Causes and Consequences of Dysregulated Gene Expression in Abnormal Hematopoiesis

Davide Monteferrario
Stellingen behorend bij het proefschrift

Causes and Consequences of Dysregulated Gene Expression in Abnormal Hematopoiesis

Davide Monteferrario

1. “How are you going to proceed?” still echoes in my ears….

2. Despite being only five, those minutes of joy are worth the effort…

3. Be happy for mistakes, you have moved forward!

4. If the quick and dirty looks good, clean it!... but not that slowly!!

5. “Stay hungry, stay foolish”. Steve Jobs

6. In the realm of ideas everything depends on enthusiasm…in the real world all rests on perseverance. Johann Wolfgang von Goethe

7. “Look deep into nature, and then you will understand everything better”. Albert Einstein

8. “The snake which cannot cast its skin has to die. As well the minds which are prevented from changing their opinions; they cease to be mind.” Friedrich Wilhelm Nietzsche

9. “Simplicity is the ultimate sophistication”. Leonardo da Vinci

10. Life is like biking. If you don’t wanna fall, keep on riding!!
Causes and Consequences of Dysregulated Gene Expression in Abnormal Hematopoiesis
D. Monteferrario

The studies described in this thesis were performed from January 2012 till August 2014 at the Laboratory of Hematology, Department of Laboratory Medicine of the Radboud University Medical Centre, Nijmegen, The Netherlands.

The illustrations used on the cover were downloaded from Shutterstock.com

Layout and printing: Off Page, Amsterdam
Causes and Consequences of Dysregulated Gene Expression in Abnormal Hematopoiesis

Proefschrift

ter verkrijging van de graad van doctor aan de Radboud Universiteit Nijmegen op gezag van de rector magnificus, volgens besluit van het college van decanen in het openbaar te verdedigen op dinsdag 12 januari 2016 om 10.30 uur precies

door

Davide Monteferrario
geboren op 3 september 1983
te Biella, Italie
Promotor:
Prof. dr. J.H. Jansen

Copromotor:
Dr. B.A. van der Reijden

Manuscriptcommissie:
Prof. dr. J.H.J.M. van Krieken
Dr. P.N. Span
Prof. dr. T.K. Sixma
to the shine of the moments we live for
Table of Contents

Chapter 1  Introduction  11

Chapter 2  A dominant-negative GFI1B mutation in the gray platelet syndrome  59

Chapter 3  High DNA-methyltransferase 3B expression predicts poor outcome in acute myeloid leukemia, especially among patients with co-occurring NPM1 and FLT3 mutations  85

Chapter 4  Improved classification of MLL-AF9-positive acute myeloid leukemia patients based on BRE and EVI1 expression  97

Chapter 5  The E3-ligase HOIP specifies linear ubiquitin chain assembly through its RING-IBR-RING domain and the unique LDD extension  105

Chapter 6  Discussion  139

Addendum  Summary  157

Nederlandse samenvatting  161

Curriculum vitae  165

List of publications  166

Dankwoord / Acknowledgments  167
CHAPTER 1

Introduction
Introduction
Blood contains a variety of short-living cells that are fundamental to the human body. These cells travel through miles of blood vessels to sustain the correct functioning of our organs. They carry out a myriad of vital activities, including transport of oxygen and nutrients to all tissues, protection against harmful pathogenic conditions and prevention of blood loss by repairing damaged blood vessels. This functional diversity is to be attributed to the presence of different, highly specialized cell types that are generally categorized in a myeloid and a lymphoid compartment. While cells belonging to the myeloid lineage are implicated in the innate immune system (granulocytes, monocytes and macrophages), oxygen supply (red cells or erythrocytes) and maintenance of hemostasis (megakaryocytes and platelets), the lymphoid compartment comprises the remaining fraction of cells that participate in the innate immune response (natural killer cells) and cells that are fundamental to the adaptive immune system (dendritic, B- and T- cells). All these different types of cell originate from a rare population of cells, denominated the hematopoietic stem cells (HSCs), during a sophisticated and highly regulated developmental process called hematopoiesis.

Hematopoiesis
Current models of hematopoiesis
Hematopoietic ontology has traditionally been described in a hierarchical fashion where HSCs sit at the bottom of a “tree”, with branches representing the different lineages of maturation that culminate in the production of all blood cell types (Figure 1). To be at the foundation of the hematopoietic system, HSCs are long-lived and possess multipotency properties that permit them to differentiate towards more committed progenitors and precursors that, subsequently, will generate all mature blood cells. Very importantly, HSCs possess a unique self-renewal capacity, which guarantees the maintenance of a pool of ancestors that support the continuous production of blood cells throughout an entire life. According to the classical model, HSCs give rise to multipotent progenitors (MPPs) that further differentiate into common myeloid progenitors (CMPs) or common lymphoid progenitors (CLPs). Through a series of lineage-restricted transitions, CMPs and CLPs generate terminally differentiated cells that constitute the myeloid and the lymphoid compartments, respectively. Accordingly, the multipotential capacity to generate other lineages is progressively lost alongside a particular differentiation route, resulting in a distinct separation between myeloid and lymphoid lineages. However, the discovery of early progenitors, such as EPLMs (early progenitors with lymphoid and myeloid potential) and LMPPs (lymphoid-primed multi potent progenitors), that showed an unexpected flexibility to differentiate towards diverse lineages, suggests that hematopoietic differentiation is not constrained to a strict linear subdivision in compartments. Rather, this supports an alternative ontological model characterized by increased plasticity in cell lineage determination of hematopoietic progenitors, where the development of a specific cell type can occur through different maturation paths.
Figure 1. Classic ontological model of hematopoiesis. In the classic model of hematopoiesis, there is a strict segregation during the formation of the myeloid and lymphoid compartments. HSCs, which sit at the top of the organizational hierarchy, give rise to common myeloid progenitors (CMPs) and common lymphoid progenitors (CLPs). CLPs differentiate further into mature lymphocytes (natural killer (NK-), B-lymphocytes and T-lymphocytes), while CMPs generate the megakaryocyte/erythroid progenitors, which give rise to megakaryocytes and erythrocytes, and the granulocyte/monocyte progenitors (GMPs), which give rise to granulocytes (neutrophils, basophils and eosinophils), monocytes and macrophages. Dendritic cells can originate either from CLPs or monocytes, while plasma cells originate from B-lymphocytes. Cell types equipped with self-renewal capacity are indicated by the circular arrows.

Regulation of hematopoiesis

The anatomical location of hematopoiesis is subjected to various transitions during development. In embryos, initial sites of hematopoietic production can be found in the yolk sac, placenta, the aorta-gonad-mesonephros region and the umbilical and vitelline arteries. HSCs derive from the development of hemangioblasts, which are multipotent precursor cells capable to generate both the hematopoietic and the endothelial tissues. These cells originate the haemogenic endothelium that lines the aortic floor in the late stages of embryonic development. Following an endothelial to hematopoietic transition (EHT), hemogenic cells lose their endothelial identity to subsequently give rise to hematopoietic progenitors. Mice studies have shown that newly formed HSCs can migrate to
Introduction

extramedullary locations, such as liver and spleen, to ultimately reach the intramedullary spaces of bones\(^{14}\). Definitive adult hematopoiesis takes place in the bone marrow, a cavity with complex architecture that provides a supportive and protective microenvironment wherein HSCs are regulated\(^{15,16}\). The bone marrow is organized in different niches that containing a variety of cell types including osteoblasts, endothelial cells, adipocytes and stromal cells, which act in concert to control HSCs survival, proliferation, migration and differentiation. The fate of HSCs and hematopoietic precursors is determined by direct interactions with surrounding cells and the stimulation of molecules that are secreted within the niche microenvironment\(^{15–17}\). Binding of hematopoietic growth factors and cytokines to their membrane receptors initiates a chain of consequential intracellular events that control the expression of specific responsive genes. The intracellular amplification of these signals is propagated by fine-tuned transduction machineries consisting of proteins that, via post-translational protein modifications, induce or inhibit downstream targets. The final receivers of this signaling cascade are generally transcriptional regulators, in particular transcription factors, which form complex networks that coordinate the expression of hematopoietic genes\(^{18–20}\). During hematopoietic differentiation, genes regulating self-renewal of progenitor cells are repressed while sets of genes that govern the differentiation are activated. The balance between self-renewal and differentiation is crucial to control the expansion of the different hematopoietic lineages and maintain tissue haemostasis. The lineage specificity is determined by a coordinated expression of cell type-specific transcription factors, which promote the differentiation towards a distinct lineage meanwhile impeding the generation of others\(^{21}\). Although the expression of some transcription factors is restricted to a particular cell type, many of them have overlapping functions in regulating the destiny of multiple lineages, thereby contributing to the high intricacy of the hematopoietic transcriptional regulatory system\(^{19–22}\).

Outlook of this thesis

As it will be discussed in this thesis, transcriptional regulation is fundamental to the development of blood cells. The function and identity of blood cells are determined by the sets of genes expressed in each cell. Acquired and inherited changes in hematopoietic transcription factors and their up-stream regulators can cause hematological disorders ranging from bleeding disorders to leukemia. These alterations disrupt the genome-wide control over gene expression resulting in the generation of pathogenic transcriptional programs that compromise proper blood cell development and function. In addition, these altered expression patterns significantly influence the biological characteristics of the various diseases they generate and therefore their clinical outcome. Because aberrant gene expression contributes to the etiology of hematological diseases, their development and their clinical responses, a better understanding of transcriptional regulation is required for the development of successful therapies.
Megakaryopoiesis and platelet production

Megakaryocytes and platelets

The integrity of the blood circulatory system is essential for survival. This is maintained through a physiological response, denominated haemostasis, which prevents significant blood loss following vascular injury. This process depends on an intricate series of events that primarily relies on the action of platelets, which are anucleate discoid cells that also play important functions in angiogenesis and innate immunity\(^\text{23–25}\). Present in the bloodstream in a quiescent state, platelets store multiple types of granules which contain a variety of molecules that mediate coagulation. Upon stimulation, platelets release their granule content and form a physical barrier at the site of injury that impedes excessive bleeding. Generation of platelets, defined as thrombopoiesis, is orchestrated by a rare population of highly specialized cells, called megakaryocytes (MKs), which are among the largest hematopoietic cells residing in the bone marrow\(^\text{26}\). According to the most recent model of platelet formation, fully mature MKs infiltrate long branching membrane protrusions, called pro-platelets, into sinusoidal blood vessels, wherein these eventually shed and release functional platelets in circulation\(^\text{27,28}\). Although being present at low frequency (only 1 in 10000 nucleated cells), MKs sustain a massive production of approximately $1 \times 10^{11}$ platelets every day, thus providing an enormous reservoir that guarantees proper control on circulating platelet levels.

Early events of megakaryopoiesis

The process underlying the formation of MKs is defined as megakaryopoiesis. According to this model, MKs derive from megakaryocyte/erythroid progenitors (MEPs) following various highly regulated developmental stages\(^\text{29}\). These multipotent cells were generally thought to be produced by CMPs, however, recent studies have proposed that MEPs can arise directly from HSCs without the CMP intermediate\(^\text{8,30}\). Differentiation along the megakaryocytic lineage begins in the osteoblastic niche where it is mainly accompanied by nuclear polyploidization\(^\text{26,31}\). This is achieved through endomitosis, which is an alternative form of mitosis consisting of repeated replications of the DNA in the absence of cytoplasmic divisions\(^\text{32,33}\). As a consequence of this, MKs accumulate genomic material resulting in the formation of their characteristic polyploid and multilobulated nuclei, with a total DNA content that can reach up to 128N. Although the precise biological consequence of this process is unclear, it is theorized that an increase in gene copy number would ensure MKs to produce large amounts of mRNA and proteins that are usually packed into platelet granules, while retaining the ability to properly perform cellular functions.

Late events of megakaryopoiesis

Following polyploidization, MK precursors migrate to the endothelial niche where they undergo maturation to ultimately generate pro-platelets\(^\text{34}\). During this stage of development, which requires an impressive rearrangement of the cytoskeleton, the cytoplasm of MKs increases in size as a result of an intense production of granules and the progressive formation
of a unique and elaborate pool of membranes denominated invaginated membrane system (IMS)\textsuperscript{26}. The latter, also referred to as demarcation membrane system (DMS), is contiguous with the plasma membrane, permeates the cytoplasm of MKs and retains contact with the extracellular environment\textsuperscript{35,36}. The IMS was believed to demarcate the cytoplasm in small territories from where platelets were originally thought to be formed\textsuperscript{37}, however, refined models support a function for the IMS as a membrane reservoir for the formation and extension of pro-platelets\textsuperscript{38,39}. Granules are essential mediators of platelet function. Upon formation in the MK body, all the different types of granules are individually transported along microtubules to the MK periphery, where they are eventually captured at the tips of forming pro-platelets\textsuperscript{40}. The multitude of roles that is attributed to platelets in various physiological processes like inflammation, angiogenesis, microbial response, malignancy and wound healing mostly derives from the heterogeneous and distinct cargo transported in their granules. Each granule population can be distinguished based on the role played by the released constituents\textsuperscript{26,41,42}. Dense (δ-) granules, for example, contain membrane transporters and small bioactive molecules like adenosine nucleotides and calcium that stimulate platelet aggregation. α-granules, the most abundant platelet organelles, transport large adhesive and healing proteins including crucial hemostatic factors such as fibrinogen, PF4 (platelet factor 4), BTG (beta thromboglobulin) and VWF (von Willebrand factor), that mediate blood coagulation and tissue repair\textsuperscript{26}. As proteins present in α-granules can have opposing functions, this indicates that platelets release different sub-populations of α-granules in a regulated manner, depending on the type of activity to perform. In addition, platelet granules also include lysosomes, which contain degradative enzymes that eliminate circulating platelet aggregates, and the newly identified t-granules\textsuperscript{41}, whose function is still unknown.

**Pro-platelet formation and platelet release**

Maturation of MKs culminates with the formation of pro-platelets. These are long, thin extensions of the megakaryocyte cytoplasm that serve as assembly lines for the production of platelets. As platelet release occurs in circulation, MKs migrate to the bone marrow sinuses to infiltrate and extend pro-platelets into the underlying blood vessels\textsuperscript{27,28}. Pro-platelet production begins with the formation of a primitive membrane structure, called pseudopod. Powered by the essential activity of microtubules, this pro-platelet precursor elongates further to ultimately became a pro-platelet and reach the bloodstream\textsuperscript{43,44}. Subsequently, this process continues until the complete MK, with the exception of the MK cell body, is transformed in pro-platelets. At this stage, the characteristic multilobulated nucleus is excluded from the MK cell body and degraded\textsuperscript{26}. In parallel, all the granules and organelles have been concentrated within pro-platelet tips, ready to be distributed and packed into nascent platelets. Recent evidence suggests that pro-platelets are first released into the circulation to undergo additional fragmentation into platelets\textsuperscript{43}, however, the exact location and mechanisms of the final steps of thrombopoiesis are still controversial.
Transcription factors in megakaryopoiesis

As previously mentioned, mature MKs derive from MEPs, which constitute a bipotent population of hematopoietic progenitors from which both the megakaryocytic and erythroid lineages arise. During MK commitment, MEPs undergo radical changes that result in the acquisition of nuclear polyploidy, a vast cellular enlargement and the assembly of peculiar membranous structures in the cytoplasm. These profound transformations require the prevention of erythroid differentiation and the simultaneous activation of MK-related gene expression programs. In MEPs and more committed megakaryocytic cells, these are obtained through the coordinated activity of both MK- and erythroid-specific as well as widely expressed transcription factors. While some of these transcription factors are expressed throughout the entire MK differentiation, others are activated only during specific developmental stages. Many of them interact with each other and other co-factors, forming protein complexes that regulate transcriptional activity. The lineage-specificity and spatio-temporal regulation of gene expression are controlled by combinatorial binding of multiple transcription factors to gene promoters. Although many studies have proven this complexity in regulating MK gene expression, the mechanism by which these regulators collaborate to constitute fully mature MKs capable to produce platelets remains elusive. Nevertheless, precious insights have been gained from the identification of genetic alterations that cause hereditary thrombocytopenias (Table 1). As these pathologies are generally caused by the effect of a single mutation, they can unequivocally reveal which gene is required for proper control of megakaryopoiesis and platelet production. In combination with mouse studies, these investigations have uncovered important transcription factors involved in human megakaryopoiesis, such as GATA1, RUNX1, and FLI1, which will be individually discussed below.

Table 1. Summary of inherited thrombocytopenias based on the underlying mechanisms of defective platelet production. (Adapted from )

<table>
<thead>
<tr>
<th>Disease (OMIM entry)</th>
<th>Inheritance</th>
<th>Gene (chromosome location)</th>
<th>Platelet features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defective megakaryocytic differentiation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congenital amegakaryocytic thrombocytopenia (604498)</td>
<td>AR</td>
<td><em>MPL</em> (1p34)</td>
<td>Normal-sized platelets</td>
</tr>
<tr>
<td>Thrombocytopenia with absent radii (274000)</td>
<td>AR</td>
<td><em>RBM8A</em> (1q21.1)</td>
<td>Normal-sized platelets</td>
</tr>
<tr>
<td>Congenital thrombocytopenia with radio-ulnar synostosis (605432)</td>
<td>AD</td>
<td><em>HOXA11</em> (7p15-14)</td>
<td>Normal-sized platelets</td>
</tr>
</tbody>
</table>
### Table 1. (continued)

<table>
<thead>
<tr>
<th>Disease (OMIM entry)</th>
<th>Inheritance</th>
<th>Gene (chromosome location)</th>
<th>Platelet features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Defective megakaryocyte maturation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familial platelet disorder and predisposition to acute myeloid leukemia (601399)</td>
<td>AD</td>
<td>CBFA2, RUNX1 (21q22)</td>
<td>Normal-sized platelets</td>
</tr>
<tr>
<td>Paris-Trousseau thrombocytopenia (600588), Jacobsen syndrome (147791) GATA1-related diseases:</td>
<td>AD</td>
<td>Large deletion (11q-ter), including FLI1</td>
<td>Large platelets</td>
</tr>
<tr>
<td>Dyserythropoietic anemia with thrombocytopenia (300367) X-linked thrombocytopenia with thalassemia (314050) ANKRD26-related thrombocytopenia (313900)</td>
<td>XL</td>
<td>GATA1 (Xp 11)</td>
<td>Large platelets</td>
</tr>
<tr>
<td>Gray platelet syndrome</td>
<td>AR</td>
<td>NBEAL2 (3p21.1)</td>
<td>Giant gray platelets</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Defective pro-platelet formation and platelet release</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYH9-related disease (nd)</td>
<td>AD</td>
<td>MYH9 (22q12-13)</td>
<td>Giant platelets</td>
</tr>
<tr>
<td>ACTN1-related thrombocytopenia (nd)</td>
<td>AD</td>
<td>ACTN1 (14q24)</td>
<td>Large platelets</td>
</tr>
<tr>
<td>FLNA-related thrombocytopenia (nd)</td>
<td>XL</td>
<td>FLNA (Xq28)</td>
<td>Large platelets</td>
</tr>
<tr>
<td>Wiskott-Aldrich syndrome (301000)</td>
<td>XL</td>
<td>WAS (Xp11)</td>
<td>Small platelets</td>
</tr>
<tr>
<td>X-linked thrombocytopenia (313900)</td>
<td>AR</td>
<td>GP1BA (17P13)</td>
<td>Giant platelets</td>
</tr>
<tr>
<td>Bernard-Soulier syndrome (231200): Biallelic Monoallelic</td>
<td>AD</td>
<td>GP1BB(22q11) GP9(3q21) GP1BA (17P13)</td>
<td>Large platelets</td>
</tr>
<tr>
<td>ITGA2B/ITGB3-related thrombocytopenia (nd)</td>
<td>AD</td>
<td>ITGA2B (17q21.31)</td>
<td>Large platelets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ITGB3 (17q21.32)</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 1

GATA-binding factor 1 (GATA1) belongs to a family of zinc finger transcription factors that recognize and bind to specific WGATAR consensus sequences on gene promoters\(^ {51,52}\). This family is composed of six different GATA-paralogues that are differentially expressed in human tissues, with GATA1 and GATA2 being the major GATA proteins to be broadly expressed during hematopoiesis\(^ {53}\). GATA1 expression gradually increases during differentiation of both erythroid and megakaryocytic lineage. Either disruption or overexpression of GATA1 in mice leads to embryonic lethality due to the failure of erythroid maturation\(^ {54–56}\). In addition, a tissue-specific lack of GATA1 in the megakaryocytic lineage has been shown to dramatically reduce nuclear polyploidization and impair cytoplasmatic maturation of MKs, indicating that the activity of GATA1 is essential for the development of both the erythroid and the megakaryocytic lineages\(^ {57}\). The importance of GATA1 in the regulation of MK development is also underscored by the presence of GATA consensus sequences in MK-specific gene promoters\(^ {58}\). The binding to the DNA can be mediated by two highly conserved zinc finger domains, the N- and C-fingers, which also mediate protein-protein interactions\(^ {59}\). Despite the fact that these structures share sequence similarities, it has been shown that the C-finger primarily contributes to DNA-binding while the N-finger is implicated in the binding with its crucial co-regulator friend of GATA1 (FOG-1)\(^ {60–62}\). Also this transcription factor is essential for MK development. Mouse studies have shown that FOG-1 deficiency completely inhibits the formation of the megakaryocytic lineage\(^ {63}\). In addition, the expression of GATA1 mutants that are unable to interact with FOG-1 results in a similar phenotype as that caused by GATA1 deficiency\(^ {64}\). This suggests that although GATA1 and FOG-1 do not always have overlapping functions, their interaction is indispensable for proper MK development. GATA1 mutations are found in a variety of hematological diseases including anemia and thrombocytopenia\(^ {65,66}\). Depending on their location along the gene, GATA1 mutations induce distinct biochemical dys-functions on the resulting mutated proteins that are associated with distinct clinical phenotypes. Specific loss-of-function mutations abrogating the ability of GATA1 to bind to specific DNA sequences or disrupting the interaction with its co-factor FOG1, for example, are found in patients with thrombocytopenia\(^ {67,68}\). Furthermore, while mutations affecting

<table>
<thead>
<tr>
<th>Disease (OMIM entry)</th>
<th>Inheritance (chromosome location)</th>
<th>Platelet features</th>
</tr>
</thead>
<tbody>
<tr>
<td>TUBB1-related thrombocytopenia (nd)</td>
<td>AD TUBB1 (6p21.3)</td>
<td>Giant platelets</td>
</tr>
<tr>
<td>CYSC-related thrombocytopenia (612004)</td>
<td>AD CYSC (7p15.3)</td>
<td>Normal-sized platelets</td>
</tr>
</tbody>
</table>

OMIM, Online Mendelian Inheritance in Man; AD, autosomal dominant; AR, autosomal recessive; XL, linked to chromosome X; MDS, myelodysplastic syndrome; nd, not defined.
DNA binding have a mild effect on the development of the erythroid compartment, those affecting FOG1-interaction severely impact erythroid maturation causing anemia, as observed in X-linked thrombocytopenia with thalassemia and dyserythropoietic anemia with thrombocytopenia, respectively\(^{69-71}\). Commonly in these diseases, patients display prolonged bleeding time, enlarged platelets that have an abnormal spherical shape rather than discoid, and an elevated number of MKs in their bone marrow. MKs are usually small with an immature cytoplasm that contains a substantially reduced number of platelet granules and a highly disorganized membrane system, indicating that the platelet defects induced by these types of GATA1 mutations derive from aberrant MK development. Of note, GATA1 is frequently mutated in acute megakaryoblastic leukemia (AMKL), which is characterized by an uncontrolled proliferation of megakaryocytic precursor cells\(^{72,73}\). In this disease, specific GATA1 mutations that are located outside the DNA- or the FOG1-binding domains, result in the expression of a truncated GATA1 form, GATA1s, that lacks the first 83 amino acids\(^{74}\). GATA1s allows the commitment towards the MK lineage, but it prevents MK maturation, resulting in an expansion and accumulation of immature MK cells in the bone marrow\(^{75}\). Overall, the abovementioned phenotypes underscore the broad role of GATA1 in the regulation of fundamental processes underlying proper MK development.

**RUNX1**

The core binding factor (CBF) family of transcription factors comprises three homologous DNA binding proteins (RUNX1-3), of which RUNX1 covers a fundamental function in blood cell development\(^{76,77}\). Normally, RUNX1 acts by forming a heterodimer with its non-DNA binding regulatory subunit CBFβ, which confers increased stability and DNA binding affinity to the complex\(^{76,78,79}\). The importance of these transcription factors is emphasized by the fact that homozygous deficiency of either RUNX1 or CBFβ impairs the establishment of definitive hematopoiesis in mice, thus causing embryonic lethality\(^{80-83}\). These animals completely lack cells of the myeloid, erythroid and megakaryocytic lineage, indicating that the RUNX1- CBFβ complex is fundamental for the emergence of hematopoiesis\(^{82-86}\). These transcription factors are highly expressed in hematopoietic stem and progenitor cells where they coordinate the expression of crucial genes involved in proliferation, differentiation and self-renewal. In addition, RUNX1 expression is found in MEPs and retained during the differentiation along the megakaryocytic, but not the erythroid lineage\(^{87,88}\). Consistently, conditional disruption of RUNX1 after birth rapidly causes thrombocytopenia in mice, with minimal consequences on erythropoiesis\(^{89-91}\). Similarly to GATA1 deficiency, MKs from these mice fail to undergo maturation showing defective polyploidization and impaired cytoplasmatic development. Important insights into the biological functions of RUNX1 in megakaryopoiesis have been gained through investigations on patients affected by familial platelet disorder with predisposition to acute myeloid leukemia (FPD/AML)\(^{47,49,50,92-98}\). In this disease, germline heterozygous RUNX1 mutations lead to thrombocytopenia, which is associated with a higher risk to develop AML, as seen in roughly 40% of the cases. RUNX1 is a common mutational target in AML\(^{77}\) and the increased incidence of this disease among FPD/AML cases is likely to reflect its important role in balancing the homeostasis of HSCs between proliferation and...
CHAPTER 1

differentiation\textsuperscript{99}. FPD/AML patients show an increased number of morphologically abnormal micromegakaryocytes in their BM biopsies\textsuperscript{95,98}. Consistently with the phenotypes observed in knockout studies\textsuperscript{89}, these cells display a poorly developed cytoplasm and a reduced nuclear lobulation\textsuperscript{49,76,92,93}. Platelets show a marked reduction of dense granules and significant defects in aggregation response\textsuperscript{95,98}, indicating that thrombocytopenia in FPD/AML is also associated with platelet dys-functions. In combination with the fact that MKs of FPD/AML patients fail to assemble and extend proplatelets, these findings imply an important role for RUNX1 in the final stages of MK maturation\textsuperscript{98,100}. Most of FPD/AML cases exhibit similar clinical manifestations. Although recent studies have raised the possibility that some of these congenital RUNX1 mutations may act in a dominant-negative manner\textsuperscript{96,101}, many of them results in haploinsufficiency, indicating that the defects observed in FPD/AML patients derives from a RUNX1 dosage effect.

\textit{FLI1}

DNA binding sites for the ETS family of winged helix-turn-helix transcription factors are over-represented in promoters of many MK-specific genes\textsuperscript{102}. Components of this family contain a highly conserved DNA interaction domain that binds to genomic sequences harboring a core GGA(A/T) consensus motif\textsuperscript{103}, which often flank GATA recognition sites\textsuperscript{104}. An important member of this family is Friend leukemia virus integration 1 (FLI1), which has been shown to play a central role in MK development and platelet biogenesis\textsuperscript{102,105,106}. This gene is located on a particular region on chromosome 1\textsuperscript{107} which is partially deleted in patients affected by Paris-Trousseau thrombocytopenia (TCPT) and the more severe Jacobsen syndrome (JBS)\textsuperscript{106,108,109}. These disorders are associated with physical and developmental defects and mental retardation, the severity of which depends on the size of the deleted region. An increased bleeding tendency, associated to a mild thrombocytopenia, is also observed in these pathologies\textsuperscript{109}. \textit{FLI1} has been found to be deleted in almost all TCPT/JBS cases affected by bleeding complications\textsuperscript{106,108}. The majority of circulating platelets in these patients exhibit a normal morphology, but a small subset of the whole population (roughly 15\%) includes abnormally enlarged platelets with a solitary giant \(\alpha\)-granule originated by fusion\textsuperscript{110}. Bone marrow examination of TCPT/JBS patients revealed the presence of subsets of immature micromegakaryocytes with small and poorly developed \(\alpha\)-granules, mono-lobulated nuclei and abnormally distributed internal membranes\textsuperscript{109}. As \textit{FLI1} gene expression undergoes mono-allelic inactivation in early stages of differentiation, hemizygous loss of \textit{FLI1} may result in the absence of FLI1 expression, thus causing dys-megakaryopoiesis and explaining the existence of discrete populations of normal and abnormal MKs. Lentiviral overexpression of FLI1 in HSCs from TCPT patients rescued many of these abnormalities \textit{in vitro}, proving that FLI1 plays an important role at multiple levels in megakaryopoiesis\textsuperscript{110}. 
The gray platelet syndrome

Defective production or secretion of platelet granules is causative for a number of bleeding disorders that can be clinically classified as storage-pool diseases (SPDs)\textsuperscript{111–113}. The gray platelet syndrome (GPS) is a rare inherited SPD with a variable degree of severity that was originally described by Raccuglia et al. in 1971\textsuperscript{114}. This disease is characterized by a mild to moderate, occasionally severe, thrombocytopenia, splenomegaly and the presence of enlarged platelets that, upon morphological staining, appear gray under light microscopy\textsuperscript{111–120}. This typical appearance is caused by a paucity of α-granules and their constituents in platelets\textsuperscript{121}. As previously mentioned, α-granules are intracellular storage pools that contain hundreds of different proteins which are secreted to mediate adhesion and aggregation of platelets at the site of vessel damage. The biogenesis of α-granules is a complex and still poorly understood process. α-granules develop following the fusion of multivesicular bodies, which are formed upon maturation of budding vesicles in the Golgi apparatus, with endocytic vesicles, and are transported to the pro-platelet protrusions\textsuperscript{122}. In GPS, proteins that are endogenously synthetized or endocytosed by MKs fail to be incorporated into functional α-granules\textsuperscript{123}. Consequentially, it has been hypothesized that these proteins may be released in the extracellular marrow, where they may stimulate the formation of fibers and contribute to the frequently observed myelofibrosis of GPS patients\textsuperscript{114,119,120}. MKs are highly vacuolated, with large cytoplasmic areas devoid of cell organelles, and are often present in normal number\textsuperscript{119,123}. In contrast to α-granules, other platelet organelles such as lysosomes and dense granules develop normally\textsuperscript{123}. A curious phenotype that is associated to the MKs of GPS patients is emperipolesis, which is defined as the active penetration by one cell into and through larger cells\textsuperscript{117,119}. In fact, more than one neutrophil can be seen embedded within the cytoplasm of a single GPS MK. Because engulfed neutrophils can degranulate and release molecules that promote fibrogenesis\textsuperscript{124}, emperipolesis is also thought to contribute to the still unclear etiology of myelofibrosis in GPS.

Both autosomal dominant and autosomal recessive forms of the disease have been reported. However, the majority of GPS cases exhibit an autosomal recessive inheritance\textsuperscript{119}. Recently, more than 35 different \textit{NBEAL2} (neurobeachin-like-2) homozygous or compound heterozygous alterations have been identified in patients with autosomal-recessive GPS\textsuperscript{125–127}. Although \textit{NBEAL2} has been shown to localize within the endoplasmic reticulum\textsuperscript{125}, where it may be required for the sorting or packing of granule components, the precise molecular functions carried out by \textit{NBEAL2} in MKs and platelets remain elusive. The types of mutations found and their zygosity have been associated with heterogeneous clinical manifestation of the disease, ranging from moderate to severe macrothrombocytopenia with complete absence of α-granules in almost all platelets\textsuperscript{128}. Interestingly, as individuals with the same \textit{NBEAL2} mutation may have divergent clinical phenotypes, additional factors are likely to contribute to differences in bleeding severity of \textit{NBEAL2}-related GPS. Because \textit{NBEAL2} genetic variations are found in many, but not all, GPS cases, further efforts are required to identify new genes that are responsible for the pathological mechanisms underlying GPS.
Growth factor independence transcription factors

The Growth Factor Independence (GFI) family of transcription factors includes the transcriptional repressors GFI1 and its parologue GFI1B\textsuperscript{129-132}. Ablation of both genes is incompatible with HSC biology\textsuperscript{133} and deletion of either GFI1 or GFI1b in mice results in defective hematopoiesis with different lineages being affected\textsuperscript{134-141}. Whereas GFI1 is important for proper formation and function of cells from the adaptive and innate immune system\textsuperscript{135,136}, GFI1B is important for MK and red blood cell development\textsuperscript{137,138,140}. Besides these complementary functions, both proteins regulate proliferation and survival of HSCs\textsuperscript{134,139,140}. However, while loss of GFI1 results in reduced HSC self-renewal capacity\textsuperscript{139}, this phenotype is not observed upon loss of GFI1B\textsuperscript{140}. Finally, it has been recently shown that together with other transcription factors GFI1B is able to program hemogenesis in fibroblasts\textsuperscript{142}, further highlighting the broad and essential role of GFI proteins in blood cell development.

