Molecular mechanisms of acute erythroid leukemia: learning from rare chromosomal translocations in pediatric patients

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Maria Riera Piqué Borràs

aus Spanien

Basel, 2021

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Professor Dr. med. Jürg Schwaller

Professor Dr. phil. Nat. Christoph Handschin

Basel, den 18.02.2020

Prof. Dr. Martin Spiess
Dekan der Philosophisch-
Naturwissenschaftliche Fakultät

I dedicate my Ph.D thesis to Sara Díaz García, a little angel that left us too early and whose strength fighting bone marrow failure made me realize how necessary and important research is.

Dedico la meva tesis doctoral a la Sara Díaz García, un petit àngel que ens va deixar massa aviat i la seva força lluitant contra l'aplàsia medul·lar em va fer adonar el necessari i l'important és la recerca.

Maria Riera Piqué Borràs

Table of Contents

ACKNOWLEDGMENTS **7**

SUMMARY **9**

CHAPTER 1 INTRODUCTION **11**

1.1. HEMATOPOIESIS **11**
1.1.1. Primitive hematopoiesis *11*
1.1.2. Hematopoietic stem cells and definitive hematopoiesis *12*
1.1.3. Regulation of hematopoiesis by transcription factors and cytokines *15*
1.2. ERYTHROPOIESIS **17**
1.2.1. Primitive erythropoiesis *17*
1.2.2. Definitive erythropoiesis *17*
1.2.3. Regulation of erythropoiesis *20*
1.2.4. Negative regulation of erythroid differentiation *24*
1.3. MALIGNANT HEMATOPOIESIS **28**
1.3.1. Impact of the cell of origin in leukemias *29*
1.3.2. Acute myeloid leukemia *31*
1.3.3. Classification of AML *32*
1.3.4. Molecular aberrations in AML *34*
1.3.5. Aml-associated fusion genes *36*
1.3.6. Current therapies of AML *38*
1.4. FROM NORMAL TO MALIGNANT ERYTHROPOIESIS **39**
1.4.2. Classification of erythroid malignancies *39*
1.5. ACUTE ERYTHROID LEUKEMIA **40**
1.5.2. Morphologic features of erythroleukemia *42*
1.5.3. Models of erythroleukemia *42*
1.5.4. Molecular aberrations in erythroleukemia *44*
1.5.4.1. Hematopoietic transcription factors in experimental models *44*
1.5.4.2. Molecular lesions in primary patient-derived erythroleukemia cells *46*
1.5.4.3. NFIA-ETO2 ..... *47*
1.5.4.4. ZMYND8-RELA *51*

CHAPTER 2 CELLULAR AND MOLECULAR TRANSFORMATION MECHANISMS OF ERYTHROLEUKEMIA-ASSOCIATED FUSION GENES **55**

2.1. WORKING HYPOTHESIS **55**
2.1. CELLULAR AND MOLECULAR TRANSFORMATION MECHANISMS OF THE NFIA-ETO2 FUSION ASSOCIATED WITH ACUTE ERYTHROLEUKEMIA **55**

CHAPTER 3 ADDITIONAL EXPERIMENTS (not included in the manuscript) **91**

3.1. CELLULAR PROLIFERATION OF NFIA-ETO2 EXPRESSING ERYTHROBLASTS IS EPO-INDEPENDENT **91**
3.2. EXPRESSION OF A SMALL, ETO2-NHR2-DERIVED PEPTIDE (NC128) INDUCED DIFFERENTIATION OF NFIA-ETO2-EXPRESSING ERYTHROBLASTS **87**
3.3. INDUCTION OF NFIA-ETO2-INDUCED DISEASE BY TRANSPLANTATION OF BM-DERIVED HSPCS **95**
3.4. STRUCTURE-FUNCTION ANALYSIS OF NFIA-ETO2 ..... **98**

3.5. TRANSFORMING POTENTIAL OF THE ZMYND8-RELA FUSION 100
3.5.1. Cooperation of ZMYND8-RELA with TP53^{R248Q} 104
3.5.2. Transplantation of ZMYND8-RELA-transduced TP53$^{R248Q/+}$ cells did not result in a leukemic phenotype 107

CHAPTER 4 DISCUSSION 109

CHAPTER 5 HUMAN ERYTHROLEUKEMIA GENETICS AND TRANSCRIPTOMES ... 127

5.1. INTRODUCTION 127
5.2. CORRELATION OF AEL TRANSCRIPTOMES WITH ERYTHROID DIFFERENTIATION 128
5.3. CORRELATION OF AEL TRANSCRIPTOMES ACCORDING TO THE MUTATIONAL STATE . 132
5.4. POTENTIAL COOPERATION OF EPO OVEREXPRESSION AND *TP53$^{R248Q/+}$* 136
5.5. DISCUSSION 140

MATERIAL AND METHODS 145

REFERENCES. 155

LIST OF ABBREVIATIONS 172

APPENDICES 175

Acknowledgments

First of all, I would like to thank everyone who helped, supported and guided me throughout my PhD. Already at this point, I want to apologize to the ones I am not mentioning in this section, but if I wanted to include everyone, I would probably have to write another thesis just consisting of acknowledgements. As space is limited, I will here just mention the people, without whom I could have never managed to write this thesis.

Jürg, you of course have to be the first person on this list. Throughout my years as a PhD student in your lab, I always felt support and guidance and you helped me a lot to chase a goal and to achieve it. It has not always been easy dealing with my "Desigual" nature – having to realign plots for hours, changing font sizes, rewriting paragraphs – but in the end, I have to admit, that your attitude to always pursuit perfection, is what brought out the best in me. Without you, I would not be where I am today and therefore, I want to thank you for everything I learnt from you and for giving me the opportunity to do a PhD in your lab.

I would also like to thank the former and current members of the Childhood Leukemia group. **Sabine**, thank you very much for taking so much care of the lab, but more importantly, of us. You are the pillar of the lab and without you, my PhD would have been even harder. **Marwa** & **Max,** you have been there for me since the beginning of this long journey, thank you very much for all the laugh and fun during the hard times, inside and outside the lab, I will never forget what you did for me. **Matheus,** nine months sharing the lab and now, even across the ocean, you became an essential person for me, thank you so much for everything what you do for me. **Frederik**, it has been a pleasure to find another geeky person along this way and to have the possibility to share our bioinformatics jokes – pressing just "enter" is not enough – thank you for all your support. **Hugues-Étienne**, **Amber** & **Samantha**, thank you for making the PhD life funnier and easier.

I would also like to thank the Stem Cells and Hematopoiesis group, as well as the Experimental Hematology group, especially **Martina**, **Anna, Marc**, **Shivam** & **Jakub**. It's been amazing sharing the lab space with you all, not only for the scientific part, but also for the great friendship.

Special thanks to my PhD Committee members, **Prof. Christoph Handschin** and **Prof. Thomas Mercher**, for all your inputs and for taking the time to guide me during my PhD.

Finalment, vull donar gràcies als meus **Pares**, per tot el que heu fet i feu per mi, sense el vostre suport incondicional mai no hauria pogut arribar a ser doctora, així que aquest títol és vostre. Al meu germà **Robert**, que tot i la distància i la zona horària, sempre has estat apunt per ajudar-me i donar-m'ho tot. Al meu **Avi**, per totes les nous que ha trencat per mi "perquè són bones pel cervell". I al **Christoph**, trobar-te en el mig d'aquesta aventura ha fet possible que la pogués acabar, gràcies per tot el que fas per mi, ara ens toca gaudir del nostre viatge.

Summary

Two novel fusion genes, NFIA-ETO2 and ZMYND8-RELA, have been identified in pediatric patients suffering from pure erythroid leukemia (PEL). Based on the normal function of the fusion partners, we hypothesized that both fusions might initiate PEL by interfering with erythroid differentiation. I studied the role of these two fusion genes in erythropoiesis and found that ZMYND8-RELA significantly reduced the viability of erythroid progenitor cells. On the other hand, expression of the NFIA-ETO2 fusion significantly impaired terminal differentiation of murine erythroblasts and slightly increased their self-renewal capacity. However, transplantation of these cells into lethally irradiate mice did not induce any disease. Mutations of the tumor suppressor TP53 have been reported to be molecular hallmarks of PEL. I therefore investigated whether NFIA-ETO2 cooperates with one of the most prevalent PEL-associated TP53 mutation (R248Q). I found that NFIA-ETO2 not only increased self-renewal of *TP53*$^{R248/+}$ erythroblasts, but also induced a PEL-like disease upon transplantation of the cells into lethally irradiated recipients. To understand how NFIA-ETO2 expression interferes with erythroid differentiation, I performed transcriptome analysis and found that NFIA-ETO2 expression was associated with up-regulation of proto-oncogenes such as *Myb*, *Notch2, Myc* and *Stat5a* and down-regulation of GATA1 and NFIA target genes related to erythropoiesis. Surprisingly the presence of the *TP53*^{R248Q} mutation in NFIA-ETO2 erythroblasts only resulted in downregulation of few genes, mainly previously characterized TP53 targets. Chromatin analysis (ATAC-seq) showed that these genes presented a positive enrichment in their promoter and predictive transcription factor analysis suggested that they could be repressed by the cooperative activity between the *TP53*^{R248Q} and the PRC2 (Polycomb complex 2) epigenetic key regulator.

As the molecular drivers of acute erythroleukemia are mostly unknown, we decided to study the epigenomic landscape of the disease in more detail. In collaboration with Thomas Mercher, we collected patient samples worldwide for RNA-seq and Exome-seq. We found that erythroleukemia patients can be genetically grouped into at least 3 groups, those with TP53 mutations, those with mutations in epigenetic/metabolic regulators and patients with very few mutations. Interestingly, we found that the transcriptome, while poorly correlating with the mutation status, reflected the maturation stage of the erythroblasts, and was clearly different from signatures from most MDS or other AML. Overall, the results from my PhD project provides some functional insight into molecular mechanisms of acute erythroleukemia that may in the long run help to improve diagnosis and therapy for this rare but aggressive disease.

Chapter 1 : Introduction

1.1. Hematopoiesis

The bone marrow (BM) is one of the largest organs of our body occupying the central regions of all long and flat bones and provides a home for hematopoiesis (*1*). BM is the place where hematopoietic stem cells (HSCs) – the most primitive and potent cells of the blood system are formed, ultimately giving rise to over 300 billion differentiated cells every day (*2, 3*).

Alexander A. Maximow (1909), a Russian scientist, was the first person to introduce the nowadays accepted idea of hematopoiesis with HSCs sitting at the top of the hierarchically structured mammalian blood system (*4*). However, it wasn't until the late 20th century when scientist performed the first BM cell transplantation into irradiated mice to experimentally proof Maximow's theory of the need of a stem cell for the construction of a complete hematopoietic system (*5*). They discovered that BM-derived cells from transplanted mice could proliferate and form macroscopic colonies derived from one single clone. In 1957, Edward Donnall Thomas performed the first successful HSC transplantation in patients (*6*) and ever since, BM transplantation became a routine treatment for patients suffering mostly from hematological malignancies.

The process of hematopoiesis is highly conserved between vertebrates and consists of two distinct waves: The primitive and the definitive wave of hematopoiesis (*7*).

1.1.1. Primitive hematopoiesis

During early development, the oxygen need of the embryo is mainly matched by simple diffusion. However, the rapid growth and the establishment of complex tissues leads to a point, when passive diffusion is not sufficient anymore for adequate oxygen supply. A way for the embryo to overcome the increasing oxygen need, is the development of an active circulatory system (*8*). The beginning of the development of circulation not only brings along the emergence of the first endothelial cells, but also represents the starting point of hematopoiesis, with primitive nucleated erythroid cells serving as oxygen transporters (*9, 10*). Although being intensively studied, the parallel emergence of endothelial and primitive erythroid progenitor cells in the extraembryonic yolk sac blood islands remains one of the major subjects of debate in hematopoiesis. Nowadays, the theory of a bipotential common precursor named "hemangioblast", which, during early embryonic development, gives rise to the first endothelial and erythroid cells, is widely supported (*11-13*). However, the whole principle seems to be more complex than initially expected. The majority of hemangioblasts was also shown to

harbor megakaryocyte and macrophage potential, thus adding a myeloid component to primitive hematopoiesis (**Fig.1**) (*14*). Since further validation of the hemangioblast hypothesis turned out to be difficult in established *in vivo* models, the existence of such a common precursor with restricted potential as the origin of the most primitive blood cells could still not be definitely proven (*15*).

1.1.2. Hematopoietic stem cells and definitive hematopoiesis

As the vertebrate embryo is growing bigger, the transient wave of primitive hematopoiesis is replaced by another temporally limited wave called the "intermediate wave of hematopoiesis". In mammals this intermediate wave of hematopoiesis takes place in the yolk sac as well, and gives rise to myeloid-erythroid progenitors (MEPs) harboring the potential to differentiate into erythrocytes or myeloid cells. Unlike primitive erythrocytes and macrophages however, these cells are morphologically and functionally indifferent from adult blood cells. Due to their high degree of similarity, the intermediate wave of hematopoiesis is thus also often regarded as the first wave of definitive hematopoiesis (*16-18*).

Later, definitive hematopoiesis starts with the first mammalian HSCs arising in the aorta-gonad-mesonephros (AGM) region, which forms nearby the urogenital tract and ultimately generates the mesonephros, the gonad and the cortex of the adrenal gland (*19*). Similar to primitive blood cells, definitive HSCs are closely related to endothelial cells. However, unlike the hemangioblast, which is hallmarked by its bipotential nature, definitive HSCs were shown to bud directly from the endothelial layer of the dorsal aorta (also referred to as endothelial-hematopoietic transition), indicating the existence of an endothelium with hemogenic properties (**Fig.1**) (*20, 21*). Once in circulation, HSCs migrate to the fetal liver (FL), where they expand and differentiate. At the end of the embryonic development, when the invasion of bones by blood vessels and bone ossification are highly advanced, HSCs are seeding the BM. From this timepoint on, multipotent HSCs are the major source of every blood cell in the whole body.

