertrouwd! Lieve **Chris**, al sinds de 4e klas zijn we vriendinnen, 15 jaar alweer. Je bent een geweldige vriendin en samen hebben we veel lief en leed gedeeld. Ik vond het een geweldige eer dat ik getuige mocht zijn op jullie bruiloft. Het is super om te zien dat je samen met Daniël zo gelukkig bent. Ik ben trots op je dat je de laatste jaren keuzes hebt gemaakt waarvan je gelukkig wordt. Ik kan alleen maar zeggen, zet dit voort en realiseer je dat geluk niet zit in materiële dingen, maar in de liefde van de mensen die je omgeven, je familie en vrienden, en in wat je zelf van het leven maakt. Lieve **Daan**, je bent een geweldige vriend! Dank je wel voor alle leuke gesprekken, alle goede raad, je bereidheid om altijd te helpen en voor je nuchtere kijk op de wereld. En voor het delen van je mannelijke kijk op de dingen, daar heb ik heel veel van geleerd de afgelopen jaren. En dank je wel dat ik bij jullie

thuis mocht wonen toen dit nodig was! Lieve **Albert**, alweer 11 jaar geleden begon onze hechte vriendschap. Dank je wel voor de diepgaande discussies die we in de loop van de jaren hebben gevoerd en voor alle gezelligheid. Vriendschappen kennen ups en downs, dat hebben we de afgelopen jaren ervaren. Dat was een belangrijk leermoment voor mij en ik ben blij dat we nu een hechte band hebben waarop we kunnen terugvallen, al zien we elkaar misschien niet even vaak als we zouden willen. Ik heb grote bewondering voor het feit dat je je droom achterna bent gegaan en dat je nu als tropenarts werkt in Kenia! Lieve **vrienden**, dank jullie wel voor alle gezelligheid, feestjes, oudjaars-dineetjes, weekendjes, de mooie gedichten op sinterklaasavond, de vele spelletjes avonden en de goede gesprekken!

Lieve **familie**, we zien elkaar niet vaak, maar ik geniet altijd van het familieweekend! Ik weet dat het belangrijk is om je eigen keuzes te maken in het leven, maar toch voelt het heel goed om hier met jullie over te kunnen praten en te voelen dat jullie deze keuzes begrijpen. Dat is heel waardevol! Dear **Marti, Montse, David, Rosa, Carlos, Joel, Sergi, Eva** and **Daniel**, moving to Barcelona has been a big step for me the past year. Thank you for being a great family, for all your support and for making me feel at home!

Lieve **Sicco**, ik ben trots op een lieve grote broer als jij! Je gaat je eigen weg en volgt je eigen geluk, daar heb ik bewondering voor. En ik vind het leuk om de laatste jaren te merken dat we meer gemeen hebben dan dat we in eerste instantie misschien dachten. Dear **Lauren**, you are a great sister-in-law!

Lieve **pap en mam**, ik heb grote bewondering voor jullie. Jullie hebben mij en Sicco geleerd dat we zelf ons geluk moeten maken en dat we ons hart moeten volgen. Jullie hebben ons altijd gesteund in de keuzes die we maakten en jullie hebben ons alle mogelijkheden geboden om ons geluk te verwezenlijken. Voor jullie staat ons geluk altijd op nummer 1 ook al betekent dat, dat we allebei niet in de buurt wonen en dat we elkaar niet zo vaak kunnen zien als we misschien zouden willen. Lieve **mam**, ik heb bewondering voor je altijd positieve kijk op het leven, voor hoe je problemen snel aanpakt en oplost, voor je evenwichtige persoonlijkheid en voor hoe je altijd voor anderen klaarstaat. Lieve **pap**, ik waardeer je rustige persoonlijkheid, je altijd weloverwogen mening, je grote behulpzaamheid naar andere mensen toe en het feit dat je je altijd voor meer dan 100% inzet. Een mix van deze eigenschappen, de één een beetje meer en de ander een beetje minder, maakt mij mij en daar ben ik trots op! Zonder deze eigenschappen en zonder jullie hulp had dit boekje er nooit gelegen en het voelt geweldig om te weten dat jullie er net zo trots, zo niet nog trotser op zijn dan ik!

Lieve **Uri** en **Ona**, mijn laatste woorden zijn uiteraard voor jullie. Dear **Uri**, who could have ever thought that our short trip to India would change our lives forever. Thanks to you, I realized that we belong together. It took me some time to follow my heart and I highly appreciate your patience! Ever since, our lives have changed completely. Being together feels like how life should be, it makes making choices easier, it takes my doubts about the mayor questions in life away and most important of all, it makes me very happy! I would like to thank you for believing in me, for all your support and encouragement not only during my PhD-project, but far beyond that. Lieve **Ona**, al sinds dat papa en mama samen zijn, dromen we van het krijgen van een kindje. We zijn ontzettend dankbaar en gelukkig met jouw komst!