Expression, structure, regulation and function of GFI1 and GFI1B

The distinct non-overlapping hematological functions of GFI1 and GFI1B reflect the mostly mutually exclusive and tissue-specific expression of these genes\textsuperscript{143,144}. In many hematopoietic lineages, GFI1 and GFI1B have been shown to directly cross-repress each other's expression or to coordinate their own expression by auto-regulatory feedback loops\textsuperscript{143,145-149}. However, in the cellular compartments where both GFI1 and GFI1B are co-expressed, additional regulatory mechanisms are likely to be implicated. In fact, several transcription factors such as GATA1, NF-Y, Oct1, Ebf1 and HMGB2, for GFI1B\textsuperscript{146,148,150}, and Ikaros\textsuperscript{151} and Ajuba\textsuperscript{152}, for GFI1, have been reported to co-regulate the expression of GFI-genes. Additionally, also microRNAs\textsuperscript{153} and the ubiquitin proteasome-degradation system\textsuperscript{154,155} have been implicated in the regulation of GFI protein expression and stability, indicating that GFI1 and GFI1B levels are tightly controlled at multiple levels.

The introduction of the GFI1B gene in the GFI1 locus has been shown to rescue nearly all the hematopoietic phenotypes observed in GFI1-deficient mice\textsuperscript{156}, suggesting that both GFI proteins perform similar activities at the molecular level in hematopoiesis. Consistently, the protein structure of GFI proteins is largely identical and well conserved among vertebrates, with a SNAG (Snail/GFI1) domain at the N-terminal region, a central intermediate region and six zinc finger domains at the C-terminal region\textsuperscript{129-132}. Found in other transcriptional repressors, the SNAG domain is a conserved 20 amino-acid long region, which is essential for the recruitment of histone modifying proteins\textsuperscript{157}. At the other side of the protein, six C\textsubscript{2}-H\textsubscript{2} zinc fingers mediate protein-protein interactions and also the binding to the DNA, which is exclusively dependent on the integrity of zinc fingers 3 through 5\textsuperscript{129,158,159}. In contrast to the SNAG and the zinc finger domains, which are 89% identical at protein level, the intermediate part shows only 39% similarity between GFI1 and GFI1B. Recently, it has been proposed that PIAS3-mediated sumoylation of a SUMO-consensus motif present within this region induces GFI1 degradation\textsuperscript{160}. As this modification seems to be antagonized by the interaction with Notch1 intracellular domain N1-ICD\textsuperscript{161}, these findings suggest that the GFI intermediate region is likely to provide a binding platform for the interaction with specific co-factors and to function as a regulatory domain. Remarkably, the same SUMO-consensus site is absent
in GFI1B. This implies that although both GFI1 and GFI1B function in a similar biochemical manner, their activity is differentially modulated at the post-translational level, possibly depending on the cellular and molecular context and through the interaction with specific binding partners.

GFI1 and GFI1B genes encode for 422 and 330 amino acid transcription factors, respectively, that recognize specific “AATC” containing consensus-sequences within the promoters of target genes and inhibit their expression. This is achieved by remodeling chromatin through the recruitment of proteins that modify histones. Upon binding to the DNA, GFI proteins form a complex with deacetylases and (de)methylases which, in turn, apply epigenetic marks on histones that render chromatin structure inaccessible for gene transcription. Furthermore, this process is facilitated by the engagement of a variety of other co-repressor molecules that enhance the repressive activity of the GFI-complex. Although the molecular events are largely unknown, it has recently been shown, at least for GFI1, that the p150 subunit of the chaperon Caf-1 (chomatin assembly factor 1) bridges the repressive complex to histones, enabling possible co-factors like HDACs 1-2 (histone deacetylases), LSD1 (lysine specific demethylase 1), CoREST/Rcor1 (Rest co-repressor) and the major histone methyltransferases G9A and SUV39H1, to be recruited and contribute to gene repression. Interestingly, it has been reported that GFI proteins can also function as positive regulators of gene transcription, although the underlying molecular mechanisms remain to be elucidated. Moreover, GFI proteins were shown to physically interact with and reciprocally influence the activity of other transcription factors. For example, GFI1B controls the activity of GATA1 upon binding on erythrocytic gene promoters, while GFI1 can co-regulate gene expression upon DNA-dependent binding to ETS1 and DNA-independent binding to Miz-1 and Pu.1, indicating that GFI proteins can exert their transcriptional functions through different modes of action. Besides gene expression, GFI1 has been also shown to regulate alternative splicing and protein post-translational modifications. However, these latter molecular activities have not been reported for GFI1B as yet.

Biological and molecular activities of GFI1B during hematopoiesis

GFI1B is expressed in HSCs and highly expressed in MEPs and cells of the erythroid and megakaryocytic lineage. Germline deletion of a single GFI1B allele in mice does not have apparent effects on hematopoiesis. However, ablation of both GFI1B loci is lethal during embryogenesis due to the failure of erythrocyte precursors to produce fully differentiated red blood cells. Conditional knockout mouse models have been generated to investigate the biological functions of GFI1B during adult hematopoiesis. Conditional loss of GFI1B has been shown to affect dormancy and mobilization of HSCs, resulting in a large expansion of HSCs in both bone marrow and blood. Because HSCs from these mice retain the capacity to self-renew and differentiate towards all the lineages, this suggests that GFI1B inhibits their proliferation, while it is dispensable for HSC maintenance and clonogenic capacity. Besides HSCs, GFI1B is essential for proper red cell and platelet formation. Conditional inactivation of GFI1B in adult mice has been shown to affect erythroid and megakaryocytic
development at different stages. Upon *GFI1B* knockout, erythroid development is blocked at the progenitor stage, while megakaryopoiesis does not proceed beyond nuclear polyploidization. In these mice, MKs fail to undergo cytoplasmic development and thus fail to produce platelets, suggesting that *GFI1B* is required for MK maturation. *GFI1B* is also expressed in specific subsets of B- and T- cells. However, the importance of *GFI1B* in the development and function of these cell types remains to be elucidated.

Despite the molecular mechanisms controlled by GFI1B are poorly understood, all the phenotypes mentioned thus far have been associated with aberrant expression of several genes. For example, GFI1B seems to be important for the expression of cell surface molecules in HSCs, although it is not clear whether these genes are directly or indirectly controlled. Following animal and in vitro studies, it has been proposed that GFI1B targets the growth factor-β receptor III gene and genes involved in embryonic globin expression during erythroid development. Although several other genes have been reported to be directly targeted by GFI1B transcriptional activity during erythropoiesis, the putative targets of GFI1B during megakaryopoiesis are, as yet, less known. GFI1B has been shown to repress the expression of *meis1*, an important proto-oncogene associated with malignant transformation, in erythrocytic, but not megakaryocytic cells, suggesting functional and regulatory differences between the two lineages. Thus, to fully decipher the intricate fashion, through which GFI1B controls proliferation, differentiation and survival of blood cells, it will be necessary not only to determine additional target genes, but also the co-factors used in the various hematopoietic subsets and how these cooperate with GFI1B to control gene expression.

**Pathological role of GFI1B in blood-related diseases**

In spite of the fact that loss of GFI1B function causes various defects in mouse models, the direct involvement of GFI1B in the development of hematological diseases is still controversial. Early studies have demonstrated that GFI1B co-operates with known oncogenic genes in hematological malignancies and regulate proliferation of erythroid and megakaryocytic malignant cell lines. In addition, a short GFI1B isoform (GFI1B p32), lacking the zinc fingers 1 and 2, was found to be over-expressed in patients with AML and chronic myeloid leukemia (CML). However, in contrast to *GFI1*, that is mutated in patients with neutrophilia and predispose to leukemia when a specific polymorphism is present in its coding sequence, genetic variations in the *GFI1B* gene have not been found in blood related disorders. Part of this thesis, presented in chapter 2, will describe the discovery of a dominant-negative mutation in the *GFI1B* gene to be causal to an autosomal dominant form of GPS, thereby illustrating the crucial importance of GFI1B in human MK development and platelet production.

**Acute Myeloid Leukemia**

AML is a disease of the myeloid lineage characterized by the clonal expansion of dysfunctional cells that have lost their proliferative control and the ability to differentiate. These immature
cells, denominated blasts, amass within the bone marrow microenvironment and they can frequently be found in the peripheral blood, where they account for the high white blood cell count often found in AML patients. The accumulation of leukemic cells in the bone marrow results in the suppression of normal hematopoiesis with the consequential manifestation of typical clinical symptoms such as fatigue, anemia, bleeding disorders and infections. AML is an aggressive type of cancer with an incidence that increases with age\(^{178}\). In children, although the incidence is lower, AML is among the most frequently occurring malignancies. Therapeutic strategies in medically fit patients involve high-dose chemotherapy and in many cases HSC transplantation\(^{179}\). Current treatment can be divided in two phases: induction and consolidation therapy. Induction therapy aims at the eradication of the bulk of leukemic cells and the restoration of normal hematopoiesis. Complete remission is achieved when patients display normal cell count in the peripheral blood and the leukemic blast counts in the bone marrow drops from more than 20% to below 5%. Consolidation therapy is then applied to eliminate residual leukemic cells and to prevent the risk of relapse. This phase requires additional chemotherapy and allogeneic stem cell transplantation, which is usually performed in selected patients at high risk of relapse. In the past years, intensive efforts have been made to improve survival of AML patients, but still only 40% of individuals younger than 60 years survive more than 5 years from diagnosis\(^{180–182}\). This is largely due to the high relapse rate observed in patients, indicating that current therapies are unable to completely eradicate the disease in the majority of the cases. Because of this, it is believed that there is a rare subset of malignant cells that can survive treatment. These cells are responsible for relapse and are recognized as leukemia stem cells (LSCs).

Leukemia stem cells

AML arises following a stepwise accumulation of multiple genetic defects. These alterations affect critical genes involved in proliferation, differentiation and apoptosis, thus perturbing normal cellular functions. The series of transforming events is thought to start with the formation of pre-LSCs (Figure 2). These cells arise early in leukemic transformation and are considered to be HSCs that acquire initiating alterations which promote their clonal expansion. Consistently, pre-LSCs have been shown to possess a multi-lineage repopulation advantage over non-mutated HSCs in the reconstitution of xenograft models, but they are unable to induce AML\(^{183}\). Thus, although pre-LSCs can sustain blood cell development, they represent a pool of abnormal cells from which LSCs can evolve. The transition into fully transformed LSCs requires the co-occurrence of cooperating alterations, which collaborate with the initiating events to transform pre-LSCs in leukemia-generating LSCs. Xenograft studies have revealed that only a specific population of primary AML material can induce and sustain the growth of leukemia in mice\(^{184}\), showing that the leukemic clone in AML is organized in a hierarchical fashion. As in normal hematopoiesis, LSCs sit at the apex of the organization, are highly-quiescent and retain the capacity to generate more differentiated proliferating leukemic blasts. The fact that serial transplantation of LSCs in immuno-deficient mice repeatedly induces leukemia indicates that these cells also possess a self-renewing capacity. The acquisition of this feature is essential for these cells to
survive and continuously replenish the bulk of the tumor population. Because unlimited self-renewal is an intrinsic feature of normal HSCs, but not of normal hematopoietic progenitors, it was originally assumed that LSCs may be exclusively present in the CD34^+CD38^- cell subset, which defines HSCs. However, LSC activity has also been found in the more mature CD34^+CD38^+ and CD34^- cell subsets. This indicates that the cell of origin of AML may also derive from more mature cell fractions, and thus it may differ between AML patients.

In addition, consistent with the proposed hierarchical organization, immature LSCs can give rise to both immature and mature phenotypes when transplanted in secondary recipients, while more mature LSCs can only give rise to the same mature phenotype. Because mature LSCs cannot regress to an immature phenotype, this supports the existence of a different degree of differentiation between LSCs, which may reflect the differentiation status of the cell-type originally involved in LSC transformation. Classical treatment regimens are mainly targeted at the highly proliferative bulk of leukemic cells. However, the quiescent nature of LSCs may confer an important tolerance against conventional

**Figure 2. Schematic representation of leukemia transformation.** During normal hematopoiesis, hematopoietic stem cells (HSCs) give rise to committed progenitors, which in turns generate all mature blood cell types (left side). Following the acquisition of an initial genetic defect (yellow lightening symbol), pre-leukemic stem cells (pre-LSC) are formed. These acquire additional aberrations, resulting in the generation of LSCs, or leukemia initiating cells (LICs), which are characterized by increased self-renewal (circular arrow) and block of differentiation, leading to the accumulation of dys-functional leukemic blasts that form the bulk of tumor cells.
chemotherapy, explaining the high relapse rate observed in patients. Because LSCs represent a self-renewing population that is capable to resist treatment and maintain the disease, it is believed that the identification and elimination of this population is fundamental to increase survival of AML patients.

Prognostication in AML

AML is a clinically and genetically heterogeneous disease characterized by the presence of recurrent genetic defects. Some of these acquired aberrations have been shown to correlate with distinct clinical outcomes. Because these abnormalities carry important prognostic information, they are used as risk factors to predict survival and therapy response. In clinical practice, AML patients can be classified in four distinct groups with very poor, poor, intermediate and relatively good prognosis based on these and few other risk factors, such as age and white blood cell count (Table 2). Their presence strongly influences the therapeutic approach and thus their identification has a crucial relevance in the management of AML patients. Frequently observed genetic changes in AML include abnormalities at the cytogenetic level, like for instance gain or loss of genomic material and translocations, or more subtle abnormalities observable at the molecular level, such as single gene mutations and altered gene expression. Prognostically significant cytogenetic abnormalities that are currently used in diagnostic protocols include the t(8;21), inv(16), and t(15;17) translocations, which result in the RUNX1-ETO, CBFB-MYH11 and PML-RARA fusion proteins, respectively, and predict a relatively good clinical outcome. Furthermore, loss or gain of genetic material, as seen in monosomies of chromosome 5 or 7, deletion of the long arm of chromosomes 3, 5 and 7 are considered as poor prognostic markers. Complementarily, molecular lesions are particular important to improve the diagnostic stratification of otherwise undefinable risk-groups. In line with this, patients with normal cytogenetics are routinely screened for the presence of these alterations. Frequently observed molecular defects are found in the fms-like tyrosine kinase 3 (FLT3), the CCAAT/enhancer binding protein alpha (CEBPA) and nucleophosmin 1 (NPM1) genes. In combination with TP53, ASXL1, RUNX1 and c-KIT, these are the genes predominantly screened to guide risk assessment and consequential treatment decisions. With the implementation of next generation sequencing techniques, an emerging number of additional somatic mutations have been detected, particularly targeting genes involved in epigenetic regulation such as TET2, IDH1/2, DNMT3A, EZH2, and ASXL1. Although the underlying mechanisms through which these alterations contribute to leukemia are still elusive, they nonetheless have been shown to represent an additional source of prognostic information. As illustrated in chapters 3 and 4, also analysis of gene expression, either of single genes or of multiple genes may improve the stratification of patients into different risk groups. Several genes, of which expression is dys-regulated in AML, have been shown to be of clinical relevance. The most important example is high expression of EVI1, which is detected in roughly 8% of AML patients and which is associated with an extremely poor prognosis. Besides EVI1, also the differential expression of BAALC, MN1, ERG and BRE has been shown to predict disease outcome, but the incorporation of these risk factors in clinical protocols...
is still very limited. Over the last decade, several studies using gene expression profiling (GEP) have been conducted on large series of AML patients\textsuperscript{192,193,205–207}. These studies have shown that genetic aberrations associated with favorable survival, such as \textit{inv(16)}, \textit{t(15;17)} and \textit{t(8;21)} and double \textit{CEBPA} mutations, can be linked to distinct expression profiles\textsuperscript{206,207}, indicating that GEP could be useful in diagnostication. Although not all known aberrations can be tracked back to a reciprocal gene expression signature, the use of GEP has also led to the discovery of novel disease entities and expression signatures associated with disease outcome. Importantly, GEP provides transcriptome-wide information on genes of which expression is dys-regulated in the various subtypes of AML, allowing the identification of pathways that are targeted by the underlying genetic defects and thus revealing precious insights into the pathogenesis of AML\textsuperscript{205}.

The increasing availability of molecular and cytogenetic data has permitted the refinement of AML nosology and prognostication. However, many pathological events are still unknown and the majority of the currently delineated disease categories display a substantial heterogeneity. Therefore, the identification of additional risk factors that accurately predict outcome and that allow patients to be matched to the most appropriate therapies is warranted.

Towards targeted therapy
The possibility to stratify patients based on risk factors has pushed the field of hematology towards more individualized therapies, which take into account both the clinical and the genetic background of AML patients. However, although most patients achieve complete remission with first line chemotherapy, the majority of them ultimately relapse\textsuperscript{180–182}. With the exception of particular AML subclasses, such as those with \textit{t(8;21)}, \textit{inv(16)}, and \textit{t(15;17)}, it has not been possible to significantly improve the outcome of AML patients in the last three decades, particularly concerning patients older than 60-65 years. Intensification of the current therapeutic regimens or usage of higher doses of drugs is impaired by the fact that a large number of patients are not medically fit to undergo such arduous treatments (due to old age, co-morbidity, etc.) which employ toxic drugs. Thus, the identification of more selective approaches, which specifically target leukemic cells whilst minimizing side effects, is necessary to improve survival rates in AML. Important examples of successful application of targeted therapy include those interfering with the activity of the fusion oncoproteins \textit{BCR-ABL}\textsuperscript{208,209} and \textit{PML-RARA}\textsuperscript{210}, which are typically found in patients suffering from chronic myeloid leukemia (CML) and acute promyelocytic leukemia (APL), respectively. In the former case, it has been shown that the treatment of CML patients with tyrosine kinase inhibitors that interfere with the activity of BCR-ABL dramatically improves the survival of CML patients\textsuperscript{208}. In the case of PML-RARA, this abnormality induces an arrest in differentiation due to the repression of RARA signaling. Treatment with all-trans retinoic acid (ATRA) induces PML-RARA dependent transcription and subsequent release of the differentiation block. In combination with chemotherapy, the addition of ATRA to the treatment has resulted in an impressive improvement of survival\textsuperscript{211}. In contrast to CML and APL, for which the therapeutic targets are well defined, the remaining types of AML are more
heterogeneous at the molecular level and the identification of critical targets for therapy may be more difficult. Nevertheless, besides refining patient stratification, our increasing knowledge of the genetic abnormalities present in AML has also led to the identification of several biological pathways that are disrupted in leukemic cells. It has become apparent that alterations driving neoplastic transformation and progression are not confined to tumor suppressors or oncogenes involved in cell growth and survival, but also affect proteins that have functional roles in the regulation of other global processes, such as chromatin remodeling and transcriptional regulation, cell adhesion, genome methylation and signal transduction. The targeting of these pathways, using inhibitors of DNA methyltransferases, histone deacetylases, tyrosine kinases or other, new drugs that specifically interfere with mutated oncoproteins, represents a promising opportunity to develop and refine targeted therapies and to meliorate the outcome of AML patients.

Table 2. European Leukemia Network (ELN) prognostic system (Adapted from\textsuperscript{191})

<table>
<thead>
<tr>
<th>Prognostic group</th>
<th>Subsets</th>
<th>CR rate; relapse rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable</td>
<td>\textit{Inv}(16) or t(16;16); t(8;21) NK and \textit{NPM1+}/\textit{FLT3 ITD}--; NK and \textit{CEBPA} +/-</td>
<td>\ (&gt;80-90%); 35-40%</td>
</tr>
<tr>
<td>Intermediate-1</td>
<td>NK and \textit{NPM1+}/\textit{FLT3 ITD} – or \textit{CEBPA} +/-</td>
<td>50-80%; 50-60%</td>
</tr>
<tr>
<td>Intermediate-2</td>
<td>Cytogenetic abnormalities not in favorable or adverse groups; \textit{FLT3 ITD} +</td>
<td>40-60%; 70-80%</td>
</tr>
<tr>
<td>Adverse</td>
<td>-5, -7, 5q-, abn 3q, 17p, 11q (other than t(9;11), t(6;9), complex</td>
<td>&lt;50%; &gt;90%</td>
</tr>
</tbody>
</table>

NK, normal karyotype; +, abnormality present; -, abnormality absent; +/-, double mutation; abn, abnormalities; CR, complete remission

**NF-kB signaling and ubiquitination**

Post-translational modifications are reversible additions of biochemical functional groups onto specific amino acid residues present within protein structures. More than 200 types of post-translational modifications including covalent attachments of lipids, carbohydrates, particular chemical groups or even proteins, have been documented thus far\textsuperscript{212}. These transient alterations frequently affect protein function by changing the protein structure and dynamics or by modifying residues that may be implicated in the binding with a partner. Whether a protein is, for instance, (de)activated, degraded, assembled or activated for trafficking through the cell is determined by the type of modification and the amino acid residue involved. Biologically, post-translational modifications serve to coordinate the fate and function of proteins thereby governing protein signaling networks\textsuperscript{213}. Indeed,
they are involved in the regulation of a large number of cellular activities, also including gene expression control. A post-translational modification that has gained increasing relevance in the past two decades is ubiquitination. Originally believed to be solely implicated in the proteosomal degradation of proteins\textsuperscript{214,215}, ubiquitination has more recently been discovered to be involved in a plethora of cellular processes and it is nowadays regarded as an important signaling molecule\textsuperscript{216–221}. An excellent example to illustrate the signaling role of ubiquitination is represented by the NF-kB signaling pathway\textsuperscript{222}, which is essential for a large variety of biological processes including apoptosis, bone formation, viral replication, tumorigenesis, inflammation and immunity\textsuperscript{223,224}. In hematopoiesis, several hematopoietic types of cells with either myeloid or lymphoid origin rely on the activity of NF-kB signaling for their development and function\textsuperscript{225,226}. In addition, NF-kB is directly involved in the development of autoimmune and inflammatory diseases and it has been shown to be activated in leukemia\textsuperscript{227,228}. In the next section, the relevance of a particular form of ubiquitination, namely linear ubiquitination, will be discussed in the regulation of NF-kB signaling\textsuperscript{229,230}.

Ubiquitin and the ubiquitination machinery

The central molecule of the ubiquitination pathway is a 76 amino acid globular peptide called ubiquitin, which is one of the most conserved proteins among eukaryotic species\textsuperscript{231,232}. The fact that it is encoded by four different genes in humans underscores the fundamental and indispensable role that is covered by ubiquitin in nature\textsuperscript{231}. The linkage of ubiquitin to specific targets consists of the sequential transfer of a single ubiquitin moiety through three distinct types of specialized enzymes (Figure 3). First, an E1 ubiquitin-activating enzyme activates the carboxyl group of the C-terminal glycine residue of a free ubiquitin through the hydrolysis of ATP to ADP\textsuperscript{233}. As such, this residue can form a thioester bond with a specific cysteine residue that is located within the active site of the E1 activating enzyme, resulting in the loading of ubiquitin.

Figure 3. The ubiquitin system. The ubiquitination reaction begins with the ATP-dependent activation of ubiquitin (Ub, green) by the E1-enzyme, and continues with the formation of
a thioester bond between Ub and an active-site cysteine present in the E2 enzyme. The final transfer of Ub onto its target is mediated by E3 Ub-ligases, which promote a direct transfer from the E2 enzyme (as it is the case for RING ligases), or involve the formation of an intermediate thioester with Ub (HECT-, RBR-ligases).

This initial step is subsequently followed by the transfer of ubiquitin to an E2 ubiquitin-conjugating enzyme to form another intermediate thioester bond with an active site cysteine residue. Finally, an E3 ubiquitin ligase recognizes and brings in close proximity both the E2 conjugating enzyme and the target protein, catalyzing the formation of an isopeptide bond between the C-terminus of a donor ubiquitin and the ε-amino group of lysine (K) residues within the substrate protein. Lysine is not the only amino acid that may serve as a final acceptor for ubiquitin. Also serine, threonine and cysteine residues have been reported, although the underlying biochemical mechanism and the biological relevance of these modifications remain elusive. The mode of action of the E3 ligases is dependent on the motif that mediates the interaction with the E2 enzyme. E3 ligases can be subdivided in several classes because distinct structural domains that interact with E2s have been identified. All E3 ligases bind an ubiquitin-charged E2 and catalyze the transfer of ubiquitin from the E2 either directly to a substrate lysine, as this is the case for RING (really interesting new gene) E3s, or through the formation of an intermediate linkage with an active-site cysteine within the E3 itself, as it is the case for HECT (homologous with E6-associated protein C-terminus) and RBR (ring in-between fingers) E3s. Although the RING-type ligases do not physically participate to the transfer of ubiquitin, they greatly increase the efficiency of E2 activity and enhance the conjugation reaction. On the other hand, HECT- and RBR-type ligases play a direct role in substrate ubiquitination by forming a thioester intermediate bond necessary for E2-independent conjugation to the substrate.

As E3 ligases determine the substrate to be ubiquitinated, these components are responsible for the specificity of the ubiquitination machinery, while the E2-E3 combination generally determines the site of the attachment to the substrate. Although only one E1 activating enzyme and about 50 E2 conjugating enzymes have been described, a large number of approximately 1000 putative E3 ligases have been identified in the human genome.

Finally, ubiquitination is opposed by the activity of roughly 90 different de-ubiquitinating enzymes (DUBs), which selectively cleave ubiquitin from the substrate, thereby enabling a time dependent reversibility of the process. The balance between ubiquitination and de-ubiquitination determines the fate of target proteins. As DUBs are outside of the scope of this thesis, they will not be introduced further.

The fate of substrates depends on the type of ubiquitin modification

Distinct types of ubiquitin modifications can occur in nature, each of them with a specific signaling function and biological meaning (Figure 4). The attachment of a single ubiquitin molecule is referred to as mono-ubiquitination and can induce a variety of regulatory events that usually promote the recruitment or the displacement of proteins to and from...
Mono-ubiquitination

Multimono-ubiquitination

K48-linked ubiquitination

K63-linked ubiquitination

Mixed chain ubiquitination

Linear ubiquitination

Figure 4. Ubiquitin chain topologies. Different forms of ubiquitination are presented. Ub can be linked to the substrate (purple) as a monomer on either a single K residue (mono-ubiquitination) or multiple K residues (multimono-ubiquitination) of the same target. Multiple Ub molecules can be linked together through internal K residues to form chains (poly-ubiquitination), as seen, for example, in K48-linked, K63-linked and mixed chain ubiquitination. Linear chain ubiquitination occurs when the linkage between Ub molecules involves the C-terminus of the donor Ub and the N-terminus of the acceptor Ub. Abbr.: K: lysine; C: cysteine; M: methionine.

Functional complexes\(^{244-247}\). Mono-ubiquitination targeting of distinct lysine residues within the same substrate is defined as multi-mono-ubiquitination, and it has, for example, been described to be involved in the nuclear export of p53\(^{246,247}\). In addition, because ubiquitin itself contains seven lysine residues (K6, K11, K27, K29, K33, K48 and K63) that can be substrates for ubiquitination, several ubiquitin molecules can be serially linked together to generate chains\(^{248}\). All lysine residues in ubiquitin can be involved in poly-ubiquitination. A homotypic chain is formed when the same lysine residue of consecutive ubiquitin molecules are used for the assembly. When different lysine residues are implicated, a mixed chain is constituted\(^{249,250}\). Due to limitations of the current technologies applied to study ubiquitin chain topologies, only a small part of ubiquitin chain types has been investigated, with K48- and K63-linked ubiquitin chains being the most well characterized types of chains thus far\(^{251,252}\).
polyubiquitin chains, functions as a signal for proteasomal degradation of proteins\(^{253-255}\). Briefly, these structures are recognized by specific ubiquitin binding domains (UBDs)\(^{256}\) that facilitate the association of K48-ubiquitinated proteins to the regulatory 19S subunit of the proteasome. This mediates the unfolding and the transport of targets into the catalytic 20S core of the complex, which consists of a barrel-shaped subunit with protease activity that degrades cellular proteins into peptides\(^{255}\). In contrast to K48-linked chains, labeling with K63-polyubiquitin chains do not mark proteins for degradation, but rather modulates the activity of substrates. These chains are created through the specific activity of UBC13\(^{257,258}\), the only E2 enzyme known to catalyze their formation, and usually function as a scaffold to facilitate the recruitment of signaling proteins. Consistent with this, K63-ubiquitination has been implicated in the internalization of surface receptors\(^{259,260}\), DNA repair mechanisms\(^{257}\) and signal transduction\(^{261,262}\). Besides K48- and K63-linked ubiquitin chains, unconventional types of chain topology are represented by K11-linked chains\(^{248}\), that have been proposed to function as a recruiting signal\(^{261,263,264}\) and play a role in cell division\(^{265}\), and linear chains, which will be discussed next in more detail.

Linear ubiquitination and LUBAC

Linear chains represent a unique type of ubiquitination. In these chains, the linkage between ubiquitin moieties does not consist in the attachment of the C-terminus of the donor ubiquitin to one of the ε-amino group of lysine residues of the acceptor ubiquitin, but rather to the N-terminal amino-group of the final methionine. Because the end of an ubiquitin moiety is then attached to the beginning of the preceding molecule, this type of modification was defined “head-to-tail” and generates “linear” or M1-chains\(^{266}\). Linear ubiquitination is catalyzed by a tripartite complex denominated LUBAC\(^{266}\) (Linear-UBiquitin-Assembly-Complex), which is composed of HOIL-1L\(^{267}\) (haem-oxidized IRP2 ubiquitin ligase 1L, RBCK1), HOIP\(^{268}\) (HOIL-1L-interacting protein, RNF31) and SHARPIN\(^{269}\) (SHANK-associated RH domain interacting protein). The catalytic activity of LUBAC resides within HOIP, while HOIL-1L and SHARPIN binds to HOIP and act as crucial activators\(^{266–269}\). Figure 5 illustrates the domains involved in the interaction between LUBAC components. Although the existence of other E3 complexes equipped with linear ubiquitination activity cannot be excluded, LUBAC is currently the only complex described to assemble linear ubiquitin chains.

HOIP acts as RBR ubiquitin ligase

Upon LUBAC formation, HOIP generates linear chains through a mechanism that is typically found in the small family of E3 RBR (RING-in-between-RING) ligases\(^{240–242}\). A total of fourteen different RBR ligases have been identified in humans to date\(^{270}\). They are complex multi-domain proteins characterized by the presence of a highly conserved catalytic unit, which contains an IBR (in-between-RING) domain flanked by two RING structures, of which RING1 resembles the catalytic domain of RING-type ligases, while RING2 does have a peculiar structure\(^{271,272}\). These proteins function through a HECT/RING hybrid mechanism\(^{242}\). According to the current model, RBR-mediated linear ubiquitination initiates with
the association of RING1-IBR region of HOIP to an ubiquitin-charged E2 conjugating enzyme, an activity that is typically observed in RING-type E3 ligases. However, before being eventually attached onto the acceptor ubiquitin, the donor ubiquitin is transferred to a conserved cysteine located in RING2, where it forms an intermediate thio-ester bond reminiscent to that of HECT-type ligases. In RBR ligases, the linkage specificity and the length of the chain are generally determined by the E2 enzyme. In the case of LUBAC, however, the specificity for linear ubiquitination prevails over that of the E2 enzyme and is intrinsic to HOIP, which produces linear chains regardless of the type of E2 that is used in the catalysis. The region that determines the linear specificity in LUBAC is a particular domain called the linear ubiquitin chain-determining (LDD) domain, which is unique to HOIP, as it has not been found in any other RBR ligases. Situated in the C-terminal part of the protein, LDD captures the acceptor ubiquitin and positions its N-terminal amino-group in a specific orientation that facilitates the formation of the “head-to-tail” linkage, as described in chapter 5.