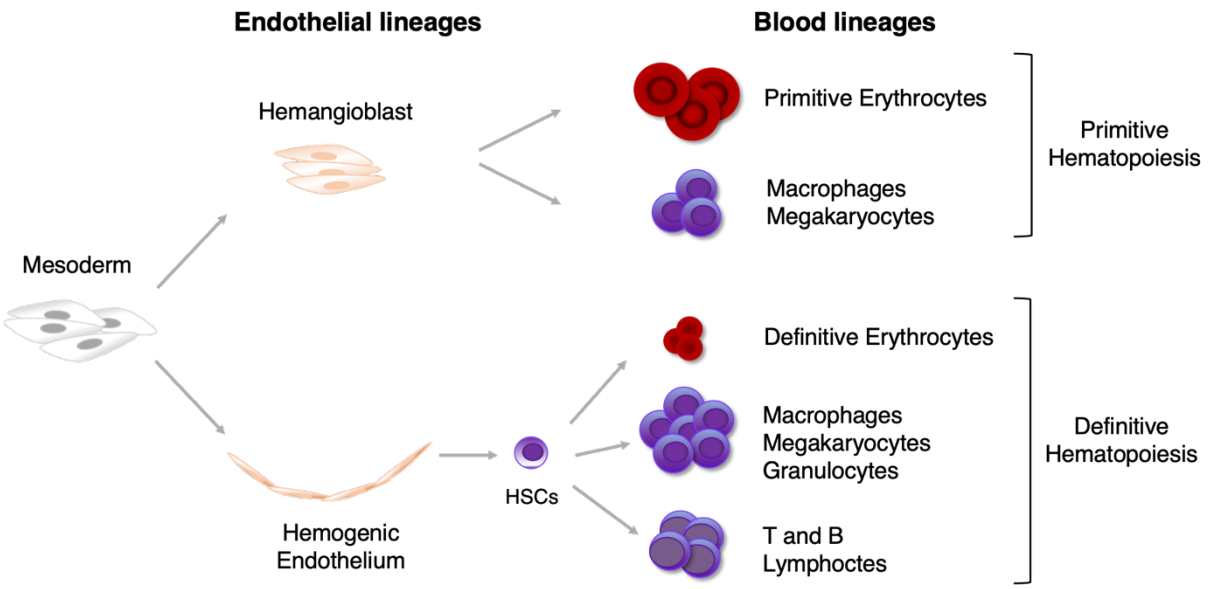


Figure 1. Cellular origin of primitive and definitive hematopoiesis. Shown is a schematic representation of hemangioblast or hemogenic endothelium emergence and the lineage contribution of primitive and definitive hematopoiesis. Adapted from (*12*).

Due to their capability of undergoing asymmetric cell division, where one of the daughter cells still remains as HSC, they are able to give rise to a novel, more differentiated daughter cell without risking complete depletion of the HSC pool (*22-24*). After several rounds of division and differentiation, HSCs become multipotent progenitors (MPPs), further losing their multipotency as they start developing characteristics of commitment (**Fig.2**) (*25*). In a first step of commitment, two lineages of hematopoietic differentiation have been proposed: The myeloid lineage, originating from Common Myeloid Progenitors (CMPs) and the lymphoid lineage, originating from Common Lymphoid Progenitors (CLPs). (*7, 26*). CMPs are able to differentiate to either Megakaryocyte-Erythroid Progenitors (MEPs) or Granulocyte-Monocyte-Progenitors (GMPs). The MEPs eventually give rise to megakaryocyte and erythroid lineages (*27*) (see chapter 1.2), whilst GMPs are the source of all granulocytes (neutrophils, eosinophils, basophils and mast cells) and macrophages. CLPs on the other hand are responsible for the generation of B lymphocytes, T lymphocytes and Natural Killer Cells (NKs). The only cell type, which is known being generated by both, the myeloid as well as the lymphoid axis, is the dendritic cell. Although their ontology might be different, the function of dendritic cells was shown to depend on external stimuli rather than on their origin (*28*).

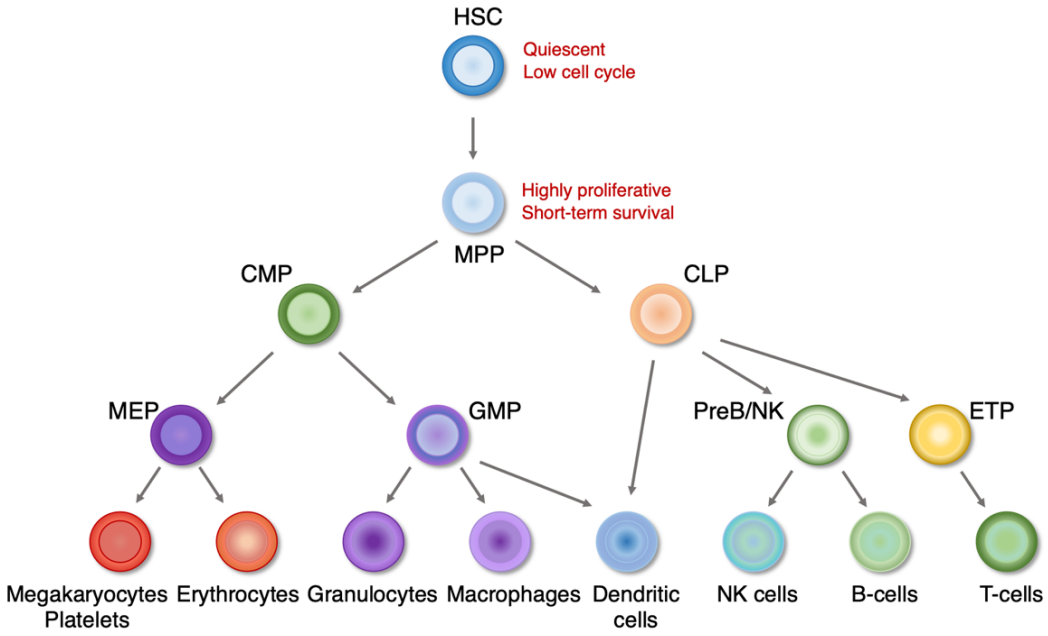


Figure 2. Representation of the classical model for HSC differentiation. Adapted from (*29*).

This classical, hierarchical model of hematopoiesis has been widely accepted for decades; however recent studies started to refine it. One of the first major modifications of the model was the separation of the MPP population into four lineage-biased subsets, termed MPP1, MPP2, MPP3 and MPP4 (*30*). However, nowadays even this adaptation is regarded as being highly oversimplified, since more recent research based on single cell methods suggested that hematopoietic progenitors might undergo lineage commitment at earlier timepoints than hitherto expected, without following the gradual routes from multipotency to unipotency (*2, 31-33*). This led to the formulation of a revised model of hematopoiesis, in which differentiation is regarded as a continuum instead of a sequence of subsequent branching decisions (**Figs.3,4**) (*25, 34, 35*). In this revised model, the first major lineage specifications acquired by HSCs prime them towards the megakaryocytic/erythroid axis or the myeloid/lymphoid axis respectively (*29, 36*). Within the myeloid/lymphoid axis, no strict separation into CMPs and CLPs can be made. This evokes the need of a population of cells, which is termed Lymphoid Myeloid Primed Progenitors (LMPPs). These cells are still pluripotent and can be primed towards lymphoid or granulocytic fate, with the exception of monocytes, whose specification seems to occur earlier in development. In general, the revised model of hematopoiesis describes blood cell formation as a much more dynamic process compared to the classical hierarchical model. Lineage priming is seen as a continuum of acquisitions of specifications, which lead to the appearance of small barriers between cell states (**Fig.3**). These barriers are not absolute and can be overcome, but still render it more likely for a cell to differentiate along a certain axis (*25*).

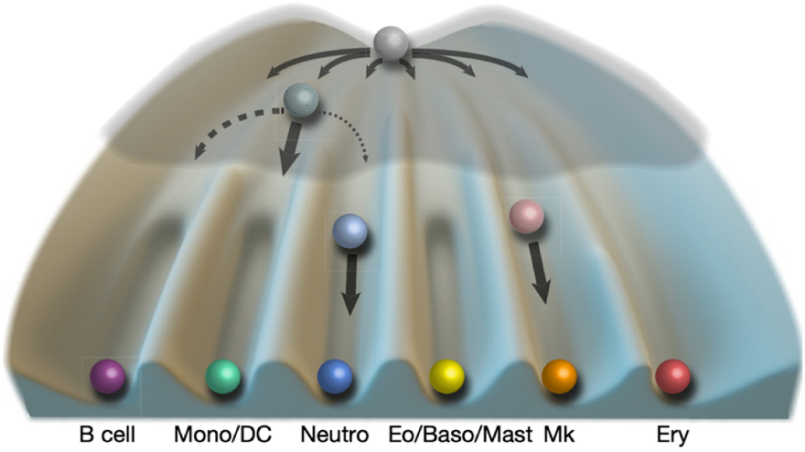


Figure 3. Graphical summary of a continuum-based model of hematopoiesis. Elevations within a Waddington landscape represent cell states and transitions that are more likely than others, due to the interactions of gene regulatory networks. At early stages, small barriers between lineages arise, thereby creating lineage biases in HSCs. At later stages of differentiation, the barriers are more pronounced, making the oligopotent stage less likely. The figure was extracted from (*25*).

1.1.3. Regulation of hematopoiesis by transcription factors and cytokines

Hematopoiesis is a tightly controlled process driven by the interplay of an immense number of regulatory molecules. Already small perturbances of this interplay may cause hematopoietic defects and eventually death (*37*). As the sites of hematopoiesis change temporally and spatially, the genetic program controlling blood cell formation must be constantly adapted.

The first event in hematopoiesis, the emergence of the hemangioblast, has been found to be regulated by transcription factors including SCL/TAL1 (T-cell acute lymphocytic leukemia 1), GATA2 (Gata-binding factor 2), LMO2 (LIM Domain Only 2), FLI1 (Friend Leukemia Integration 1 Transcription Factor) and ETV2 (ETS Variant Transcription Factor 2) (*38-42*). For the commitment of the hemangioblast towards erythroid or myeloid fate, the interplay between the transcriptional master regulators GATA1 (Gata-binding factor 1) and PU.1 (Hematopoietic Transcription Factor PU.1) was shown to be crucial (*41, 43, 44*).

As the niche changes from primitive to definitive hematopoiesis, also the network of involved signaling pathways changes. For the emergence of definitive HSCs, WNT- (Wnt Family Member 1), BMP4- (Bone Morphogenetic Protein 4), VEGF- (Vascular Endothelial Growth Factor A), SCF- (Stem Cell Factor), Hedgehog- (Sonic Hedgehog Signaling Molecule) and NOTCH- regulated signaling was shown to be of high importance (*7, 45*).The NOTCH pathway, which is regulated by the expression of RUNX1 (RUNX Family Transcription Factor 1), is specially required during early development, but also within the earliest adult hematopoietic compartment (*37, 46, 47*). Furthermore, the NOTCH pathway has been

extensively studied in the context of hematopoiesis and is regarded as one of the master regulators of HSC emergence and self-renewal (*37, 46, 47*).

Hematopoietic differentiation is nowadays considered as a continuous acquisition of different gene expression modules, priming HSCs towards a certain fate (**Fig.4**) (*25, 29*). In this context, expression modules are groups of genes which are mostly functionally related and co-regulated, thus showing similar expression profiles (*48*). Whilst the expression of genes encoding for transcriptional regulators like HOXA3 (Homeobox A3), HOXB6 (Homeobox B6) or PRDM16 (PR/SET Domain 16) has been associated with the least-primed state of HSCs, the first priming transcriptional motifs observed in HSCs and MPPs are regulated by NFE2 (Nuclear Factor, Erythroid 2) and GATA2 and by CBPE (Choline binding protein E), STAT1 (Signal Transducer And Activator Of Transcription 1) and TCF4 (Transcription Factor 4) (*25, 29, 34*), which were proposed to guide differentiation along the megakaryocyte/erythroid and lymphoid/myeloid axis respectively (**Fig.4**). After this first specification, the cells acquire more and more lineage-associated expression modules to differentiate towards mature hematopoietic cells. Whilst the CEBPA/CEPBD (CCAAT Enhancer Binding Protein Alpha, CCAAT Enhancer Binding Protein Delta) expression module for example is a characteristic of the neutrophil lineage, IRF8 (Interferon Regulatory Factor 8) is a hallmark of monocytic/dendritic lineage and EBF1/ID3 (EBF Transcription Factor 1, Inhibitor of Differentiation 3) is specific for B-cell differentiation. Within the megakaryocyte/erythroid axis, the GATA1/KLF1 (Kruppel Like Factor 1) expression module was proposed to drive erythropoiesis and the PBX1/GPI1BB (PBX Homeobox 1, Glucose-6-Phosphate Isomerase 1BB) module to control megakaryopoiesis (*25, 34*) (**Fig.4**).

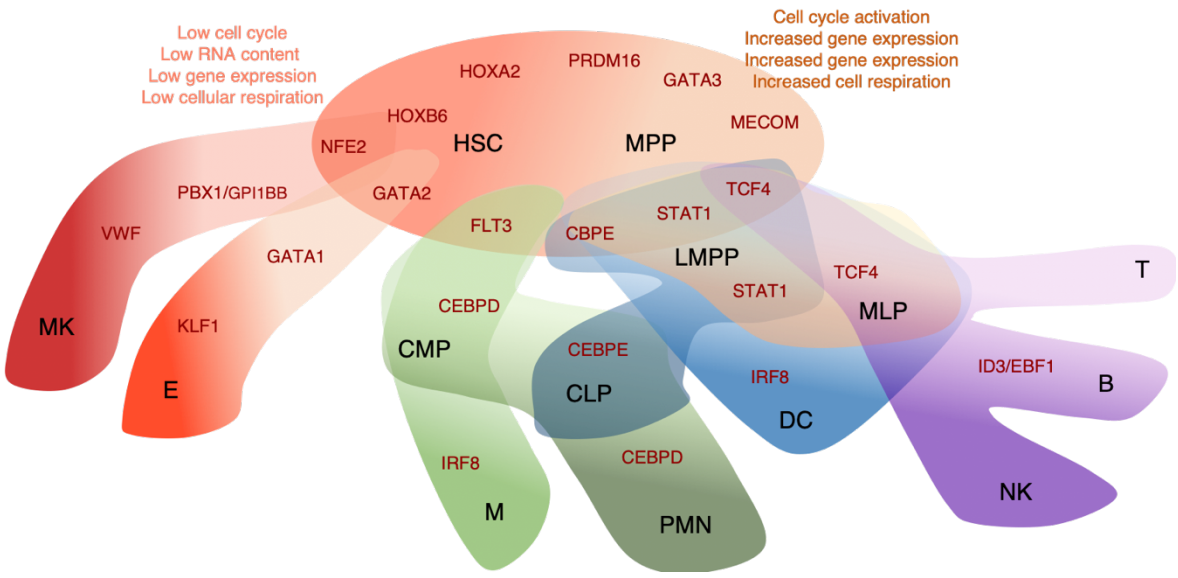


Figure 4. Transcriptional control of HSC differentiation toward mature lineages. The name of hematopoietic populations is indicated in black; the differentiation-driving transcription factors are given in dark red. Adapted from (*29*).