Renée

CURRICULUM VITAE, AWARDS AND LIST OF PUBLICATIONS

PERSONAL DATA

- **Name:** Renée Beekman
- **Date of birth:** December 17th, 1981
- **Place of birth:** Vlaardingen
- **Nationality:** Dutch
- **Titles:** MD, MSc

ACADEMIC ACHIEVEMENTS

- **Medical Degree (2000 - 2007)**
  Erasmus University Rotterdam (cum laude)
- **Master of Science degree in Molecular Medicine (2001 - 2005)**
  Erasmus University Rotterdam
- **Secondary education (1994 - 2000)**
  SG Spieringshoek, Schiedam. Subjects: Dutch, English, Latin, Mathematics, Physics, Chemistry and Biology

PROFESSIONAL ACTIVITIES

- **PhD-student in training (February 2008 - June 2012)**
  Department of Hematology, Erasmus University Rotterdam, The Netherlands
  Supervisor: Prof. Ivo P. Touw
  Thesis title: Genome-Wide Analysis of Severe Congenital Neutropenia and Leukemia; Implications for Leukemogenesis
- **Master student in training (September 2004 - September 2005)**
  Department of Hematology, Erasmus University Rotterdam, The Netherlands
  Supervisor: Prof. Ivo P. Touw
  Project title: Large-scale identification of tumor suppressor genes in acute myeloid leukemia
- **Master student in training (February 2004 - July 2004)**
  Department of Hematology, Erasmus University Rotterdam, The Netherlands
  Supervisor: Dr. Ruud Delwel
  Project title: Analysis of the protein interaction and mechanism of transcriptional repression of the binding partners MBD3 and EVI1

MEMBERSHIPS

- **PhD-committee (2009 - 2012)**
  Erasmus Postgraduate school Molecular Medicine
- **Erasmus Postgraduate school Molecular Medicine (2008 - present)**

- **European Hematology Association (2008 - present)**

AWARDS

- **Translational Research Training in Hematology (TRTH) Award** (American Society of Hematology (ASH) and European Hematology Association (EHA), 2011)
- **Travel grant** for the 14th, 15th and 17th annual EHA congress (June 2009, 2010 and 2012)
- **Best Erasmus Medical Center graduate student of 2007** (Batavian Society for Experimental Philosophy in Rotterdam, September 2008)

LIST OF PUBLICATIONS

Peer-reviewed

- **Beekman R**, Valkhof MG, Sanders MA, van Strien PMH, Haanstra JR, Broeders L, Geertsma-Kleinekoort WM, Veerman AJ, Valk PJ, Verhaak RG, Löwenberg B, Touw IP. Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. Blood 2012 May 31;119(22):5071-7 (Plenary paper)
- **Beekman R**, Valkhof M, Erkeland SJ, Taskesen E, Rockova V, Peeters JK, Valk PJ, Lowenberg B., Touw IP: Retroviral Integration Mutagenesis in Mice and Comparative Analysis in Human AML Identify Reduced PTP4A3 Expression as a Prognostic Indicator. PLoS One. 2011 Oct 20;6(10):e26537.
- Palande K, **Beekman R**, van der Meeren LE, Beverloo HB, Touw IP. The antioxidant protein peroxiredoxin 4 is epigenetically down regulated in acute promyelocytic leukemia. PLoS One. 2011 Jan 20;6(1):e16340.
- **Beekman R**, Touw IP. G-CSF and its receptor in myeloid malignancy. Blood. 2010 Jun 24;115(25):5131-6.
- Taskesen E, **Beekman R**, de Ridder J, Wouters BJ, Peeters JK, Touw IP, Reinders MJT, Delwel HR. HAT: hypergeometric analysis of tiling-arrays with application to promoter-GeneChip data. BMC Bioinformatics. 2010 May 21;11:275.
- Spensberger D, Vermeulen M, Le Guezennec X, **Beekman R**, van Hoven A, Bindels E, Stunnenberg H, Delwel R. Myeloid transforming protein Evi1 interacts with methyl-CpG binding domain protein 3 and inhibits in vitro histone deacetylation by Mbd3/Mi-2/NuRD. Biochemistry. 2008 Jun 17;47(24):6418-26.

Invited contributions

- **R. Beekman** and I.P. Touw. The role of G-CSF in leukemic progression of severe congenital neutropenia. Oncologie up-to-date. 2010 Jun 24;1(3):19-20.