LUBAC-mediated linear ubiquitination of NEMO is crucial for NF-kB activation

NF-kB is implicated in an important signal transduction route that translates various extracellular signals into transcriptional events. Stimulation by cytokines, oxidative stress or pathogens results in the activation of the Rel family of NF-kB transcription factors. Depending on the nature of the stimuli, these proteins determine the expression of specific sets of genes that mediate cellular responses. Two distinct pathways are responsible for the activation of NF-kB, the canonical and the non-canonical. The canonical pathway (Figure 6), which induces the expression of several genes involved in proliferation, differentiation and survival, is highly regulated at the post-translational level. As depicted

Figure 5. Schematic representation of LUBAC. LUBAC is composed of HOIL-1L, HOIP and SHARPIN. Domains involved in LUBAC formation are shown by double arrows. The interactions between HOIP NPL4 zinc finger 2 (NZF2) and the ubiquitin-like (UBL) domain of SHARPIN has also been proposed to be involved in LUBAC formation (not shown). The PUB domain (peptide:N-glycanase/UBA or UBX-containing proteins) interacts with the de-ubiquitinating enzyme OTULIN (also not shown).
in Figure 7, the activation of this signaling route requires an intricate cross talk between different types of ubiquitin chains and other post-translational modifications\textsuperscript{222,229,263,274}. LUBAC and related linear ubiquitination play a prominent and exclusive role in this process, as they contribute to the activation of NEMO (NF-kappa B Essential Modulator), which is a crucial mediator of the NF-kb signaling cascade\textsuperscript{267,269}. In short, following the activation of this pathway by pro-inflammatory cytokines, such as TNF-α and IL-1β, LUBAC is recruited to the TNF-α receptor where it interacts with and linearly ubiquitinates specific components of the NF-kb signaling cascade such as NEMO\textsuperscript{275} and, to a lesser extent, RIP1 (Receptor-Interacting Protein 1)\textsuperscript{276}. LUBAC-mediated linear ubiquitination of NEMO is important for the activation of the IKK complex, which is required for the degradation of IkBs and subsequent translocation of NF-kb transcription factors to the nucleus\textsuperscript{267,268,277}. The molecular mechanism underlying the involvement of LUBAC has been partially resolved. Upon receptor activation, LUBAC is recruited to the receptor together with IKK, and this process has been shown to depend on the activity of TRADD, TRAF2 and c-IAP proteins\textsuperscript{277}. At the receptor complex, the recognition of NEMO by LUBAC is mediated by the ZF domains of HOIP and HOIL\textsuperscript{267}. Linear ubiquitination of NEMO occurs within its UBAN domain, which is required for IKK activation, and it may induce conformational changes required for NEMO activation. In addition, the UBAN domain functions as a specific binding region for linear chains. Consistently, it has been shown that NEMO-bound linear chains serve as a binding

Figure 6. Schematic representation of the canonical NF-kB pathway. In resting cells, NF-Kb transcription factors (p50 and p65) are sequestered in the cytoplasm by inhibitory proteins, known as inhibitors of NF-kB (IkBa in this case), which prevent their translocation to the nucleus. Upon binding of TNF-α (or IL-1β) to its receptor, LUBAC mediated-linear ubiquitination activates NEMO, resulting in the phosphorylation of IKKβ and the activation of the IKK complex. Activated IKK is
platform for the recruitment of additional NEMO proteins. Multimerization of NEMO is thought to stabilize the NEMO-IKK complex, thereby promoting NF-κB activation. Of note, as depletion of LUBAC does not completely impair, but rather attenuates, canonical NF-κB signaling in response to pro-inflammatory cytokines, this indicates that LUBAC is dispensable for canonical NF-κB activation, but it operates to enforce and prolong the intensity of NF-κB activity. Finally, control over the balance between formation and degradation of linear chains is essential for proper signaling. Consistently, it has been reported that LUBAC is counteracted by the activity of specific DUBs, which interact with LUBAC, cleave linear polyubiquitin chains and consequentially prevent an excessive activation of NF-κB signaling. These include OTULIN (ovarian tumor DUB with linear linkage specificity /Gumby/ FAM105B) and CYLD (cylindromatosis). The roles of these proteins in the regulation of linear ubiquitination will not be discussed.

Biological functions of LUBAC

Consistent with the fact that canonical NF-κB is crucial to a variety of biological processes, LUBAC has been implicated in the regulation of various cellular functions besides TNF-α response. Important data on the physiological role of LUBAC has been obtained from animal studies. With the exception of HOIP, whose deletion in mice is embryonic lethal, ablation of the other two LUBAC components has been shown to cause several types of defects. Absence of SHARPIN activity is found in cpdm (chronic proliferative dermatitis in mice) mice, which lack SHARPIN due to germline mutations. These animals develop a spontaneous inflammatory syndrome, which has been associated to reduced NF-κB activation and increased TNF-α-induced cell death. In addition to dermatitis, cpdm mice exhibit several abnormalities including defective organ development and various immunodeficiencies. Unlike cpdm mice, ablation of HOIL does not result in apparent overt phenotypes. However, cells derived from HOIL-deficient mice do not respond properly to NF-κB stimulation. In cpdm and HOIL-deficient mice, many cellular pathways implicated in innate and adaptive immunity have been identified to rely on the linear activity of LUBAC at various levels. Known examples include the TLR (Toll-like receptors), RLR (retinoic acid-inducible gene (RIG)-I-like receptors) and NLR (nucleotide-binding oligomerization domain (NOD)-like receptors) pathways, which are known to modulate the response of monocytes and macrophages to microbial and viral infections. In adaptive immunity, LUBAC is indispensable for CD40 signaling, which is required for the activation of B-cells and the humoral immunological memory response. Finally, LUBAC-mediated activation of NF-κB seems also to be involved in the response to genotoxic agents, indicating that LUBAC may activate NF-κB to protect cells from DNA damage.
LUBAC and human disease

In human disease, aberrant NF-κB signaling has been implicated in a plethora of inflammatory, infectious and autoimmune disorders such as rheumatoid arthritis, atherosclerosis, inflammatory bowel disease, systemic inflammatory response syndrome, diabetes, septic shock and psoriasis. Given its central role in NF-κB regulation, future studies are warranted to determine the relevance of LUBAC in the development of these pathologies. Constitutive activation of NF-κB signaling is found in leukemia. Inherited mutations in either both HOIP and HOIL-1L have been identified in patients with a rare fatal disorder characterized by autoinflammation, infections and muscular amylopectinosis. Two germ line polymorphisms in HOIP are frequently observed in patients with B cell-like (ABC) subtype of diffuse large B-cell lymphoma (DLBCL). These findings have provided evidence that aberrations affecting components of LUBAC can contribute to the etiology of a broad variety of disorders. Because of the specificity to uniquely generate linear chains and consequentially activate NF-κB signaling, LUBAC represents a very attractive target for the development of new therapies to fight cancer, inflammatory and autoimmune diseases. In this respect, it will be important to fully understand how LUBAC is regulated and how it governs the formation of linear chains.

Figure 7. Involvement of ubiquitination in the canonical NF-κB pathway. In addition to the linear chains and K-48 linked polyubiquitin chains, ubiquitination is essential for early steps of NF-κB activation. The binding of TNF-α to its receptor induces the recruitment of the TNF receptor signaling complex (TNF-R1), which is composed of receptor interacting kinase 1 (RIP1), TNF-R1 associated death domain (TRADD), TNF receptor-associated factor 2 (TRAF2), and the cellular inhibitors of apoptosis protein 1 and 2 (c-IAP1/2). The latter two initiate the NF-κB pathway.

Legend
- K48-ubiquitin
- K11-ubiquitin
- K63-ubiquitin
- Linear-ubiquitin
signaling cascade by labeling RIP1 with K11- and K63-linked chains. These provide a platform for
the recruitment of the adaptor complexes, such as TAK1 (TGFβ-activated kinase 1)-TAB1-TAB2/3
complex, which are required for the phosphorylation of IKKβ. The activity of TRADD, c-IAPs and
TRAF2 is also important for the recruitment of LUBAC at the receptor, where it mediates linear
ubiquitination of NEMO, thus activating NF-kB transcription factors.

Reference List


CHAPTER 1


CHAPTER 2

A dominant-negative GFI1B mutation in the gray platelet syndrome

Davide Monteferrario1,9, Nikhita A. Bolar2,9, Anna E. Marneth1,9, Konnie M. Hebeda3, Saskia M. Bergevoet1, Hans Veenstra1, Britta A.P. Laros-van Gorkom4, Marius A. MacKenzie4, Cyrus Khandanpour5, Lacramioara Botezatu5, Erik Fransen2,6, Guy Van Camp2, Anthonie L. Duijnhouwer7, Simone Salemink8, Brigith Willemsen3, Gerwin Huls1,4, Frank Preijers1, Waander Van Heerde1, Joop H. Jansen1, Marlies J.E. Kempers8, Bart L. Loeys2, Lut Van Laer2,10 and Bert A. Van der Reijden1,10

1Department of Laboratory Medicine, Laboratory of Hematology, Radboud University Nijmegen Medical Centre, Nijmegen Centre for Molecular Life Sciences, Nijmegen, The Netherlands
2Department of Medical Genetics, Antwerp University Hospital and University of Antwerp, Antwerp, Belgium
3Department of Pathology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
4Department of Hematology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
5Department of Hematology, University Hospital, Essen, Germany
6StatUa Center for Statistics, University of Antwerp, Antwerp, Belgium
7Department of Cardiology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
8Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
9These authors contributed equally to this work
10Co-corresponding authors

Adapted from:
A dominant-negative GFI1B mutation in the gray platelet syndrome
NEJM 2014 Jan; 370(3):245-53
Abstract

The gray platelet syndrome is a hereditary, usually autosomal recessive bleeding disorder caused by a deficiency of alpha granules in platelets. We detected a nonsense mutation in the gene encoding the transcription factor GFI1B (growth factor independent 1B) that causes autosomal dominant gray platelet syndrome. Both gray platelets and megakaryocytes had abnormal marker expression. In addition, the megakaryocytes had dysplastic features, and they were abnormally distributed in the bone marrow. The GFI1B mutant protein inhibited nonmutant GFI1B transcriptional activity in a dominant-negative manner. Our studies show that GFI1B, in addition to being causally related to the gray platelet syndrome, is key to megakaryocyte and platelet development.

Introduction

Platelets are formed through fragmentation of megakaryocytes that reside in the bone marrow\(^1,2\). Platelet alpha granules, which are by far the most abundant platelet organelles, store proteins that stimulate platelet adhesiveness, hemostasis, and wound healing\(^3,4\). The gray platelet syndrome is an inherited bleeding disorder characterized by defective production of alpha granules\(^5,6\). Patients with this syndrome have reduced numbers of larger-than-normal platelets, and on light microscopy these platelets have a typical gray appearance caused by the lack of alpha granules. For a final diagnosis, the lack of alpha granules must be confirmed by means of electron microscopy\(^7\). Clinically, patients with the gray platelet syndrome have bleeding tendencies with variable severity\(^8\). Additional features associated with this syndrome are bone marrow fibrosis and the capture of neutrophils within megakaryocytes in a process known as emperipolesis (the presence of an intact cell within the cytoplasm of another intact cell)\(^6,8\). It has been proposed that the alpha-granule deficiency in the gray platelet syndrome can be attributed to the failure of megakaryocytes to efficiently route endogenous proteins into alpha granules, thereby hampering the maturation of these granules\(^9,10\). Although the gray platelet syndrome is most often an autosomal recessive disease, families with an autosomal dominant inheritance pattern have been described as well\(^6,11\). Recently, biallelic NBEAL2 missense mutations have been identified in an autosomal recessive form of the gray platelet syndrome\(^12–14\). NBEAL2 localizes to the dense tubular system in platelets\(^12\), but how NBEAL2 mutations contribute to defective alpha-granule formation remains unknown. No genes have been implicated in autosomal dominant gray platelet syndrome. Here we report on the identification of a nonsense mutation in the transcription-factor gene, growth factor independent 1B (GFI1B), in a family with autosomal dominant gray platelet syndrome. GFI1B and its paralog GFI1 function as transcriptional repressors and play fundamental roles during hematopoiesis\(^15–20\). Both proteins contain a highly conserved C-terminal region that includes three zinc fingers that are necessary for interaction with DNA and an identical N-terminal region through which epigenetic modifiers required to inhibit transcription are recruited\(^16,21–23\). In addition, GFI1 and GFI1B interact directly with other transcription factors to coregulate gene expression during blood-cell development\(^15\).
Materials and Methods

Study family

We studied a large family with an autosomal dominant form of thrombopathic thrombocytopenia characterized by mild-to-severe bleeding complications, a diagnosis originally described in 1968. The thrombopathic disorder in this family consisted of a low platelet count, an increased platelet size, and few alpha granules. An updated evaluation of affected family members according to the current disease classification allowed us to redefine the clinical diagnosis as the gray platelet syndrome (number 139090 in the Online Mendelian Inheritance in Man database). The study was approved by the institutional review board of Radboud University Medical Center. Written informed consent was obtained from all family members who underwent evaluation.

Laboratory analyses

The bleeding score was calculated, and morphologic and immunohistochemical studies, electron microscopy, flow cytometry, GFI1B sequencing, and platelet-related analyses were performed according to standard procedures (see Supplementary Material and Methods). Measurements of plasma glycocalcin (Cusabio) and megakaryocyte colony-forming assays (StemCell Technologies) were performed according to the manufacturers’ instructions. Linkage analysis based on genomewide single-nucleotide polymorphism array with the use of Linkage 24 DNA Analysis BeadChips (Illumina), GFI1 and GFI1B gene-reporter assays, the method of retroviral transduction, and primary cell cultures are described in the Supplementary Information.

Results

Clinical findings and laboratory analyses

Morphologic analysis of blood smears from the six affected family members who were evaluated showed a ghostlike, gray appearance of enlarged platelets, in contrast to the normal platelets observed in the two unaffected family members (Figure 1A and Figure 1B). Electron-microscopic analysis confirmed a marked reduction in the number of alpha granules (Figure 1C, Figure 1D and Figure 1E). Affected family members had moderate-to-severe bleeding tendencies (Table S1). In addition to reduced platelet counts, diminished platelet factor 4 and β-thromboglobulin levels were observed in platelets obtained from the affected family members (Table S1).

Histopathological findings

Histopathological examination of a bone marrow biopsy specimen obtained from family member III.2 (Figure 2A) showed stage I myelofibrosis (Figure 1F) and cellular marrow with an increased number of megakaryocytes that were pleomorphic in size and shape (Figure 1G and Figure S1). Emperipolesis was frequently observed.
A dominant-negative GFI1B mutation in the gray platelet syndrome

Figure 1. Characteristics of autosomal dominant gray platelet syndrome. (A) May–Grünwald-Giemsa staining showed normal platelets (arrows) obtained from unaffected family member II.4. (B) May–Grünwald-Giemsa staining showed large gray platelets (arrows) obtained from affected family member II.2. (C) Electron-microscopic analysis showed a normal platelet with numerous alpha granules (arrows) in a specimen obtained from unaffected family member II.4. (D-E) Electron-microscopic analysis shows gray platelets with few alpha granules, which are poorly developed (arrows), in a specimen obtained from family member III.2 (D), and no alpha granules in a specimen obtained from family member II.2 (E). (F-I) Abnormalities in a bone marrow–biopsy specimen obtained from affected family member III.2. Laguesse staining (F) revealed stage I bone marrow fibrosis, consisting of a focal network of fine, dark-staining reticulin fibers that extend from the sinuses. Hematoxylin and eosin staining (G) showed megakaryocytes that are both larger and smaller than normal, with shapes that deviate from the normal round form (arrows). Some nuclei are irregular and hyperchromatic. Periodic acid–Schiff staining (H) showed the capture of neutrophils (arrows) by megakaryocytes (emperipolysis). β3 integrin–CD61 staining (I) showed stretched megakaryocytes abnormally positioned in tandem along the sinuses (arrows).
(Figure 1H and Figure S1). Megakaryocytes were clustered along bone marrow sinuses and had stretched features (Figure 1I and Figure S1).

Linkage and sequence analyses
To determine the disease-causing mutation, we first performed linkage analyses in 14 family members and identified a candidate locus on chromosome 9q34 with a maximum LOD score of 3.9 (Figure S2 and Figure S3). Within the linked region, we considered GFI1B to be an excellent candidate gene because it encodes a transcriptional repressor that has been implicated in megakaryopoiesis. Sequence analysis detected a nonsense mutation in exon 6 (c.859C→T, p.Gln287*) of GFI1B that completely cosegregated with the gray platelet syndrome (Figure 2A, Figure 2B, Figure S4 and Table S2). The mutation introduces a premature stop codon that is predicted to lead to the formation of a truncated protein (GFI1B<TR>) that lacks 44 carboxy-terminal amino acids (Figure 2B). Since premature stop codons can induce nonsense-mediated messenger RNA (mRNA) decay, we determined whether mutated GFI1B was expressed at the mRNA level. Sequencing of complementary DNA generated from CD34+ progenitor cells obtained from two affected family members showed that mutated and nonmutated GFI1B transcripts were expressed, indicating that the mutated transcript is not targeted for decay (Figure S4).

Molecular and functional characterization of mutant GFI1B
GFI1B functions as a transcriptional repressor. The truncating mutation is located within zinc finger 5, which is required for DNA binding. This zinc finger is 100% identical between humans and rodents (rats and mice), and it is identical to zinc finger 5 of the paralog GFI1 (Figure 2B). Murine Gfi1 and Gfi1b bind the same DNA consensus sequence, and Gfi1 zinc finger 5 directly interacts with the major groove of the DNA core sequence AATC through four amino acids. Because the truncated protein GFI1B<TR> lacks all these DNA-interacting amino acids (Figure 2B), we hypothesized that it would be unable to repress gene expression. To test this hypothesis, we introduced the truncating mutation in a GFI1B expression vector and performed transcriptional repression assays using the Gfi1 promoter as a validated Gfi1B target. In line with the lack of an intact DNA-binding zinc finger, we observed that GFI1B<TR> did not repress the Gfi1 promoter, whereas the nonmutant GFI1 and GFI1B did (Figure 2C and Figure S5). When coexpressed, GFI1B<TR> inhibited repression mediated by nonmutant GFI1B, indicating that the mutant interferes with nonmutant GFI1B in a dominant-negative fashion (Figure 2C).

To validate that GFI1B<TR> adversely affects normal GFI1B, we expressed GFI1B<TR> in mouse bone marrow cells and then induced megakaryocytic differentiation. GFI1B<TR>-positive megakaryocytes had several dysplastic features, including hypolobulation of the nuclei, irregular contours, and multiple separated nuclei; these features were not observed in control cells (Figure 2D and Figure S6). These abnormalities are very similar to those observed in cells obtained from affected family members with the gray platelet syndrome (Figure 3E), indicating that expression of GFI1B<TR> is sufficient to cause megakaryocytic abnormalities and that it functions in a dominant-negative manner.
Figure 2. *GFI1B* nonsense mutation in the gray platelet syndrome, inhibiting the function of nonmutant *GFI1B*. (A) The pedigree of the studied family with the gray platelet syndrome is shown. Squares denote male family members, circles female family members, solid symbols affected family members, open symbols unaffected family members, and slashes deceased family members. Mutation status is indicated below the number of each family member, with −/− indicating the absence and +/− the presence of the *GFI1B* c.859C→T, p.Gln287* mutation. (B) On top, human *GFI1B* (h*GFI1B*) contains an N-terminal Snail/Gfi-1 (SNAG) domain through which epigenetic modifiers are recruited. At the C-terminal, *GFI1B* contains six zinc fingers, of which three through five are essential for DNA binding. The *GFI1B* nonsense mutation (c.859C→T, p.Gln287*) is located in zinc finger 5 (arrow).
Cellular analyses

Studies involving Gfi1b knockout mice have shown that Gfi1b plays an important role in the development of several hematopoietic lineages\textsuperscript{15,17,19}. In humans, GFI1B is required for in vitro differentiation of erythrocytic and megakaryocytic lineages\textsuperscript{18}. We observed that megakaryocyte colony-forming cells were significantly more frequent in a bone marrow specimen from affected family members than in a specimen from a healthy person. In addition, the colonies from the affected family member were larger than those from a healthy control (Figure 3A and Figure S7); this is in line with the increased number of megakaryocytes in the biopsy specimen from affected family member III.2 (Figure 1I).

To determine whether GFI1B\textsuperscript{T} was associated with other hematopoietic abnormalities, we performed immunophenotypic analyses of peripheral-blood specimens obtained from eight family members and five healthy controls. Myeloid and erythroid lineages were unaffected (data not shown). The surface expression of characteristic platelet markers, such as $\beta_3$ glycoprotein–CD36, $\alpha_{II} \beta_3$ integrin–CD41, and $\beta_3$ integrin–CD61, was also unperturbed in all tested persons. However, as compared with unaffected family members and healthy controls, five of six affected members had a marked decrease in the level of platelet surface-membrane glycoprotein 1bα–CD42B within the $\alpha_{II} \beta_3$ integrin–CD41 platelet compartment (Figure S8). In addition, strong expression of CD34, which is usually confined to immature hematopoietic progenitors, was detected on platelets from all affected family members in the study (Figure 3B).

Because platelets are derived from megakaryocytes\textsuperscript{1,2}, we evaluated whether aberrant expression of surface molecules was also present on megakaryocytes. Immunostaining (with $\beta_3$ integrin–CD61) of a bone marrow–biopsy specimen obtained from affected family member III.2 confirmed the presence of megakaryocytes that had high levels of CD34 expression (Figure 3C). Electron-microscopic analysis of the same biopsy specimens showed megakaryocytes with few alpha granules, which were small, irregularly shaped, and centrally located; the megakaryocytes were characterized by an extensive peripheral cytoplasm with irregular proplatelets, largely devoid of cell organelles (Figure 3D and Figure S9). Consistently, bone marrow smears obtained from the same patient (family member III.2) and her mother (family member II.2) showed dysplastic megakaryocytes that were pale, as a result of diminished alpha granules (Figure 3E). To determine whether these
A dominant-negative GFI1B mutation in the gray platelet syndrome

Abnormalities were intrinsic to the cell, we stimulated CD34+ cells obtained from affected family members II.2 and III.2 to differentiate along the megakaryocytic lineage in culture. Megakaryocytic cells had dysplastic features such as those observed in the bone marrow aspirates (Figure 3E). In addition, altered expression of CD34 and glycoprotein 1bα–CD42B was observed in megakaryocytic cells that were positive for αIIβ integrin–CD41 (Figure S10).

Figure 3. GFI1BTr-associated platelet and megakaryocyte abnormalities. (A) αIIβ integrin–CD41 staining showed the results of a megakaryocyte colony-forming assay of bone marrow cells from family member III.2, as compared with normal bone marrow cells from a healthy control. The cells from affected family member III.2 developed into more and bigger colonies. CFU denotes colony-forming units, and GPS the gray platelet syndrome. (B) Platelets obtained from affected family members II.2, II.3, II.6, II.8, II.10, and III.2, as compared with platelets obtained from unaffected family members II.4 and II.7 and healthy controls, had high levels of CD34 expression. MFI denotes mean fluorescence intensity. (C) CD34 staining of a bone marrow–biopsy specimen obtained from affected family member III.2 (left) showed a high level of CD34 expression on
Discussion

The gray platelet syndrome is a hereditary bleeding disorder characterized by platelets that lack alpha granules. Recently, mutations in NBEAL2 have been shown to cause an autosomal recessive form of the gray platelet syndrome. We found that a mutation in GFI1B was causally related to autosomal dominant gray platelet syndrome. In addition to large gray platelets that are almost devoid of alpha granules, other phenomena associated with the gray platelet syndrome, such as thrombocytopenia, emperipolesis, mild myelofibrosis, and low expression of platelet factor 4, were observed in affected persons. However, we also found new features of the gray platelet syndrome such as megakaryocytes and platelets that were strongly positive for the stem-cell and progenitor-cell marker CD34, and had a diminished expression of glycoprotein 1α–CD42B. Diminished glycoprotein 1α–CD42B platelet expression may contribute to the pathogenesis of the gray platelet syndrome, since glycoprotein 1α–CD42B plays an important role in the initial adhesion of platelets to vascular subendothelium after injury. Low platelet expression of glycoprotein 1α–CD42B may be related to the large platelets observed in affected persons, since large platelets are present in the Bernard–Soulier syndrome, a bleeding disorder caused by GP1BA/CD42B mutations. In addition, and possibly consequential to aberrant marker expression, we observed that megakaryocytes were abnormally distributed within the bone marrow, and they were frequently aligned in clusters along the sinuses in a stretched fashion. These features, in combination with the presence of morphologic and nuclear abnormalities in megakaryocytes, indicate that GFI1B<sup>Tr</sup> affects the terminal maturation of megakaryocytes.

GFI1B silencing and Gfi1b ablation studies have shown that the protein is fundamental to the development of several blood-cell lineages. In the affected members of the family we studied, the presence of GFI1B<sup>Tr</sup> was associated with major abnormalities only in the megakaryocytic lineage. The mutation in the gray platelet syndrome introduces a stop codon that results in a truncated GFI1B protein. The deleted region includes four amino acids that directly interact with DNA. Indeed, GFI1B<sup>Tr</sup> expression did not inhibit gene expression, but it inhibited the function of nonmutant GFI1B. This indicates that the gray platelet syndrome may be caused not by GFI1B haploinsufficiency but rather by dominant-negative inhibition of the nonmutant protein. Indeed, forced GFI1B<sup>Tr</sup> expression in normal bone marrow cells resulted in severely dysplastic megakaryocytes that were similar to those observed in affected
A dominant-negative GFI1B mutation in the gray platelet syndrome

persons. These findings are consistent with mouse models showing that only complete Gfi1b ablation disturbs in vitro development of megakaryocytes, whereas the loss of a single Gfi1b allele is not sufficient to yield a clear phenotype17. GFI1BTr retains several corepressor and transcription factor–interacting domains and may act in a dominant-negative manner by quenching proteins that would normally cooperate with nonmutant GFI1B. The identification of GFI1B target genes that may be disturbed by GFI1BTr would provide relevant insight into the molecular pathways that are key for megakaryopoiesis and platelet production. In summary, our data indicate that GFI1B is a causative gene in autosomal dominant gray platelet syndrome. The truncated GFI1B mutant acts in a dominant-negative manner over the nonmutant protein and affects the normal development of megakaryocytes and platelets, reflecting the pivotal role of GFI1B in governing normal megakaryopoiesis and platelet production.

References


Supplementary Information
Supplementary Materials and Methods

Genome wide linkage analysis and GFI1B sequencing
DNA was extracted using standard methods. Fourteen family members (Figure 2A, Table S1) were genotyped using the Linkage-24 DNA Analysis BeadChip, containing 6000 Single Nucleotide Polymorphisms (SNPs), according to the manufacturers’ guidelines (Illumina). The initial (fast) linkage analysis was carried out using Merlin version 1.1.2 (1) with SNP data from twelve family members (omitting cases II:11 and IV:3 because of the limitations of the Merlin program) (Figure S2). To use the full potential of the family and to reach the maximum lod score, linkage analysis was subsequently carried out using SIMWALK2 on a limited region from chromosome 9, from 103 to 140 kb, with the data from all genotyped family members (Figure S3). Since the SIMWALK2 algorithm is non-deterministic, the analysis was repeated six times. The mean and standard error of the maximal LOD scores from all six runs were calculated.

The coding exons and intron-exon borders of GFI1B were amplified using the primers listed in Table S2 and standard PCR conditions. PCR products were bidirectionally sequenced using the BigDye Terminator Cycle Sequencing kit (Applied Biosystems) and separated on an ABI 3130XL Genetic Analyzer (Applied Biosystems). Sequence numbering was based on Ensembl transcript ENST00000450530, with the A nucleotide of the start-codon ATG indicated as position +1.

Total RNA was isolated from bone marrow FACS-sorted CD34+ cells using RNA-Easy Mini-Kit according to the manufacturer’s protocol (Qiagen). RNA was transcribed into cDNA using M-MLV reverse transcriptase according to manufacturer’s instructions (Invitrogen). cDNA was amplified using primers listed in Table S3. The same primers were used to bidirectionally sequence purified PCR products.

GFI1B gene reporter assays
As readout for GFI1BTr transcriptional activity we performed a Dual Luciferase Reporter Assay (Promega). HEK293T cells were cultured in Dulbecco modified Eagle medium (DMEM; Gibco) supplemented with 10% non-heat-inactivated fetal calf serum ( Gibco), 1% penicillin/streptomycin (MP Biomedical), 1% non-essential amino acids (GIBCO) and 1% L-glutamine (MP Biomedical). Cells were cultured in 24-well plates at 37°C supplied with 5% CO₂. For transient expression, in total 2 µg of DNA were used. Amounts of used pcDNA3.1-GFI1-FLAG, pcDNA3.1-GFI1B-FLAG, pcDNA3.1-GFI1BTr-FLAG are indicated in Figure 2C. Empty vector pcDNA3.1 was used to compensate for differences in DNA amounts. In all experiments, 800 ng of luciferase reporter construct (pGfi1-pGL3) and 200 ng of Renilla luciferase vector (pRL-TK) were used. Transfections were performed at 60% cell confluency with Lipofectamine 2000 (Invitrogen). Forty-eight hours after transfection, cells were washed with phosphate buffered saline (PBS) and lysed in 100 µl of passive lysis buffer (Promega) for 1 h. Luciferase assays were performed according to the protocol provided by the manufacturer (Promega). Luciferase signals were measured on a Lumat LB 9507 (EG&G Berthold). Luciferase activity
was normalized to renilla activity. Each condition was experimentally tested in triplicate. Western blot analysis was performed to confirm protein expression. Proteins of total lysates generated as described above were separated on 10% polyacrylamide gels and transferred onto polyvinylidenefluoride (PVDF) membranes (Bio-Rad). PVDF membranes were probed with anti-FLAG (Abcam) primary antibodies, followed by probing with HRP-conjugated secondary antibodies. Antibody signal was visualized by chemiluminescence using a Bio-Rad ChemiDox XRS⁺.

**Electron microscopy**

Blood was collected in a tube with 3.8% sodium citrate and centrifugated at room temperature for 10 minutes at 1800 rpm. The plasma was removed and the buffy coat was fixed with 2% glutaraldehyde in phosphate buffer. After 4 hours fixation at 4°C, the buffy coat was isolated and cut in small blocks and further fixed in the same fixative for at least 24 hours. The blocks were washed with phosphate buffer and fixed in 1% osmic acid in Pallade buffer containing 1% potassium ferrocyanide for 1 hour. A fresh bone marrow biopsy fragment was fixed in 2% glutaraldehyde in phosphate buffer for 24 hours at 4°C. After fixation, the tissue was decalcified in 10% EDTA and 5% glutaraldehyde in phosphate buffer (pH 7.4). After six days, a X-Ray photograph was taken to check if the bone marrow decalcification was complete. The tissue was rinsed in phosphate buffer during 24 hours. Post fixation treatment was performed in 1% osmic acid in pallade buffer containing 1% potassium ferrocyanide for 1 hour. The samples were dehydrated in a graded series of ethanol and propylene oxide, followed by propylene with a decreasing concentration epon until pure epon resin embedding. Ultrathin sections were cut on a Leica Ultracut and stained with uranyl acetate and lead citrate. Examinations were made on a Jeol 1200 EX/II electron microscope at 60 kV.

**Flow cytometry**

For flow cytometry, expression of surface molecules on peripheral blood and ex vivo cultured cells was analyzed by staining with the following fluorochrome-conjugated antibodies: CD11b, CD13, CD34, CD41, CD42b, CD45, CD61, CD105, CD117, CD235a, HLA-DR (Beckman Coulter), CD15, CD16, CD36 (BD Biosciences) and CD71 (Dako). After staining, cells were analyzed using a Coulter FC500 flow cytometer (Beckman Coulter). Peripheral blood from healthy donors was used as control.

**Microscopy**

Images were obtained using a Leitz microscope DMRB (Leitz) with Leica DFC320 camera (Leica) and a Zeiss Axiolmager M1 with a Zeiss Plan Apochromat 63 / 1,4 Oil objective (Zeiss, MGG staining only).

**Ex vivo generation of megakaryocytes**

CD34+ cells were cultured for 1 day in Iscove's Modified Dulbecco's Medium (IMDM;GIBCO) supplemented with 10% human serum, recombinant human stem cell factor (100 ng/mL) and
recombinant human Flt3-ligand (100 ng/mL). Subsequently, megakaryocyte differentiation was induced by culturing cells in IMDM supplemented with 10% human serum, recombinant human stem cell factor (20 ng/mL, Immunotools) and recombinant human Thrombopoietin (20 ng/mL, Immunotools). The medium was refreshed every 4 days. At day 12, cells were harvested and analyzed.