1.2. Erythropoiesis

As outlined before, the first blood cells arising during the primitive wave of mammalian hematopoiesis are immature red blood cells, also called erythroblasts, highlighting the fundamental role of erythropoiesis. Red blood cells have the essential function of transporting oxygen. Therefore, high numbers of erythrocytes are required for the maintenance of health of all vertebrate organisms. To match the high demand of mature red blood cells, erythroid progenitors undergo an enormous expansion to fulfill the daily requirement of 2 x 10^{11} erythrocytes (*49*). During development, two types of erythrocytes were identified – embryonic and adult, which develop from different hematopoietic progenitors in different anatomical sites.

Developmentally, erythropoiesis in humans occurs in two waves: primitive erythropoiesis within the embryonic yolk sac and definitive erythropoiesis in the FL and postnatal bone marrow.

1.2.1. Primitive erythropoiesis

During embryonic development, newly formed mesodermal cells migrate towards the yolk sac and contact closely with endodermal cells (*50, 51*). It has been shown that the interaction of these two primary germ layers is required for the initiation of erythropoiesis (*52*).

The hemangioblast, the common bipotential mesodermal precursor, is located in the posterior primitive streak (*12*) and migrates to the yolk sac, where cells begin their commitment to not only primitive erythroblasts, macrophages and megakaryocytes, but also endothelial cells (**Fig.1**) (*11, 12, 53, 54*). These cells form blood islands, which proliferate and merge, developing yolk sac vascular networks (*8*).

Primitive erythrocytes are relatively large and characterized by the expression of embryonic globins ε, γ, and ζ. These cells are the only erythroid cells in mammals released in the bloodstream retaining their nucleus (*55*). However, they disappear very quickly during the embryonic-to-fetal transition period.

1.2.2. Definitive erythropoiesis

Whilst the primitive erythroid cells directly emerge from hemangioblasts, definitive erythropoiesis is much more complex with the erythroid differentiation following a more hierarchical structure of hematopoiesis. For many years, it has been thought that definitive erythropoiesis starts in the FL (*17*). However, a unique β-globin gene expression profile of definitive erythroid progenitors or burst forming units-erythroid (BFU-E), was found in maturing definitive erythroblasts isolated from the yolk sac at E8.25, just before the onset of circulation.

BFU-E proliferate rapidly in the yolk sac and are subsequently found in the blood stream, later colonizing the FL (**Fig.5A**) (*56*). Interestingly, human BFU-Es were also first detected in the yolk sac at 4-5 weeks of gestation and were subsequently found in the FL (*57*). These observations suggested that definitive erythropoiesis emerges in the yolk sac before colonization of the FL.

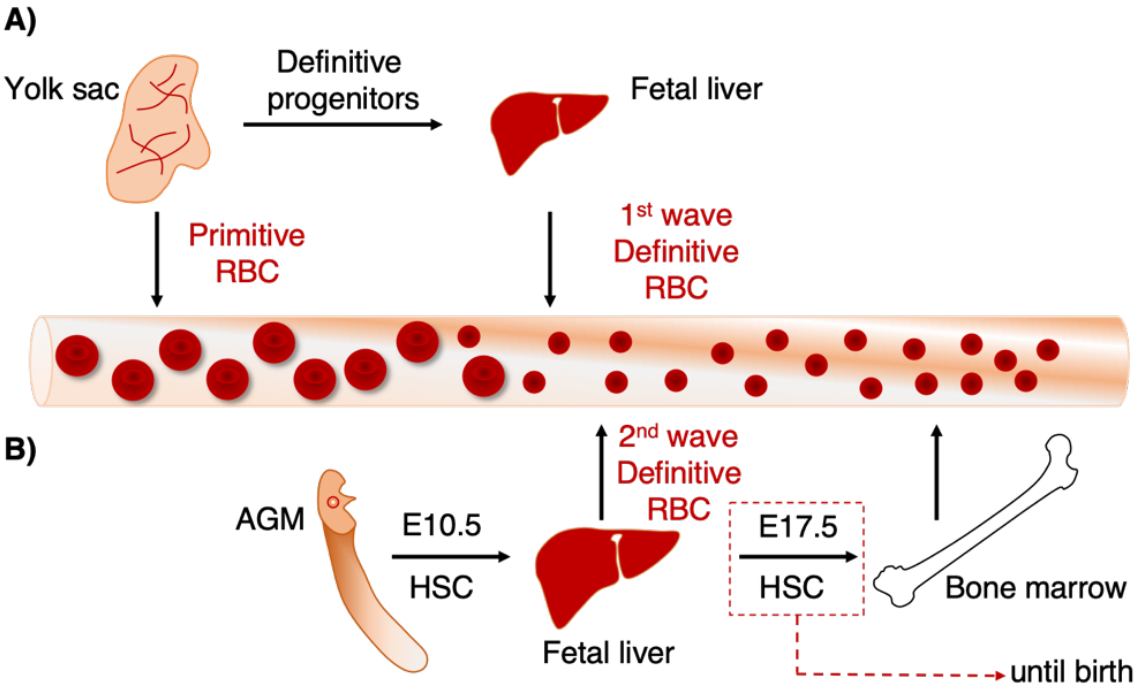


Figure 5. Ontogeny of the erythroid lineage cells in the circulation. Primitive red blood cells (RBC) are made by the yolk sac at E7.5 and are found in the circulation until E11/12. At E9, the yolk sac generates definitive progenitors that migrate to the FL, where they differentiate to definitive red blood cells (RBCs) and enter the circulation. At E10.5, the AGM generates the first HSCs that migrate to the FL and differentiate to the erythroid lineage, and these definitive RBCs enter the circulation. FL HSCs migrate to the BM, completing the colonization after birth, where they provide lifelong production of definitive RBCs for the circulation. Adapted from (*49*).

This first wave of definitive erythropoiesis is a transient system, which is part of the intermediate wave of hematopoiesis and ultimately replaced by a second one, where cells are derived from HSCs, providing permanent erythrocyte production (**Fig.5A,B**). As outlined before, the first HSC emerges not in the yolk sac, but from major arterial vessels, such as the aorta (AGM), at the beginning of E10.5. HSCs expand in the FL, where the second wave of definitive erythropoiesis begins (**Fig.5B**). This event is marked by an immediate switch from embryonic hemoglobin to the expression of fetal hemoglobin HbF (*58*). Subsequently, by E17.5 and 10-11 weeks in mouse and human embryos respectively, the transition of HSCs to the BM starts, which is still ongoing at the time of birth (**Fig.5B**) (*59-61*). The second wave of definitive erythropoiesis is characterized by the differentiation of lineage-committed cells from progenitors to precursors and finally to mature enucleated and smaller erythrocytes. Most

immature erythroid progenitors in the erythroid lineage are the slowly proliferating BFU-E, in which the receptor for erythropoietin (EPOR) starts to be expressed. However, the erythroid-stimulating hormone erythropoietin (EPO) (*62*), seems not to be required at this stage (**Fig. 6**), as BFU-E derived from FL lacking both alleles of EPO or EPOR could still differentiate towards colony-formation unit-erythroid (CFU-E) (*63, 64*). BFU-E progenitors then appear in the BM and to a lower degree also in the peripheral blood. BFU-E cells divide and develop into rapidly proliferating erythroblasts called CFU-E (*65*). Growth of these cells is fully EPO-dependent, but under normal conditions, they do not circulate in the peripheral blood (*66, 67*). CFU-E progenitors then undergo a series of rapid cell divisions allowing them to progress through different morphologically distinguishable stages from proerythroblasts (Pro-E) via basophilic (Baso-E) to polychromatophilic (Poly-E) and finally to orthochromatic erythroblast (Ortho-E) (**Fig.6**).

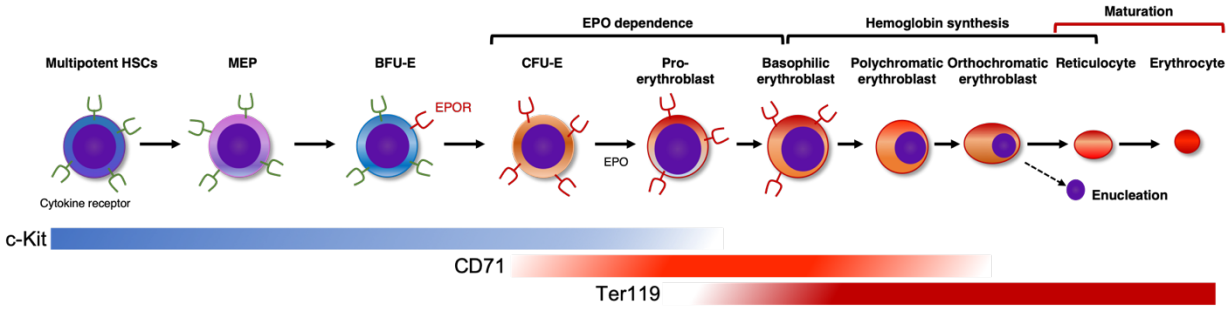


Figure 6. Progressive stages of erythroid differentiation. BFU-E expressing c-Kit differentiated CD71 expressing CFU-E and into Ter119 expressing basophilic, polychromatic and orthochromatic erythroblasts undergo enucleation and become reticulocytes and erythrocytes. Development from CFU-E to basophilic erythroblasts is fully EPO-dependent. Adapted from (*49, 65, 68*).

Erythroid lineage maturation is characterized by an accumulation of hemoglobin and a decrease in cell size and RNA content. It is associated with typical expression changes of surface molecules like increased Ter119 or CD235a in mouse and human cells respectively, as well as decreased levels of c-Kit (CD117) and transferrin receptor CD71 (Transferrin Receptor) (*69-71*). Hemoglobin production starts in the basophilic erythroblast stage and continues until the reticulocytes stage. Late erythroid precursors enucleate, followed by the release of reticulocytes in the bloodstream. These cells consist mostly of cytoplasm and express high levels of hemoglobin. The differentiation of reticulocytes into uniform, biconcave discoid erythrocytes is another complex process, characterized by a reduction of 20% of cell volume and loss of all residual cytoplasmic organelles and reticulin (**Fig.6**) (*72-74*). Notably, reticulocytes as well as erythrocytes do not express EPOR and stop being responsive to EPO (*67*).

In the BM, erythroid differentiation takes place in specific structural units that are called "erythroblastic islands" (*75*). Erythroblastic islands are composed of a central macrophage, also called "nurse cell", which is surrounded by erythroid cells of all different stages, except for mature erythrocytes (**Fig.7**). Within the erythroblastic island, the extruded nuclei from the orthochromatic erythroblasts are phagocytosed and degraded by the central macrophage. Reticulocytes then enter the circulation where they lose their residual internal organelles and mature into erythrocytes.

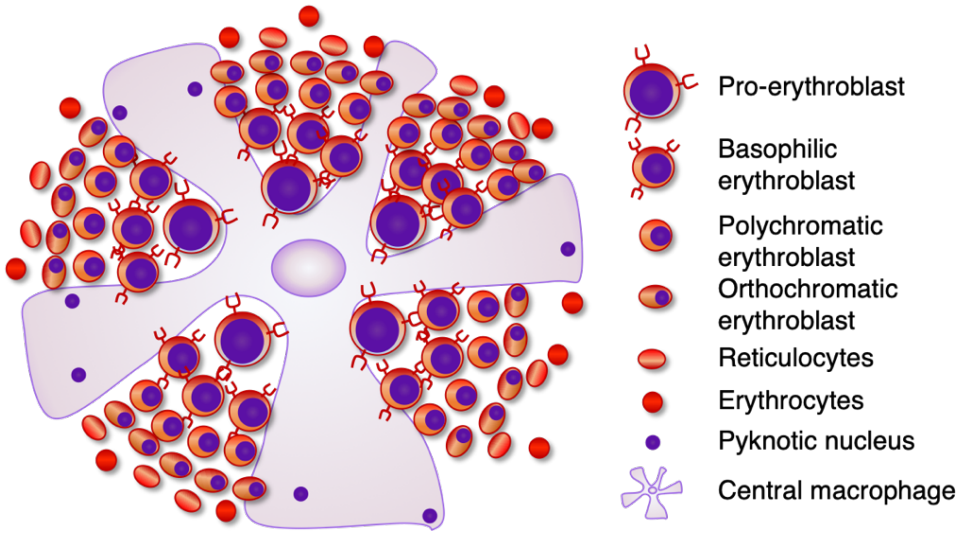


Figure 7. Schematic representation of an erythroblastic island. Erythroid cells from the CFU-E through to the reticulocyte stage are attached to a central macrophage. Reticulocytes detach from the macrophage before leaving the BM and entering the bloodstream. Adapted from (*49, 68*).

1.2.3. Regulation of erythropoiesis

Productive erythroid differentiation is controlled by functional cooperation of transcriptional regulators and cytokines. Overall, one can distinguish transcription factors that control commitment of pluripotent precursors to an erythroid fate, from those that regulate the erythroid progenitor differentiation into erythrocytes (*76*). The commitment of HSCs towards the erythroid fate is mainly determined by transcription factors PU.1 and GATA2. PU.1 is a member of the ETS transcription factor family exclusively expressed by hematopoietic cells and has a crucial role for erythroid homeostasis (*77*).