PARTICIPATION IN CONFERENCES

Invited contributions

- **Annual Spring Meeting Dutch Foundation for Quality Assessment in Clinical Laboratories (MODHEM/SKML), April 2012.** Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia.
- **Annual meeting Center Translational and Molecular Medicine (CTMM), March 2011.** Identification of novel disease genes in acute myeloid leukemia
- **Annual meeting Center Translational and Molecular Medicine (CTMM), February 2010.** Large-scale identification of tumor suppressor genes and prognostic factors in human AML based on retroviral insertion mutagenesis

Oral presentations at national and international conferences

- **Annual European Hematology Association (EHA) conference, June 2012.** Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia.
- **Annual American Society of Hematology (ASH) meeting, December 2011.** Methylated Retroviral Integration Mutagenesis (MRIM) in Mice and Comparative Analysis in Human AML Identify Reduced PTP4A3 Expression as a Prognostic Indicator.
- **Dutch Hematology Congress, January 2010.** Retroviral integration mutagenesis in mice as a platform for identification of critical tumor suppressor genes in human myeloid disorders.
- **Dutch Hematology Congress, January 2009.** A new strategy to identify methylated viral integration sites in murine acute myeloid leukemia (AML).
- **Annual European Hematology Association (EHA) conference, June 2006.** A new strategy to identify tumor suppressor genes in acute myeloid leukemia (AML).

Poster presentations at national and international conferences

- **The European Molecular Biology Organisation (EMBO) meeting, September 2010.** Identification of critical tumor suppressor genes (tsgs) and new prognostic factors in human myeloid disorders using retroviral insertion mutagenesis.
- **Annual European Hematology Association (EHA) conference, June 2010.** Identification of critical tumor suppressor genes (tsgs) and new prognostic factors in human myeloid disorders using retroviral insertion mutagenesis.
- **The European Molecular Biology Organisation (EMBO) meeting, September 2009.** Retroviral integration mutagenesis in mice as a platform for identification of critical tumor suppressor genes (tsg) in human myeloid disorders.
- **Annual European Hematology Association (EHA) conference, June 2009.** Retroviral integration mutagenesis in mice as a platform for identification of critical tumor suppressor genes in human myeloid disorders.

- **Summerschool on Chromatin Plasticity, September 2008.** A new strategy to identify methylated viral integration sites in murine acute myeloid leukemia (AML).
- **Annual Molecular Medicine Day, February 2006.** Large-scale identification of tumor suppressor genes in murine acute myeloid leukemia.

Erasmus MC
Universitair Medisch Centrum Rotterdam

SUMMARY PHD TRAINING AND TEACHING ACTIVITIES

Name PhD student: R. Beekman	PhD period: February 2008 - June 2012
Erasmus MC Department: Hematology	Promotor: Prof. I.P. Touw
Research School: Molecular Medicine	Supervisor: Prof. I.P. Touw

1. PhD training	**Year**	**ECTS**
General academic/research skills		
Talentday Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)	2008	0,3
Get-out-of your lab days	2009	0,6
Workshop "in je eigen tempo naar de top"	2011	0,1
Grant proposal writing	2010	0,5
Research management	2010	1
In-depth courses		
Chromatin plasticity	2008	2
Epigenetic mechanism in health and disease	2008	0,6
Ensembl gene browsing workshop	2008	0,6
Partek workshop	2008	1,3
Basic and translational oncology	2008	1,8
Ingenuity workshop	2008	0,1
Matlab fundamentals and statistical methods	2009	1
Photoshop workshop	2010	0,3
Annual course Molecular Medicine	2010	0,7
Translational Research Training in Hematology (TRTH)	2011	2
Oral presentations		
Algemene vergadering Bataafsch genootschap	2008	1
AIO/postdoc meeting (department of hematology, 3x)	2008-2012	1,5
Journal club (department of hematology, 4x)	2008-2012	2
Workdiscussion (department of hematology, 12x)	2008-2012	6
Dutch hematology congress	2009-2010	2
Annual conference American Society of Hematology (ASH)	2011	1
Annual MODHEM/SKML Spring Meeting	2012	1
Annual conference European Hematology Association (EHA)	2012	1
National and international conferences		
Molecular Medicine day	2008-2011	1,2

SUMMARY PHD TRAINING AND TEACHING ACTIVITIES (continued)

Dutch stem cell meeting	2009	0,3
Annual meeting European Molecular Biology Organisation (EMBO)	2009-2010	1,2
Dutch hematology congress	2009-2011	1,8
Annual conference European Hematology Association (EHA)	2009-2012	4
Annual conference American Society of Hematology (ASH)	2011	1
Jornada de cromatina i epigenètica	2011	0,3
Seminars		
Erasmus hematology lectures	2008-2012	3
Scientific meetings department of hematology		
Workdiscussions	2008-2012	4
AIO/postdoc meetings	2008-2012	1,6
Journal club	2008-2012	3
Other		
Writing application René Vogels stichting	2010	0,3
Writing application Translational Research Training in Hematology (TRTH)	2010	1
2. Teaching activities	**Year**	**ECTS**
Supervising students		
Master student ("Clinical Research" master)	2008-2009	3
Supervising practical training and excursions		
Medical school hematology practical training	2008-2009	0,4
Organisation and supervision invited speaker lunch	2008-2009	0,2
Medical school Fanconi anemia practical training	2009	0,4
Total		**55,1**