**Platelet factor 4 and β-thromboglobulin in platelets**
Platelet factor 4 (PF4) and β-thromboglobulin (β-TG) levels were measured in theophylline, adenosine and dipyridamole (TAD) containing platelet rich citrated plasma. Theophylline, adenosine and dipyridamole were added to avoid any platelet activation. Additionally TAD containing citrated platelet poor plasma was obtained to correct for potential circulating PF4 and β-TG levels, either due to *in vivo* platelet activation or artificial phlebotomy. Platelet rich plasma was subjected to three cycles of freeze (-80°C) and thawing (37°C). PF4 levels were measured with the Zymotest PF4-ELISA kit (Hyphen Biomed) and β-TG levels were analyzed using the Asserachrom β-TG ELISA kit (Diagnostica Stago). Both PF4 and β-TG levels were normalized for platelet count measured in platelet rich plasma. Plasma of a healthy volunteer, treated similarly as described above, was used as control for both determinations.
Supplementary Figures

**Figure S1.** Characteristics of autosomal dominant GPS. (A) HE staining of a bone marrow biopsy from patient III:2 showing pleomorphic megakaryocytes in size and shape (dysmorphic, a-c). Emperipolesis is also visible (c). (B) PAS staining of a bone marrow biopsy from patient III:2 showing dysmorphic and stretched megakaryocytes abnormally positioned in tandem along sinus (a,b). Emperipolesis is also visible (b). (C) PAS staining of a bone marrow biopsy from patient III:2 showing stretched megakaryocytes abnormally positioned in tandem along sinus (a,b). Emperipolesis (c) and hyperchromatic nuclei are also visible (b). (D-F) CD61 staining (brown) showing dysmorphic megakaryocytes abnormally positioned in tandem along sinuses.
Figure S2. Linkage analysis on family with autosomal dominant GPS. LOD scores on chromosomes 1 to 22 obtained using Merlin version 1.1.2 with SNP-array data from twelve family members.
Figure S2. (continued)
A dominant-negative $GFI1B$ mutation in the gray platelet syndrome

Figure S3. Linkage analysis on chromosome 9. LOD scores using SIMWALK2 on a limited region from chromosome 9 with the data from all genotyped family members. The analysis was repeated six times since the SIMWALK2 algorithm is non-deterministic.

Figure S4. $GFI1B$ mutation in autosomal dominant GPS. Representative $GFI1B$ c.859C>T mutation in DNA (pink shading) and RNA (cDNA, arrow) obtained from affected family member II:2.
**Figure S5. Protein expression validation for GFI1B gene reporter assay.** Western blot analysis confirming proper protein expression of GFI1 (upper intense band), GFI1B and GFI1BTr in 293T cells used for luciferase reporter assay. A GAPDH staining was performed to show equal loading.

**Figure S6. Exogenous GFI1BTr expression results in abnormal megakaryocytes.** Retroviral expression of GFI1BTr in murine Linkit progenitor cells followed by in vitro megakaryocytic differentiation resulted in abnormal megakaryocyte development characterized by a pale cytoplasmatic appearance and dysplastic features such as multiple separated nuclei. Magnification used is indicated below the pictures. When using Lin- cells similar results were observed (data not shown).
A dominant-negative GFI1B mutation in the gray platelet syndrome

Figure S7. GPS derived megakaryocyte colonies are larger compared to control colonies. Representative examples of megakaryocyte colonies showing that bone marrow cells from patient III:2 (GPS) develop into bigger colonies than bone marrow cells from a healthy individual. Magnification used is indicated below the pictures.

Figure S8. Reduction of GP1BA/CD42B expression on GPS platelets. (A) Platelets from affected family members (II:2, II:3, II:6, II:8, II:10, III:2) have low GP1BA/CD42B expression. (B) GP1BA/CD42B may be cleaved resulting in the formation of glycocalcin in plasma. Extensive cleavage may result in low GP1BA/CD42B levels observed on affected platelets. Glycocalcin levels were not increased in plasma from affected individuals indicating that the low expression on platelets is not caused by increased cleavage of GP1BA/CD42B.
Figure S9. GF11B<sup>Tc</sup>-associated megakaryocyte abnormalities (family member III:2). (A-B) GPS megakaryocytes with a cytoplasm largely devoid of α-granules, that showed a very irregular disintegration of the cytoplasm into large irregular proplatelets (cytoplasmic extensions that will detach to form platelets). Also the proplatelets lack alpha granules. (C-D) GPS megakaryocytes with a central perinuclear cytoplasmic area with few alpha granules, surrounded by an extensive system of membranes (demarcation membrane), believed to mark the future platelets. The surrounding cytoplasm is abnormally wide and lacking cell organelles as α-granules and mitochondria. (E) Additional example of megakaryocyte with a diminished number of α-granules.
A dominant-negative GFI1B mutation in the gray platelet syndrome

**Figure S10. Reduction of GP1BA/CD42B expression and CD34 expression in GPS cultured cells.** FACS analysis showing high CD34 and low GP1BA/CD42B expression in cultured megakaryocytes from patients II:2 and III:2 (green and red, respectively) compared to a healthy control (blue).
## Supplementary Tables

### Table S1: Hematological and genetic data of analyzed family members

<table>
<thead>
<tr>
<th>Family member</th>
<th>GFI1B Mutational status</th>
<th>Sex</th>
<th>Platelet count (10^9/L) (150-450)</th>
<th>PF4 in 10^6 platelets (µg) (0.7-1.5)</th>
<th>BTG in 10^6 platelets (IU) (18.6-47.1)</th>
<th>Bleeding score (F&lt;5) (M&lt;3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>II:4</td>
<td>wt</td>
<td>F</td>
<td>275</td>
<td>1.3</td>
<td>42.4</td>
<td>2</td>
</tr>
<tr>
<td>II:7</td>
<td>wt</td>
<td>M</td>
<td>254</td>
<td>1.3</td>
<td>41.5</td>
<td>0</td>
</tr>
<tr>
<td>II:2</td>
<td>mut</td>
<td>F</td>
<td>85</td>
<td>0.2</td>
<td>10.6</td>
<td>19</td>
</tr>
<tr>
<td>II:3</td>
<td>mut</td>
<td>M</td>
<td>49</td>
<td>0.3</td>
<td>12.9</td>
<td>10</td>
</tr>
<tr>
<td>II:6</td>
<td>mut</td>
<td>M</td>
<td>42</td>
<td>0.4</td>
<td>19.3</td>
<td>25</td>
</tr>
<tr>
<td>II:8</td>
<td>mut</td>
<td>M</td>
<td>56</td>
<td>0.4</td>
<td>17.2</td>
<td>7</td>
</tr>
<tr>
<td>II:10</td>
<td>mut</td>
<td>F</td>
<td>116</td>
<td>0.2</td>
<td>9.8</td>
<td>12</td>
</tr>
<tr>
<td>III:2</td>
<td>mut</td>
<td>F</td>
<td>80</td>
<td>0.3</td>
<td>10.2</td>
<td>11</td>
</tr>
</tbody>
</table>

wt = wild type, mut = mutated, PF4 = Platelet factor 4, BTG = β-Thromboglobulin. Reference values are indicated between brackets in italics. Bleeding scores were determined as described elsewhere.

### Table S2: GFI1B PCR and sequencing primers (DNA)

<table>
<thead>
<tr>
<th>Exon</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>ATTCACAGGTTCTG</td>
<td>CTTCATCTTGCAACC</td>
</tr>
<tr>
<td>3</td>
<td>CAGTCTAAGGGCT</td>
<td>CAAGCAGAAGCCACAG</td>
</tr>
<tr>
<td>4</td>
<td>GTTGGTGCGCTG</td>
<td>CCCACATTCACAGACAGA</td>
</tr>
<tr>
<td>5</td>
<td>TCCCCCTCTCTACA</td>
<td>ACCCCTCCAAATCTG</td>
</tr>
<tr>
<td>6</td>
<td>AGACTCCAAGCC</td>
<td>GCAATTCCTTGGACT</td>
</tr>
<tr>
<td>7</td>
<td>ATCCAGTAGGG</td>
<td>CTCCCTGCGGGGT</td>
</tr>
</tbody>
</table>

Exons were determined from transcript ENST00000450530.

### Table S3: GFI1B sequencing primers (cDNA)

<table>
<thead>
<tr>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
</tr>
<tr>
<td>Reverse</td>
</tr>
</tbody>
</table>
Supplementary References


High DNA-methyltransferase 3B expression predicts poor outcome in acute myeloid leukemia, especially among patients with co-occurring NPM1 and FLT3 mutations

Davide Monteferrario\textsuperscript{1,4}, Sylvie M. Noordermeer\textsuperscript{1,2,4}, Saskia Bergevoet\textsuperscript{1}, Gerwin Huls\textsuperscript{1,3}, Joop H. Jansen\textsuperscript{1} and Bert A. van der Reijden\textsuperscript{1,5}.

\textsuperscript{1}Department of Laboratory Medicine, Laboratory of Hematology, Radboud UMC, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands.
\textsuperscript{2}Current address: The Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, M5G1X5, Canada
\textsuperscript{3}Department of Hematology, Radboud UMC, Nijmegen, The Netherlands.
\textsuperscript{4}These authors contributed equally to this work.
\textsuperscript{5}Corresponding author.

Adapted from:

High DNA-methyltransferase 3B expression predicts poor outcome in acute myeloid leukemia, especially among patients with co-occurring NPM1 and FLT3 mutations.

DNA methyltransferases (DNMTs) are epigenetic regulators targeted to the treatment of hematological malignancies. Mutations in the DNA methyltransferase DNMT3A and high expression of its parologue DNMT3B have been associated with inferior outcome in acute myeloid leukemia (AML) and other hematological malignancies. Using a publicly available gene expression data set, we studied whether DNMT3B expression correlates with outcome in genetically well-defined AML subgroups. We first validated the expression data from the microarray by quantitative PCR, using 39 patient samples (Figure S1A). DNMT3B micro-array analyses showed that the expression was not normally distributed among AML patients; in the quartile with highest expression, a larger variation in expression was observed compared to the three quartiles with relatively lower expression (Figure S1B). The median DNMT3B expression in AML samples was significantly lower compared with that observed in normal bone marrow (NBM)-derived CD34+ cells (P<0.0001), while it was higher compared to NBM cells, but the latter difference was not statistically significant (Figure S1B). Subsequently, we investigated the correlation between DNMT3B expression and overall survival (OS) and event-free survival (EFS). In univariate Cox regression analyses, continuous DNMT3B expression was significantly associated with poor survival (P<0.001 for both OS and EFS, data not shown). To visualize the prognostic impact, we performed Kaplan–Meier analyses on the four quartiles based on expression levels. The quartile including the patients with the highest DNMT3B expression, exhibiting the largest variation in expression, showed a significantly reduced OS and EFS compared to the other quartiles (Figure S2). As the survival between the lower three quartiles did not differ significantly, we grouped these patients together as having lower DNMT3B expression, whereas the remaining patients were ranked as having higher DNMT3B expression. Using these criteria, the 5-year OS and EFS were 17.2%±3.3% and 13.6%±3.0% for patients with higher DNMT3B expression compared to 43.8%±2.5% and 34.4%±2.4% for patients with lower DNMT3B levels (P<0.001, Figure 1A). We next performed a multivariate Cox regression analysis including known prognostic factors (including age >60 years; white blood cell counts >100 × 10⁹/l; transplantation status; karyotypes t(8;21), t(15;17) and inv(16); nucleophosmin 1 (NPM1), FLT3-ITD, DNMT3A and double CEBPA mutations, and ecotropic viral integration site 1 (EVI1) overexpression), which revealed that higher DNMT3B expression carried an independent prognostic risk for both OS and EFS (hazard ratio (HR): 1.768, 95% confidence interval (CI): 1.384–2.260; P<0.001 and HR: 1.706, 95% CI: 1.342–2.168; P<0.001, respectively, Table 1), in line with a recently published study. In fact, higher DNMT3B expression showed a higher hazard ratio for OS than that of well-known adverse prognostic factors such as internal tandem duplications of the fms-related tyrosine kinase 3 (FLT3-ITD, HR: 1.675, 95% CI: 1.287–2.179; P<0.001) and overexpression of the EVI1 gene (HR: 1.430, 95% CI: 0.999–2.047; P=0.051).

We then studied the association of higher DNMT3B expression with specific AML subcategories. Higher DNMT3B expression was under-represented in FAB-M4 and mutually exclusive with the favorable karyotypes t(8;21) and inv(16), as shown in Table S1. In contrast, higher DNMT3B expression was over-represented in the FAB-M1 subcategory (Table S1), and predicted poor OS and a trend toward poor EFS in this group (data not shown).
Figure 1. Higher DNMT3B expression correlates with inferior OS and EFS in AML. (A) Kaplan-Meier plots for OS and EFS showing that higher DNMT3B expression correlated significantly with a poor OS and EFS among AML patients (5-year OS of 17.2% ± 3.3% vs 43.8% ± 2.5% and 5-year EFS of 13.6% ± 3.0% vs 34.4% ± 2.4% for patients with higher and lower DNMT3B expression, respectively). (B) Higher DNMT3B expression predicted a very poor OS and EFS among patients with normal karyotype carrying NPM1 and FLT3-ITD mutations compared to patients with lower DNMT3B expression (5-year OS: 16.7 ± 6.2% vs 47.6 ± 8.8%, P = 0.001 and 5-year EFS: 16.7 ± 6.2% vs 39.0 ± 8.6%, P = 0.005 for patients with higher and lower DNMT3B expression, respectively). (C) Within the group of patients with normal karyotype that carry NPM1 mutations with high FLT3-ITD allelic burden, higher DNMT3B expression correlated with an extremely poor OS and EFS compared to patients with lower DNMT3B expression (5-year OS: 0% ± 0.0% vs 38.9% ± 12.9%, P < 0.001 and 5-year EFS: 0% ± 0.0% vs 32.0% ± 12.4%, P < 0.001 for patients with higher and lower DNMT3B expression, respectively). P values were determined with the log-rank test. In agreement with Jonge et al., high FLT3-ITD allelic burden was defined as an allelic FLT3-ITD/FLT3 ratio higher than 1.
association between higher DNMT3B expression and EVI1 overexpression and IDH2 mutations was also observed (Table S1), but not with IDH1 mutations. DNMT3B expression did not predict clinical outcome in these subgroups (data not shown). We also observed that higher DNMT3B expression was associated with a normal karyotype (NK, \( P=0.017 \)), and strongly associated with mutations in NPM1 (NPM1\(^+\), \( P<0.001 \)) and FLT3-ITD (FLT3-ITD\(^+\), \( P<0.001 \), Table S1). Patients with NK can be classified based on NPM1 and FLT3-ITD mutational status. NPM1 mutations, particularly in the absence of FLT3-ITD, display a favorable disease outcome, whereas the presence of a FLT3-ITD mutation is generally considered as an adverse prognostic factor\(^{10} \). Remarkably, among patients with NK, higher DNMT3B expression did not significantly associate with NPM1\(^−\)/FLT3-ITD\(^−\) status and was under-represented in the NPM1\(^+\)/FLT3-ITD\(^+\) and NPM1\(^−\)/FLT3-ITD\(^+\) groups, but was significantly over-represented in the NPM1\(^+\)/FLT3-ITD\(^+\) group (Table S2). Within the latter group, patients with higher DNMT3B expression showed a significantly worse outcome than patients with lower DNMT3B expression (Figure 1B). The 5-year OS and EFS of patients with higher DNMT3B expression were 16.7\%±6.2\% (\( P=0.001 \)) and 16.7\%±6.2\% (\( P=0.005 \)) compared to 47.6\%±8.8\% and 39.0\%±8.6\% in patients with lower DNMT3B expression. The survival of patients with co-occurring NPM1 and FLT3-ITD mutations is negatively influenced by high FLT3-ITD allelic burden\(^{11–13} \). Higher DNMT3B expression was observed in both patients with low and those with high FLT3-ITD allelic burden (Table S2). We next analyzed whether DNMT3B expression had an effect on survival among these subgroups. Higher DNMT3B expression did not exhibit a significant effect on OS and EFS among patients with low FLT3-ITD allelic burden (data not shown). However, patients with higher DNMT3B expression showed an extremely poor OS and EFS (5-year OS: 0.0\%±0.0\%, 5-year EFS: 0.0\%±0.0\%, \( P<0.001 \)), compared to patients with lower DNMT3B expression (5-year OS: 38.9\%±12.9\%, 5-year EFS: 32.0\%±12.4\%) within the subgroup with high FLT3-ITD allelic burden (Figure 1C). In multivariate analysis, higher DNMT3B expression showed an independent prognostic value for OS and EFS, with a high hazard ratio both in the NPM1\(^+\)/FLT3-ITD\(^+\) subgroup (HR: 4.850, 95% CI: 1.980–11.880) and among those patients with high FLT3-ITD allelic burden, indicating that the latter subgroup can be separated into two groups, one with an intermediate and the other with an extremely poor survival, based on DNMT3B expression (Table 1).

In conclusion, these data show that higher DNMT3B expression is a strong independent predictive factor for poor disease outcome in AML in general, and especially among NPM1\(^+\)/FLT3-ITD\(^+\) patients with normal karyotype. Drugs that inhibit DNA methylation are tested for clinical efficacy in AML\(^{1–4} \). Because DNMT3B catalyzes DNA methylation, its expression level may affect the therapeutic sensitivity to these drugs. Thus, it will be interesting to investigate whether DNMT3B expression predicts therapy responses to treatments affecting DNA methylation.
**Table 1.** DNMT3B expression is a prognostic factor in AML, particularly among NPM1⁺/FLT3-ITD⁺ with normal karyotype

<table>
<thead>
<tr>
<th></th>
<th>OS</th>
<th></th>
<th>EFS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95%CI)</td>
<td>P</td>
<td>HR (95%CI)</td>
<td>P</td>
</tr>
<tr>
<td><strong>AML</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher DNMT3B exp.</td>
<td>1.768 (1.384-2.260)</td>
<td>&lt;0.001</td>
<td>1.706 (1.342-2.168)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age &gt; 60 y</td>
<td>1.523 (1.111-2.087)</td>
<td>0.009</td>
<td>1.339 (0.985-1.821)</td>
<td>0.063</td>
</tr>
<tr>
<td>WBC count &gt; 100*10⁹/L</td>
<td>1.448 (1.098-1.909)</td>
<td>0.009</td>
<td>1.531 (1.173-1.997)</td>
<td>0.002</td>
</tr>
<tr>
<td>FLT3-ITD mutations</td>
<td>1.675 (1.287-2.179)</td>
<td>&lt;0.001</td>
<td>1.649 (1.274-2.136)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>NPM1 mutations</td>
<td>0.397 (0.291-0.541)</td>
<td>&lt;0.001</td>
<td>0.381 (0.280-0.518)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Favorable karyotype</td>
<td>0.388 (0.266-0.565)</td>
<td>&lt;0.001</td>
<td>0.457 (0.323-0.646)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CEBPA double mutation</td>
<td>0.338 (0.177-0.645)</td>
<td>0.001</td>
<td>0.361 (0.199-0.653)</td>
<td>0.001</td>
</tr>
<tr>
<td>DNMT3A mutations</td>
<td>1.631 (1.204-2.210)</td>
<td>0.002</td>
<td>1.531 (1.136-2.063)</td>
<td>0.005</td>
</tr>
<tr>
<td>EVI1 overexpression</td>
<td>1.430 (0.999-2.047)</td>
<td>0.051</td>
<td>1.711 (1.208-2.423)</td>
<td>0.002</td>
</tr>
<tr>
<td>Transplantation status</td>
<td>0.700 (0.610-0.806)</td>
<td>&lt;0.001</td>
<td>0.758 (0.665-0.865)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**Normal karyotype**

<table>
<thead>
<tr>
<th></th>
<th>OS</th>
<th></th>
<th>EFS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95%CI)</td>
<td>P</td>
<td>HR (95%CI)</td>
<td>P</td>
</tr>
<tr>
<td><strong>NPM1⁺/FLT3-ITD⁺</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher DNMT3B exp.</td>
<td>3.090 (1.623-5.883)</td>
<td>0.001</td>
<td>2.414 (1.309-4.451)</td>
<td>0.005</td>
</tr>
<tr>
<td>Age &gt; 60 y</td>
<td>2.703 (1.134-6.438)</td>
<td>0.025</td>
<td>1.951 (0.831-4.583)</td>
<td>0.125</td>
</tr>
<tr>
<td>WBC count &gt; 100*10⁹/L</td>
<td>2.365 (1.270-4.404)</td>
<td>0.007</td>
<td>1.763 (0.958-3.246)</td>
<td>0.069</td>
</tr>
<tr>
<td>High FLT3-ITD A.B.</td>
<td>3.791 (1.971-7.294)</td>
<td>&lt;0.001</td>
<td>3.130 (1.685-5.813)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**NPM1⁺/FLT3-ITD⁺ with high FLT3-ITD A.B.**

<table>
<thead>
<tr>
<th></th>
<th>OS</th>
<th></th>
<th>EFS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95%CI)</td>
<td>P</td>
<td>HR (95%CI)</td>
<td>P</td>
</tr>
<tr>
<td>Higher DNMT3B exp.</td>
<td>4.850 (1.980-11.880)</td>
<td>0.001</td>
<td>4.428 (1.824-10.749)</td>
<td>0.001</td>
</tr>
<tr>
<td>Age &gt; 60 y</td>
<td>2.007 (0.660-6.101)</td>
<td>0.219</td>
<td>1.478 (0.489-4.464)</td>
<td>0.489</td>
</tr>
<tr>
<td>WBC count &gt; 100*10⁹/L</td>
<td>1.356 (0.634-2.901)</td>
<td>0.433</td>
<td>0.954 (0.446-2.043)</td>
<td>0.904</td>
</tr>
</tbody>
</table>

Multivariate Cox regression model for probability of Overall Survival (OS) and Event Free Survival (EFS). Favorable karyotype include: inv(16), t(8;21) and t(15;17). Transplantation status includes either no transplantation, autologous transplantation, or allogeneic transplantation. Age, WBC-count are dichotomized. DNMT3B expression is dichotomized in multivariate analysis. P-values (bold) indicate whether differences are significant at the level of 0.05. HR=hazard ratio. 95%CI: 95% confidence interval. WBC = white blood cell. A.B. = allelic burden.

**References**


High DNA-methyltransferase 3B expression predicts poor outcome in AML


**Supplementary Information**

**Supplementary Materials and Methods**

For the present study, a database from previously published cohort of 525 AML patient samples, 11 CD34+ donor samples, and 5 normal bone marrow (NBM) control samples was used. Details on clinical data and experimental procedures of the microarray profiling on the Affymetrix HG-U133 plus 2.0 arrays can be found elsewhere. The array data are accessible online at the Gene Expression Omnibus (GSE14468). *DNMT3B* expression was determined using the probe 220668_s_at. This probe-set targets all the functional *DNMT3B* isoforms 1, 2, 6 (ENST00000353855, ENST00000328111, ENST00000201963), the truncated *DNMT3B* isoform 3 (ENST00000348286) and the coding *DNMT3B* isoform 201 (ENST00000344505). *DNMT3B* expression was validated by q-PCR on cDNA samples from 39 patients using a commercially available primer/probe set (Applied Biosystems, probe Hs00171876_m1), and data were normalized to β-actin gene expression levels. Data regarding *FLT3-ITD* allelic burden were described elsewhere. SPSS version 20.0 software (SPSS inc, Chicago, IL) was
used for statistical analyses. Clinical and molecular characteristics were compared using the Pearson’s chi-squared test. Survival curves were calculated by the Kaplan-Meier method and compared using the log-rank test. Univariate and multivariate survival analyses were carried out using the Cox proportional hazards model; covariates included in the analyses are indicated in the corresponding tables. We considered $P<0.05$ as statistically significant.

Supplementary Figures

**Figure S1. DNMT3B expression in AML.** (A) DNMT3B expression from microarray data was validated by q-PCR. DNMT3B expression measured through the HGU133 plus2.0 array was compared with q-PCR data of the same samples using β-ACTIN expression for normalization. Data were represented as the ΔCt of DNMT3B and β-ACTIN (2ΔCt). Included for analysis were 20 samples with higher DNMT3B expression and 19 samples with lower DNMT3B expression according to the array data. Correlation was found using the Spearmann’s test ($P < 0.001$). (B) Whisker-box plots showing micro-array based DNMT3B expression distribution among AML patients in comparison with CD34+ and normal bone marrow (NBM) cells. Each box represents a quartile of the AML cohort based on DNMT3B expression.
Figure S2. High DNMT3B expression correlates with inferior OS and EFS in AML. (A-B) Kaplan-Meier plots for OS (A) and EFS (B) showed that patients included in the higher quartile of DNMT3B expression (5-year OS: 17.2 ± 3.3 % and 5-year EFS: 13.6 ± 3.0%) exhibited a significantly poor OS and EFS when compared to patients classified within the lower (5-year OS: 51.5 ± 4.4 and 5-year EFS: 39.1 ± 4.3%), lower intermediate (5-year OS: 38.0 ± 4.3% and 5-year EFS: 32.6 ± 4.2%) and higher intermediate (5-year OS: 41.8 ± 4.4% and 5-year EFS: 31.6 ± 4.1%) quartiles. P values were determined with the log-rank test. Survival data were available for 518 patients. The number of patients included in each group is shown in brackets.

Supplementary Tables

Table S1. Characteristics of patients with lower and higher DNMT3B expression

<table>
<thead>
<tr>
<th></th>
<th>Lower DNMT3B expression</th>
<th>Higher DNMT3B expression</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (N=518), no. (%)</td>
<td></td>
<td></td>
<td>0.289†</td>
</tr>
<tr>
<td>Female</td>
<td>191 (49.4)</td>
<td>69 (52.7)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>196 (50.6)</td>
<td>62 (47.3)</td>
<td></td>
</tr>
<tr>
<td>Age (N=518), no. (%)</td>
<td></td>
<td></td>
<td>0.482†</td>
</tr>
<tr>
<td>&lt; 60 y</td>
<td>331 (85.5)</td>
<td>113 (86.3)</td>
<td></td>
</tr>
<tr>
<td>&gt; 60 y</td>
<td>56 (14.5)</td>
<td>18 (13.7)</td>
<td></td>
</tr>
<tr>
<td>WBC x 10⁹/L (N=516), no. (%)</td>
<td></td>
<td></td>
<td>0.227†</td>
</tr>
<tr>
<td>&lt; 100</td>
<td>316 (82.1)</td>
<td>103 (78.6)</td>
<td></td>
</tr>
<tr>
<td>&gt; 100</td>
<td>69 (17.9)</td>
<td>28 (21.4)</td>
<td></td>
</tr>
<tr>
<td>Transplantation status (N=504), no. (%)</td>
<td></td>
<td></td>
<td>0.886‡</td>
</tr>
<tr>
<td>No transplantation</td>
<td>219 (58.7)</td>
<td>80 (61.1)</td>
<td></td>
</tr>
<tr>
<td>Autologous transplantation</td>
<td>53 (14.2)</td>
<td>17 (13.0)</td>
<td></td>
</tr>
<tr>
<td>Allogeneic transplantation</td>
<td>101 (27.1)</td>
<td>34 (26.0)</td>
<td></td>
</tr>
<tr>
<td>FAB classification (N=525), no. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>12 (3.1)</td>
<td>6 (4.5)</td>
<td>0.285†</td>
</tr>
<tr>
<td>M1</td>
<td>57 (14.5)</td>
<td>43 (32.6)</td>
<td>&lt;0.001†</td>
</tr>
<tr>
<td>M2</td>
<td>94 (23.9)</td>
<td>37 (28.0)</td>
<td>0.203†</td>
</tr>
<tr>
<td>M3</td>
<td>19 (4.8)</td>
<td>4 (3.0)</td>
<td>0.272†</td>
</tr>
<tr>
<td>M4</td>
<td>78 (19.8)</td>
<td>15 (11.4)</td>
<td>0.016†</td>
</tr>
<tr>
<td>M5</td>
<td>94 (23.9)</td>
<td>22 (16.7)</td>
<td>0.051†</td>
</tr>
<tr>
<td>M6</td>
<td>6 (1.5)</td>
<td>1 (0.8)</td>
<td>0.440†</td>
</tr>
<tr>
<td>Other/Unknown</td>
<td>33 (8.4)</td>
<td>4 (3.0)</td>
<td>0.023†</td>
</tr>
<tr>
<td>Cytogenetics, no. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(15;17) (N=522)</td>
<td>22 (5.6)</td>
<td>3 (2.3)</td>
<td>0.086†</td>
</tr>
<tr>
<td>t(8;21) (N=522)</td>
<td>38 (9.7)</td>
<td>0 (0.0)</td>
<td>&lt;0.001†</td>
</tr>
<tr>
<td>inv(16) (N=522)</td>
<td>42 (10.8)</td>
<td>0 (0.0)</td>
<td>&lt;0.001†</td>
</tr>
</tbody>
</table>
Table S1. (continued)

<table>
<thead>
<tr>
<th></th>
<th>Lower DNMT3B expression</th>
<th>Higher DNMT3B expression</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (N=525)</td>
<td>153 (38.9)</td>
<td>66 (50.0)</td>
<td>0.017†</td>
</tr>
<tr>
<td>11q23 (N=513)</td>
<td>21 (5.4)</td>
<td>6 (4.8)</td>
<td>0.490†</td>
</tr>
<tr>
<td>Complex (N=525)</td>
<td>14 (3.6)</td>
<td>10 (7.6)</td>
<td>0.052†</td>
</tr>
</tbody>
</table>

Other genetic aberrations, no. (%)

<table>
<thead>
<tr>
<th></th>
<th>Lower DNMT3B expression</th>
<th>Higher DNMT3B expression</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPM1 mutations (N=525)</td>
<td>101 (25.7)</td>
<td>57 (43.2)</td>
<td>&lt;0.001†</td>
</tr>
<tr>
<td>FLT3-ITD mutations (N=525)</td>
<td>91 (23.2)</td>
<td>52 (39.4)</td>
<td>&lt;0.001†</td>
</tr>
<tr>
<td>FLT3-TKD mutations (N=523)</td>
<td>43 (11.0)</td>
<td>10 (7.6)</td>
<td>0.177†</td>
</tr>
<tr>
<td>CEBPA single mutations (N=525)</td>
<td>8 (2.0)</td>
<td>4 (3.0)</td>
<td>0.355†</td>
</tr>
<tr>
<td>CEBPA double mutations (N=525)</td>
<td>20 (5.1)</td>
<td>6 (4.5)</td>
<td>0.507†</td>
</tr>
<tr>
<td>IDH1 mutations (N=522)</td>
<td>31 (7.9)</td>
<td>7 (5.3)</td>
<td>0.210†</td>
</tr>
<tr>
<td>IDH2 mutations (N=522)</td>
<td>31 (7.9)</td>
<td>18 (13.6)</td>
<td>0.042†</td>
</tr>
<tr>
<td>DNMT3A mutations (N=512)</td>
<td>78 (20.5)</td>
<td>34 (26.0)</td>
<td>0.118†</td>
</tr>
<tr>
<td>EVI1 overexpression (N=518)</td>
<td>33 (8.5)</td>
<td>19 (14.5)</td>
<td>0.039†</td>
</tr>
</tbody>
</table>

WBC indicates white blood cell count; and FAB, French-American-British. P-values (bold) indicate whether differences are significant at the level of 0.05. †P values are based on Fisher exact tests. ‡P values are based on Pearson Chi-Square.

Table S2. Characteristic of patients with normal karyotype (n=219)

<table>
<thead>
<tr>
<th></th>
<th>Lower DNMT3B expression</th>
<th>Higher DNMT3B expression</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal karyotype patients, n. (%)</td>
<td>153 (69.9)</td>
<td>66 (30.1)</td>
<td>&lt;0.001†</td>
</tr>
<tr>
<td>NPM1+ / FLT3-ITD</td>
<td>51 (33.3)</td>
<td>19 (28.8)</td>
<td>0.309†</td>
</tr>
<tr>
<td>NPM1+ / FLT3-TID</td>
<td>47 (30.7)</td>
<td>8 (12.1)</td>
<td>0.002†</td>
</tr>
<tr>
<td>NPM1+ / FLT3-ITD+</td>
<td>22 (14.4)</td>
<td>3 (4.5)</td>
<td>0.025†</td>
</tr>
<tr>
<td>NPM1+ / FLT3-TID+</td>
<td>33 (21.6)</td>
<td>36 (54.5)</td>
<td>&lt;0.001†</td>
</tr>
</tbody>
</table>

NPM1+ / FLT3-ITD+ patients, n. (%) 0.144†

without high FLT3-ITD A.B. 18 (54.5) 14 (38.9)
with high FLT3-ITD A.B. 15 (45.5) 22 (61.1)

A.B. indicates allelic burden. P-values (bold) indicate whether differences are significant at the level of 0.05. In line with de Jonge et al., high FLT3-ITD allelic burden was defined as an allelic FLT3-ITD/FLT3 ratio higher than 1. †P values are based on Fisher exact tests. ‡P values are based on Pearson Chi-Square.
Supplementary References


Improved classification of *MLL-AF9*-positive acute myeloid leukemia patients based on *BRE* and *EVI1* expression

Sylvie M. Noordermeer¹,⁴, Davide Monteferrario¹,⁴, Mathijs A. Sanders², Lars Bullinger³, Joop H. Jansen¹ and Bert A. van der Reijden¹,⁵

¹Department of Laboratory Medicine, Laboratory of Hematology, Radboud University Nijmegen Medical Centre, Nijmegen Centre for Molecular Life Sciences, Nijmegen, The Netherlands.
²Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands.
³Department of Internal Medicine III, University of Ulm, Ulm, Germany.
⁴These authors contributed equally to this work.
⁵Corresponding author.