In a first step of erythroid commitment HSCs and their progeny lose the ability to differentiate into cells of the lymphoid and granulocytic-monocytic lineages, and become MEPs (*68*). MEP growth depends on multiple cytokines including stem cell factor (SCF), interleukin 3 (IL-3), interleukin 9 (IL-9), thrombopoietin (TPO) and granulocyte-macrophage colony-

stimulating factor (GM-CSF). Their differentiation is promoted by increased activity of the KLF1 resulting in the formation of BFU-Es (**Figs.4&8**) (*78-80*). Proliferation of BFU-E growth mainly depends on IL-3, interleukin 11 (IL-11), insulin-like growth factor 1 (IGF-1) and in particular on SCF (*63, 64*). SCF, also known as KIT-ligand, binds to its receptor KIT, which subsequently activates the EPOR by tyrosine phosphorylation, and induces proliferation and maturation towards the next stage of erythroid progenitors, the CFU-Es (*81*). Compared to MEPs and BFU-Es, CFU-Es and proerythroblasts lose responsiveness to SCF and IGF-1 and become dependent on EPO for proliferation and survival (*76*).

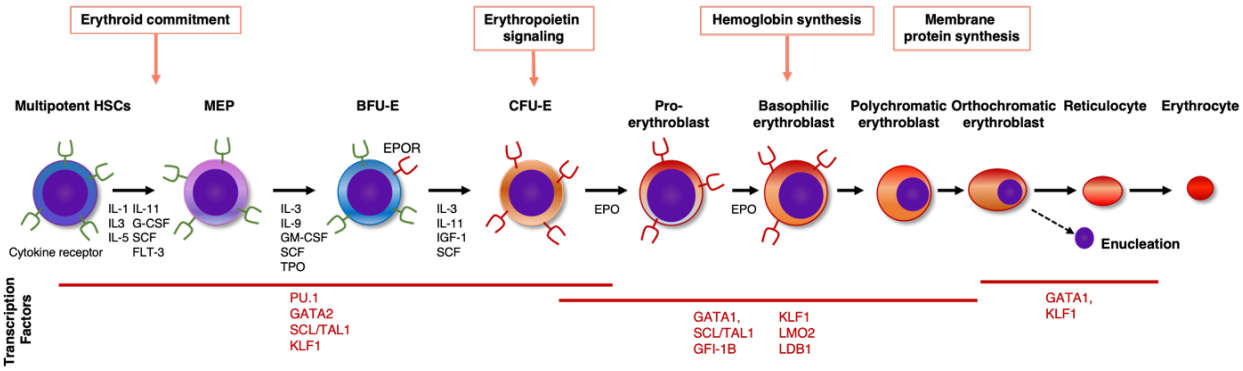


Figure 8. Progressive stages of erythroid differentiation. PU.1 and GATA2 transcription factors (red) are important to determine whether HSCs will progress towards an erythroid or a non-erythroid fate, whereas KLF1 is important to determine whether MEPs will progress towards an erythroid or a megakaryocytic fate. GATA1 and SCL/TAL1 complex members induces differentiation of CFU and pro-erythroblasts. Multipotent HSCs, MEP and BFU-E proliferation and differentiation depend on different cytokines (black). Adapted from (*49, 65, 68, 82*).

In addition to PU.1 and GATA2, differentiation of MEPs into mature RBCs requires the activity of several transcription factors including GATA1, SCL1/TAL1, LMO2, LDB1 (LIM domain-binding protein 1) and KLF1 (*83*). PU.1 has a crucial role for erythroid homeostasis and it has been shown that its deficiency undermines the capacity to sustain an immature phenotype, resulting in impaired proliferation and premature differentiation (*77*). Interestingly, aberrant PU.1 activity is also the consequence of integration of the *Friend virus complex* in vicinity of the *Spi1* gene locus resulting in murine erythroleukemia (*84*). Remarkably, a modest elevation of PU.1 expression is enough to sustain immature progenitors and block differentiation of murine Friend erythroleukemia-derived cells (*85*). PU.1 is essential to keep the balance between self-renewal and differentiation of HSPCs, whereas in erythroid lineage it specifically acts to maintain self-renewal and prevent differentiation. Mechanistic studies proposed that overexpression of PU.1 inhibits erythroid differentiation mainly by physically interacting with GATA1 (*86*). On the other hand, GATA2, exhibits a critical function in the emergence and maintenance of HSCs, and in the specification of early erythroid progenitors

(*87*). In HSCs, GATA2 directly activates its own gene. During early stages of erythropoiesis, however, GATA2 induces *Gata1* expression (**Fig.8**). Once the *Gata1* gene is activated, it is regulated by itself. Interestingly, ChiP-seq analysis revealed that GATA1 and GATA2 share many chromatin sites (*88*). This indicates, that GATA1 replaces GATA2, resulting in repression of *Gata2* transcription and thus also GATA2 targets (*89*). Consequently, this interplay, often referred to as the "GATA-switch", initiates erythropoiesis.

GATA1 activity is indispensable for the expression of erythroid-specific genes at multiple stages of erythroid differentiation (*90, 91*). Furthermore, GATA1 deficiency impaired proerythroblast differentiation leading to anemia and embryonic death in mice (*92, 93*). Interestingly, early erythroid progenitors lacking GATA1 could differentiate into non-erythroid lineages, suggesting that GATA1 expression is not only required for proerythroblast differentiation, but also for erythroid-lineage commitment in general (*94*). Expression of GATA1 is tightly regulated increasing at late BFU-E stage, being highest at CFU-E and proerythroblastic stage (*95*) correlating with the activation of EPOR (**Fig.8**). In fact, GATA1 directly activates the expression of the EPOR gene to support the proliferation and differentiation of CFU-E (*96*). Additionally, GATA1 also activates a large number of erythroid-specific genes in more mature cells, such as the genes encoding α- and β-globin, erythroid membrane proteins and heme biosynthesis enzymes (**Fig.8**) (*97, 98*).

More recent studies have shown that GATA1 initiates an extensive program of gene activation and repression (*90*). Comparison of GATA1 whole-genome occupancy and gene expression identified around 300-700 genes which could be defined as direct GATA1 targets. Most of these genes are involved in the development and the maintenance of the red cell cytoskeleton or belong to the groups of hematopoietic transcription factors and signaling molecules (*88, 99*). Bioinformatic analysis of transcription factor motifs revealed that binding sites of the transcription factor SCL/TAL1 were greatly enriched in the vicinity to GATA1 binding sites (*88, 100*), suggesting that GATA1 may cooperate with SCL/TAL1 to activate gene expression. SCL/TAL1, which is expressed in pluripotent HSCs, multipotent myeloid and lymphoid progenitors, as well as erythroid and megakaryocytic cells (*101*), is a member of the basic helix-loop-helix (bHLH) family of transcription factors that specifically binds to canonical DNA sequences (CANNTG) named E-box. These sequences are found in cis-regulatory elements of erythroid-specific genes (*102*). Precisely, the bHLH transcription factor family is formed by two different groups: the ubiquitously expressed class I bHLHs, also known as E-proteins, and tissue-specific class II bHLHs such as SCL/TAL1. Class II transcription factors heterodimerize with class I members, thereby promoting binding specificity for tissue-specific gene regulatory regions (*103*). In erythroid cells, SCL/TAL1 forms a pentameric complex with class I bHLH E proteins, such as E2A, and also with the cofactors LMO2 and LDB1 (*104*).

LMO2/LDB1 complex acts to bridge the SCL/E2A heterodimer, which is bound to an E-box sequence and GATA1 bound to a GATA element in close vicinity (**Fig.9B**) (*105*). In HSCs, GATA2 interacts also with SCL/TAL1 complex to control expression of genes responsible for HSC maintenance, such as c-Kit (*106, 107*) (**Fig.9A**). Thus, the SCL/TAL1 complex provides the core structure by which the "GATA-switch" controls the regulation of erythroid lineage commitment and differentiation.

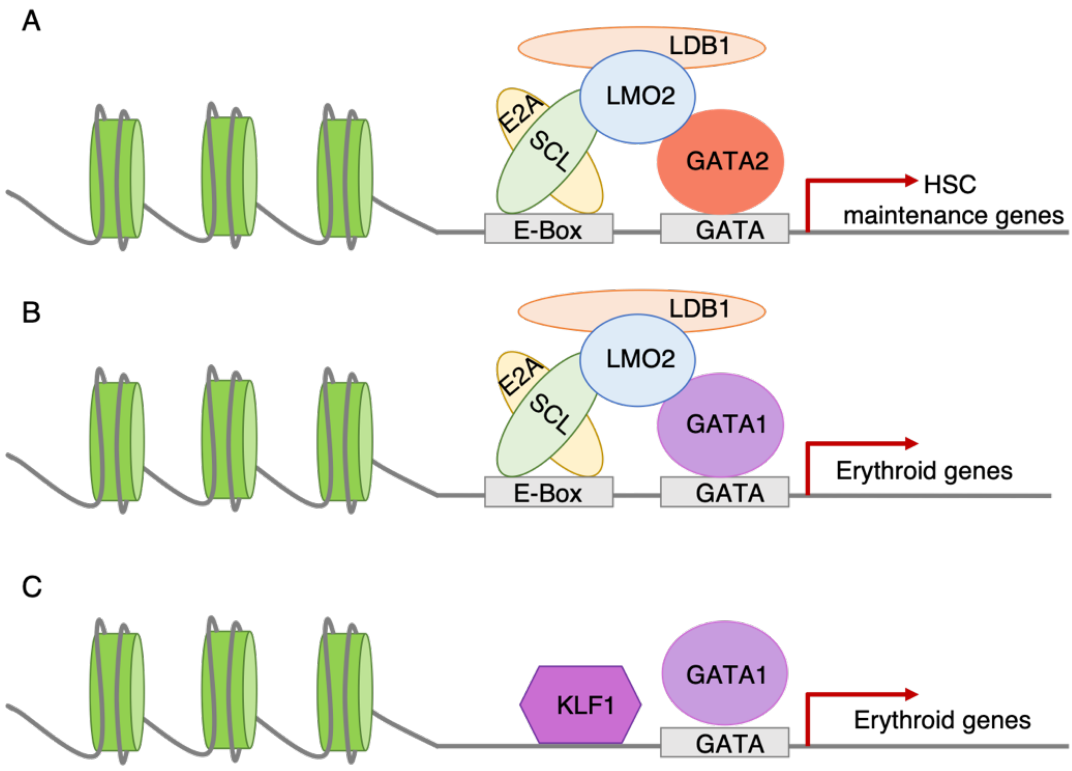


Figure 9. Effect of the GATA-switch on the composition of pentameric complex formed between SCL/TAL1, E2A, LMO2, LDB1 and GATA. Schematic representation of cis-acting regulatory DNA from an erythroid target gene containing an E-Box motif and GATA-binding motif. (A) In HSCs, SCL/TAL1 complexes contain GATA2 and regulate expression of genes related to HSC maintenance. (B) As result of the GATA-switch, GATA1 is incorporated into the SCL/TAL1 complex to activate expression of erythroid genes. (C) GATA1 also activates genes expression in coordination with KLF1. Adapted from (*83, 104, 108*).

Another important erythroid-specific transcription factor, whose expression is activated by GATA1, is KLF1 (aka EKLF) (*109*). KLF1 is a zinc finger transcription factor, which is restricted to the erythroid lineage (*110*). In MEPs, where KLF1 is upregulated, it represses the development of megakaryocytes and promotes differentiation into erythroid lineage (*111*). KLF1 is also important for the regulation of β-globin gene expression; the DNA binding site consensus sequence of KLF1 corresponds to an important motif within the adult β-globin gene promoter (*112*). By comparing KLF1 and GATA1 chromatin binding sites, Tallack et al. discovered that 48% of the KLF1 sites were within 1kb of GATA1 sites, strongly suggesting *in*

vivo cooperation of these two factors (*110*). They also compared GATA1/KLF1-occupied regions with the regions bound by TAL1/GATA1. Rather surprising, they minimally overlapped, suggesting that GATA1 and KLF1 might regulate genes in a different complex from the one containing TAL1 (**Fig.9C**).

Other molecular key players in the regulation of erythropoiesis are enzymes called caspases. Caspases are a family of cysteine proteases, which regulate apoptotic cell death, dismantling most cellular structures (*113*). It has been shown that activated caspase-3 proteolytically cleaves GATA1 resulting in increased cell death and blockage of erythroid differentiation (*114, 115*). Surprisingly, caspases seem activated during normal erythroid differentiation and absolutely necessary to prepare the cells for enucleation and hence enable formation of mature erythrocytes. This indicates, that in this process, GATA1 needs to be protected from caspase-mediated cleavage, guaranteeing that terminal maturation remains conserved (*115*). In fact, GATA1 was shown to be safeguarded by binding to the heat shock protein 70 (HSP70) chaperone. Contrarily, during apoptosis upon EPO deprivation, HSP70 is exported from the nucleus, which allows caspase-mediated cleavage of GATA1 (**Fig.10**) (*115, 116*).

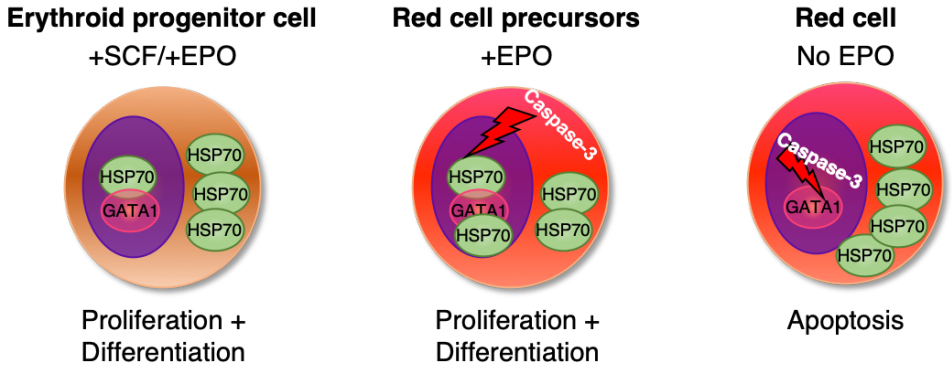


Figure 10. Regulation of terminal erythroid maturation by interaction of GATA1 with HSP70. SCF/EPO activate the proliferation and differentiation of erythroid progenitor cells. In erythroid progenitor cells EPO induces maturation as HSP70 translocates into the nucleus to protect GATA-1 from caspase 3-induced degradation. In the absence of EPO, caspase-3 induces the cleavage of GATA-1 as HSP70 cannot translocate to the nucleus, and as a result, apoptosis is induced. Adapted from (*117*).

1.2.4. Negative regulation of erythroid differentiation

The SCL/TAL1 complex can also repress gene expression by interacting with the transcriptional co-repressors. ETO2 and MTGR1 (*88, 118*). ETO2 (aka CBFA2T3, Core-Binding Factor, Runt Domain, Alpha Subunit 2; Translocated To, 3, or MTG16, Myeloid Translocation Gene On Chromosome 16 Protein) and its related proteins, MTG8 (aka RUNX1T1, ETO, MTG8 or CBFA2T1) and MTGR1 (aka CBFA2T2), define a small family of

transcriptional co-repressors known as the “ETO family” (Eight Twenty One). ETO2 is a regulator of normal hematopoiesis and plays key roles in the maintenance of HSCs (*119*) and productive erythropoiesis (*120*). During erythropoiesis, its expression levels diminish as erythroid progenitors undergo terminal differentiation (*121*). Proteomics-based interaction screens identified several ETO2-binding proteins including the interferon regulatory factor 2-binding protein 2 (IRF2BP2), growth factor-independent 1B (GFI1B) and lysine-specific demethylase 1 (LSD1), which are also transcriptional repressors (*122*). Stadhouders and colleagues revealed that IRF2BP2 is a novel component of the SCL/TAL1 complex and strongly enhances ETO2-mediated transcriptional repression. Additionally, proteomic analysis showed that IRF2BP2 recruits the very well-known NCoR1 co-repressor (Nuclear Receptor Corepressor 1), which binds ETO2/IRF2BP2 erythroid target genes. Taken together, Stadhouders *et al.* defined a novel complex of multiple co-repressor proteins, which represses the SCL/TAL1 complex at the erythroid progenitor stage maintaining erythroid-specific genes in a primed state before terminal differentiation (*122*).

ETO2 was also shown to interact with class II bHLH proteins such as SCL/TAL1 and zinc finger-containing transcription factors. The consequence of this interaction is the suppression of the activator function of SCL/TAL1 on genes that were shown to be crucial for hematopoietic differentiation (**Fig.11A**) (*123*). Examples for ETO2-regulated genes during erythroid and megakaryocytic differentiation are GPA (Glycophorin A) and α-globin, as well as cell-cycle inhibitor genes such as CDKN1A (Cyclin Dependent Kinase Inhibitor 1A). In addition, it has been reported that ETO2 interacts with the GFI-1 and GFI-1B zinc finger transcription factors (*124*), thereby mediating the repression of their targets. Upon terminal erythroid differentiation, the ratio of SCL/TAL1 to ETO2 expression rapidly increases, which leads to higher numbers of active, non-ETO2-bound SCL/TAL1 complexes (*123, 125*). However, this rapid depletion of ETO2 seems not to be permanent and not all target genes are affected, as ETO2 associated to LDB1 represses fetal γ-globin expression in mature adult erythroid cells (*125*).

In addition to their beforehand described function in erythroid development, ETO2 and SCL/TAL1 are also involved in HSC maintenance. In particular, SCL/TAL1 co-expressed with GATA2 regulates ID1 (Inhibitor of DNA Binding 1, HLH Protein) and CDKN1A expression in HSCs and thereby promotes quiescence. Upon ETO2 interaction however, the quiescent function of ID1 and CDKN1A is lost and cells undergo proliferation (**Fig.11B**) (*126*).

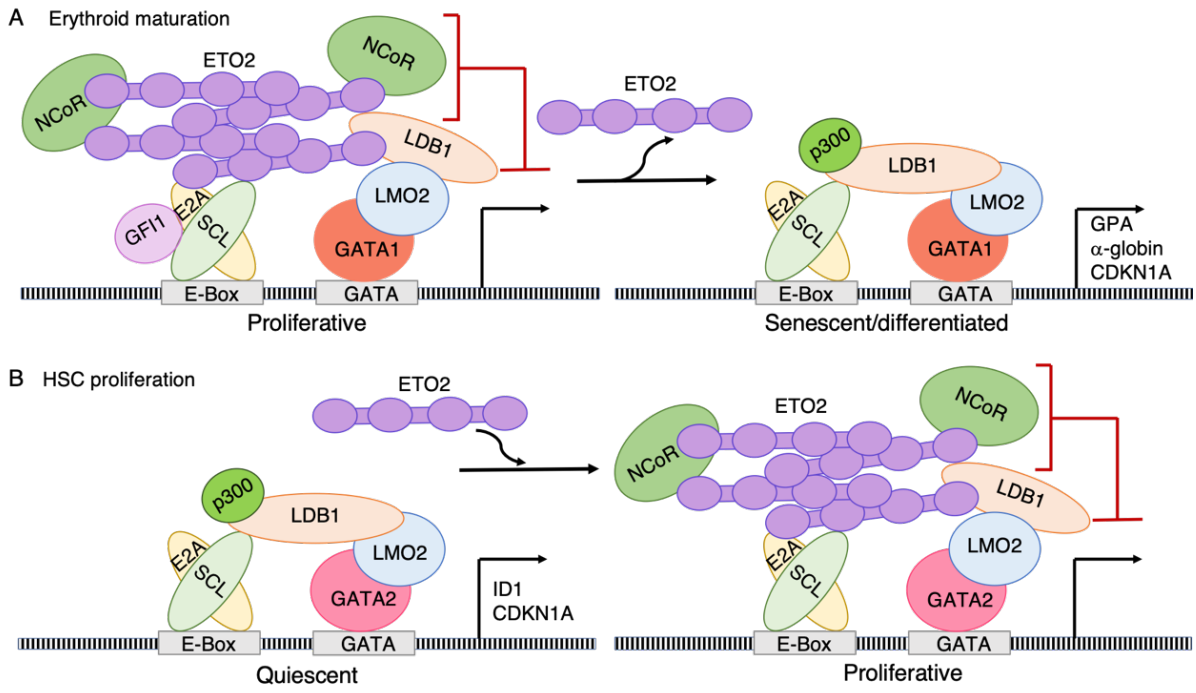


Figure 11. Repression of the SCL/TAL1:E2A complex by ETO2. (A) ETO2 (in purple) interacts with the SCL/TAL1 complex in erythroid progenitors to repress the expression of erythroid genes, such as *GPA, a-globin* and *CDKN1A*. In erythroid progenitors, the SCL/TAL1:E2A complex interacts with ETO2 and Gfi1, but upon erythroid terminal differentiation, both proteins are dissociated from the complex, and the CBP/EP300 coactivator complex is recruited and promotes gene transcription. (B) In HSPCs, the SCL/TAL1 complex preserves quiescence by activating several genes such as *ID1* and *CDKN1A*. ETO2 expression can repress those genes and induce proliferation in HSPCs. Adapted from (*123, 127*).

Other studies have reported that NOTCH signaling plays an important role in erythropoiesis. In particular, SCF was found to exert its function in erythroblasts by inducing expression of NOTCH2. Thus, blocking NOTCH2 neutralized the effect of SCF on erythroid differentiation (*128*). In addition, it was shown that mice with factor X deficiency, a protein that converts GDP-mannose to GDP-fucose, known to be essential for the NOTCH signaling pathway, had a 40% decrease of BM erythroid cells (*129*). Another study found that SCF together with JAGGED1 (Jagged Canonical Notch Ligand 1), the NOTCH2 ligand, increased erythroid colony formation, furthermore supporting a mechanistic link between SCF and NOTCH signaling (*130*). However, others suggested that the activation of NOTCH signaling in embryonic stem (ES) cells inhibits primitive erythropoiesis (*131*), indicating that NOTCH signaling has different roles in different phases of erythroid development. Zeuner *et al.* found that human $CD34^+$ HSPCs transduced with constitutively active NOTCH2 undergo apoptosis upon erythroid differentiation when cultured in an EPO-containing medium. However, in more mature erythroblasts, over-expression of NOTCH2 resulted in increased proliferation and differentiation suggesting a complex role of NOTCH signaling pathway in regulating the

erythroid compartment, probably by restraining the expansion of immature erythroblast, but also by enhancing the production of more mature erythroid progenitors and erythrocytes (*128*).

1.3. Malignant hematopoiesis

Like in most cancers, hematological malignancies are characterized by continuous acquisition of oncogenic lesions leading to the development of transformed hematopoietic cells, which are accumulating in an uncontrolled way thereby sustaining tumor growth at the expense of normal adult blood cell production (*132*). Hematologic malignancies are a very heterogeneous group of disorders, which can be broadly categorized into acute and chronic leukemias, myelodysplastic syndromes (MDS), myeloproliferative neoplasms (MPN) and lymphomas (**Table 1**) (*133*).

Table 1. WHO myeloid neoplasm and acute leukemia classification. Adapted from (*134*).

WHO myeloid neoplasm and acute leukemia classification
Myeloproliferative neoplasms (MPN)
Chronic myeloid leukemia (CML)
Chronic neutrophilic leukemia (CNL)
Polycythemia vera (PV)
Primary myelofibrosis (PMF)
Essential thrombocythemia (ET)
Chronic eosinophilic leukemia, not otherwise specified (NOS)
MPN, unclassifiable
Mastocytosis
Myeloid/lymphoid neoplasms with eosinophilia and rearrangement of PDGFRA, PDGFRB, or FGFR1, or with PCM1-JAK2
Myelodysplastic/myeloproliferative neoplasms (MDS/MPN)
Chronic myelomonocytic leukemia (CMML)
Atypical chronic myeloid leukemia (aCML), BCR-ABL1-
Juvenile myelomonocytic leukemia (JMML)
MDS/MPN with ring sideroblasts and thrombocytosis (MDS/MPN-RS-T)
MDS/MPN, unclassifiable
Myelodysplastic syndromes (MDS)
Myeloid neoplasms with germ line predisposition
Acute myeloid leukemia (AML) and related neoplasms
AML with recurrent genetic abnormalities
AML_ with myelodysplasia-related changes
Therapy-related myeloid neoplasms
AML, NOS
AML with minimal differentiation
AML without maturation
AML with maturation
Acute myelomonocytic leukemia
Acute monoblastic/monocytic leukemia
Pure erythroid leukemia
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
Blastic plasmacytoid dendritic cell neoplasm
Acute leukemias of ambiguous lineage
B-lymphoblastic leukemia/lymphoma
T-lymphoblastic leukemia/lymphoma

Most leukemia patients present with non-specific symptoms such as fever, fatigue, weight loss, bone pain, bruising, bleeding, an increased infection rate and anemia due to the loss of normal hematopoietic cells (*135*). There are two types of leukemias: acute leukemia, which develops clinically very fast and leads to early death when not treated; and chronic leukemia, which develops slowly and patients may live with this disease for several years. Acute leukemia can arise from either the myeloid or the lymphoid compartment and is separated into acute myeloid leukemia (AML) and acute lymphoid leukemia (ALL) respectively.

1.3.1. Impact of the cell of origin in leukemias

In contrast to many solid cancers, the overall number of genetic lesions is rather low in newly diagnosed leukemias, but the functional and phenotypic heterogeneity represents a major challenge when it comes to developing efficient therapies (*136, 137*).

In addition to the genetic heterogeneity, the different biological behavior may also be based on the cell of origin. Studies on acute leukemia found that HSCs or lineage restricted hematopoietic progenitors transduced with particular AML-associated oncogenic mutations like MOZ-TIF2 (*138*), MLL-AF9 (*139*) and MLL-ENL (*140*) were both able to fully initiate AML upon transplantation into irradiated mice. In particular, MLL-AF9 was able to transform and develop leukemia if derived from both HSCs or committed progenitor cells (*139, 141*). Disease latency, aggressiveness and treatment resistance were different depending on the cell of origin (e.g. an HSC origin was associated with earlier and more aggressive disease than when originating from GMPs) (*139, 141*). These findings were somehow conflictive to the beforehand widely accepted idea that leukemia originates from HSCs (*140*) and eventually led to the emergence of the so called “cell-of-origin model” (*142*) (**Fig.12**). This model suggests, that differences in phenotypic manifestation of comparable genetic lesions could be caused by different “cells of origins”. The term “cell of origin” hereby describes the cell that was hit by the first oncogenic event and thus represents the starting point in the hierarchical development of cancer. Depending on the degree of commitment of the cell of origin, similar oncogenic mutations can lead to the development of different disease subtypes (*142*). However, the situation appears more complex: MLL- or MOZ-fusions seem particular as not all genetic driver mutations are able to transform HSCs or more committed progenitors. The tumor cell fate and pathology are influenced by the nature of the cellular target. Only mutations, which confer aberrant self-renewal, comparable to HSCs, are able to reprogram committed progenitors in a way, that they gain the potential to become cells of origin of malignant transformation (*138*). For instance, BCL-ABL fusion associated with chronic myeloid leukemia (CML), compared to MLL-AF9, MLL-ENL and MOZ-TIF2 fusions, was not able to transform committed progenitor

cells (CMPs and GMPs) and transformation and disease development was only achieved from uncommitted HSCs (*138*). Furthermore, it also has been shown that oncogenes expressed in specific lineage-restricted hematopoietic progenitors can give rise to a mixed-phenotype leukemia. Expression of *Myc and Bcl2* (*v-myc avian myelocytomatosis viral oncogene homolog. B-cell CLL/lymphoma 2*) transgenes or the MLL-AF9 fusion in T-cell progenitors resulted in a mixed leukemia consisting mainly of myeloid and partially of T-, and biphenotypic tumor cells. This is especially interesting, as it proposes leukemic plasticity as a driver mechanism of malignant transformation and reveals, that also tissues harboring more committed progenitors can be sources of cells of origin (*143*).