Adapted from:
*Improved classification of MLL-AF9-positive acute myeloid leukemia patients based on BRE and EVI1 expression. Blood 2012 May; 119(18):4335-7*
**Letter to the editor:**

The 5-year overall survival (OS) rate of patients with acute myeloid leukemia (AML) containing an *MLL-AF9* fusion gene is approximately 40%\(^1,2\). We recently showed in 2 independent cohorts that the prognosis among *MLL-AF9* positive patients can be refined based on *BRE* mRNA expression\(^3\). *MLL-AF9* positive patients with outlier high *BRE* expression exhibited a superior outcome (5-year OS of 80% and 64% for the 2 cohorts, respectively) while patients with normal *BRE* expression exhibited a very poor outcome (5-year OS of 0% and 7%, respectively). Thus, *BRE* expression may be used for refined risk stratification among *MLL-AF9* positive cases. However, the identification of patients with high *BRE* expression with routinely available techniques such as qPCR is difficult as the fold-difference between normal and high *BRE* expression is small\(^3\).

High *EVI1* expression occurs in approximately 10% of AML cases and is associated with an inferior outcome. High *EVI1* expression has recently been associated with *MLL* rearrangements including *MLL-AF9*\(^4-6\). To determine the correlation between *BRE* and *EVI1* expression, we reanalyzed the 2 *MLL-AF9* cohorts for which we reported high *BRE* expression\(^3,7,8\). This showed that high *BRE* and high *EVI1* expression are mutually exclusive (Figure 1A). Hence, the poor prognosis of the patients lacking high *BRE* expression could be explained by *EVI1* expression (Figure 1B). Of note, in both *MLL-AF9* cohorts we identified a few cases without high *BRE* or *EVI1* expression (Figure 1A). As the number of these patients is low, additional studies are required to reliably determine their prognosis.

To study whether the mutually exclusive expression of *BRE* and *EVI1* is accompanied by distinct expression profiles, we performed unsupervised genome-wide cluster analysis. The results showed that among *MLL-AF9* patients, high *BRE* expressing patients clustered apart from *EVI1*-positive patients (Figure 1C). Although the *EVI1*-positive patients were separated from the *BRE*-positive patients, they did not share highly similar expression profiles, while *BRE*-positive patients did. Indeed, in an unsupervised cluster analysis of the total AML cohort, *MLL-AF9* positive patients with high *EVI1* expression clustered only partially, showing modest similarities in expression profiles (Figure 2). This was in contrast to patients with high *BRE* expression that were almost completely confined to one distinct AML cluster, as described before\(^3\). Thus, *MLL-AF9* positive patients with high *BRE* expression seem to represent a specific AML subclass with highly similar expression profiles, while the *MLL-AF9* positive patients with high *EVI1* expression do not.

In our previous study, 40% of the *MLL-AF9* positive cases were missed by routine cytogenetics\(^3\). In addition, *MLL-AF9* positive cases with high *EVI1* expression lack chromosomal rearrangements encompassing the *EVI1* locus on chromosome 3\(^4,5\). Therefore, risk stratification of these patients could be improved by molecular screening for *MLL-AF9*–positivity and *EVI1* overexpression, in addition to routine cytogenetics.
Figure 1. High **BRE** and high **EVI1** expression are mutually exclusive in **MLL-AF9** leukemia.

(A) **BRE** expression was plotted against **EVI1** expression for **MLL-AF9** positive cases in 2 separate **MLL-AF9** cohorts. In the first cohort (left plot), 33% (6/18) of the samples showed high **EVI1** expression, and 55.6% (10/18) showed high **BRE** expression. In the second cohort (right plot), 27.3% of the samples showed high **EVI1** expression (3/11), and 54.5% showed high **BRE** expression (6/11). Both cohorts contained 2 patients with neither high **BRE** nor high **EVI1** expression. High **BRE** expression was defined as described before. High **EVI1** expression was defined as the expression of the upper 10% of the total cohort. P values for negative correlations were calculated using Spearman correlation tests. (B) Kaplan-Meier plots showed that patients with high **BRE** expression had significantly better overall survival than patients with high **EVI1** expression among **MLL-AF9**.

---

**MLL-AF9 cohort 1**

![Graph showing **BRE** and **EVI1** expression](image1)

**MLL-AF9 cohort 2**

![Graph showing **BRE** and **EVI1** expression](image2)
positive cases. P-values were determined with the log-rank test. The number of patients included in the analyses is shown in brackets. (C) EVI1-positive patients cluster apart from high BRE expressing patients among MLL-AF9 positive patients in unsupervised clustering analysis. However, EVI1-positive patients show less similar expression profiles among each other compared with high BRE expressing patients (indicated by faint red color compared with bright red color, respectively). Unsupervised clustering was performed on the first cohort as described elsewhere and clustering is represented as pairwise correlations between samples with a gradient from red to blue indicating degree of correlation (bright red: high correlation, blue: poor correlation). Black bars represent relative BRE (212645_x_at) and EVI1 (221884_at) expression, as indicated.

Figure 2. Unsupervised clustering of 525 AML patients. Unsupervised cluster analysis of the first AML cohort shows that MLL-AF9 patients with high BRE expression cluster together almost completely, in contrast to EVI1-positive MLL-AF9 patients that show partial clustering. This graph was adapted from a previously published figure (Noordermeer et al.), by including EVI1 expression (221884_at). MLL-AF9 and 11q23 status (red: positive, green: negative) and relative BRE and EVI1 expression are indicated along the pairwise correlation graph by the numbers 1-4, respectively.
References


The E3-ligase HOIP specifies linear ubiquitin chain assembly through its RING-IBR-RING domain and the unique LDD extension

Judith J. Smit¹, Davide Monteferrario², Sylvie M. Noordermeer², Willem J. van Dijk¹, Bert A. van der Reijden² and Titia K. Sixma¹,³

¹Division of Biochemistry, The Netherlands Cancer Institute, Amsterdam, The Netherlands
²Department of Laboratory Medicine, Laboratory of Hematology, Nijmegen Centre of Molecular Life sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
³Corresponding author

Adapted from: The E3 ligase HOIP specifies linear ubiquitin chain assembly through its RING-IBR-RING domain and the unique LDD extension. EMBO J 2012 Oct; 31: 3833-44
Abstract

Activation of the NF-κB pathway requires the formation of Met1-linked ‘linear’ ubiquitin chains on NEMO, which is catalysed by the Linear Ubiquitin Chain Assembly Complex (LUBAC) E3 consisting of HOIP, HOIL-1L and Sharpin. Here, we show that both LUBAC catalytic activity and LUBAC specificity for linear ubiquitin chain formation are embedded within the RING-IBR-RING (RBR) ubiquitin ligase subunit HOIP. Linear ubiquitin chain formation by HOIP proceeds via a two-step mechanism involving both RING and HECT E3-type activities. RING1-IBR catalyses the transfer of ubiquitin from the E2 onto RING2, to transiently form a HECT-like covalent thioester intermediate. Next, the ubiquitin is transferred from HOIP onto the N-terminus of a target ubiquitin. This transfer is facilitated by a unique region in the C-terminus of HOIP that we termed ‘Linear ubiquitin chain Determining Domain’ (LDD), which may coordinate the acceptor ubiquitin. Consistent with this mechanism, the RING2-LDD region was found to be important for NF-κB activation in cellular assays. These data show how HOIP combines a general RBR ubiquitin ligase mechanism with unique, LDD-dependent specificity for producing linear ubiquitin chains.

Introduction

Ubiquitin conjugation is a highly versatile system for conferring post-translational modifications, since this 76-amino acid protein can make a variety of chains that signal to different downstream effectors. The ubiquitins in these chains are usually linked between the ubiquitin C-terminus of the donor ubiquitin and any of the seven lysines in the acceptor ubiquitin, but the donor ubiquitin can also link to the amino group in the N-terminal methionine of the acceptor ubiquitin, leading to the formation of linear ubiquitin chains. Linear ubiquitin chains are assembled by the Linear Ubiquitin Chain Assembly Complex (LUBAC), which plays a critical role in the activation of the NF-κB pathway that is involved in various functions, including cell survival and inflammation. NF-κB activation can be induced by e.g. cytokines or DNA-damage, which lead to LUBAC-mediated ubiquitination of NEMO with linear ubiquitin chains. This linear ubiquitination of NEMO causes IKKβ phosphorylation and activation. Subsequently, IkBα is degraded and free NF-κB translocates to the nucleus to activate the transcription of target genes. The LUBAC complex consists of at least three different proteins, HOIP (RNF31), HOIL-1L (RBCK1) and Sharpin. HOIP and HOIL-1L belong to the RING-in-between-RING (RBR) class of E3-ligases. However, only the RBR domain of HOIP and not HOIL-1L is required for linear ubiquitin chain formation by LUBAC and subsequent IKKβ phosphorylation. Nevertheless, a combination of HOIP with either HOIL-1L or Sharpin is the minimal requirement for linear ubiquitin chain catalysis. The RBR class of E3-ligases, also known as the TRIAD class (two RING fingers and DRIL (double RING linked)), was first described in 1999. The structures of the separate RING domains and the in-between RING (IBR) have been solved (PDB entry 1WIM), however it remains unclear how the RBR forms a functional unit. RING1 has a classical RING fold, which is typically used for E2-E3 interactions. Also RING2 interacts with different E2s in yeast-two-hybrid studies and the cysteine and histidine distribution of RING2 suggests that it forms a RING domain.
However, even though Zn\(^{2+}\) stochiometry analysis indicates that all RING domains in Parkin coordinate two zinc ions\(^{19}\), the solution structure of HHARI RING2 does not have a classical RING fold and coordinates only one zinc ion per monomer. Furthermore, the HHARI RING2 domain was recently shown to form a thioester adduct with ubiquitin (HHARI~ubiquitin) on a free cysteine as an intermediate step in the ubiquitin transfer\(^{20}\), similar to that found in HECT domains. Although the thioester adduct could not be visualized on the RBR protein Parkin in the same study, mechanistic analysis indicated that both RBR proteins include a HECT-like step in the ubiquitin transfer. An intact RBR domain is necessary for efficient E3-ligase functioning, however Parkin IBR-RING2 can mediate the formation of ubiquitin linkages in the absence of RING1\(^{21,22}\). In addition to the interaction of both RING domains with E2 enzymes, the RBR of Parkin also interacts non-covalently with ubiquitin during chain formation\(^{23}\). The specificity for ubiquitin chain types is regulated completely at the level of the E3-ligase in HECT domains\(^{24,25}\). In contrast, with RING domain E3-ligases the E2 enzymes contribute to the chain types that are formed. Some E2s directly mediate the formation of specific ubiquitin chains via the non-covalent binding of an acceptor ubiquitin, positioning a particular lysine residue to attack the thioester bond between the E2 and the donor ubiquitin\(^{26–28}\). A single RING E3 can recruit several of these E2s and makes different chains dependent on the E2 specificity\(^{29,30}\). Occasionally the chain type that is being formed by a combination of a RING E3 and a less specific E2, such as Ube2D3 (UbcH5c), is determined by the specific E2-E3 combination\(^{31,32}\). So far LUBAC is the only E3-ligase complex that is known to promote linear ubiquitin chain formation. Although it contains RING domains, its ubiquitin chain forming specificity overrides that of the collaborating E2 enzymes. Thus even highly specific E2s that are known to catalyze the formation of very specific chain types, such as Ube2K (E2-25K) that forms K48-linked chains\(^{33}\), will form linear ubiquitin chains in the presence of LUBAC\(^{5}\). Therefore, chain type specificity is thought to be embedded within the LUBAC complex, but it is unclear how this specificity is organized. We performed an in-vitro analysis of HOIP ubiquitin chain assembly activity to investigate the mechanism underlying linear ubiquitin chain formation by the LUBAC complex. We show that a truncated form of HOIP is active in in-vitro linear chain formation in the absence of HOIL-1L and Sharpin. The catalytic activity and specificity for linear ubiquitin chain assembly of the LUBAC complex is completely embedded within HOIP RING2 and a newly identified Linear ubiquitin chain Determining Domain (LDD) in the C-terminus of HOIP. Furthermore, we show that the ubiquitin thioester is first transferred from the E2 onto HOIP and is subsequently linked to a target ubiquitin that is docked on the LDD. This study strengthens the knowledge on the general mechanism for RBR mediated ubiquitin chain formation and provides novel mechanistic insights in linear ubiquitin chain assembly by HOIP.
Materials and methods

Construction of plasmids

Codon optimized cDNA for *E. coli* expression of HOIP and HOIL-1L was obtained from Genscript. The cDNA was subcloned into pGEX-6P-1 vectors (GE Healthcare) with an N-terminal GST-tag for *E. coli* expression. HOIP<sub>RBR-LDD</sub>, HOIP<sub>R2-LDD</sub> and HOIP<sub>LDD</sub> were cloned into a pETNKI-His-3C-LIC-amp vector for *E. coli* expression<sup>34</sup>. Mammalian expression constructs pcDNA3.1-HOIL-1L-His, pcDNA3.1-Myc-HOIP and pcDNA3.1-Myc-HOIP-ΔUBA<sup>563-616</sup> were kindly provided by Dr. K. Iwai (Osaka University, Japan). UbiquitinΔGly76, Ubiquitin single and triple point-mutations and point-mutations in HOIP were introduced by using the QuickChange Mutagenesis Kit from Stratagene (La Jolla, CA, USA). The luciferase NF-κB reporter construct, pNF-κB-Luc, and the positive control pFC-MEKK were obtained from Agilent Technologies. Renilla luciferase vector, pRL-null, was obtained from Promega.

General, proteins and antibodies

Ubiquitin, hUba1, Ube2D3 and Ube2L3 were expressed and purified as described previously<sup>18,35–37</sup>. TAMRA-Ubiquitin, ubiquitinTAMRA, His<sub>6</sub>ubiquitin and biotin-ubiquitin (inhibition assay) were generously provided by Remco Merkx, Dharjath Hameed and Huib Ovaa<sup>36</sup>. Biotin-Ubiquitin (di-ubiquitin formation assays) and ubiquitin lysine mutants were obtained from Boston Biochem.

Protein expression and purification

Full-length HOIP and HOIL-1L were expressed in *E. coli* BI21 (DE3) pLysS cells by induction with 0.8 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and 0.2 mM ZnSO<sub>4</sub> over night at 18°C. Cells were resuspended in 20 mM Tris/HCl pH 8.0, 100 mM NaCl, 5 mM βME and Complete EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland). Cells were lysed by a high-pressure EmulsiFlex-C5 device (Avestin, Mannheim, Germany). Initial purification was achieved by binding the proteins to glutathione beads, washing the beads with buffer supplemented with 0.5 M NaCl and elution in buffer containing 50 mM GSH. The GST-tag was cleaved by incubation with 3C protease, followed by gel filtration (Superose6) online with a GST-column in 20 mM Hepes/HCl pH 8, 100 mM NaCl, 1 µM ZnCl<sub>2</sub> and 5 mM βME. HOIP<sub>RBR-LDD</sub> (699-1072) was produced in *E. coli* BI21 (DE3) pLysS cells. Expression was induced by the addition of 0.4 mM IPTG and 10 µM ZnCl<sub>2</sub> at an OD600 of 0.8 in LB medium supplemented with 50 µg/ml carbenicillin and chloramphenicol. Expressions were further cultivated over night at 16°C. The cells were lysed in 50 mM Tris/HCl pH 8, 150 mM NaCl, 2 mM Imidazole, 1 µM ZnCl<sub>2</sub>, 5 mM β-mercapto-ethanol (βME) in the presence of complete EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland), DNase and 1 mM MgCl<sub>2</sub> by a high-pressure EmulsiFlex-C5 device (Avestin, Mannheim, Germany). Cleared lysate was incubated with Talon beads. The protein was eluted from the beads in buffer containing 200 mM imidazole and was subsequently loaded on a Resource-Q column. The His-tag was cleaved in solution with 3C protease. Further purification was achieved by a Heparin column followed by size-exclusion chromatography (Superdex 200) in buffer containing
25 mM Hepes/HCl pH 7.0, 150 mM NaCl, 1 µM ZnCl₂, and 5 mM βME. HOIP<sub>RBR-LDD</sub> point-mutants, HOIP<sub>R2-LDD</sub> and HOIP<sub>LDD</sub> were expressed and purified as described for HOIP<sub>RBR-LDD</sub>, excluding the cleavage of the His-tag and size-exclusion chromatography. For comparison with the HOIP<sub>RBR-LDD</sub> point-mutants, wild type HOIP<sub>RBR-LDD</sub> was prepared following the same protocol.

**In-vitro ubiquitin chain formation**

*In-vitro* ubiquitin chain formation reactions were performed in standard conditions, unless specified otherwise. Standard conditions for ubiquitin chain formation were 100 nM hUba1, 600 nM of the indicated E2, 1 µM E3, 15 µM ubiquitin and 10 mM ATP in buffer containing 20 mM Hepes/HCl pH 8, 150 mM NaCl, 10 mM MgCl₂, 0.5 mM DTT. Reactions were performed at 37°C and stopped by the addition of protein loading buffer containing βME. Samples were separated on 4-12% NuPAGE gels (Invitrogen) in MES-buffer and analyzed by western blot using ubiquitin antibody (P4D1, Santa Cruz biotechnology) and HRP conjugated anti-Mouse antibody (BioRad).

**Single-cycle turnover assays**

Single-cycle turnover assays were performed in the same buffer conditions as described for the ubiquitin chain formation. TAMRA-ubiquitin (500 nM) was loaded onto E2 (600 nM) in an ATP (1 mM) dependent manner via hUba1 (100 nM) in 120 µL final reaction volume for 20 min at 37°C. The charging reaction was terminated by depleting the ATP with 2U apyrase. After 5 min incubation at room temperature, the sample was divided into smaller aliquots to compare the effects of the addition of HOIP<sub>RBR-LDD</sub> (1 uM) and ubiquitinΔGly76 (500 nM). Biotin-Ubiquitin was added simultaneously with HOIP<sub>RBR-LDD</sub> and wtUbiquitin in the acceptor ubiquitin competition assays. Reactions were performed at 37°C and stopped by the addition of non-reducing loading buffer on ice. Samples were analyzed on 4-12% NU-PAGE gels (Invitrogen) in MES buffer and the TAMRA signal was visualized on a ChemiDoc XRS (BioRad). Band quantification of Ube2D3~<sub>TAMRA</sub>ubiquitin was done with the ImageJ program (http://imagej.nih.gov/ij). Loading differences were accounted for by measuring the total amount of TAMRA signal per lane. Percentages normalized for the total amount of Ube2D3~<sub>TAMRA</sub>ubiquitin at T=0.

**Western blot analysis** was performed using anti-HOIP (ab85294, Abcam) and HRP-conjugated anti-Rabbit (BioRad) antibodies.

**Covalent HOIP~ubiquitin intermediate formation**

E2~ubiquitin was prepared as described for the single-cycle turnover assays in buffer containing 20 mM Hepes pH 8.5 and 5 mM βME. After the addition of Apyrase, HOIP (2 uM) was added to the mixture. Reactions were performed for 5 min on ice. The TAMRA signal was visualized on a ChemiDoc XRS (BioRad) and HOIP<sub>RBR-LDD</sub> was visualized on non-reducing western blots with anti-HOIP antibody (ab85294, Abcam). Sample loading buffer was supplemented with 1 M Urea to partially unfold the proteins.
Fluorescence polarization assays

The fluorescence anisotropy of the C-terminal TAMRA-labeled ubiquitin (1nM) in binding buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM βME and 1 g/l chicken ovalbumin) was measured on a PerkinElmer EnVision 2010 Multilabel Reader. The binding was measured in 75 µL reactions. Serial 1:1-dilutions, starting at 220 µM HOIP, were performed in three repeats. Reactions were incubated for 20 min at 4°C before the measurements. The samples were excited at 531 nm and emission was measured at 579 nm, with correction for both the buffer background and G-factor of the instrument. The assays were performed in “non binding surface flat bottom” black 96-well plates (Corning) at room temperature. The resulting binding isotherms (anisotropy vs. HOIP concentration) were fit to a 1:1 non-linear binding model (Y=Bmax*X/(K_D+X)). All experimental data were processed using Ms Excel and Prism 4.03 (GraphPad Software, Inc.).

Multi-angle laser light-scattering (MALLS)

MALLS experiments were performed on a Mini-Dawn light scattering detector (Wyatt Technology) inline with a Superdex S200 10/30 column at 4°C in buffer containing 25 mM Hepes/HCl pH 7, 150 mM NaCl, 1 µM ZnCl₂ and 5 mM βME. Refractive index and light scattering detectors were calibrated against toluene and BSA. Data were analyzed using the Astra software.

Cell culture and transient transfection

HEK293FT cells were cultured in Dulbecco modified Eagle medium (DMEM; Gibco) supplemented with 10% non-heat-inactivated fetal calf serum (GIBCO), 1% penicillin/streptomycin (MP Biomedical), 1% non essential amino acids (GIBCO) and 1% L-glutamine (MP Biomedical). Cells were cultured in 24-wells plates at 37°C supplied with 5% CO₂. For transient expression, 400 ng plasmid DNA (pcDNA3.1-HOIL-1L-His, pcDNA3.1-Myc-HOIP, pcDNA3.1-Myc-HOIP-mutants, pFC-MEKK) was used. Empty vector pcDNA3.1 was used to compensate for differences in DNA amounts. Furthermore, 400 ng of luciferase NF-kB reporter construct and 200 ng of Renilla luciferase vector were added to the transfection medium. In total 2 µg of DNA was transfected in each condition. Transfection was performed at 60% confluence with lipofectamine 2000 (Invitrogen). Each condition was experimentally tested in triplicate.

NF-kB transactivation assay

As readout for NF-kB activation we performed a Dual luciferase™ reporter assay (Promega). 48 Hours after transfection cells were washed with PBS and lysed in 100µL of passive lysis buffer (Promega) for 1 hour. Luciferase assays were performed according to the protocol provided by the manufacturer (Promega). Luciferase signals were measured on the Lumat LB 9507 (EG&G Berthold). Western blot analysis was performed to confirm protein expression. Proteins of total lysates generated as described above were separated on 10% polyacrylamide gels and transferred to polyvinylidenefluoride (PDVF) membranes (Bio-Rad, Hercules, CA).
PVDF membranes were probed with anti-Myc (Santa Cruz) and anti-His (Abcam) primary antibodies, followed by probing with HRP-conjugated secondary antibodies. Antibody signal was visualized by chemiluminescence using the Biorad ChemiDox XRS+.

Results

Linear ubiquitin chain formation specificity is embedded within HOIP

To study linear chain formation, we expressed full-length human HOIL-1L, full-length HOIP and a series of HOIP deletion constructs in E. coli. We used synthetic genes that are optimized for bacterial expression (Figure 1A) and used the purified proteins for in-vitro reactions, analyzing free ubiquitin chain formation. As expected, full-length HOIP alone was not active in forming ubiquitin chains, but in the presence of HOIL-1L robust chain formation was observed (Figure 1B). Since previous published data were derived from assays performed in the absence of sodium chloride \(^5,6\) under conditions that are far from physiological (~150 mM NaCl), we tested the influence of NaCl and pH on the in-vitro reactions. In the absence of salt the reactions were more active and it was easier to visualize detailed chains (Figure S1A, S1B), but the overall pattern of the bands on gel remained the same. Furthermore, the proteins were only active in conditions above pH 7 and raising the pH up to pH 9.5 caused a minor extra activation of the reactions (Figure S1C). We mainly used reaction conditions with 150 mM NaCl at pH 8; however, conditions without NaCl are used in some of our experiments as a tool to improve visualization of the activity of the LUBAC proteins.

Next, we used an N-terminally truncated form of HOIP, which includes only the RBR domain and the C-terminal region that we have named Linear ubiquitin chain Determining Domain (HOIP\(^{RBR-LDD}\), Figure 1A). The sequence of the LDD is not conserved between RBR proteins, and Psi-BLAST searches and a Phyre threading analysis on this region reveal that it is exclusive to HOIP. Nevertheless, between HOIP orthologues the LDD is highly conserved (Figure S1D), which suggests that the LDD functions specifically in the context of the upstream RBR domain in HOIP. When we tested HOIP\(^{RBR-LDD}\) for in-vitro activity we found, surprisingly, that this construct does not require HOIL-1L and Sharpin for in-vitro activity (Figure 1B, Figure 1C). HOIP\(^{RBR-LDD}\) does not contain the UBA domain that is needed for the interaction with HOIL-1L and Sharpin \(^2,5,8\), explaining why the activity of HOIP\(^{RBR-LDD}\) is hardly increased by the addition of HOIL-1L in the reactions (Figure 1D). As HOIL-1L and Sharpin have been shown to be important for HOIP activity, we wondered whether the short RBR-LDD construct of HOIP retained the specificity for making N-terminally linked ubiquitin chains. We tested chain formation, using either Ube2D3 (UbcH5c) or Ube2L3 (UbcH7) as E2 enzymes. In both cases, HOIP\(^{RBR-LDD}\) forms ubiquitin chains with lysine-less ubiquitin (K0) and mutated ubiquitins that contain either a single lysine or a lysine point mutation (Figure S1E). In addition, when the ubiquitin N-terminus is blocked with a His-tag, a biotin or a TAMRA-label, the ubiquitin chain formation is eliminated (Figure 1E, Figure S1F and Figure S1G), indicating that the accessibility of the N-terminus is critical for this reaction. Combinations of any of the N-terminally blocked ubiquitins with ubiquitin\(^{ΔGly76}\), which can only function as an acceptor ubiquitin, produces solely di-ubiquitins (Figure 1E, Figure S1F and Figure S1G),
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly

Figure 1. HOIP mediates linear ubiquitin chain formation. (A) HOIP and HOIL-1L constructs used in this study. Ubiquitin Like domain (UBL), Npl4 Zinc Finger (ZF), Ubiquitin Associated domain (UBA), Linear chain Determining Domain (LDD) and a RING-in-Between-RING domain (RBR) consisting of two RING domains (R1, R2) and an in-between RING domain (IBR). The domain borders of the UBL, ZF, UBA and RBR-domains are drawn to scale according to Uniprot definitions (www.uniprot.org). (B) Ubiquitin chain formation with Ube2L3 in combination with 2 µM of the different E3s after 0, 10, 20 and 40 min. Standard reaction conditions are described in the materials and methods. Reactions were performed without NaCl. (C) Increasing amounts of ubiquitin chain formation with increasing concentrations of HOIP RBR-LDD. Reactions were performed in the presence of Ube2D3 and were stopped after 1 hour. (D) HOIP RBR-LDD cannot be further activated by HOIL-1L in a 1 hour reaction with Ube2L3. Reactions were performed without NaCl. (E) Ubiquitin chain formation by HOIP RBR-LDD with N-terminally blocked ubiquitin (biotin ubiquitin) and C-terminally truncated ubiquitin (UbiquitinΔGly76). Reactions were stopped after 90 min at 32°C. (F) HOIP RBR-LDD activity with and without Ube2D3 and the Ube2D3 active site mutant (Ube2D3 C85A). Reactions were stopped after 15, 30, 60 and 120 min. (G) Di-ubiquitin linkage formation with TAMRA ubiquitin in the presence and absence of ubiquitinΔGly76 and Ube2D3 after 2 hours.
confirming that a free ubiquitin N-terminus is essential for ubiquitin chain formation by HOIP. Consequently, the RBR-LDD in the C-terminus of HOIP is sufficient for the linear ubiquitin chain formation specificity of the LUBAC complex and does not require the presence of other LUBAC-proteins. Since HOIL-1L and Sharpin are essential for full-length HOIP activity, but not for the HOIP\textsuperscript{RBR-LDD}, it seems that the catalytic center is not available for catalysis in the full-length protein. The UBL domains of either HOIL-1L or Sharpin have to bind to the UBA domain of HOIP, which lies N-terminally of the catalytic RBR-LDD, to activate the proteins in the NF-Kb pathway\textsuperscript{38,39}. This could suggest some level of auto-inhibition within HOIP, similar to that seen in the RBR protein Parkin, where the N-terminal UBL is binding to the C-terminal ubiquitin binding domain to block the catalytic center\textsuperscript{23}. Therefore, we tested if the N-terminus of HOIP can inhibit HOIP\textsuperscript{RBR-LDD} \textit{in trans}. Full-length HOIP, HOIP\textsuperscript{N-term} or HOIP\textsuperscript{UBA} were added to the reaction with HOIP\textsuperscript{RBR-LDD}, but the constructs did not inhibit the HOIP\textsuperscript{RBR-LDD} mediated chain formation (Figure S1H). Apparently the covalent linkage of the N-terminal domains to the RBR is required for the inhibition, either by increasing the local concentration or by arranging some position specific conformational change that can be released by the Sharpin or HOIL-1L interaction. Consequently, the exact mechanism by which the catalytic domain is kept in an inactive state in full-length HOIP remains to be resolved. The active LUBAC complex mediates the specific formation of linear ubiquitin chains in cooperation with many different E2 enzymes that are normally highly specific in the formation of different types of ubiquitin chains\textsuperscript{5}. This ability to override the E2 specificity is retained in HOIP\textsuperscript{RBR-LDD}. It specifically catalyses the formation of linear ubiquitin chains in the presence of the E2s Ube2D3, which can mediate the formation of many different types of lysine-linked ubiquitin chains\textsuperscript{30}, and Ube2L3, which targets to cysteines\textsuperscript{20}, indicating that the E2s are important to deliver the activated ubiquitin to the complex, but do not contribute to the chain type specificity.

HOIP has E2-independent linear chain forming activity

Interestingly, we observed very weak chain formation activity with HOIP\textsuperscript{RBR-LDD} even with an inactive Ube2D3 mutant (C85A) (Figure 1F). Therefore, we analyzed the HOIP\textsuperscript{RBR-LDD} activity in the absence of E2 enzymes and still observed HOIP dependent activity (Figure 1F), confirming that the E3 does not require an E2 for activity. However, in the absence of the E1 no activity is observed (Figure 1F). The chains formed in the E2-independent reaction are exclusively linear ubiquitin chains (Figure 1G). A similar E2-independent activity was recently described for the RBR-protein Parkin\textsuperscript{21}, indicating that this may be a general feature of RBR proteins. Nevertheless, E2-independent activity is unlikely to reflect a physiological activity, since the reaction is much more efficient in the presence of an E2~Ub thioester. However, these data emphasize that linear chain specificity does not rely on E2s, but is completely embedded within HOIP.

HOIP RING1 and IBR are involved in E2 mediated activity

We next examined how HOIP promotes linear ubiquitin chain formation. To address this point, we made a series of point mutations and deletion constructs to unravel the contributions of
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly (Figure S2). The activities of all point mutants that are used in this study are shown in Figure 2 and are summarized in supplementary Figure S2. The effect of the mutations in the different domains of HOIP will be discussed throughout this article. First, the importance of RING1 and the IBR were analyzed. RING and IBR domains coordinate two zinc-ions via 8 Cys/His residues, whereby each zinc-ion is coordinated by 4 Cys/His residues. Cysteine mutations in RING1 that disrupt the coordination of the zinc-ions caused reduced E2 dependent activity with both Ube2D3 and Ube2L3 (Figure 2). Also HOIP\textsubscript{RBR-LDD} V701A, which was designed to interfere directly with the E2-E3 interaction but not to disrupt the RING fold\textsuperscript{40}, inhibited the ubiquitin chain formation. Interestingly, the C171,179A mutant solely disrupted Ube2L3 dependent activity and not Ube2D3 mediated chain formation, revealing a difference in the binding-interface between HOIP and different E2s. Nevertheless, the complete set of mutants reveals that RING1 is essential for the activity with both E2s. The E2-independent activity of HOIP\textsubscript{RBR-LDD} was not affected by the RING1 mutations, indicating a classical RING-type role for RING1 where the RING domain catalyzes the transfer from the E2 onto a target site. The IBR cysteine mutants also influenced the E2 dependent ubiquitin chain assembly, but not for E2-independent activity (Figure 2). Therefore, both RING1 and the IBR are important for E2 mediated ubiquitin chain formation by HOIP.

**Figure 2.** All domains in HOIP\textsubscript{RBR-LDD} are involved in ubiquitin chain formation. Ubiquitin chain formation with HOIP\textsubscript{RBR-LDD} mutants in the presence and absence of the E2 Ube2L3 or Ube2D3. Reactions were stopped after 6 h. The molecular weight marker is indicated by the asterisk (*). The solid bars indicate the location (RING1 (R1), IBR, RING2 (R2), LDD) of the mutations within HOIP\textsubscript{RBR-LDD}.
Figure 3. HOIPR2-LDD is the catalytic centre for in cis linear ubiquitin chain assembly. (A) Di-ubiquitin formation with HOIPRBR-LDD RING1 mutants with TAMRA ubiquitin in the presence and absence of ubiquitinΔGly76 and Ube2D3. Reactions were stopped after 60 min. (B) HOIPR2-LDD mediated ubiquitin chain assembly after 2 hours cannot be activated by Ube2D3. (C) Activity of different HOIP constructs after 1 hour in the presence of Ube2D3 reveals that HOIPLDD is not active. (D) S200 10/300 gel filtration elution profile of HOIPRBR-LDD (black line) and HOIPR2-LDD (dotted line). Maximum absorbance has been set at 100 mAU. The according calculated molar mass from the MALLS for HOIPRBR-LDD (gray dots) is plotted against the elution volume. Three peaks are identified, correlating to HOIPRBR-LDD monomers (42.9 kDa), dimers (85.8 kDa) and multimers (> 85.8 kDa). (E) Ubiquitin chain formation by HOIPRBR-LDD monomers and multimers. Monomers and multimers were taken from corresponding single fractions from the middle of the different peaks of the gel filtration profile and used immediately (without concentrating the samples) in ubiquitin chain formation reactions. HOIPRBR-LDD was assayed at 0.5 µM in the presence of Ube2D3. Time points were taken after 0, 10, 20, 40, 60 min. (F) Combinations of HOIP mutants do not rescue ubiquitin chain formation activity with Ube2D3 (assay performed in absence of NaCl). Mutations were introduced in RING1 (C699A, K873A), RING2 (C871,874A), and LDD (C998A).
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly

**HOIP**\[^{R2-LDD}\]** forms the minimal domain for linear ubiquitin chain formation

The linear ubiquitin chain assembly specificity of HOIP is preserved in HOIP RING1 mutants (Figure 3A) and RING1/IBR mutations do not affect the E2-independent activity (Figure 2), indicating that these domains are not used in the actual linkage formation between two ubiquitins. Accordingly, when the RING1 and IBR domains are deleted (HOIP\[^{R2-LDD}\], Figure 1A), the ability to form ubiquitin linkages in an E2 independent manner is retained (Figure 3B). HOIP\[^{R2-LDD}\]** cannot be further activated by Ube2D3 (Figure 3B, Figure S3), showing the importance for RING1-IBR in E2 dependent activity. Thus a completely intact HOIP\[^{RBR-LDD}\]** is needed for efficient ubiquitin chain formation that is facilitated by the E2, but the intrinsic ubiquitin chain assembly activity is located more C-terminally in the RING2 and the LDD. We aimed at mapping the regions in HOIP\[^{R2-LDD}\]** that are essential for ubiquitin chain catalysis. The importance of RING2 was explored by comparing the activity of HOIP\[^{R2-LDD}\]** and a construct that lacks all of the RBR-domain (HOIP\[^{LDD}\], Figure 1A). Although HOIP\[^{R2-LDD}\]** can still form E2 independent di-ubiquitin linkages, HOIP\[^{LDD}\]** is catalytically inactive even at high concentrations (Figure 3C). In addition, single cysteine to alanine mutants of HOIP\[^{RBR-LDD}\]** RING2 are catalytically inactive (Figure 2). Therefore, RING2 is essential for ubiquitin chain assembly. Next, the relevance of the LDD in ubiquitin chain formation was investigated. We were unable to express constructs of HOIP that lack the LDD and all LDD mutants are catalytically inactive (Figure 2). Nevertheless, the LDD alone is not sufficient for catalysis. Consequently, the integrity of both RING2 and the LDD are needed for linear ubiquitin chain assembly by HOIP.

**HOIP mediates ubiquitin chain formation in cis**

The presence of multiple copies of HOIP within the LUBAC complex\[^{5}\] suggests that HOIP might assemble ubiquitin chains *in trans*. Therefore we next examined if the ubiquitin chain formation reaction is catalyzed by single HOIP molecules (*in cis*) or by the cooperation of multiple copies of HOIP (*in trans*). The gel filtration profile and multi angle laser light scattering (MALLS) of HOIP\[^{RBR-LDD}\]** show that the protein is purified as a mixture of monomers, dimers and multimers (Figure 3D). Nevertheless, the different fractions of the gel filtration profile show equal activity in free ubiquitin chain formation assays (Figure 3E), implying that the multimerization of HOIP\[^{RBR-LDD}\]** is not a requirement for activity. To confirm these data, we combined inactive HOIP\[^{RBR-LDD}\]** mutants in chain formation assays to test whether they would collaborate to rescue the ability to form ubiquitin linkages. RING1 mutants are affected in E2 dependent chain formation and LDD mutants cannot support the formation of the isopeptide bond between two ubiquitins. Consequently, a combination of a RING1 mutant and a LDD mutant is expected to be effective in chain formation if the reaction occurs *in trans*.

The combination of a RING1 and a LDD mutant did not lead to effective ubiquitin chain assembly, showing that the mutants do not complement each other (Figure 3F). Furthermore, a combination of a RING1 mutant and a RING2 mutant, or a RING2 mutant and a LDD mutant did not result in chain formation (Figure 3F), suggesting again that HOIP\[^{RBR-LDD}\]** proteins act individually and do not collaborate in ubiquitin chain formation. Finally, HOIP\[^{R2-LDD}\]** is purified
mainly as a monomer (Figure 3D) and is still active in E2-independent chain formation. Therefore we conclude that multimerization of HOIP is not a requirement for activity and linear ubiquitin chain formation is catalyzed within single HOIP molecules.

HOIP forms a reversible covalent intermediate with ubiquitin
In light of the HECT-like character of RING2 in other RBR proteins, we tested whether HOIP could make a covalent thioester intermediate. We used single-cycle turnover assays, with pre-charged E2~TAMRA-ubiquitin thioester and were able to trap an E3-ubiquitin intermediate with HOIP<sub>RBR-LDD</sub>. A covalent E3~TAMRA-ubiquitin complex could be visualized on non-reducing gels using anti-HOIP western blotting or, more clearly, by the TAMRA signal of ubiquitin (Figure 4A, Figure 4B, Figure S4A (Ube2D3), FigureS4B (Ube2L3)). The HOIP<sub>RBR-LDD</sub>~ubiquitin intermediate could be disrupted by the addition of reducing loading buffer, which illustrates that HOIP<sub>RBR-LDD</sub> forms a reversible covalent bond with ubiquitin in cooperation with both Ube2D3 and Ube2L3. Here we show for a second RBR protein the presence of an E3~ubiquitin thioester bond. The covalent HOIP~ubiquitin is transient, as indicated by the low signals, however the bond could be detected in the RING1 mutants as well as in the C916A LDD mutant (Figure 4A, Figure S4B). The RING2 mutants were completely impaired in forming this intermediate (Figure 4A, Figure S4B). RING2 has been suggested in literature as the actual site for the E3~ubiquitin thioester in RBR proteins, although visualization of the E3~thioester has only been successful for HHARI<sup>20</sup>. We could not assign the thioester forming cysteine, since several cysteines in RING2 are impaired in thioester formation. The HOIP Cys885 that aligns with the thioester-forming Cys357 in HHARI, could not form an oxyester HOIP~ubiquitin intermediate, when mutated to serine (Figure S4C). However, this could be due to detection limits of the assay, since the reaction is less favorable. Unlike the LDD, RING2 is conserved between RBR proteins (Figure S4D) and it is essential for E2~ubiquitin discharge and E3~ubiquitin formation. Therefore, it seems likely that RING2 provides the actual site on which the E3~ubiquitin is formed.

HOIP mediated ubiquitin transfer from the E2 onto a target is a two-step mechanism
To understand how the different domains within HOIP<sub>RBR-LDD</sub> contribute to the assembly of ubiquitin chains, we monitored the in-vitro E2~ubiquitin discharge and di-ubiquitin formation in single-cycle turnover assays with TAMRA-ubiquitin and the selected purified HOIP<sub>RBR-LDD</sub> mutants. The amino-terminus of TAMRA-ubiquitin is not available for linear ubiquitin chain formation and can only be linked to a ubiquitin with a free N-terminus by HOIP<sub>RBR-LDD</sub>. This feature allowed us to uncouple the discharge of ubiquitin from the E2 active site cysteine (in the absence of ubiquitinΔGly76) and the formation of the isopeptide bond between the N- and C-terminus of two ubiquitins (in the presence of ubiquitinΔGly76). HOIP<sub>RBR-LDD</sub> completely discharged TAMRA-ubiquitin from the E2 over time in the single-cycle turnover assays and formed di-ubiquitins when ubiquitinΔGly76 was added to the reaction (Figure 4C, Figure S4E). The E2~ubiquitin discharge is less efficient when RING1 mutants are used in the reaction and also the amount of di-ubiquitin that is formed declines.
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly.

**Figure 4.** E2-ubiquitin discharge and chain formation are two separate events. (A) Formation of a reversible covalent intermediate between HOIP and TAMRA-ubiquitin with different
HOIPmutants. Ube2D3 was used as the E2 enzyme. The TAMRA-signal is visualized on a reduced gel and at two contrast levels on a non-reduced gel. (B) Single-cycle turnover assay monitoring Ube2D3~TAMRA ubiquitin discharge and HOIP~TAMRA ubiquitin formation after 0, 4, 16 min at 37°C. (C) Single-cycle turnover assays showing Ube2D3~TAMRA ubiquitin discharge by HOIP mutants (left half of each gel), and di-ubiquitin formation upon the addition ubiquitinΔGly76 (right half of each gel). Discharge reactions were stopped after 0, 2, 4, 8, 16, 32, 64 min. RING1 (R1), RING2 (R2). (D) Ube2D3~TAMRA ubiquitin discharge rates in the presence and absence of an acceptor ubiquitin. Standard error for the Ube2D3~TAMRA ubiquitin were calculated over three independent experiments.

![Graph showing FP assay results](image)

**Figure 5. HOIP LDD interacts with the acceptor ubiquitin** (A) FP assay of ubiquitin\(^{\text{TAMRA}}\) binding to HOIP, showing increase in FP as a function of [HOIP], HOIP\(^{\text{RBR-LDD}}\) KD= 118 ± 8 µM, HOIP\(^{\text{RBR-LDD C930A}}\) KD= 83 ± 9.2 µM, HOIP\(^{\text{LDD}}\) KD= 97 ± 7 µM, HOIP\(^{\text{LDD C930A}}\) KD= 734 ± 395 µM. Standard deviations were calculated over three repeats. (B) HOIP\(^{\text{LDD}}\) and HOIP\(^{\text{LDD C930A}}\) inhibition on ubiquitin chain formation by HOIP\(^{\text{RBR-LDD}}\) in a concentration series of 0, 1, 2, 4, 8, 16, 32, 64 µM. Control reactions at the highest concentration of HOIP\(^{\text{LDD}}\) do not contain either HOIP\(^{\text{RBR-LDD}}\) (-E3) or Ube2D3 (-E2). The molecular weight marker is indicated by the asterisk (*). (C) Single-cycle turnover assays in the presence and absence of the acceptor ubiquitin-competitor: biotin ubiquitin. The TAMRA-signal visualizes di-ubiquitin formation by HOIP\(^{\text{RBR-LDD}}\) and Ube2D3~ ubiquitin discharge by HOIP\(^{\text{RBR-LDD C930A}}\) after 0, 2, 4, 8, 16 minutes. (D) Di-ubiquitin formation between TAMRA ubiquitin and different ubiquitin mutants visualized by the TAMRA-signal on a non-reduced gel. T= 10 minutes. Ubiquitin\(^{\text{Tmale}}\) = L8A, I4A, V70A triple mutant.
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly

(Figure 4C, Figure S4E). This confirms the role of RING1 in E2 mediated activity. Nevertheless, the E2-independent activity of the RING1 mutants is hardly affected (Figure 2), showing that RING1 is less important for the E2 independent driven activity and the ubiquitin linkage formation. The discharge of the ubiquitin from the E2 on HOIP and the linkage of the ubiquitin to a target ubiquitin by RING2-LDD were uncoupled in the single-cycle turnover assays. Although LDD mutants do not have any ubiquitin linkage formation activity in ubiquitin chain formation reactions, they are capable of efficiently discharging ubiquitin from the E2 (Figure 4C, Figure S4E). This indicates that the LDD is not involved in the destabilization of the E2~ubiquitin thioester, but is critical for ubiquitin chain assembly. Apparently the trans-thiolation of the ubiquitin from the E2 onto HOIP is independent of ubiquitin chain assembly. Accordingly, the E2~ubiquitin discharge efficiency is not dependent on the presence of an acceptor ubiquitin to which the donor ubiquitin can be transferred (Figure 4D). Interestingly, RING2 mutants are impaired in both the E2~ubiquitin discharge and the ubiquitin linkage formation (Figure 4C, Figure S4E), suggesting that RING2 is required for both steps of the ubiquitin transfer. This central role for RING2 in the transfer of the ubiquitin further supports its role as acceptor site for the E3~ubiquitin intermediate. Consequently, efficient ubiquitin chain formation is initiated by the E2-dependent delivery of ubiquitin to HOIP RING2 and is completed by subsequent LDD mediated ubiquitin chain assembly.

HOIP LDD catalyzes the final step of the ubiquitin transfer

The binding of ubiquitin to ubiquitin docking sites in E3-ligases is suggested to play a role in ubiquitin chain formation specificity by bringing a specific ubiquitin lysine residue in close proximity of the ubiquitin thioester bond\(^4\). HOIP has several known ubiquitin interaction motifs (UBA and the NFZs), but since these are not present in HOIP\(^{RBR-LDD}\) they cannot explain the linear ubiquitin chain formation. We found that the LDD is important for linking ubiquitins together, but not for E2~ubiquitin discharge and HOIP~ubiquitin intermediate formation. Therefore we wondered whether the LDD could function as a ubiquitin docking site. We measured the affinity of ubiquitin for HOIP\(^{RBR-LDD}\) and HOIP\(^{LDD}\) by fluorescent polarization (FP) with TAMRA labeled ubiquitin (TAMRA ubiquitin and ubiquitin\(^{TAMRA}\)). Both HOIP\(^{RBR-LDD}\) and HOIP\(^{LDD}\) interacted with ubiquitin\(^{TAMRA}\) with an affinity of approximately 100 µM, and do not bind to the free TAMRA-dye, showing that the LDD does interact with ubiquitin (Figure 5A, Figure S5A, Figure S5B). Then we analyzed the effect of the C930A mutant of the LDD, which is impaired in ubiquitin chain formation, on the affinity for ubiquitin in the FP assays. Unexpectedly, in the context of HOIP\(^{RBR-LDD}\), the C930A mutation did not affect affinity for ubiquitin\(^{TAMRA}\). In contrast, in the LDD alone, the HOIP\(^{LDD}\) C930A has a greatly reduced affinity for ubiquitin\(^{TAMRA}\) (Figure 5A, Figure S5A). The loss of activity of the HOIP\(^{RBR-LDD}\) C930A was not caused by unfolding of the protein as was shown by the gel filtration profile (Figure S2D). Therefore, the loss of binding of the LDD C930A mutant indicates interference with ubiquitin binding. In the longer construct, the mutation is possibly not strong enough to disrupt the complete interaction and just interferes with the proper ubiquitin orientation for chain formation or a second ubiquitin interaction site may be present elsewhere outside the LDD. The interaction between the LDD domain and ubiquitin was verified in in-vitro
ubiquitin chain formation assays. First, HOIP<sup>LDD</sup> was titrated into the ubiquitin chain reaction with HOIP<sup>RBR-LDD</sup>. The increasing amounts of HOIP<sup>LDD</sup> inhibited HOIP<sup>RBR-LDD</sup> mediated ubiquitin chain formation, presumably by competing away the freely available ubiquitin (Figure 5B). Importantly, addition of the HOIP<sup>LDD</sup> C930A to linear ubiquitin chain formation assays did not inhibit HOIP<sup>RBR-LDD</sup> mediated ubiquitin chain formation (Figure 5B). The loss of inhibition by the LDD C930A mutant indicates that this site is indeed important for ubiquitin interaction. We then tested if the interaction between the acceptor ubiquitin and HOIP is needed for di-ubiquitin formation in a single-cycle turnover assay. Ube2D3 was loaded with TAMRA ubiquitin, after which the HOIP<sup>RBR-LDD</sup> dependent discharge of the E2~<sub>TAMRA</sub>ubiquitin, and the linkage of TAMRA ubiquitin to wild type ubiquitin was monitored. To test if the acceptor ubiquitin interacts with HOIP during the di-ubiquitin formation, a<sup>biotin</sup> ubiquitin, which cannot act as an acceptor, was added during the discharge reaction to compete with the wild type ubiquitin (Figure 5C). Under these conditions the E2~<sub>TAMRA</sub>ubiquitin discharge and HOIP~ubiquitin intermediate formation were unaffected by the presence of <sup>biotin</sup>ubiquitin, showing that the transfer of the donor ubiquitin from the E2 onto HOIP was not affected (Figure 5C, Figure S5C). In contrast, the di-ubiquitin formation was inhibited by the <sup>biotin</sup>ubiquitin, suggesting that the N-terminally blocked ubiquitin competes with the wild type ubiquitin for binding to HOIP in the final step of the ubiquitin transfer. These results are in line with the fact that the LDD mutants are impaired in ubiquitin-ubiquitin linkage formation but not in E2~ubiquitin discharge (Figure 4C, Figure S4E), showing that the LDD does not interact with the donor ubiquitin, but rather with the acceptor ubiquitin. Interestingly, the LDD/ubiquitin interaction does not require the ubiquitin hydrophobic patch, which is used by many ubiquitin interaction motifs<sup>41</sup>, since point mutants of the hydrophobic patch still accept TAMRA ubiquitin via HOIP<sup>RBR-LDD</sup> (Figure 5D, Figure S5D). Consequently, the interaction between the LDD and ubiquitin is specific to HOIP, which is in line with the unique linear ubiquitin specificity that is evoked by HOIP.

The HOIP RING2 and LDD domains are crucial for NF-κB activation

Both RING2 and the LDD of HOIP are essential for linear ubiquitin chain formation in vitro. To determine the biological relevance of this finding, we tested wild type HOIP and the RING2 C885A, LDD C916A, LDD C930A, single and LDD C916,930A double mutants in HEK293FT cells. First, we verified the expression levels of the mutants and their capacity to interact with HOIL-1L in pull-down assays. All of the mutants retained the capacity to bind to HOIL-1L (Figure S6A), showing that the mutations did not cause major deficiencies in the folding of the full-length protein. Next, we tested the activity of the mutants in NF-κB luciferase reporter assays. In line with published data<sup>2</sup>, the combined expression of HOIP and HOIL-1L resulted in NF-κB activation to levels approaching those when using MEK-kinase 1 (Figure 6)<sup>42</sup>. However, the combined expression of HOIL-1L with any of the HOIP cysteine mutants did not activate NF-κB. As expected, the observed NF-κB signaling activity was dependent on the presence of the presence of the HOIL-1L interaction domain, the ubiquitin-associated-domain (UBA domain) in HOIP, as well as on the co-expression with HOIL-1L, since neither wild-type nor HOIP mutants were active without it (Figure S6B). These
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly. Data show that RING2 and the LDD are essential for linear ubiquitin chain formation and LUBAC mediated NF-κB activation in cells.

![Figure 6. HOIP RING2 and LDD are essential for NF-κB pathway activation.](image)

**Discussion**

NF-κB activation is an important signal during immune and DNA-damage responses. Upon stimulation, the formation of linear ubiquitin chains on NEMO forms an essential early event in the activation of the pathway. The linear ubiquitin chains are assembled by the LUBAC complex, which contains the two RBR E3-ligases HOIP and HOIL-1L. It is poorly understood how RBR proteins form functional units to mediate the assembly of ubiquitin chains or how linear ubiquitin chain specificity is determined by the LUBAC complex. Our results give a detailed insight in the molecular mechanism by which linear ubiquitin chains are formed by HOIP (Figure 7). We show that removing the N-terminal 698 residues of HOIP unmasks the full linear ubiquitin chain formation activity and bypasses the need for HOIL-1L and Sharpin. Within this active region of HOIP, the first two domains (RING1-IBR) are needed for the catalysis of the E2-mediated delivery of ubiquitin to the E3, but the mechanism that directs the linkage to the N-terminus of a target ubiquitin is embedded within HOIP RING2-LDD. The relevance of this activity in the context of the full-length protein within the LUBAC complex was verified by experiments in HEK293FT cells, showing that single
point mutations in either RING2 or the LDD impair NF-kB signaling. The transfer of ubiquitin from the E2 onto an acceptor ubiquitin is mediated by HOIP in a two-step mechanism. First, the ubiquitin thioester is transferred from the E2 onto HOIP, most likely on RING2, to form a reversible covalent intermediate (Figure 7C). This step can be catalyzed by the RING1-IBR mediated interaction with an E2–Ub thioester. Second, the ubiquitin is transferred from HOIP onto the N-terminus of the target ubiquitin to form an isopeptide bond. This uncoupling of the E2 catalyzed step from the transfer step to the acceptor ubiquitin explains why E2 enzymes do not affect the chain type specificity of HOIP. The LDD is essential for the specific transfer of the donor ubiquitin from HOIP onto the acceptor ubiquitin (Figure 7). We have shown that the interaction between the LDD and the acceptor ubiquitin is important during this process, suggesting that it functions as an ubiquitin docking domain for the acceptor ubiquitin. The need for a C-terminal ubiquitin interaction-domain within HOIP is likely to reflect a general feature for ubiquitin chain catalysis of RBR proteins, since Parkin also contains a recently identified ubiquitin interaction domain, which is located just before RING2 of the RBR domain, that is used in ubiquitin chain formation. Among RBRs, the RING2-LDD uniquely promotes linear ubiquitin chain formation. This selectivity for the amino terminus is exquisite since the ubiquitin N-terminus and K63 are located close to each other, indicating that precise positioning of the acceptor ubiquitin by the LDD is very important. It seems plausible that RING2-LDD provides additional contributions to selective targeting, possibly by preferring the chemical constellation of the ubiquitin amino-terminus over a lysine amino group.

Figure 7. Model for HOIP\textsuperscript{RBR-LDD} mediated ubiquitination. (A-B) Linear ubiquitin chain assembly requires both (A) the binding and correct orientation of an acceptor ubiquitin by
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly (A) the LDD and (B) the recruitment of an E2–ubiquitin to RING1-IBR. The ubiquitin is transferred from the E2 onto the acceptor ubiquitin in two independent steps. (C–D) First the ubiquitin thioester is transferred form the E2 onto RING2 (C) and second the ubiquitin is covalently linked to the N-terminus of the acceptor ubiquitin (D) that is oriented by the LDD. E. The E2 dependent activity can be bypassed by a less pronounced E2-independent activity.

References


Supplementary information
Supplementary Materials and Methods

Multi-sequence alignment
Multi-sequence alignments were made in ClustalW and the figures were produced manually.

Co-immunoprecipitation (CO-IP)
For co-immunoprecipitation experiments, approximately 2.5 million cells were plated in standard round 9 cm culture petri dishes 24 hours prior to transfection. 16 hours post transfection the culture medium was refreshed and 40 hours post transfection, cells were harvested and lysed using RIPA buffer supplemented with complete protease inhibitors cocktail (Roche diagnostic) and benzonase (Novagen). Cell lysates were incubated with mouse anti Myc-antibody overnight at 4°C. The remaining cell lysates were used as control for protein expression on immunoblot. The immunoprecipitation was performed using Protein G Sepharose beads (GE Health care). Beads were washed with RIPA buffer and incubated with the antibody-containing-lysates for 2.5 hours at 4°C. After incubation, beads were washed with RIPA buffer to remove all non-specifically bound proteins. Proteins
bound to the beads were released by adding SDS loading buffer and boiling at 95°C for five minutes. Western blot analysis was run to confirm protein expression. Proteins were separated on 10% polyacrylamide gels and transferred to polyvinylidene fluoride (PDVF) membranes (Bio-Rad, Hercules, CA). PVDF membranes were probed with anti-Myc (Santa Cruz) and anti-His (Abcam) primary antibodies, followed by probing with HRP-conjugated secondary antibodies. Antibody signal was visualized by chemiluminescence using the Biorad ChemiDox XRS+. 
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly.
**Figure S1.** (A) Ubiquitin chain formation by HOIP$^{RBR-LDD}$ and Ube2D3 in the presence and absence of 150 mM NaCl. Time points were taken after 0, 10, 20, 40, 80 min. (B) Ubiquitin chain formation activity of HOIP and HOIL-1L in the presence and absence of 150 mM NaCl. Ube2D3 was used as the E2 enzyme. Reactions were stopped after 0 and 30 min. (C) Ubiquitin chain formation by HOIP and HOIL-1L with Ube2D3 at increasing pH-values in 50 mM MMT-buffer. Reactions were stopped after 40 min. (D) Multi-sequence alignment of the LDD-domain of different HOIP orthologues. UniprotID: Human (Q96EP0), Mouse (Q924T7), Rat (E9PU29), Xenla (A0JMU3), Drome (Q8IPJ3), Ciona (Q1RL47). The asterisks indicate the sites of the cysteine mutations that have been introduced in this study. (E) Ubiquitin chain formation by HOIP$^{RBR-LDD}$, Ube2L3 or Ube2D3 and ubiquitin mutants. Single lysine ubiquitin point-mutants (K*R), all lysines mutated to arginine except for one lysine (K* only), all lysines mutated to arginine (K0). Control reactions lack either hUba1 (-E1); Ube2L3 and Ube2D3 (-E2); or HOIP$^{RBR-LDD}$ (-E3). Reactions were stopped after 1 hour at 32°C. (F, G) Ubiquitin chain formation by HOIP$^{RBR-LDD}$ with N-terminally blocked ubiquitin ($^{\text{H6}}$Ubiquitin or $^{\text{TAMRA}}$ubiquitin) and C-terminally truncated ubiquitin (ubiquitin$^{\Delta\text{Gly76}}$). 2 Hour reactions contained hUba1 (E1), Ube2D3 or Ube2L3 (E2) and HOIP$^{RBR-LDD}$ (E3) unless indicated otherwise. (H) Ubiquitin chain formation with 600 nM HOIP$^{RBR-LDD}$ in the presence of 10 µM HOIP$^{\text{N-term}}$, 100 µM HOIP$^{\text{UBA}}$ or 2 µM HOIP. The lane with the asterisk shows the reaction with HOIP in the absence of HOIP$^{RBR-LDD}$. Reactions were stopped after 0, 10, 20, 40, 80 min at 37°C.
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly

### Chain formation

<table>
<thead>
<tr>
<th></th>
<th>E2 dep</th>
<th>E2 indep</th>
<th>E3 thioester</th>
</tr>
</thead>
<tbody>
<tr>
<td>+++</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>nd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>nd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>++</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>++</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+++</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
</tbody>
</table>

### Diagram

#### A

- **HOIP<sub>RBR-LDD</sub>**
  - C699A/K873A
  - C717A/C719A
  - C799A/C802A
  - C825A/C828A
  - C871A/C874A
  - C885A/H887A
  - C890A/C893A
  - C898A/C901A
  - C885A
  - C885S
  - C911A
  - C916A
  - C930A
  - C969A
  - C986A
  - C998A
  - V701A
  - V701I
  - C719H
  - H729A
  - D751A
  - P750A/D751A
  - E814A
  - I869A
  - K873A
  - YAL878IEK
  - Q896E

- **Chain formation**
  - E2 dep
  - E2 indep
  - E3 thioester

#### B

- **R1**
  - C699A/K873A
  - C717A/C719A
  - C799A/C802A
  - C825A/C828A
  - C871A/C874A
  - C885A/C893A
  - C898A/C901A
  - C911A
  - C916A
  - C930A
  - C969A
  - C986A
  - C998A
  - V701A
  - V701I
  - C719H
  - H729A
  - D751A
  - P750A/D751A
  - E814A
  - I869A
  - K873A
  - YAL878IEK
  - Q896E

- **R2**
  - C699A/K873A
  - C717A/C719A
  - C799A/C802A
  - C825A/C828A
  - C871A/C874A
  - C885A/C893A
  - C898A/C901A
  - C911A
  - C916A
  - C930A
  - C969A
  - C986A
  - C998A
  - V701A
  - V701I
  - H729A
  - E814A
  - K873A
  - C885A
  - I869A
  - YAL878IEK
  - Q896E

#### Diagram

- **kDa**
  - 150
  - 75
  - 37
  - 25
  - 15

---

*Figure captions and additional information not included in the image.*
Figure S2. (A) Schematic representation of the HOIP^{RBR-LDD} point-mutants used in this study. Cysteine to alanine mutations in the RING domains and the IBR were designed to disrupt proper folding of the separate domains. The additional mutations were designed to explore RBR functioning without disrupting the fold of the domains, by interfering with possible E2-E3 interaction sites in the RING domains (V701A/I, K873A), changing residues that are likely to be surface exposed in RING1 (H729A) and the IBR (E814A), or residues that are located just outside the RING domains (P750A, D751A, I869A). All other RBR mutations were introduced at random throughout the protein to enhance our changes to identify the importance of the separate domains in linear ubiquitin chain formation. The cysteine mutations in the highly conserved cysteines of the LDD were introduced to cause local changes in the LDD to help unravel the importance of this domain in linear ubiquitin chain formation. The table lists the E2 dependent activity (E2 dep.), E2 independent activity (E2 indep.), discharge of ubiquitin from an E2 (E2 discharge) and E3~ubiquitin thioester formation (E3 thioester) of the mutants in the different experiments in this study. (B) Final purification product of the HOIP^{RBR-LDD} mutants shown on SDS-PAGE gel. Solid bars indicate the location (RING1 (R1), IBR, RING2 (R2), LDD) of the mutations within HOIP^{RBR-LDD}. The molecular weight marker is marked with the asterisk (*). (C) Final purification product of HOIP^{LDD}, HOIP^{LDD} and HOIP^{LDD} C930A shown on SDS-PAGE gel. The molecular weight marker is indicated by the asterisk. (D) Analytical gel filtration profiles of HOIP^{LDD} and HOIP^{LDD} C930A on a S200 5/150 column.
**Figure S3.** Control reactions with HOIP\textsuperscript{RBR-LDD} for E2 dependent activity in figure 3B. The E2 increases the amount of ubiquitin chain formation with HOIP\textsuperscript{RBR-LDD}, but not with HOIP\textsuperscript{R2-LDD}. Reactions were stopped after 0, 1, 1.5, 2, 3, 4 hours.
Figure S4. (A) Formation of a reversible covalent intermediate between HOIP\textsuperscript{RBR-LDD} and TAMRA-ubiquitin. An anti-HOIP western blot and the according reduced and non-reduced gels for the TAMRA-signal are shown. The reversible covalent bonds are indicated as E1~Ub (hUba1), E2~Ub (Ube2D3) and E3~Ub (HOIP\textsuperscript{RBR-LDD}). (B) Formation of a reversible covalent intermediate between HOIP\textsuperscript{RBR-LDD} and TAMRA-ubiquitin with different HOIP\textsuperscript{RBR-LDD} mutants. Same assay as in Figure 4A, but with UBE2L3 as the E2 enzyme. The TAMRA-signal is visualized on a reduced gel and at two contrast levels on a non-reduced gel. The reversible covalent bonds are indicated as E1~Ub (hUba1), E2~Ub (Ube2L3) and E3~Ub (HOIP\textsuperscript{RBR-LDD}). (C) Formation of a reversible covalent intermediate between HOIP\textsuperscript{RBR-LDD} and TAMRA-ubiquitin with RING2 mutants in the presence of Ube2D3. The reversible covalent bonds are indicated as E1~Ub (hUba1), E2~Ub (Ube2D3) and E3~Ub (HOIP\textsuperscript{RBR-LDD}). (D) Multi-sequence alignment of the C-terminus of human RBR protein sequences, including the RBR domain. UniprotID: HOIP (Q96EP0), HOIL-1L (Q9BYM8), ARIH1 (Q9Y4X5), Parkin (O60260), Triad1 (O95376). RING domains are underlined with the black line, the IBR is underlined with the dotted line and the LDD of HOIP is indicated by the gray line. Asterisks show the positions of the mutations in HOIP\textsuperscript{RBR-LDD} that have been introduced in this study. (E) Single-cycle turnover assay with the E2 Ube2L3, corresponds to the assay with Ube2D3 in Figure 4C (main text). Assay shows Ube2L3~TAMRA-ubiquitin discharge by HOIP mutants (left half of each gel) and di-ubiquitin formation upon the addition ubiquitinΔGly76 (right half of each gel). Discharge reactions were stopped after 0, 2, 4, 8, 16, 32, 64 min.
Figure S5. (A) FP-assay for TAMRA-ubiquitin binding to HOIP<sub>LDD</sub>, as a control for the ubiquitin<sub>TAMRA</sub> binding to HOIP in figure 5A. The increase in FP is shown as a function of [HOIP]. HOIP<sub>LDD</sub> KD = 170 ± 21 µM, HOIP<sub>LDD</sub> C930A KD = 1218 ± 1437 µM. Standard deviations were calculated over three repeats. (B) FP-assay for free TAMRA-dye binding to HOIP<sub>RBR-LDD</sub> and HOIP<sub>LDD</sub>. The increase in FP is shown as a function of [HOIP]. (C) Formation of a reversible covalent intermediate between HOIP<sub>RBR-LDD</sub> and TAMRA-ubiquitin. Wild type HOIP<sub>RBR-LDD</sub> and HOIP<sub>RBR-LDD</sub> C930A were tested in the absence and presence of biotin-ubiquitin. (D) Single-cycle turnover assay with TAMRA-ubiquitin. The formation of di-ubiquitin between the TAMRA-ubiquitin and different ubiquitin mutants was followed over time. Reactions were stopped after 0, 5 and 15 min at 37°C.
The RBR and LDD domains of HOIP specify linear ubiquitin chain assembly

Figure S6. (A) Co-immunoprecipitation of full-length HOIP, HOIP mutants and HOIL-1L. Western blots are shown for the immunoprecipitated protein (IP) and whole cell extracts (WCE). (B) Activity of HOIP and HOIP mutants in the absence of HOIL-1L in a Dual Luciferase™ reporter assay for NF-kB activation (control conditions for Figure 6). Full-length HOIL-1L, HOIP or HOIP cysteine mutants (see lower panel) were expressed separately in the presence of a luciferase reporter construct, containing 5 NF-kB binding sites. A luciferase renilla construct was used as transfection control. Firefly luciferase values were normalized to renilla luciferase values. Normalized luciferase activity of 5x NF-kB reporter vector (upper panel) is shown as mean ± s.e.m. (*P < 0.001, Student’s t-test, representative experiment of n=4).
Discussion
Discussion

Thesis objective

Transcriptional regulation is fundamental to the development of functional blood cells. Acquired and inherited changes in transcription factors and their up-stream regulators cause hematopoietic diseases ranging from bleeding disorders to leukemia. These alterations disrupt the control over gene expression resulting in the generation of pathogenic transcriptional programs that compromise proper blood cell development and function. Because aberrant gene expression contributes to disease formation, a better understanding of transcriptional regulation may provide important information for the design of more successful therapies. In this light, the objective of this thesis was addressed to broaden the knowledge on the pathways and molecular mechanisms that regulate gene expression and the consequences of their alteration on hematopoiesis. A part of our studies was dedicated to the identification and characterization of a dominant-negative mutation in the transcription factor GFI1B that appeared to be causal to the gray platelet syndrome, a rare bleeding disorder of the megakaryocytic lineage. Chapter 2 describes the deleterious effects of this mutation on the development of megakaryocytes, defining GFI1B as a critical regulator of platelet production. Another part of this work focused on translational aspects of gene expression analysis in acute myeloid leukemia (AML). By correlating gene expression data with clinical characteristics of patients, we defined novel AML subclasses based on the expression levels of the BRE, EVI1 and DNMT3B genes, thereby contributing to the refinement of AML patient stratification (chapters 3 and 4). Finally, the last part was aimed to investigate the role of ubiquitination in the control of NF-κB signaling, whose altered activity has been associated with the pathogenesis of blood diseases. In chapter 5, we performed a functional characterization of the E3-ligase LUBAC complex, whose exclusive ability to generate linear ubiquitin chains is required to fully activate the NF-κB pathway. Our results illustrate the identification of an essential domain involved in LUBAC mediated-linear ubiquitination and reveal precious insight in the biochemical mechanisms that dictate the formation of linear ubiquitin chains.