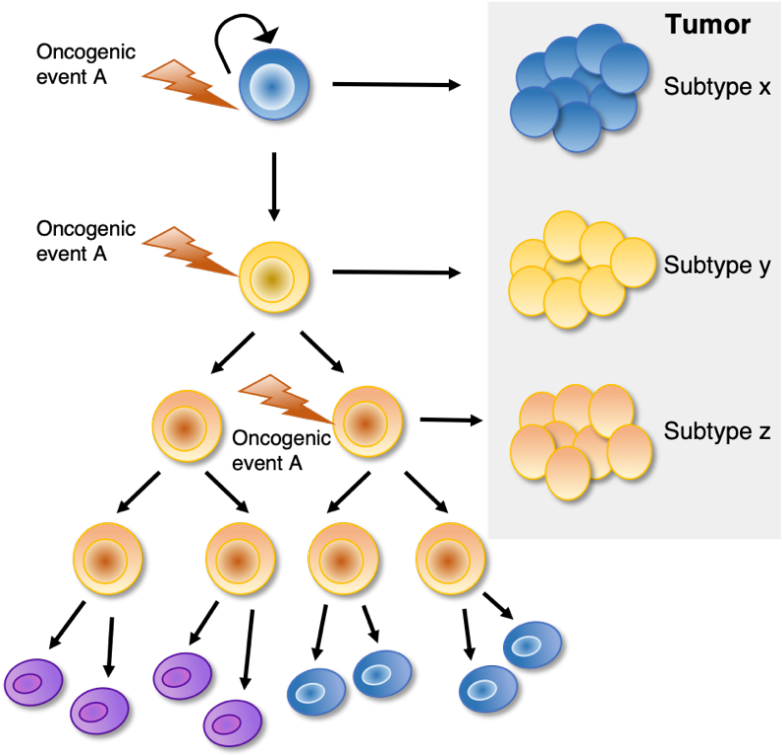


Figure 12. **Cell-of-origin model.** Different populations of cells hit by the same oncogenic event serve as cells of origin for different cancer subtypes arising within the organ or tissue. Adapted from (*142*).

The "cells of origin" acquire further oncogenic lesions and eventually develop into leukemic stem cells (LSCs) (*144, 145*). LSCs are comparable to HSCs, as they express similar surface markers (*146, 147*) and harbor unlimited self-renewal potential. Whether these cells originate from normal HSCs or from more mature progenitors that gained stemness features seems mostly dependent on the nature of the initiating genetic lesion. Furthermore, due to their therapy-resistant characteristics, LSCs are thought to be the origin of relapse in patients, and as such, novel therapies with the capacity to target this population are needed to improve

patients' outcomes. LSCs, that are often also called leukemia-initiating cells (LICs) due to their ability to initiate leukemia after transplantation and give rise to so called leukemic blasts, which are not able to differentiate terminally and thus start occupying the spleen, lymph nodes and bone marrow, thereby impairing the normal function of the hematopoietic system (*148, 149*) (**Fig.13**).

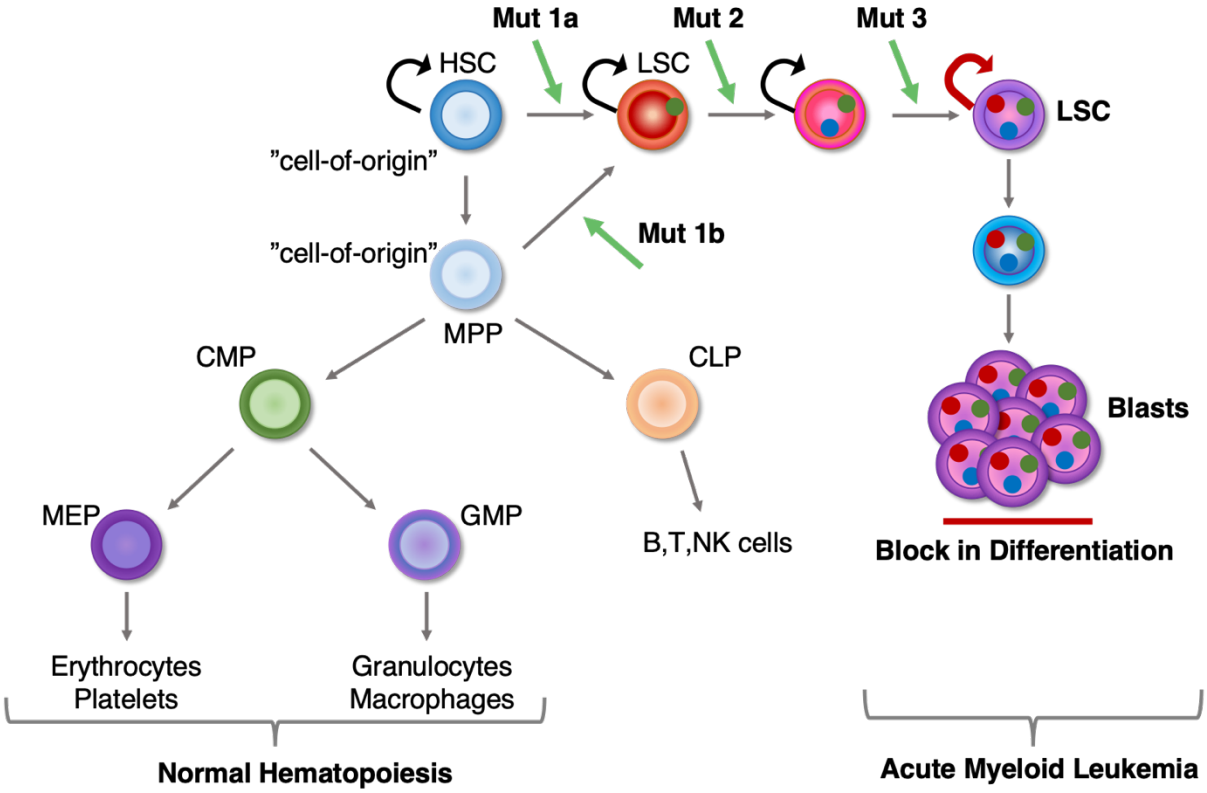


Figure 13. Schematic illustration of normal hematopoiesis and the development of AML. In a first step, either an HSC or a more committed progenitors like MPP is hit by an oncogenic event (Mut1a or Mut1b) and thus represents the cell of origin of leukemic transformation. As it subsequently acquires other disease driving mutations, the "cell-of-origin" develops into an LSC with unlimited self-renewal potential. LSCs eventually give rise to blast cells, which are blocked in differentiation and accumulate in the major hematopoietic organs. Adapted from (*144*).

1.3.2. Acute Myeloid Leukemia

AML is the most frequent form of leukemia in adults in the Western world, accounting for approximately 25% of all patients, and despite improved therapies for particular subtypes, it continues to have the lowest survival rate (*150*). The incidence of AML increases with age and age-related comorbidities clearly limit the options with current therapies resulting in much poorer outcome than in younger patients (*151*). Statistics from the UK showed that 42.8% of patients diagnosed with AML were over 65 years old (*150*). Even though AML is rarely diagnosed before the age of 40 years (*151*) it can affect individuals at any age. In fact, although rare, AML can occur in newborns in their first year of life, in children and in young adults.

Importantly, in pediatric and adolescent populations, leukemia (with ALL more prevalent than AML) comprises almost 50% of all cancers, compared to only 5% to 8% in adults (http://www.seer.cancer.gov). It has been suggested that these relatively high numbers of blood neoplasms in pediatric patients are a consequence of the intensive proliferation phase during fetal development, making it more probable to acquire oncogenic abnormalities (*152*).

1.3.3. Classification of AML

Traditionally, AML is classified by the French-American-British (FAB) system. It was established in 1976 and it defines eight subtypes (M0 through M7) according to the morphology, type, maturation, cytochemical and immunophenotypic behavior of the leukemic blasts (**Table 2**) (*153*).

Table 2. FAB classification of AML subtypes. Adapted from (*154*).

Types	Name	Impaired differentiation
M0	AML, with minimal differentiation	Granulocytic
M1	AML, without maturation	Granulocytic
M2	AML, with maturation	Granulocytic
M3	Acute promyelocytic leukemia (APL)	Granulocytic
M4	Acute myelomonocytic leukemia	Granulocytic and monocytic
M4eos	Myelo-monocytic together with BM eosinophilia	Granulocytic and monocytic
M5	Acute monoblastic leukemia (M5a) or acute monocytic leukemia (M5b)	Monocytic
M6	Acute erythroid leukemias	Erythroid and granulocytic
M7	Acute megakaryoblastic leukemia (AMKL)	Megakaryocytic

These FAB-based definitions were refined by a classification system that was introduced by the World Health Organization (WHO) in 2001, followed by a revised version in 2008 (*155*) and 2016 (*134*).This classification incorporated the genetic abnormalities into the diagnosis of AML. Most cases of AML are subdivided into either “AML with recurrent genetic abnormalities” or “AML, not otherwise specified” (**Table 3**).

Table 3. WHO classification of AML and related neoplasms. Adapted from (*155*).

AML with recurrent genetic abnormalities	**Therapy-related myeloid neoplasms**
AML with t(8;21)(q22;q22.1); RUNX1-RUNX1T1	**AML, Not Otherwise Specified**
AML with inv(16)(p13.1q22)ort(16;16)(p13.1;q22); CBFB-MYH11	AML with minimal differentiation
APL with PML-RARA	AML without maturation
AML with t(9;11)(p21.3;q23.3); KMT2A-MLLT3	AML with maturation
AML with t(6;9)(p23;q34.1); DEK-NUP214	Acute myelomonocytic leukemia
AML with inv(3)(q21.3q26.2)ort(3;3)(q21.3;q26.2); GATA2,MECOM	Acute monoblastic/monocytic leukemia
AML (megakaryoblastic) with t(1;22)(p13.3;q13.3); RBM15-MKL1	Pure erythroid leukemia
Provisional entity: AML with BCR-ABL1	Acute megakaryoblastic leukemia
AML with mutated NPM1	Acute basophilic leukemia
AML with biallelic mutations of CEBPA	Acute panmyelosis with myelofibrosis
Provisional entity: AML with mutated RUNX1	**Myeloid proliferations related to Down syndrome**
Myeloid sarcoma	Transient abnormal myelopoiesis (TAM)
AML with myelodysplasia-related changes	Myeloid leukemia associated with Down syndrome

In the last years, technological advances have introduced genome-wide expression and sequencing analysis. These new approaches have allowed to identify potential biomarkers, which may significantly improve the diagnostic criteria as well as allow targeted therapies to be developed for this previously highly lethal disease (*134*). Therefore, in their latest update 2016, the WHO further modified the classification system for AML, defined principally by clinical features and history, by cytogenetic abnormalities, molecular genetics, morphology and immunophenotype (**Table 4**).

Table 4. The 2016 WHO classification of AML. Adapted from(*156*).

AML with recurrent genetic abnormalities	**Therapy-related myeloid neoplasms**
AML with t(8;21)(q22;q22.1); RUNX1-RUNX1T1	**AML, Not Otherwise Specified**
AML with inv(16)(p13.1q22)ort(16;16)(p13.1;q22); CBFB-MYH11	AML with minimal differentiation
APL with PML-RARA	AML without maturation
AML with t(9;11)(p21.3;q23.3); KMT2A-MLLT3	AML with maturation
AML with t(6;9)(p23;q34.1); DEK-NUP214	Acute myelomonocytic leukemia
AML with inv(3)(q21.3q26.2)ort(3;3)(q21.3;q26.2); GATA2,MECOM	Acute monoblastic/monocytic leukemia
AML (megakaryoblastic) with t(1;22)(p13.3;q13.3); RBM15-MKL1	Pure erythroid leukemia
Provisional entity: AML with BCR-ABL1	Acute megakaryoblastic leukemia
AML with mutated NPM1	Acute basophilic leukemia
AML with biallelic mutations of CEBPA	Acute panmyelosis with myelofibrosis
Provisional entity: AML with mutated RUNX1	**Myeloid proliferations related to Down syndrome**
Myeloid sarcoma	Transient abnormal myelopoiesis (TAM)
AML with myelodysplasia-related changes	Myeloid leukemia associated with Down syndrome

In my PhD project I studied erythroleukemia, whose division into subtypes significantly changed after the 2016 WHO Classification. More details of the erythroleukemia classification will be explained in the chapter 1.4 Acute Erythroid Leukemia.

1.3.4. Molecular aberrations in AML

Earlier cytogenetic studies revealed that about 40% of *de novo* diagnosed AML carry mostly balanced chromosomal abnormalities (*157, 158*). Later, improved molecular tools detected mutations and aberrant gene expression programs that might drive the development of AML. At that time, known molecular drivers of AML were classified in two complementation groups. “Class I mutations” that often result in constitutive activation of signal transduction pathways causing a proliferation and survival advantage of HSPCs. Activating mutations in KIT, FLT3 or N/K-RAS fall into this group. On the other hand, “class II mutations” affect transcriptional regulators or components of the cell cycle machinery and impair differentiation. Alterations (often translocations) affecting RARA, RUNX1, KMT2A (aka MLL1), BAALC (brain and acute leukemia gene), WT1, CEBPA, or NPM1 fall into this group (*159-162*). Mutations

within both groups rarely occur in the same AML patient, whereas mutations between both complementation groups are often found together, indicating the necessity of co-occurrence of class I and class II mutations to drive leukemic transformation (*159*). It has been proposed that class II mutations are the initiating aberrations ("1st hit"), whereas class I mutations are typically appearing later ("2nd hit") (*163*) (**Fig.14**). The initial mutation "1st hit" may occur in HSC and or more committed progenitor cells leading to the blockage in differentiation and formation of LSCs. In the pre-leukemic stage, cells acquire a second mutation "2nd hit", developing leukemia with characteristic influenced by their "origin".