GFI1B is a master transcriptional regulator of megakaryopoiesis

Our incomplete understanding of thrombopoiesis hampers diagnosis and treatment of patients with thrombocytopenia, and the ex vivo generation of thrombocytes for transfusion. Recently, we observed that the transcription factor Growth Factor Independence 1B (GFI1B) plays a broad role in the proliferation and differentiation of megakaryocytes and the formation of platelets. In a family with an inherited autosomal dominant bleeding disorder, we observed a heterozygous truncating mutation in GFI1B. Affected family members showed increased proliferation of megakaryocytes, severe morphological megakaryocyte abnormalities, abnormal expression of the stem cell marker CD34 and the membrane receptor glycoprotein 1ba–CD42B on megakaryocytes, and their dislocation in the bone marrow. In addition, platelet numbers were reduced, platelets were increased in size and, importantly, lacked α-granules. As α-granules are main platelet constituents contributing to wound healing, their paucity may contribute to the bleeding episodes
observed in affected individuals. The \textit{GFI1B} mutation completely co-segregated with the disease in the analyzed family and predicted a C-terminally truncated protein lacking crucial DNA-binding sequences located in zinc finger 5. In a luciferase gene reporter assay, truncated GFI1B was not solely unable to repress gene transcription, but also inhibited the activity of the wild-type protein in a dominant-negative manner. Expression of the dominant-negative GFI1B mutant in normal hematopoietic stem cells, followed by megakaryocyte differentiation recapitulated MK defects as observed in patients. This confirmed that the heterozygous \textit{GFI1B} mutation does not cause a loss of function, but it functions in a dominant-negative manner. Together, these data define a broad role for the transcription factor GFI1B in thrombopoiesis as it controls the expansion of megakaryocytes, the production of α-granules and the shedding of platelets.

**Mode of action of GFI1B mutant**

Conditional deficiency or silencing of \textit{GFI1B} expression causes megakaryocyte abnormalities that are clearly distinct from those caused by truncated GFI1B\textsuperscript{2–4}, indicating that the dominant-negative mutant is able to install an aberrant transcriptional program that drives the pathogenesis of the disease. The molecular mechanism through which truncated GFI1B interferes with the wild type protein is unclear. Although sequences essential for the binding to DNA are lost, the domains responsible for the interaction with co-repressor molecules are still intact. Thus, non-DNA bound GFI1B mutant may act in a dominant-negative manner by quenching essential factors that are normally used by the wild-type protein, resulting in aberrant expression of GFI1B target genes. In addition, truncated GFI1B may titrate away other transcription factors thereby deregulating the expression of genes normally not regulated by GFI1B. Mutant GFI1B also lacks zinc finger 6. This region has been shown to be responsible for the proteasomal degradation in GFI1\textsuperscript{5}, the closely related paralogue of GFI1B. Whether ubiquitination regulates the stability of GFI1B during erythrocytic and megakaryocytic differentiation is unknown and deserves further investigation. Nevertheless, because the zinc finger domain of GFI1B is highly homologous (97%) to GFI1, it is conceivable that GFI1B protein levels are regulated in a similar manner. The specific zinc fingers implicated in the proteasomal targeting of GFI1 proteins have not been mapped. If zinc finger 6 is involved, it is tempting to speculate that its absence may counteract proteasomal degradation, resulting in a more stable truncated protein. Increased stability of the mutant may enhance its potential quenching effect, exacerbating the competition for the wild-type protein to interact with co-repressor molecules. Importantly, during the preparation of this thesis, six additional \textit{GFI1B} variants in cases with thrombocytopenia were reported\textsuperscript{6}. Interestingly, as these alterations are spread across the N-terminal region outside the DNA-binding domain, it will be interesting to determine whether these variants cause similar defects as those caused by the \textit{GFI1B} mutation reported by us. A detailed molecular and phenotypic characterization of all the reported \textit{GFI1B} variants is required to identify through which molecular mechanisms GFI1B mutants interfere with normal megakaryopoiesis and shed more light on the still elusive mode of action of GFI proteins.
GFI1B mutants: promising tools to zoom in on megakaryopoiesis

Mutant GFI1B represents a precious tool to decipher the transcriptional programs that drive thrombopoiesis. Gene expression programs installed by GFI1B are orchestrated in a hierarchical order, with genes early and later during differentiation that, in turn, may have impact of downstream changes in gene expression. Chromatin immunoprecipitation (CHIP) studies are required to map the genes that are directly bound by GFI1B at different stages of megakaryocyte development. By correlating gene expression and CHIP-profiles induced by normal and mutant GFI1B, it would be possible to identify those genes whose differential expression may contribute to the GPS and how these genes would normally function. These studies should be definitely supported by proteomic analysis as changes in RNA expression do not necessarily imply changes in protein expression. Depending on differences in gene expression, the expression of the identified genes may be corrected in rescue experiments, allowing the identification of players relevant for megakaryocyte proliferation, nuclear organization, α-granules biogenesis and platelet shedding.

In addition to define the function of genes implicated in megakaryopoiesis and thrombopoiesis, mutant GFI1B can also be used to determine which epigenetic marks are applied by GFI1B to repress transcription, which protein complexes are associated to GFI1B at distinct stages of megakaryopoiesis, their relation to gene expression and for which process these interactions are required. For example, it would be particularly interesting to determine whether megakaryocytic abnormalities caused by GFI1B depend on the interaction with the histone H3K4/K9 demethylase LSD1. This co-repressor is required for proper megakaryopoiesis and is recruited through a conserved proline in the SNAG domain of GFI1B. An alanine substitution at this position abrogates this interaction and inhibits the repressive function of GFI1B. The introduction of this mutation in dominant-negative GFI1B, followed by expression of this “double-mutant” in megakaryocytic cells, should reveal whether quenching of LSD1 contributes to disease pathogenesis and potentially uncover specific functions for this protein during megakaryocyte development.

In conclusion, the discovery of dominant-negative GFI1B offers an excellent opportunity to shed light on functional and molecular aspects of megakaryocyte development and platelet production. This increased knowledge would provide a better understanding of thrombopoiesis ultimately contributing to a better classification of platelet disorders, and more efficient diagnostic and treatment strategies, including the generation of thrombocytes for transfusion.

High **DNMT3B** expression levels predicts poor outcome in AML

The DNA methyltransferases DNMT1, DNMT3A and DNMT3B are epigenetic regulators that regulate gene expression through methylation of specific cytosines in the DNA. Because aberrant methylation profiles have been consistently found in patients with hematological malignancies, pharmacological targeting of these enzymes with hypomethylating agents is currently being evaluated for therapeutic efficacy. Notably, mutations in **DNMT3A** occur in 25-30% of AML patients and predict poor prognosis, in particular within the cytogenetically
normal, intermediate risk group\textsuperscript{12}. In contrast to \textit{DNMT3A}, genetic variations in \textit{DNMT1} and \textit{DNMT3B} have not been reported in leukemia as yet. Instead, a recent study associated high \textit{DNMT3B} expression with an inferior outcome in AML\textsuperscript{13}. However, the patient cohort used for this study was cytogenetically and molecularly not well characterized and included only 195 de novo cases. In the work described in chapter 3, we examined the prognostic significance of \textit{DNMT3B} expression in the context of several established molecular prognostic factors in a well-characterized cohort of 525 AML patients. Our results confirmed the association of high \textit{DNMT3B} expression (corresponding to the top 25\% highest \textit{DNMT3B}-expressing patients) with an inferior disease outcome. Importantly, the adverse effect on survival was particularly evident in cases with normal karyotype, especially among those with co-occurring \textit{NPM1} and \textit{FLT3-ITD} mutations. Within this subgroup, survival is known to be negatively influenced by the presence of a high \textit{FLT3-ITD} allelic burden\textsuperscript{14,15}, a factor which is clinically used for risk stratification. Observed both in patients with high and low \textit{FLT3-ITD} allelic burdens, high \textit{DNMT3B} expression predicted poor survival especially among patients with high allelic burden. This allowed us to further separate this group into two subgroups: one with a survival similar to that of the established intermediate risk group and one with an extremely poor prognosis. In combination with the fact that high \textit{DNMT3B} expression displayed a strong independent prognostic value in multivariate analyses, these data therefore defined high \textit{DNMT3B} expression as a poor prognostic factor in AML, and in particular among patients with normal karyotype and co-occurring \textit{NPM1} and \textit{FLT3-ITD} mutations. Thus, \textit{DNMT3B} expression may be used to refine risk stratification. However, as the fold-difference between high and low \textit{DNMT3B} expression is small, the definition of a cut-off to identify patients with high \textit{DNMT3B} expression may prove to be challenging with routinely available techniques.

High \textit{DNMT3B} expression: innocent by-stander or more?

Very recently, an independent study by Niederwieser et al. corroborated our results, also showing a significant adverse effect of high \textit{DNMT3B} expression on survival of patients with normal karyotype older than 60 years of age\textsuperscript{16}. Excluding our work, the adverse prognostic significance of high \textit{DNMT3B} expression has been demonstrated by two other independent studies\textsuperscript{13,16}, putting forth that high \textit{DNMT3B} expression may predict poor response to current therapeutic regimens. Methylation studies have shown that global DNA methylation levels in leukemic blasts from AML cells with normal karyotype are not dependent on changes in \textit{DNMT3B} expression, as inactive \textit{DNMT3B} isoforms are primarily expressed in these cells\textsuperscript{17}. In addition, a recent report has reported that changes in \textit{DNMT3B} expression do not perturb the methylation status of putative \textit{DNMT3B} target genes\textsuperscript{18}. Although these findings suggest that patients with high \textit{DNMT3B} expression may not benefit from hypomethylating therapies, it will be anyway interesting to determine whether \textit{DNMT3B} levels influence the sensitivity of AML patient to such treatments. Further investigations are also necessary to establish the exact mechanisms underlying high \textit{DNMT3B} expression in AML and its effect on disease outcome. Importantly, high \textit{DNMT3B} expression was not observed in all of the AML subclasses, but only in specific subgroups, suggesting the existence of functional associations between high \textit{DNMT3B} expression and specific molecular aberrations\textsuperscript{16}. 
Recent findings have associated the presence of specific gene polymorphisms in the *DNMT3B* gene locus with the predisposition for AML development\(^\text{19}\). As polymorphisms can influence the enzymatic activity of the protein, this suggests that DNMT3B functions need to be tightly regulated to prevent increased susceptibility to malignant transformation. However, conditional *DNMT3B* ablation does not cause leukemia in mice\(^\text{20}\), but only minor hematopoietic defects, suggesting that changes in *DNMT3B* expression may not be directly implicated in malignant transformation, but rather co-operates with other oncogenes. In this respect, ectopic over-expression in mouse models would be helpful to gather precious insights on the contribution of increased *DNMT3B* expression to leukemogenesis. As current data are insufficient to exclude the possibility that high *DNMT3B* expression is an innocent by-stander or a surrogate, additional investigations are therefore granted to understand the cryptic pathological role of *DNMT3B* and whether it directly affects treatment response in AML.

**AML patients with *MLL-AF9* rearrangements can be sub-divided based on *BRE* and *EVI1* expression**

The Mixed Lineage Leukemia (*MLL*) gene, encoding for a H3K4-methyltransferase, is recurrently involved in translocations with a large variety of fusion partners, which are found in about 10% of patients with acute myeloid leukemia\(^\text{21}\). In particular, the chromosomal rearrangement with the *AF9* gene, resulting in the *MLL-AF9* fusion gene, is associated with a five-year overall survival (OS) rate of approximately 40%\(^\text{21–23}\). Recently, we refined the prognosis of this intermediate risk group based on the mRNA expression levels of *BRE*\(^\text{24}\), a gene involved in DNA-damage response. High expression of *BRE* predicted a superior survival while normal expression levels predicted a very poor outcome in two independent cohorts (5-year overall survival of 80% vs 0%, respectively), defining high *BRE* expression as a favorable prognostic marker for patients with the *MLL-AF9* translocation. As roughly 50% of *MLL-AF9* positive patients displayed normal expression, high *BRE* expression associated only partially to this cytogenetic abnormality. Chapter 4 illustrates that virtually all *MLL-AF9* patients without high *BRE* expression show high expression of *EVI1*. High expression of the latter gene strongly associates with an inferior disease outcome\(^\text{25}\). As high expression levels of *BRE* and *EVI1* predicted opposite disease outcomes, their mutually exclusive expression explains the poor prognosis of patients lacking high *BRE* expression. Thus, *BRE* expression may be used for refined risk stratification of *MLL-AF9* positive patients. However, due to the small difference between normal and high *BRE* expression, it may be challenging to create diagnostic tools that are sensitive enough to allow the identification of high *BRE* expressing patients. In contrast, as screenings for *EVI1* overexpression are currently feasible using conventional q-PCR techniques\(^\text{26}\), this could facilitate their potential incorporation into diagnostic protocols for *MLL-AF9* patients. Importantly, as many of the *MLL-AF9* positive patients with high *EVI1* expression do not harbor chromosomal rearrangements encompassing the *EVI1* locus on chromosome 3\(^\text{25}\), a molecular screening for *EVI1* over-expression would be therefore strongly recommended to improve *MLL-AF9* patient risk stratification.
High BRE expression defines a distinct group of patients with a unique gene expression signature

In unsupervised genome-wide expression clustering analysis, a large fraction (85%) of patients with high BRE expression displayed highly similar expression profiles and commonly associated with a specific expression signature, suggesting that these patients share a common mechanism that govern gene expression\textsuperscript{24}. Consistently, the BRE-associated gene expression clustered separately also within MLL-AF9 positive patients, completely separated from that of the remaining EVI1 positive MLL-AF9 patients. Interestingly, the latter presented only subtle similarities in gene expression profiles; suggesting that, in contrast to BRE positive patients, transcription programs are not commonly dys-regulated in EVI1 positive patients. In combination with the fact that high EVI1 expression was not found in all patients, despite being reported as a direct target of the MLL-AF9 fusion protein\textsuperscript{27}, these findings suggest that global changes in gene expression are not solely attributable to the activity of the fusion protein. Possibly, they depend on the molecular context or the differentiation status of the transformed cells. Chromatin immunoprecipitation experiments using patient derived material would be helpful to understand which genes are direct targets of MLL-AF9 and which co-factors are used by the fusion protein. Also epigenetic profiling of chromatin modifications may reveal important insight in the molecular mechanisms underlying dys-regulated gene expression by MLL-AF9. Interestingly, recent data have shown that MLL-AF9 positive patients display aberrant methylation marks at lysine 79 on histone H3 (H3K79)\textsuperscript{28}, an epigenetic modification that is associated with gene activation. Therefore, it would be interesting to determine whether, and to which extent, aberrant H3K79 methylation is implicated in the dys-regulated gene expression profiles that we observed in MLL-AF9 patients. The methyltransferase DOT-1L is the only enzyme known to catalyze H3K79 methylation and it has been shown to interact with the MLL-AF9 fusion\textsuperscript{29}. Recruitment of DOT1L to the MLL-AF9 fusion protein causes chromatin hypermethylation and induces aberrant transcription of MLL-AF9 target genes, thereby promoting leukemic transformation\textsuperscript{30–32}. DOT1L inhibitors, which are currently under clinical investigation, could be helpful to study the contribution of DOT1L to the de-regulated activity of the fusion protein more specifically. Nevertheless, how the expression of the genes clustering along with high BRE expression in MLL-AF9 patients is similarly de-regulated remains unclear. Since BRE positive patients partially co-occurred with the MLL-AF9 translocation in the total AML cohort, it cannot be excluded that other (epi) genetic regulators, like for instance transcription factors, other than MLL-AF9, are exclusively responsible for the highly similar gene expression profiles observed in high BRE expressing patients. In conclusion, our data suggest that MLL-AF9 is likely to participate in malignant transformation through different molecular mechanisms, therefore warranting additional studies to determine how the malignant activity of the MLL-AF9 onco-protein is modulated.
HOIP governs linear ubiquitin specificity of LUBAC

The biological relevance of LUBAC is emerging as it has recently been implicated in the regulation of other cellular processes beyond NF-kB-mediated transcriptional regulation\textsuperscript{33,34}. For most of these pathways, it remains to be defined which substrates are recognized, whether they are targeted for linear ubiquitination and its molecular implications. Nevertheless, because of its exclusive ability to generate linear polyubiquitin chains, LUBAC represents a potential target for the development of targeted therapies for blood-related disorders that rely on the dys-regulated activity of the NF-kB signaling. Therefore, understanding how LUBAC dictates the formation of linear chains, the specific functions of the various domains, and how these are regulated could be very beneficial for the development of treatments capable to modulate LUBAC activity.

The R2-LDD region of HOIP is required for linear ubiquitination

The specificity of LUBAC for linear ubiquitination activity is embedded within HOIP, the catalytic subunit of the complex. This E3 ligase belongs to the family of RBR ligases which is characterized by a conserved tri-partite region, called RBR domain, comprising a RING1, an in-between RING (IBR) and an atypical RING2 domain\textsuperscript{35,36}. As shown for other RBR ligases, like Parkin and HHARI\textsuperscript{37}, these domains act in concert to mediate the transfer of ubiquitin from E2 onto its target using a hybrid RING/HECT like mechanism that involves the formation of a thioester intermediate between the C-terminus of the donor ubiquitin and an active-site cysteine available on RING2. In line with recently published studies\textsuperscript{38}, we were able to trap an intermediate transient interaction between ubiquitin and a catalytically active HOIP mutant, denominated HOIP RBR-LDD, which lacks the amino terminal region upstream of the RBR domain, proving that linear chains are also assembled through a two-steps mechanism. As described in chapter 5, the catalytic mechanism of HOIP involves the transfer of ubiquitin to form a covalent intermediate with RING2. This domain is essential for the RBR ligase activity of HOIP because it serves as acceptor for the transient linkage of the donor ubiquitin, and contains the recently identified residues responsible for its transfer onto the acceptor ubiquitin. Despite being similar to RING1, RING2 does not conform to the canonical RING structure. As observed in parallel studies\textsuperscript{37–40}, RING2 of HOIP does not coordinate two zinc ions, but only one, leaving a cysteine residue, which is also conserved in the ubiquitin ligase HHARI, available for the formation of the thioester intermediate with the donor ubiquitin. The reaction initiates with the recruitment of the ubiquitin charged-E2. In HOIP, the RING1-IBR region is involved in this process since mutations or deletions of these domains disrupt the E2-dependent chain-forming capacity of the protein. RING1 is likely responsible for the interaction with E2s, as it closely resembles classical RING E3 ligases which are known to allosterically activate the ligase activity of E2s to build ubiquitin chains\textsuperscript{41}. However, RING1 does not discharge ubiquitin from the E2 enzyme in absence of an active site cysteine on RING2. This suggests that, unlike classical RING enzymes, RING1 does not enhance the reactivity of E2s. Thus, the generation of the covalent intermediate within RING2 may take place via a different mechanism than that classically described for
RING ligases, warranting additional studies in this direction. Also the IBR domain seems to be involved in this process. Because this domain lacks ubiquitination activity and the active site cysteine necessary for the intermediate transfer of ubiquitin, it was originally thought to function as a mere linker between RING1 and RING2. However, the crystal structures of the RBR ligase Parkin revealed that this domain interacts with multiple sites of RING1\textsuperscript{42}, possibly contributing to regulate the accessibility and the extension of the docking surface for ubiquitin-charged E2s. Furthermore, it has recently been shown that IBR of Parkin adopts the same folding as that observed for RING2\textsuperscript{43}. Whether this implies a functional involvement of IBR in the catalytic activity is unclear. Nevertheless, as its high conservation suggests that the IBR may be more than a structural requirement of the RBR domain\textsuperscript{42}, further investigations are obviously needed to decipher the enigmatic function of this region.

**HOIP acts as atypical RBR ligase in linear ubiquitination**

LUBAC selectively produces linear polyubiquitin chains regardless of the type of E2s that is engaged, indicating that the E2 does not contribute to the chain formation specificity of HOIP. This is achieved through the overruling activity of the LDD domain, which is located at the carboxyl-terminal end of the protein and positions the amino-terminus of the acceptor ubiquitin facing the thioester ubiquitin intermediate in RING2. The recent structural resolution of RING2-LDD in complex with ubiquitin\textsuperscript{40} confirmed our results and revealed that LDD co-operates with RING2 to guarantee the accurate orientation between the two reactive groups. In addition, these studies identified a previously unpredicted zinc finger, adjacent to the end of RING2, which interacts with the acceptor ubiquitin. Despite the fact that this may indicate that RING2 forms a functional unit with LDD that determines the specificity for linear chains, it will be interesting to determine whether the particular positioning of the acceptor ubiquitin is sufficient to overrule the specificity intrinsic to E2s or other mechanisms are involved. Nevertheless, both this newly identified zinc finger and the LDD domain are not found in other RBR ligases, likely explaining the unique ability of HOIP to assemble linear polyubiquitin chains.

In contrast to HOIP, the specificity for chain assembly of the majority of RBR ligases as yet characterized is determined by the E2 involved in the reaction. Parkin and Triad1, for example, have been shown to generate different ubiquitin signals by co-operating with a variety of conjugating enzymes\textsuperscript{35,37,44}. Given the HECT-like mechanism used by this class of enzymes, the dependence on the recruitment of the E2 to determine the chain specificity is controversial. Possibly, the interaction with E2 enzymes may provide an additional level of regulation to the RBR mechanism or it may indicate that the E2 actively takes part to the reaction. Consistently, it has been shown for some RBR ligases that also RING2 is capable to bind to E2s\textsuperscript{44–46}, despite the functional significance of this interaction remains to be evaluated. Although enormous advances in the understanding of the molecular mechanisms underlying the functioning of RBR ligases have been made, important aspects of catalysis as mentioned above necessitate further elucidation.
Auto-inhibition regulates the functional state of HOIP

An interesting observation derived from our study is that the removal of the N-terminus from HOIP unleashes the ubiquitination activity of the protein. The addition of HOIL-1 did not increase the efficiency of the reaction mediated by our constitutively active HOIP mutant, indicating that the activity of HOIP is auto-inhibited by the domains present within its N-terminus. Consistently, the UBA- and NZF2 domains, which are implicated in the activation of HOIP through the interaction to the UBL-domain of HOIL-1L and Sharpin, respectively, are located in this region. Whether these domains inhibit HOIP by direct interaction with its RBR domain needs to be verified. Interestingly, the N-terminal region also did not activate HOIP in trans in our assays, indicating that the direct binding is not sufficient for activation and conformational changes may possibly be involved in the release of HOIP activity. With respect to these assumptions, the crystal structure of HOIP in its inactive state will definitely provide important information about the mechanism governing auto-inhibition. This auto-regulatory mechanism is also a typical feature of Parkin and HHARI, although HHARI is inhibited by the C-terminal part of the protein. This reflects the structural heterogeneity observed outside the RBR domains in these enzymes, indicating that different mechanisms exist to activate the catalytic center. Because RBR ligases have been shown to interact with several partners and undergo a variety of post-translational modifications, it is conceivable that their activation and activity are regulated at multiple levels. Additional efforts are therefore required to identify co-regulatory molecules that interact with LUBAC and characterize their functional contribution to the linear chain formation capacity.

References


ADDENDUM

Summary

Nederlandse samenvatting

Curriculum vitae

List of publications

Dankwoord / Acknowledgments
Summary
Vital activities required for the integrity and functionality of our body, such as oxygen transport, wound healing or protection against harmful pathogens, are performed by highly specialized cells present in our blood. These cellular blood components amount to more than 10 different cell types, each of them with a specific function. Leukocytes, for example, encompass several specialized cell types involved in innate and acquired immunity. Megakaryocytes produce platelets for blood clotting and wound healing, while erythrocytes are crucially implicated in $O_2$ and $CO_2$ transport. As many of these cell types are short-lived, the blood needs to be one of the most regenerative and plastic tissues in our body, with an enormous number of cells being constantly replaced with newly formed ones. The process of blood cell formation is called hematopoiesis and takes place, under normal circumstances and in the adult, within the bone marrow (BM), which consists of a soft spongy tissue filling the cavities of bones. Under the influence of the BM microenvironment, all blood cells originate from hematopoietic stem cells (HSCs) through a series of developmental stages that generally culminate with the release of mature cells into the circulation. HSCs represent a unique pool of long-lived quiescent cells endowed with multipotency and self-renewal capacity. Thanks to these essential stem-cell characteristics, HSCs are meant to preserve a constant pool of undifferentiated cells, capable of generating all the different types of mature blood cells, throughout the lifetime of the organism. Control over the balance between self-renewal and differentiation as well as over the various transitions through consecutive differentiation stages is required for proper blood-cell development. Regulation of gene expression plays an essential role in this process. During hematopoiesis, specific sets of genes are differentially regulated to generate cell type-specific gene expression patterns that are characteristic of each stage of differentiation. These transcriptional programs coordinate cellular functions and can guide the progression of hematopoietic cells through subsequent developmental phases. Acquired and inherited changes in factors involved in gene transcription, or their upstream regulators, can cause hematological disorders ranging from bleeding disorders to leukemia. These alterations disrupt the genome-wide control of gene expression resulting in the generation of pathogenic transcriptional programs that contribute to disease onset. In addition, these altered expression patterns significantly influence the biological characteristics of the various diseases that they generate and thereby the severity of their clinical outcome. Because aberrant gene expression contributes to the pathogenesis of hematological diseases, a better understanding of how gene transcription is regulated during normal and abnormal hematopoiesis would be beneficial for the advancement of disease treatments. The scope of this thesis is to broaden our knowledge on the implications of dysregulated gene expression on the development of blood disorders, with a particular focus on the identification of dysregulated pathways that can serve as potential targets for the improvement of therapies.

The conversion, or transcription, of the genetic information contained in the DNA sequences of genes into messenger RNA is an essential step for the expression of genes. Transcription factors play a crucial role in this process as they determine whether a gene is transcribed or silenced. Genetic variations in these regulatory factors are often implicated
in the etiology of blood disorders. In chapter 2, we describe the identification of a novel genetic alteration in the transcription factor GFI1B among patients affected by a rare bleeding disorder called the gray platelet syndrome (GPS). This disease encompasses a broad spectrum of clinical conditions characterized by low platelet counts and by platelets appearing gray due to a shortage of α-granules. We investigated a large family with an autosomal dominant form of GPS, the affected members of which showed an increased number of dysplastic megakaryocytes (MKs) that were abnormally located in clusters along the walls of the BM sinusoids. In addition, we also observed a reduced surface membrane expression of platelet glycoprotein Ib α-chain (CD42b) and an increased expression of the blood stem-cell marker CD34, in both MKs and platelets of affected individuals. Genetic linkage analysis and targeted sequencing identified a nonsense mutation in the GFI1B gene locus which completely co-segregated with the disease in the analyzed family. Using a gene repression reporter assay, mutant GFI1B was shown to inhibit wild-type GFI1B in a dominant negative manner. In addition, ectopic expression of mutant GFI1B in murine HSCs followed by ex vivo megakaryocytic differentiation recapitulated some of the disease characteristics, demonstrating that the identified GFI1B mutation is causal to the presented form of GPS. Taken together, these findings reveal a key role for GFI1B in MK cell growth and development, synthesis of α-granules, and platelet formation. These studies open new avenues to identify GFI1B-driven transcriptional programs, which will contribute to enlarge our understanding of megakaryopoiesis, platelet production and bleeding disorders.