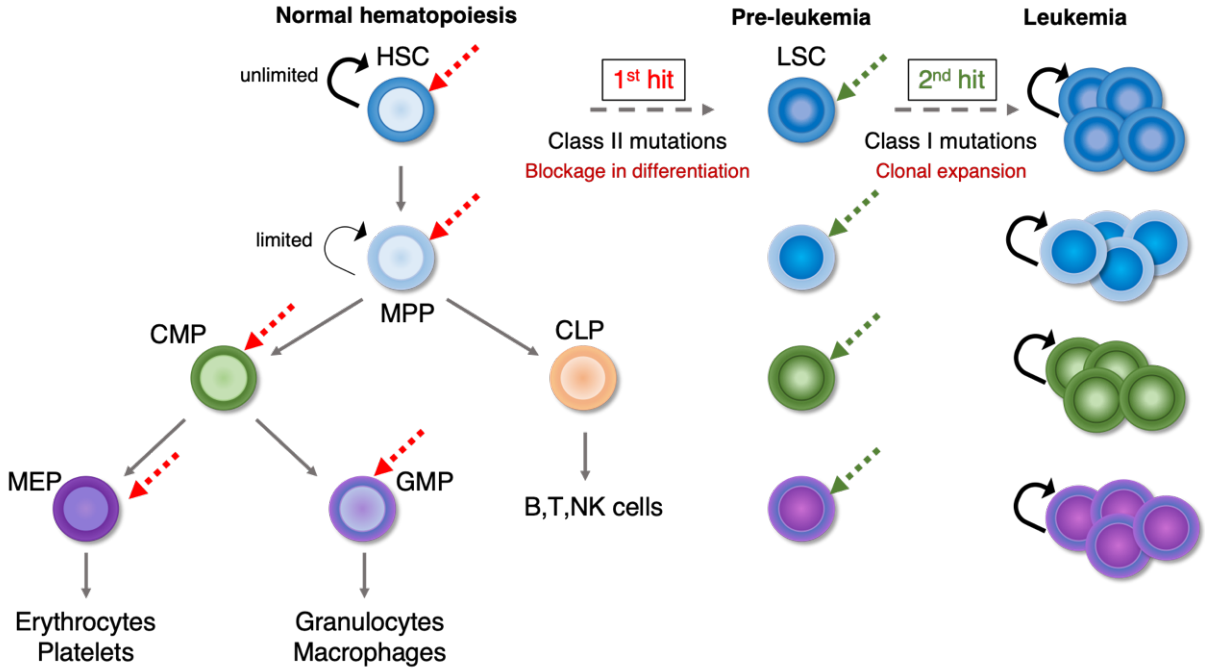


Figure 14. The leukemic stem cell model. In normal hematopoiesis, HSCs with unlimited self-renewal give rise to cells with limited self-renewal capacity (MPPs). MPPs further differentiate towards different cell linages (CLP, CMP) which have lost their self-renewal capacity. Progenitor cells in different differentiation stages can acquire a 1st hit (red arrows), class II mutations that result into blockage in differentiation and formation of LSCs. In the pre-leukemic stage, LSCs self-renew and expand, facilitating the acquisition of further mutations or 2nd hit, finally developing leukemia arising from different clones. Adapted from (*164*).

More recent next generation sequencing (NGS) studies unraveled a more complex genetic landscape of AML. Papaemmanuil *et al*. identified 5235 driver mutations across 76 genes, and 86% of 1540 patients carrying 2 or more drivers (*136*). From this analysis, at least three novel heterogeneous genomic groups for the already defined AML subgroups emerged: AML with mutations in genes encoding chromatin, RNA-splicing regulators, or both (in 18% of patients) (RUNX1, KMT2A^{PTD}, BCOR, SRSF3, ASXL1, STAG2, PHF6, EZH2); AML with TP53 mutations, chromosomal aneuploidies, or both (in 13%); and, provisionally, AML with *IDH2*^{R172}, *DNMT3A* or +8/8q mutations (in 1%).

A public funded project, based on high-throughput sequencing, which aimed to catalogue and discover major cancer-causing genome lesions, is the so called “The Cancer Genome Atlas” (TCGA) project (*165*). Meyer and Levine proposed that AML mutations identified by the TCGA project fell into nine new groups (*166*): transcription factor fusions, NPM1 mutations, tumor suppressor genes (TP53, WT1), DNA-methylation-related genes (TET2, IDH1/2), activated signaling genes (FLT3, KIT, RAS), chromatin-modifying genes (ASYL1, EZH2), myeloid transcription factor genes (CEBPA, RUNX1), cohesion-complex genes (SATG2, RAD21) and spliceosome-complex genes (SRSF2). Even though the TCGA project has revealed a comprehensive mutational landscape in *de novo* AML, the functional characterization of these genes, mutated or alone, is still necessary (*167*).

1.3.5. AML-associated fusion genes

In about 40% of all cases, AML is driven by chimeric fusion proteins, which usually develop through chromosomal translocations. Many of the fusion partners are genes encoding for transcription factors (or co-regulators) and the formed chimeric fusion often retains the DNA-binding motif of the respective wild-type protein. Importantly, the wild-type function of those transcription factors often consists of interacting with co-repressor complexes, meaning that the fusion proteins alter those interactions and the expression of target genes necessary for normal blood cell development (*158*). Overall, chimeric fusion proteins mostly affect common molecular mechanisms potentially increasing the transformation capacity of the cells, such as transcriptional regulation, ETS factor homeostasis, chromatin modification and signaling (*168*).

Ley and colleagues (*167*) showed that in a cohort of 179 AML samples, 118 chimeric fusions were identified in 80 samples (average of 1.5 per sample). 74 of the identified fusion genes were in-frame and many of them were previously described, such as PML-RARα, RUNX1-RUNX1T1, BCR-ABL1 and multiple KMT2A-fusions (Lysine Methyltransferase 2A). Only 15 hitherto unknown in-frame fusion genes were identified, which, however, were not highly prevalent in this group of AML samples. The most common AML aberrations are listed in **Table 5**.

Table 5. AML-associated Oncofusion Proteins. Adapted from (*161, 168*).

Translocation	Prognosis	FAB	Oncofusion protein	Frequency of occurrence (% of AML)
t(8;21) aa	Favorable	M2	AML1-ETO	10%
t(15;17)	Favorable	M3	PML-RARa	10%
Inv(16)	Favorable	M4	CBFb-MYH11	5%
der(11q23)	Variable	M4/M5	MLL-fusions	4%
t(9;22)	Adverse	M1/M2	BCR-ABL1	2%
t(6;9)	Adverse	M2/M4	DEK-CAN	<1%
t(1;22)	Intermediate	M7	OTT-MAL	<1%
t(8;16)	Adverse	M4/M5	MOZ-CBP	<1%
t(7;11)	Intermediate	M2/M4	NUP98-HOXA9	<1%
t(12;22)	Variable	M4/M7	MN1-TEL	<1%
Inv(3)	Adverse	M1/M2/M4/M6/M7	RPN1-EVI1	<1%
t(16;21)	Adverse	M1/M2/M4/M5/M7	FUS-ERG	<1%

Recently, an international consortium analyzed a large cohort of 1023 pediatric AML cases (*169*). They found that infants, children and young adults had highly different cytogenetics and clinical risk classifications. Furthermore, structural abnormalities were also different. In fact, only five structural aberrations were found in more than 5% of samples, including RUNX1, CBFB (Core-Binding Factor Subunit Beta) and KMT2A. In particular, in infants, they found higher numbers of ETO2-GLIS2, KMT2A-AFF1 and MNX1-ETV6 fusion genes. Lastly, in young adult samples, most of the fusions were also found in adults.

Many AML-associated fusion proteins target key hematopoietic regulators resulting in the disruption of the normal differentiation program by different mechanisms. For instance, the RUNX1-RUNX1T1 or NUP98-HOXA9 fusion genes contain a DNA-binding domain from one fusion partner and the transactivation activity from the other one and act as aberrant transcription factors with novel functions, inducing proliferation and blocking differentiation of hematopoietic precursors (*170, 171*). Other fusion proteins, like those involving the histone acetyltransferases CBP/EP300 or KAT6A (aka MOZ) act as aberrant transcriptional co-activators. Essentially, most of these aberrant activities directly regulate transcriptional programs affecting differentiation and related lineage fate decisions, which have been suggested to be the principal mechanisms of leukemogenesis (*172, 173*).

On the other hand, fusion genes can also epigenetically dysregulate DNA and posttranslational modifications of histone proteins, which have an important role in the regulation of gene transcription through chromatin accessibility (*174*). Furthermore, oncogenic fusions can mediate aberrant recruitment of the epigenetic machinery to the promoters of key

differentiation genes. One example is the PML-RARα fusion which drives acute promyelocytic leukemia (APL) (*175*) via recruitment of DNMTs (DNMT1 and DNMT3A, DNA Methyltransferase) and establishes a repressive chromatin environment with low acetylation and high DNA methylation, like H3K27me3 at the fusion binding sites (*176*). Another example is the RUNX1-RUNXT1 fusion, which recruits the NCoR/SMRT/HDAC co-repressor complex to DNA through the RUNX1T1 partner, causing transcriptional impairment by deacetylating histones and creating repressive chromatin structures (*177*). Other AML-associated fusion oncogenes themselves act as epigenetic modifiers such as KMT2A-fusions. Even though they lose the SET domain that mediates H3K4me activity, activation of their target genes is achieved by aberrant recruitment of enzymes that mediate histone H3K4 and H3K79 methylation (*178*).

1.3.6. Current therapies of AML

Despite the epi/genomic AML heterogeneity, the treatment of most patients is uniform, consisting of intensive chemotherapy followed (if possible) by allogeneic hematopoietic stem cell transplantation (*179*). Nonetheless, increasing cytogenetic analyses and mutational screenings help to define prognostic risk groups and offer novel possibilities for improved, more personalized treatment of patients (*180*). An example for the increasing use of targeted therapies are AML cases carrying an activating mutation of the FMS-like tyrosine kinase 3 (FLT3). FLT3 mutations are found in more than 30% of all newly diagnosed AML cases and are, especially in the case of the common FLT3 internal-tandem duplication (ITD), associated with increased risk of relapse and inferior overall survival if treated with conventional strategies (*181, 182*). Recently, potent small molecule FLT3 inhibitors such as midostaurin (*183*), sorafenib (*184, 185*) and gilternitinib (*186*) were developed, which were well tolerated and significantly improved the outcomes of FLT3-driven leukemias (*181*). However, numerous questions regarding FLT3 inhibitor-based treatment strategies, such as the integration of these inhibitors into combination regimens or the emergence of resistant clones, are still a subject to intensive research (*181, 187*).

The best example of successful treatment for a specific group of AML patients is acute promyelocytic leukemia (APL). In >90% of the patients, the disease is driven by the PML-RARα fusion, and a combination of all-*trans* retinoic acid (ATRA) with chemotherapy was able to cure the majority of the patients (*188*). In the last years, several studies even proved, that no traditional chemotherapy is needed for the successful cure of most APL patients. Combinatorial treatment with ATRA and arsenic trioxide (ATO), which bind directly to PML-RARα in a collaborative manner and lead to degradation of the fusion. This degradation then

triggers PML to cooperate with p53; a process definitely clearing the malignant clone. The simplicity, yet astonishing effectiveness of this targeted therapy gives hope for the discovery of similar treatments for other leukemia subtypes (*189-191*).

1.4. From normal to malignant erythropoiesis

As outlined in details before, erythropoiesis is regulated by a complex network of transcription factors, cytokines, oxygen sensors and other factors. Genetic alterations disturbing this network can lead to malignant disorders with aberrant erythroid differentiation such as chronic myeloproliferative neoplasms (MPN), myelodysplastic syndromes (MDS) or acute erythroleukemia (*117*).

Polycythemia vera (PV) is a subtype of MPN that is characterized by clonal expansion of erythroid precursor cells in the BM, resulting in a central and peripheral accumulation of normally differentiating erythropoiesis. Some PV patients eventually progress to erythroleukemia or other AML forms (*192*). In contrast to MPN, MDS or erythroid leukemias are clonal diseases, characterized by a complete or nearly complete block of differentiation in early erythropoiesis (*193*). MDS is characterized by red cell dysplasia, resulting in the combination of excessive proliferation of erythroid precursors (erythroid BM hyperplasia) and peripheral anemia. On the other end of the spectrum is pure erythroid leukemia (PEL), in which the BM is taken over by ≥80% of erythroid progenitor cells.

1.4.2. Classification of erythroid malignancies

Red cell neoplasms can be classified according to their occurrence (either *de novo* or secondary), WHO criteria and the presence or absence of specific molecular markers (*117*). This group of diseases includes PV, which is characterized by the presence of $JAK2^{V617F}$ mutation, MDS with a prominent erythroid compartment, before known as AML M6a, and PEL (**Table 6**).

Table 6. Overview of red cell neoplasms. Adapted from (*117*).

Neoplasm	Key Features / Criteria
Polycythemia vera (PV)	JAK2 V617F and WHO criteria for PV fulfilled
Myelodysplastic syndrome (MDS) with erythroid predominance	BM cell dysplasia >50% erythroid cells and myeloblasts <20%
Pure erythroid leukemia	Proerythroblasts ≥30% and >80% of all BM cells are erythroid cells; myeloblasts <20%
WHO: World Health Organization; BM: bone marrow; PV: polycythemia vera; JAK2: Janus kinase 2	

The classification of erythroid malignancies has changed several times over the last years (*134, 194*). PV has always been, and still is, characterized by erythrocytosis in presence of an activating $JAK2^{V617F}$ mutation and in around 40% of patients, some degree of leukocytosis and thrombocytosis. However, the thresholds to classify PV changed over time. Recently for instance, the revised WHO definition adapted the threshold levels of hemoglobin according to sex specific differences in hemoglobin expression, so that the PV diagnosis is now different for male and female patients (*134, 194*).

Erythroid leukemias are a heterogenous group of malignancies which can emerge as primary (*de novo*) or secondary leukemia, following an MDS or MPN. MDS and erythroid leukemia classification has also been improved and adapted over time (see details below).