Dysregulated gene expression is also one of the hallmarks of acute myeloid leukemia (AML). This is a clinically heterogeneous disease characterized by a premature block in the differentiation of one or multiple myeloid lineages. In AML, malignant cells also gain a proliferative advantage that leads to their accumulation in the peripheral blood and BM, where they interfere with the formation of normal blood cells. The five-year overall survival for AML patients younger than 65 years of age is only 40% and it declines with age. Underlying genetic defects are frequently observed and often involve signal transduction molecules, transcription factors and epigenetic regulators. Because these molecular alterations contribute to the pathogenesis of the disease, they represent potential targets for the development of targeted therapies. In addition, some of these aberrations have also been shown to correlate with clinical outcome, allowing the stratification of AML patients into risk-based prognostic groups. Despite the development of individualized risk-adapted treatment strategies, survival within well-defined sub-groups fluctuates widely, indicating that additional prognostic markers are required to ameliorate the prediction of disease outcome. Chapters 3 and 4 illustrate the potential of gene expression analysis in the identification of genes and pathways that are dysregulated in the various subcategories of AML and their implications on the refinement of patient stratification. Using a publicly available gene expression data set, chapter 3 investigates the correlation between survival of patients and expression level of the DNA methyl-transferase 3B (DNMT3B) gene. Together with its parologue DNMT3A, which is frequently mutated in AML, these genes are crucially involved in the silencing of gene expression through the methylation of specific DNA sequences. As no genetic changes in DNMT3B have been reported in AML as yet,
the contribution of this gene to the disease remains elusive. In line with independent studies, our data showed that higher expression of DNMT3B is a very strong independent predictive factor for poor disease outcome in AML in general, and especially in cases with normal karyotype and co-occurring NPM1 and FLT3-ITD mutations. Importantly, the prognosis of this sub-category is strongly influenced by the FLT3-ITD mutational burden, which usually associates with a poor survival. Based on DNMT3B expression level, the latter sub-group could be further separated into an intermediate risk group and an extremely poor risk group, further indicating that DNMT3B expression can be used to refine patient stratification. From the clinical diagnostic testing point of view, however, the fold-difference between high and low DNMT3B expression is small, thereby rendering the definition of a discriminative cut-off for the identification of high DNMT3B expressing patients rather challenging. Although our data support the prognostic relevance of high DNMT3B expression in AML, additional studies are required to determine whether high DNMT3B expression is just an innocent bystander or is implicated in the pathogenesis of the disease. Chapter 4 describes the effort to improve the classification of AML patients carrying the MLL-AF9 translocation, who are currently classified as intermediate risk, with a five-year overall survival rate of roughly 40%. The prognosis of MLL-AF9-positive patients has been refined recently on the basis of the expression levels of BRE, a gene involved in DNA-damage response. High expression of BRE is associated with a superior survival while normal expression levels predict a very poor disease outcome. We revealed that the poor survival of patients with normal BRE expression can be associated with the high expression of EVI1, which is generally considered as an adverse prognostic factor in AML. In line with the fact that high expression of BRE and EVI1 were almost completely mutually exclusive, MLL-AF9-positive patients could be subdivided into a very good risk group and a very poor risk group based on the expression of these genes, respectively. However, as the fold-difference in BRE expression among the patients is limited, the identification of an expression cut-off to discriminate patients with high BRE expression using routinely available techniques may be challenging. Notwithstanding, as screening for EVI1 overexpression is feasible with current diagnostic techniques, it could be incorporated into diagnostic protocols for MLL-AF9 patients. Interestingly, despite the fact that EVI1 has been described to be a direct target of the MLL-AF9 fusion protein, patients with high EVI1 expression did not exhibit a common transcriptional profile in our study. In combination with the fact that MLL-AF9 cases with high BRE expression have been previously shown to constitute a new AML subclass in terms of global gene expression profile, which co-occurred only partially with MLL-AF9 positivity, these findings suggest that dysregulated gene expression in MLL-AF9-positive patients also depends on other mechanisms than only those installed by the MLL-AF9 translocation itself. Further investigations are therefore warranted to identify the underlying pathways that co-operate with MLL-AF9 in malignant transformation.

Various hematopoietic disorders are characterized by altered activity of the NF-kB transcription factors, which are crucial regulators of a large number of genes implicated in cell differentiation, proliferation and apoptosis. LUBAC, a multi-domain protein complex with ubiquitin ligase activity, is a critical component of the NF-kB pathway. Consisting of
two regulatory subunits, HOIL-1L and SHARPIN, and the catalytic subunit HOIP, LUBAC is the only protein complex that has been reported thus far to generate linear poly-ubiquitin chains, which work as signaling molecules for the activation of the pathway. Because of this exclusive ability, LUBAC may represent a potential target for the development of targeted therapies for blood disorders that rely on dysregulated activity of NF-κB signaling. Therefore, understanding how LUBAC dictates the formation of said linear chains, the specific functions of its various domains, and how these are regulated would be very helpful for the development of compounds capable of modulating functions of LUBAC, and possibly improving current treatments. Chapter 5 sheds light on important aspects of the biochemical mechanism that dictates the catalysis of linear poly-ubiquitin chains. In that section, we showed that both the catalytic activity and the specificity for linear chains of LUBAC reside within the ubiquitin ligase HOIP. This subunit is characterized by a conserved tripartite region, called RBR domain, comprising a RING1, an in-between RING (IBR) and an atypical RING2. According to our model, the formation of linear chains involves a two-step mechanism in which the RING1 and the IBR domains are responsible for the formation of a covalent thioester intermediate of ubiquitin with RING2. Subsequently, RING2 catalyzes the linkage between the C-terminus of the donor ubiquitin with the N-terminus of an acceptor ubiquitin. The formation of the linear” linkage is promoted by a unique C-terminally located region, exterior to the RBR domain and called the “Linear ubiquitin chain Determining Domain” (LDD). This area was shown to interact with the acceptor ubiquitin, suggesting that the LDD domain could engage the target ubiquitin to position it in a specific orientation that facilitates the “linear” linkage. Interestingly, we also found that the catalytic activity of HOIP is auto-inhibited by its own N-terminal region, which is required for the interaction with HOIL-1L and SHARPIN. Binding of HOIP to its regulatory subunits is required to release the auto-inhibition and to induce the catalytic activity of the RBR domain.
**Nederlandse samenvatting**


De veranderde genexpressiepatronen bepalen in belangrijke mate de ernst van de ziekte en de respons op therapie. Omdat afwijkende genexpressie bijdraagt aan de pathogenese van hematologische aandoeningen, is een beter begrip van genexpressieregulatie tijdens normale en abnormale bloedcelaanmaak belangrijk voor de verbetering van de behandeling van de ziektes. Dit proefschrift richt zich op het verbeteren van onze kennis over verstoorde regulatie van genexpressie bij het ontstaan van bloedziekten. De omzetting of transcriptie van de genetische informatie van genen naar messenger RNA is een essentiële stap bij het tot expressie komen van genen. Transcriptiefactoren spelen hierbij een cruciale rol omdat zij bepalen of een gen wel of niet afgelezen wordt. Genetische variaties in transcriptiefactoren spelen vaak een belangrijke rol bij het ontstaan van bloedziekten. In hoofdstuk 2 beschrijven we de identificatie van een nieuwe genetische verandering in de transcriptiefactor *GFI1B*. Deze werd gevonden bij een familie met een autosomaal dominante vorm van het grijze bloedplaatjessyndroom (GPS). GPS wordt gekenmerkt
door een verhoogde bloedingsneiging veroorzaakt door een tekort aan - en niet goed functionerende - bloedplaatjes. De bloedplaatjes zien er grijs uit onder de microscoop omdat ze een tekort aan α-granules hebben. Normaliter kleuren in bloeduitstrijkanalyses deze granules paars aan. Bij de door de GFI1B mutatie aangedane familieleden werd verder een toename gevonden in de bloedplaatjes producerende megakaryocyten. Daarnaast vertoonden de megakaryocyten dysplastische kenmerken en vertoonden ze een afwijkende locatie in het beenmerg; de megakaryocyten lagen namelijk in clusters langs de bloedvaten in het beenmerg, een fenomeen dat niet gevonden wordt in niet aangedaan beenmerg. Daarnaast vonden we bij aangedane familieleden op bloedplaatjes een verlaagde expressie van het glycoproteïne 1bα-CD42B. Verder werd op bloedplaatjes en megakaryocyten een verhoogde expressie van de bloedstamcelmarker CD34 gevonden. Met genetische linkage- en gerichte sequentie-analyse werd een nonsens-mutatie in het GFI1B gen gevonden. Deze mutatie cosegregateerde met de aandoening. Met genrepressie reporter assays werd van het gemuteerde GFI1B eiwit aangetoond dat het de functie van het normale GFI1B eiwit remt op een dominant negatieve manier. Daarnaast werd aangetoond dat expressie van het gemuteerde GFI1B eiwit in HSCs van muizen, gevolgd door megakaryocyten-differentiatie, afwijkingen veroorzaakten die ook gevonden werden in de patiëntencellen. Op basis van bovenstaande bevindingen werd een oorzaakelijk verband gelegd tussen de nonsens-mutatie in GFI1B en de autosomaal dominante vorm van GPS. Uit deze studies blijkt dus dat GFI1B een belangrijke rol speelt bij de ontwikkeling en groei van megakaryocyten, de vorming van α-granules en de vorming van bloedplaatjes. Op basis van dit werk kan nu onderzocht worden welke transcriptieprogramma’s tijdens de megakaryopoëse door GFI1B gereguleerd worden om zodoende meer inzicht te krijgen in de vorming van α-granules en bloedplaatjesproductie.

Een verstoorde regulatie van genexpressie is ook een van de kenmerken van acute myeloïde leukemie (AML). AML is een klinisch heterogene ziekte, die wordt gekenmerkt door een blokkade in de differentiatie van onrijpe myeloïde cellen in het beenmerg. Daarnaast groeien de aangedane cellen sneller, waardoor normale onrijpe bloedcellen overgroeid en verdrongen worden. Door de enorme expansie in het beenmerg komen de onrijpe kwaadaardige bloedcellen uiteindelijk ook in de bloedbaan terecht. De 5-jaars overleving voor volwassen AML patiënten jonger dan 65 jaar is slechts 40% en daalt sterk met de leeftijd. Veel van de genetische defecten die ten grondslag liggen aan het ontstaan van AML zijn bekend en worden vaak gevonden in signaaltransductie-eiwitten, transcriptiefactoren en eiwitten die een belangrijke rol spelen bij epigenetische veranderingen. Omdat de gemuteerde eiwitten een belangrijke rol spelen bij de pathogenese van de ziekte zijn ze belangrijk voor de ontwikkeling van doelgerichte therapie waarbij de rol van het gemuteerde eiwit teniet wordt gedaan. Daarnaast correleren bepaalde genetische afwijkingen sterk met de respons op therapie. Op basis hiervan worden patiënten momenteel geclassificeerd met een relatief gunstige -, intermediaire -, slechte - of zeer slechte prognose. Patiënten met een relatief goede prognose ontvangen een minder intensieve behandeling dan patiënten met een slechte tot zeer slechte prognose. Ondanks de introductie van deze meer geïndividualiseerde therapie blijkt de overleving binnen de verschillende subgroepen
Addendum

Hoofdstuk 3 en 4 laten zien dat genexpressie-analyse bij AML tot een betere voorspelling van de overleving van patiënten met AML kan leiden. Met behulp van genexpressie datasets die publiekelijk toegankelijk zijn werd in hoofdstuk 3 de correlatie tussen overleving van AML-patiënten en de mate van expressie van het DNA-methyl-transferase 3B (DNMT3B)-gen onderzocht. Samen met de paralog DNMT3A, die vaak gemuteerd is bij AML, spelen deze genen een cruciale bij het onderdrukken van genexpressie door methylering van specifieke DNA-sequenties. Omdat er in tegenstelling tot DNMT3A geen genetische veranderingen in DNMT3B bij AML gevonden zijn, was een potentiële rol van DNMT3B bij deze ziekte onduidelijk. Echter, in overeenkomst met onafhankelijke studies, bleek uit onze studies dat verhoogde expressie van DNMT3B een onafhankelijke voorspellende waarde had op therapierespons. Patiënten met een verhoogde DNMT3B-expressie lieten namelijk een significant slechtere overleving zien, in het bijzonder bij patiënten met een normaal karyotype in combinatie met een NPM1 en FLT3-ITD-mutatie. Bij deze categorie patiënten wordt met name een slechte overleving gevonden wanneer het gemuteerde FLT3-ITD allel homozygoot aanwezig is. In hoofdstuk 3 wordt beschreven dat de subgroep met een homozygoot FLT3-ITD allel gesplitst kan worden in een groep met een zeer slechte overleving en een intermediaire overleving op basis van DNMT3B-genexpressie. Dit betekent dus dat DNMT3B-genexpressie-analyse kan bijdragen aan een verbetering van de classificatie van AML patiënten. Omdat het verschil tussen lage en hoge DNTM3B-genexpressie echter klein is, is dit met standaard moleculair diagnostische methodes nog niet betrouwbaar te meten. Alhoewel hoge DNMT3B-genexpressie sterk correleert met een slechte overleving, betekent dit niet per definitie dat de hoge expressie ook een rol speelt bij het ontstaan van de ziekte. Om te bepalen of dat zo is dient vervolgonderzoek uitgevoerd te worden.

Hoofdstuk 4 beschrijft een verbetering van de classificatie van AML-patiënten met een MLL-AF9 translocatie. Deze patiënten worden momenteel geclassificeerd met een intermediair risico, met een vijf-jaars overleving van 40%. De prognose van MLL-AF9-positieve patiënten werd verfijnd op basis van de expressieniveaus van BRE, een gen dat betrokken is bij DNA-schadeherstel. Hoge expressie van BRE bleek geassocieerd te zijn met een betere overleving terwijl de normale expressieniveaus een zeer slechte overleving voorspelden van MLL-AF9-positieve patiënten. In hoofdstuk 4 hebben we aangetoond dat de slechte overleving van patiënten met normale BRE-expressie geassocieerd worden met hoge expressie van EVI1. Van de hoge EVI1-expressie was al bekend dat het een slechte prognostische factor is bij AML. Omdat patiënten met hoge BRE-expressie geen hoge EVI1-expressie vertoonden, en omgekeerd, kunnen MLL-AF9-positieve patiënten dus onderverdeeld worden in een groep met een zeer goede en zeer slechte overleving. Wegens het kleine verschil tussen normale en hoge BRE-expressie is het momenteel met routine moleculaire diagnostiek niet mogelijk de verschillen betrouwbaar te meten. Voor EVI1-genexpressie bestaan wel gevalidateerde moleculaire assays om de expressie betrouwbaar te meten. Voor een betere risico-indeling van MLL-AF9-positieve patiënten is het dus raadzaam te screenen op EVI1-genexpressie. Wat verder uit onze studies bleek is dat, ondanks het feit


toch nog sterk te verschillen. De identificatie van nieuwe prognostische markers zou de respons op therapie mogelijk nog beter kunnen voorspellen.
dat geclaimd wordt dat \textit{EVI1}-genexpressie direct door het MLL-AF9 fusie-eiwit beïnvloed wordt, er geen uniform genexpressieprofiel aan te wijzen is waarmee deze patiënten geclassificeerd kunnen worden. Een uniek genexpressieprofiel bleek wel aantoonaar voor MLL-AF9-positieve patiënten met hoge \textit{BRE}-expressie, maar ook voor patiënten met hoge \textit{BRE}-expressie maar zonder de MLL-AF9 mutatie. Dit suggereert dat niet MLL-AF9, maar een andere, nog te identificeren, factor verantwoordelijk is voor het unieke genexpressieprofiel dat patiënten met hoge \textit{BRE}-expressie kenmerkt. Vervolgonderzoek naar de onderliggende paden die samenwerken met MLL-AF9 bij maligne ontaarding is daarom belangrijk.

Verschillende hematopoëtische aandoeningen worden gekenmerkt door een veranderde activiteit van NF-kB-transcriptiefactoren. Deze factoren reguleren genen die betrokken zijn bij celdifferentiatie, -proliferatie en apoptose. Een essentieel onderdeel van de NF-kB-route wordt geregeld door Lubac, een multi-domein eiwitcomplex met ubiquitineligase-activiteit. Lubac bestaat uit twee regulatoire subeenheden, HOIL-1L en SHARPIN en de katalytische subeenheid HOIP. Lubac is tot op heden het enige bekende eiwitcomplex dat de vorming van lineaire poly-ubiquitine-ketens vormt. Deze ketens werken als signaalmoleculen bij de activering van NF-kB. Omdat Lubac het enige eiwitcomplex is dat deze ketens vormt, is het mogelijk een belangrijk doelwit voor de ontwikkeling van therapieën voor bloedziekten die gekenmerkt worden door een verhoogde NF-kB-activatie. Het is daarom belangrijk op moleculair niveau op te helderen hoe Lubac precies lineaire ketens vormt om zodoende specifiek geneesmiddelen te ontwikkelen die dit proces tegengaan. Hoofdstuk 5 beschrijft een aantal belangrijke biochemische mechanismen die bijdragen aan de vorming van lineaire ubiquitineketens. In dit hoofdstuk laten we zien dat zowel de katalytische activiteit als de specificiteit voor de vorming van lineaire ubiquitineketens van Lubac binnen het ubiquitine ligase HOIP liggen. In het HOIP-eiwit ligt een domein dat verantwoordelijk is voor de vorming van lineaire ubiquitine ketens. Dit domein is opgebouwd uit drie geconserveerde subdomeinen, te weten een RING1-domein, een atypisch RING2-domein en een domein dat tussen RING1 en RING2 ligt. In Hoofdstuk 5 tonen we aan dat de vorming van lineaire ubiquitineketens in twee stappen wordt uitgevoerd waarbij RING1 en het tussenliggende domein verantwoordelijk zijn voor de vorming van een thioesterverbinding tussen ubiquitine en het RING2-domein. Vervolgens katalyseert het RING2-domein de binding tussen de C-terminus van een donorubiquitine met de N-terminus van een acceptorubiquitine. De vorming van de “lineaire” koppeling wordt bevorderd door een domein dat C-terminaal van RING2 ligt. Dit zogenaamde “Lineair ubiquitineketen vormend domein” gaat op een dusdanige manier een interactie met het acceptor-ubiquitine aan dat het goed gepositioneerd wordt om een lineaire ubiquitineketen te maken. Wat verder gevonden werd is dat de katalytische activiteit van HOIP geremd wordt door de N-terminale kant van het eiwit. Dit gedeelte van het eiwit gaat een interactie aan met HOIL-1L en SHARPIN. Juist de binding van deze eiwitten gaat de remming van dit domein op de katalytische activiteit van HOIP tegen. Hierdoor dragen HOIL-1L en SHARPIN bij aan de vorming lineaire ubiquitineketens.
**Curriculum vitae**

Davide Monteferrario was born on September 3rd, 1983, in Biella, Italy. He attended a science- and technology-oriented high-school, where he learnt fundamental notions of life sciences. Following graduation, Davide continued his education at the Faculty of Mathematical, Physical and Natural Sciences of the University of Pavia “Università degli studi di Pavia” in 2001. As he developed a strong interest for biological sciences, he chose to major in Biotechnology. In 2006, he obtained his BSc degree in Pharmaceutical Biotechnology and subsequently enrolled in a master’s program of the same faculty, entitled “Industrial Biotechnology”. During that time, he realized his aptitude for molecular life sciences in particular. So as to broaden his knowledge and gain experience in the field of molecular life sciences, he moved to the University of Utrecht (The Netherlands) within the framework of the ERASMUS exchange program. During that time, he performed an internship at the department of molecular microbiology under the supervision of Prof. Dr. G Riccardi, Prof. Dr. J.P.M. Tommassen and Dr. Peter van Ulsen. His research, focusing on immune-evasive mechanisms adopted by *Neisseria meningitidis*, yielded the results required for the completion of his master thesis. Immediately after having obtained his MSc diploma with a *summa cum laude* in 2008, he started his PhD studies at the Laboratory for Hematology of the Radboud university medical center (Nijmegen, the Netherlands). Under the supervision of Dr. B. A. van der Reijden and Prof. Dr. J. H. Jansen, his research focused on the identification and understanding of the molecular mechanisms involved in the development of blood disorders. Subsequently, he continued his career as a researcher under the supervision of Dr. G. Huls at the Department of Hematology of the Radboud university medical center, where he is currently investigating the role of the bone marrow microenvironment in the pathogenesis of myelodysplastic syndromes.
List of publications

High expression of transcription factor 4 (TCF4) is an independent adverse prognostic factor in acute myeloid leukemia that could guide treatment decisions. *Haematologica* 2014 Dec; 99(12):e257-9. in ’t Hout FE, van der Reijden BA, Monteferrario D, Jansen JH and Huls G.


Dankwoord / Acknowledgments

Though only my name appears on the cover of this dissertation, a great many people have contributed to its making. I owe my gratitude to all of them, for their direct, or indirect, contribution, not only to the production of this book, but also to my professional development; and because of whom my graduate experience has been one that I will cherish forever.

Dear Bert, it was a big pleasure and a great honor working with you all these years. Your precious guidance, through my really first working experience in the world of science, was fundamental for my professional formation. In the early days, when you put me in the “cold water” (as you used to say, remember?), I really appreciated that you gave me the freedom to explore on my own and, at the same time, had the patience to wait for my slow progresses. You taught me many important aspects regarding the scientific approach and guided to recover when my steps faltered. Unfortunately, this was not enough to fully decipher the elusive hematopoietic role of our beloved, almost legendary, Triad1 ligase, but it certainly laid the foundations of my professional growth. Luckily, the good results did not wait long to come. Your dynamism and dedication were essential to over-come many obstacles on the way and obtain the manuscripts for my thesis. Thank you for your constant willingness in always listening, for the unlimited number of advices you helped me with planning my research, and for giving me words of encouragement when needed. Your profuse joy and contagious enthusiasm for science have always represented a special motivational drive for me, while your strong determination is what impressed me the most of you, and I particularly tried to treasure it. The old Louis Pasteur used to infer that fortune favors prepared minds. “Hitting the jackpot“ with the GFI1B paper was the fortunate culmination of a long and fruitful nurture, for which I would like to address my deepest gratitude to you. I really wish that you will fulfill your ambitions at best and keep on riding the wave of the brilliant career you are on very far…buona fortuna!!

Dear Joop, after seven rich years, the”asino” is finally leaving the stable…and I will leave very much grateful to have been supervised by such a “shepherd”! From you, one of the many qualities that I particularly tried to learn was your capacity to engage science from a broad perspective. Your contributions to my projects were always helpful to keep the focus on the famous “bigger picture” and realistically approach the scientific research questions, without drowning in minor details. I enjoyed a lot your relaxed approach during our meetings, which always ended-up instilling a positive and propositional vibe in the discussion. As head responsible of the lab, I admire your efforts in trying to meet the needs of people, and the drive you have in caring also beyond work-related matters. I will definitely miss the good humor and jokes when we sometimes bumped into each other around the corridors of the lab, as they were always fun moments to cheer up with the long days at work. Besides many precious opinions and suggestions, I would like to particularly thank you for stepping-in sometimes with occasional pep-talks of yours that spurred me to take on some intricate situations. Many sincere thanks, Don!
Dear Gerwin, I would like to share my gratitude for the high consideration and faith you expressed in me since your arrival in Nijmegen. Thanks to your guidance, I could further implement my skills, hone my criticism and gain more confidence in my abilities. Although setting-up the complete research program took quite some time, we slowly managed to put it on track and produce interesting preliminary data. Unfortunately, I will no longer be directly involved in the experiments, but I will surely keep an eye on how the projects will further disclose. I would like to thank you for the way you entirely dealt with the "Groningen affair" in my regards, especially for the comprehension and understanding you demonstrated with respect to my final, very difficult, choice. I will certainly miss our frequent talks at the end of the day, your openness in confrontation and your genuine points of view. I enjoyed and learnt a lot by teaming-up with you. Good luck with your brand new job-position, although I am sure you will easily rise up to the challenge that comes with it!

Lieve Sylvie, after lots of pipetting and writing since the time you left, I am hopefully about to make it too! However, I am a bit sorry that we won’t probably be able to share those moments together, as you played a very big part in it. Thanks for taking care of me at the beginning of my adventure in the lab, which definitely facilitated my integration in a new environment, and for constantly showing your interest in my projects, which particularly benefited from your constructive criticism and numerous thought-provoking remarks. Your brilliance and dedication have always evoked my admiration and respect. Good luck with your career in Canada, the Netherlands, or wherever...in any case, I am confident that you will leave your mark!!

Dear Saskia, I really would like to express thankful words for your enormous contribution to my projects. Your advices helped me tremendously to sort out grueling technical details and to carefully evaluate the feasibility of many experiments that I planned. Without your support, life in the lab would have been very hard!

Dear Anne, many thanks for your decisive support for the execution of the final experiments of the GFI1B project. I wish you to have a longstanding career in science and I count on you, and of course on Rinske too, to further reveal the magic around GFI1B!!

In the past years, I had the pleasure to cooperate and share nice publications with many brilliant and skilled collaborators. Judith, Titia, Nikita, Lut, Bart, Konnie, Lacramioara, Cyrus, Hans, Frank, Wander and all the others, whom I didn’t have the chance to directly work with, I would like to warmly thank you for the dedication and professionalism with which you contributed to my thesis.

I obviously would like to spend thankful words for all the other members of the molecular unit, for the technical and intellectual support and, most of all, for the nice and enthusiastic atmosphere that accompanied each and every single day of my stay in the lab. Laurens and Saskia, I would like to thank you for all the occasions you assisted me with many tips that
definitely eased my life, also outside work, and sometimes even helped me to get a house where to live!! Good luck for your new adventure as parents!! Lieve Gorica, many thanks for the times you were there to advise me, with your long and clear explanations. However, I will gladly remember you more for the wine-toasts, with which we have often enjoyed life after work. As we miss the two most important ones, I am looking forward to toast them after our PhD defenses! Good luck with your forthcoming one!! Lieve Thessa, without your “sorting skills” my experiments would have not gone that far! Besides thanking you for always trying to squeeze time for me in the very busy “sorting-schedule”, I would like to praise your spontaneous care, with which you naturally support people. You are one of the most honest and altruist persons that I ever crossed, and with whom I am very happy to have created a genuine bond in time. Lieve Leonie, your ceaseless blushes makes you as the most colorful members of the living AIO-office! ☺ Thanks a lot for revealing the secrets of bisulfite sequencing and keeping the room always so warm and cosy (although sometimes I seemed to be roasting inside an oven!!). Also for you, good luck with your promotion!! Dear Ruth, although we never directly collaborated in any project, I would like to thank you for the good care of the lab and for the precious help in sorting out the messy database of viral vectors. I’ll be waiting to meet you soon at the lago Maggiore!! Dear Leilei, it was an immense pleasure to work with such a gentle as well as strong-minded girl as you. Thanks for introducing me so well to the lentiviral business and for always being so kind with me. I wish we will shortly enjoy a dive together in the fantastic sea of Sardegnia! Dear Maaike, despite it took longer than the others lab-members to get to know you, it was worth the wait! I definitely enjoyed our nice talks in the lab and also the sharp comments you often provoked me with, I had much fun with teasing each other! I am confident you will become an excellent pediatric hematologist, good luck!! Lieve Jolanda, it was really a pleasure to work with such a skilled and organized technician, as much as a hearted person, as you. Thanks a lot for the administration of the huge MSC database and the tremendous help with students and experiments in the last couple of years. Since I know how much you miss it, I hope you will get back soon to your first love, the mice-work!! Dear Florentien, it always amazes me to see how much vitality and energy there is in such a small girl, but you know… the good wine is kept in the small barrel!!! Thanks for the unlimited enthusiasm you can uniquely cheer me up with. Good luck with TCF4 (I expect big things from it, don't disappoint me, Billy!!) and, most of all, with all your endeavours!!

I also would like to praise my fantastic students, who I had the luck to supervise, and with whom I gladly shared the many frustrations and joys that naturally come with doing research. Pascal, Vicky, Frederique, Secil, Sandra, Ugur, Reneè, Cenna, Chloè and Annemarie: having hopefully contributed to the development of each and every one of you is something that really makes me proud of. Thanks a lot for the hard work and the enthusiasm you brought into my projects.

My dearest Mariam and Jimmy, my original “partners in crime”, although our professional routes diverged in the past years, I could not have been blessed with a better couple of
cheerful travel buddies than you!! Dear Mariam (S., O.), I strongly doubt that I will be able to fully convey the joy I feel to have you in my life, as few words can’t simply be enough to express it. Nevertheless, I would like to deeply thank you for always being there for me, to console me with your “warm” hugs when I fell, and to cheer me up with your “you can do it!” when things were tough. What an amazing paraninfen I am proud to have! Dear Jimmy, since the day we set foot together for the first time into the lab, your sharp humor has delighted countless moments of my “dutch” experience. Besides the many brilliant jokes you privileged me with, I would like to thank you for being the hearted and caring friend to count on you are. I will definitely miss the savor of our fat laughs and our (scientifically) deep conversations, but I really feel gifted to have picked such amazing friends, as both of you are, up in my way.

A huge praise goes for the members of the “Pool-club”, of whom I enjoyed the foolish attitude, but not the talent for sure!!! My dear friend Louis, thank you for being my best “football-talks” buddy, for your magic touch in fixing all those PCR’s that hindered my experiments, and foremost, for your unique jovial spirit, which cheered-up uncountable fun moments of my long stay in the Netherlands. Caríssimo Pedro, és de facto um personagem mais único do que raro, pelo qual me sinto deveras lisonjeado por ter como leal amigo. Ficarei-te para sempre agradecido por nunca teres perdido a esperança de poder melhorar o meu nível de Inglês (e claro também de Japonês!), corrigindo sem piedade qualquer tipo de erro gramatical ou de sintáx. De entre as pequenas e as grandes coisas pelas quais te quero agradecer, devo realçar os mexericos oitocentistas e a fantástica companhia com os quais alegraste e enriqueceste os meus dias. 大好き. Mio caro Niccolò, grazie di essere...Niccolò!! Oltre alla tua talvolta unica visione del mondo, alle numerose discussioni con Mustaschfide (inclusa anche quella, a volte, animata) e ad avermi introdotto al rito del fischio “risolutore”, mi mancherai per l’allegria e onestà con cui sai essere un ottimo collega, e soprattutto un grande amico. Sebbene saresti stato bello che fosse durata un pò di più, sono sicuro che non avrei potuto trovare un socio migliore con cui condividere due tra gli anni più belli della mia esperienza a Nijmegen. F..es!!

I am definitely thankful to all the other colleagues that animate the lab of hematology. In all this time, I always felt welcome and really happy to be part of it. There are some people that I would like to particularly mention for their precious help. First of all, the dames of the molecular diagnostic unit, thanks a lot for the prompt helps with PCRing (Evelyin in particular) and, especially, for the joyful atmosphere I had the pleasure to enjoy any time I stepped into the office. Dear Rob, no one can squeeze every single FACS tube down to the last cell as you do!! Many thanks for the great support at the cell-sorter. All the girls of the morphology unit, I am very thankful for the numerous GIEMSA stainings, the fun talks and the gezelligheid you always welcomed me with whenever I needed to use the microscopes. I am also indebted to the “immuno-phenotyping people”, for all the times I requested antibodies, and the BM transplantation unit, for reserving primary material for my experiments. I would like to thank you, Eugenie, for the patience in teaching me
all the tricks to recognize the different types of white colonies. Thank you, Basav, for the nice collaboration on the Triad1 project and the fruitful scientific talks in the corridors. Last, but not least of course, a big hug goes to the most “lekker-ding” in the lab, Nelleke!!! Many thanks for the unlimited parent-like care you had in my regards. I have been super lucky to have had a super “lab-mum” such as you!

I need to absolutely praise all the good friends that helped me to stay sane throughout all these years. Matilde, Antoine, Anchel, Angela, Ganesh, Sip, Nuria and many others, you have been an incredible source of love and energy ever since we first encountered. Thanks for the fun time, the unlimited support and your belief in me. I could not have wished for a better “family”!

Grazie, di cuore, anche a te Stefania per l’aiuto e il supporto dedicatomi durante la preparazione di questo (quasi) interminabile lavoro.

A special word goes to my paraninfen André. Besides the fact that it is always handy to have an engineer around, the pleasure of sharing with you all the phases of my life in Holland is something that can’t be simply presented in few lines. Nevertheless, I just simplistically would like to thank you for the great time I owe you and your amazing “mentalità” as friend!

Un grazie immenso a tutti i “polli” di Cossato, per essermi sempre stati vicini nonostante la distanza, e per il calore con cui mi avete sempre accolto quando tornavo al pollaio. Sono orgoglioso di avervi con me.

Concludo i miei ringraziamenti con i membri della mia famiglia, che più di tutti mi hanno sostenuto e a cui devo praticamente tutto. Nonna Ada, anche se non sei mai riuscita a venirmi a trovare, non ho mai smesso di sentire la tua presenza. Grazie per come sai essere così orgogliosa di me, e soprattutto per tutto l’affetto e la bontà con cui mi hai sempre accompagnato in questi anni. Mamma e papà, non esiste alcun dizionario al mondo che contenga una parola adatta per ringraziarvi, per tutti i valori che mi avete trasmesso e per il grande sostegno, sia morale che materiale, con cui mi avete permesso di crescere e migliorare. Seppur non sia assolutamente abbastanza per esprimere il mio senso di gratitudine nei vostri confronti, questa tesi la dedico a voi e a tutti i vostri sacrifici e la fiducia che sono serviti a portarmi davanti a questo traguardo. Siete la parte migliore che ho.