1.5. Acute erythroid leukemia

Acute erythroid leukemia (AEL) is a rare disease accounting for less than 1% of all *de novo* occurring AML (*195*). AEL was first described by the Italian hematologist Giovanni Di Guglielmo in the early 1920s, and it was separated from myeloid leukemia based on the accumulation of erythroid cells in the BM. Di Guglielmo distinguished two subtypes of the disease: a pure acute (aka. Di Guglielmo disease) and more chronic form (Di Guglielmo syndrome) (*196, 197*). The FAB working group revised the criteria of AEL classification and re-defined AEL (aka AML-M6) as requiring at least 30% of the non-erythroid compartment to be myeloblasts and >50% of erythroid lineage cells present in the BM (*153*). However, this classification did not appear to be precise enough for pure erythroblastic forms of AEL (*198*). In 2001, the WHO categorized AEL into two subtypes taking into consideration the pure erythroblastic forms: erythroid/myeloid leukemia (FAB-M6a) and pure erythroid leukemia (FAB-M6b or PEL). This new classification characterized M6a by ≥50% of erythroid cells of total nucleated BM cells and ≥20% of myeloblasts in the non-erythroid cells compartment. M6b or PEL was described by more than 80% of all nucleated BM cells corresponding to primitive erythroblasts (*199*). However, the morphological and immunophenotypic features of AEL are very similar to other diseases such as MDS and AML with MDS-related changes (AML-MRD). Thus, it has been suggested that these diseases might represent a continuum of the same disease (*200*). The 2008 WHO classification described AEL as an exclusion diagnosis and categorizes them into three different groups: AEL with multilineage dysplasia, therapy-related and MDS; AEL divided into AML, not otherwise specified (NOS) (erythroid/myeloid leukemia type) and AML, NOS (pure erythroid leukemia type) (*155*). Yet this new classification did not clearly distinguish MDS from AEL, which have shared molecular and cytogenetic

characteristics. Thus, the WHO came up with yet another classification in 2016 based on the percentage of erythroid blasts among all cells, eliminating the nonerythroid blast cell count as a disease definer. The 2016 WHO classification in fact eliminates the AML (erythroid/myeloid leukemia type) group and categorized it under MDS. PEL now remains the only type of AEL, characterized by >80% erythroid precursors, >20% erythroblasts and minor myeloblast populations in the BM (*134*) (**Fig15**). With this new classification, the WHO tried to avoid diagnostic overlap between PEL and MDS, since at least 20% myeloblasts must be present for a diagnosis of AML with myelodysplasia-related changes.

PEL accounts for less than 5% of acute erythroid leukemia, and patients with this disease have a very poor prognosis, with a median survival of about 2-3 months. PEL occurs as "*de novo*", arises secondary from other diseases such as MDS or arises therapy-related (*195*).

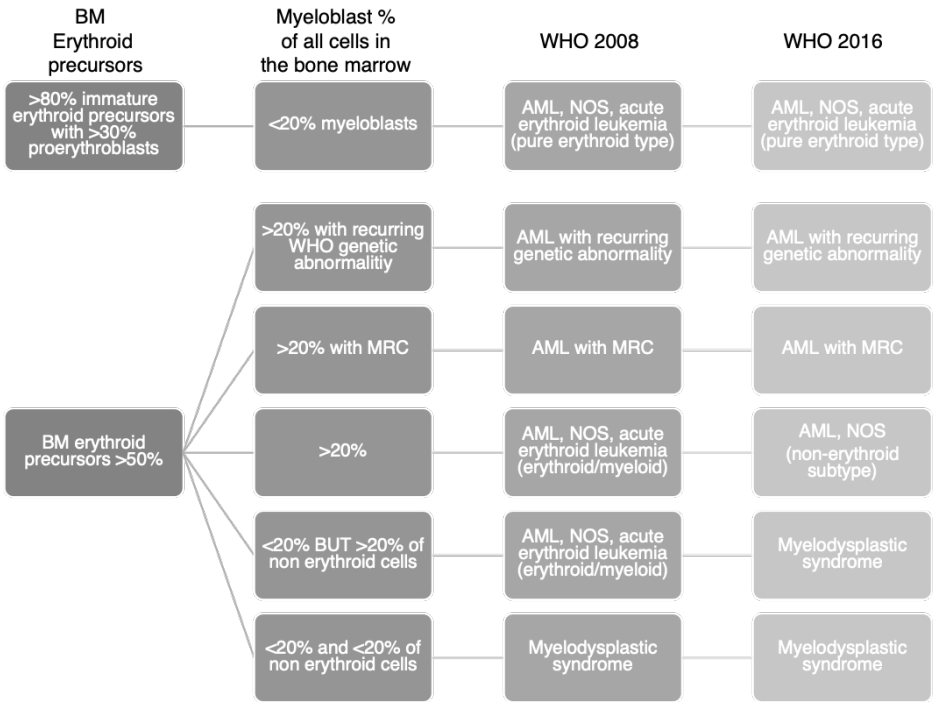


Figure 15. Classification of erythroleukemia. PEL is the "true" form of erythroleukemia based on the WHO 2016 classification. MRC-myelodysplasia related changes, NOS-not otherwise specified. Adapted from (*201*).

1.5.2. Morphologic features of erythroleukemia

In PEL, impaired differentiation and aberrant proliferation result in the accumulation of immature erythroid precursors and absence of expanding myeloblastic cells. Characterization of a cohort of 16 patients showed that peripheral blood from PEL patients had circulating erythroblasts (**Fig.16A**), with the BM containing a median of 80% erythroblasts compared to total nucleated cells (**Fig.16B**) and myeloblasts representing a median of 3% of all BM cells (*195*).

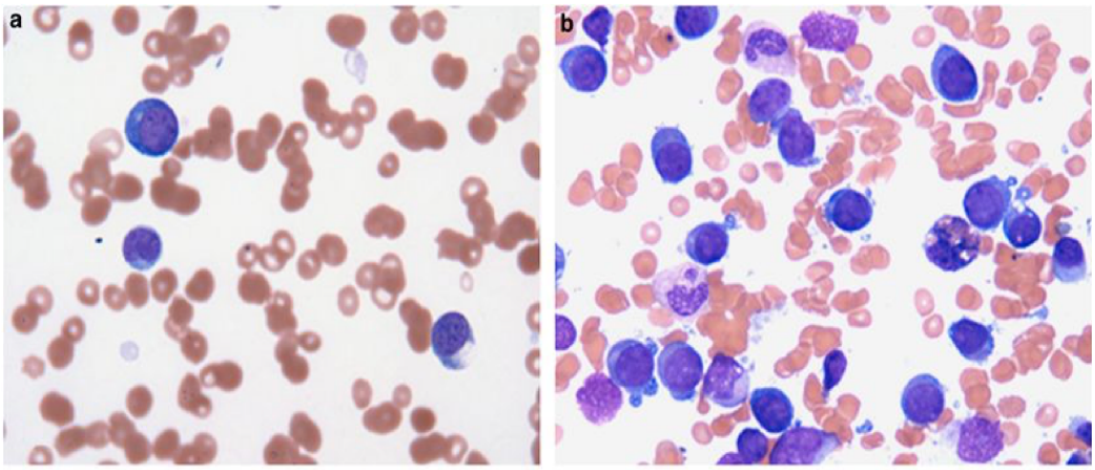


Figure 16. Morphologic characteristics of PEL. (A) Peripheral blood presenting circulating erythroblasts. (B) Bone marrow smear with many erythroblasts. Pictures were taken from (*195*).

Proerythroblasts are the earliest morphologically recognizable form of erythroid progenitor cells characterized by their relatively large size, blue cytoplasm and a central, round nucleus, non-condensed chromatin and eventually several nucleoli (*193, 195, 202-204*). Interestingly, it seemed that the erythroid differentiation block associated with AEL could occur at different stages. In some cases, it affects the most immature erythroid cells such as BFU-E, while in others more mature cells such as CFU-E are predominant. Apart from the number of erythroblastic cells in the peripheral blood or BM, diagnosis of PEL is also based on immunohistochemical analysis or flow cytometry, which help to identify cells based on cell markers expressed in immature erythroid cells, such as CD71 or E-cadherin (*193, 204*).

1.5.3. Models of erythroleukemia

The earliest models of erythroleukemia emerged from studies addressing the activity of putative tumor viruses in chicken and mouse cells.

The avian erythroblastosis retrovirus (AEV) was found to induce erythroleukemia in young chicken. AEV encodes two oncogenes: *v-erbA*, an oncogenic variant of the thyroid hormone

receptor alpha (TRα), and *v-erbB*, an oncogenic variant of epithelial growth factor (EGF) receptor (*205, 206*). It has been shown that AEV blocks terminal differentiation of committed erythroid progenitor cells. As a mechanism driving this blockage, it has been suggested that *v-erbA* interacts with SCF, the receptor c-KIT, and *v-erbB*, thereby hampering erythroid differentiation and driving the development of erythroleukemia (*207, 208*). Due to the beforehand described characteristics, a chicken erythroblastic cell line transformed by AEV retrovirus (HD3) has been widely used as a model to study erythroid differentiation.

In 1957, the virologist Charlotte Friend observed induction of an erythroleukemia-like disease by transfer of cell-free extracts from Ehrlich carcinoma cells, derived from an undifferentiated mouse tumor, in which cells carried particular cellular inclusions that suggested a viral origin. Later work by Friend and Moreau-Gachelin and colleagues revealed that the disease was induced by a mouse leukemia virus (Friend virus complex) that results in aberrant expression of the transcription factor PU.1 due to integration close to the *Spi1* gene locus. However so far, no functional link has been found between human erythroleukemia and viral infection. Proerythroblasts derived from the spleen of susceptible mice infected with Friend virus have been immortalized and established as a cell line model known as murine erythroleukemia (MEL) cell line and extensively used to study erythroid differentiation *in vitro* (*209*). Maturation of these cells is arrested at the proerythroblast stage. However, MEL cells can differentiate into more mature and hemoglobinized precursors upon chemical induction with polar compounds like DMSO (dimethyl sulfoxide) or HMBA (hexamethylene bisacetamide) (*210*), associated with decreased PU.1/Spi1 expression (*211*). The transformed erythroblastic state of MEL cells is based on impaired differentiation by deregulated expression of PU.1; and aberrant proliferation due to the acquisition of additional genetic events such as Tp53 mutations (*211*). These characteristics are shared with the already described double-hit model in AML, suggesting that the molecular oncogenesis of erythroleukemia might be also driven by two consecutive collaborating mutations (*163*). In fact, an activating mutation of the KIT receptor tyrosine kinase was found to cooperate with PU.1/Spi1 overexpression in progression of erythroleukemia in mice (*212*).

Besides the beforehand mentioned HD3 and MEL cell lines, numerous other erythroleukemic cell lines exist. Most of these, such as K562, HEL, TF1, KMOE, F36P, OCIMI or OCM2, were derived from patients with CML, *de novo* or secondary AEL or PEL and are mostly characterized by an erythroid morphology and immunophenotype (*213*). Despite their heterogeneity, these cell lines provide a platform to study cellular and molecular mechanisms of acute erythroleukemia.

Another example of an erythroleukemia model was developed when studying progression of $JAK2^{V617F}$ carrying MPN (*214, 215*). In some MPN patients, progression to AML is

associated with somatic alterations such as TET2, IDH2 and TP53. Kurokawa and colleagues retrovirally transduced mouse *Tp53*$^{-/-}$ HSPCs with JAK2^{V617F} and transplanted them into lethally irradiated recipients to demonstrate potential cooperation between JAK2^{V617F} and loss of *Tp53*. Mice developed disease between 50 to 100 days after transplantation, and showed hepatosplenomegaly, infiltration of CD71^{+}/Ter119^{+} erythroid progenitor cells and erythroblast-like cells in the peripheral blood. Furthermore, erythroleukemic cells from diseased mice could be serially transplanted. These results showed that loss of Tp53 is enough to induce leukemic transformation in JAK2^{V617F} MPN.

Another rather unexpected erythroleukemia model emerged from mice with engineered reduced expression of GATA1. Yamamoto and colleagues (*216, 217*) targeted the murine Gata1 promoter and thereby reduced the expression of Gata1 erythroid transcripts (Gata1$^{1.05/x}$) by 95%. Mutant embryos showed a reduction in the number of erythroid cells and CFU-E formation in the FL and E9.5 embryos already showed an accumulation of primitive erythroid progenitors. At the age of 5 months, some mice presented with splenomegaly, anemia, thrombocytopenia and accumulation of proerythroblasts and megakaryocytes in the spleen. Importantly, the development of leukemia in Gata1$^{1.05/x}$ mice was completely rescued by transgenic expression of Gata1. This model suggests that GATA1 depletion is important to maintain a transformed phenotype of erythroid progenitors (*216*).

BCR/ABL is a fusion gene found in more than 90% of patients with CML and in up to 15% of adult patients with *de novo* ALL (*218*). BCR-ABL+ CML can eventually also progress to AEL. Tenen and colleagues (*219*) retrovirally expressed this fusion in FL-derived HSCs lacking one or both copies of the myeloid transcription factor C/EBPa (C/EBP$\alpha^{+/-}$ or C/EBP$\alpha^{-/-}$). Mice carrying at least one allele of C/EBPα developed the classical MPN-like phenotype. In contrast, mice lacking C/EBPα developed transplantable erythroleukemia with erythroid progenitor on peripheral blood smears and infiltration of BM and spleen (*219*). Interestingly, the human erythroleukemia cell line BCR-ABL^{+} K562 was shown to express high levels of GATA1 but no C/EBPα (*220*).

1.5.4. Molecular aberrations in erythroleukemia

1.5.4.1. Hematopoietic transcription factors in experimental models

Like in other AML forms, aberrant differentiation of erythroleukemia cells seems to be the consequence of aberrant activity of transcriptional hematopoietic master regulators. As briefly mentioned before, aberrant PU.1 expression (e.g. by integration of Friend virus) represses the activity of GATA1 and results in blocked erythroid differentiation of MEL cells (*211*). PU.1 expression decreases during normal erythroid maturation, however, in AEL cells, PU.1 levels

seem to be maintained, suggesting that GATA1 cannot fulfill its function as a PU.1 repressor (*221*). Therefore, it has been speculated that the *Spi-1* gene is an erythroid proto-oncogene and that its aberrant expression in erythroid cells, together with an additional event such as inactivation of the *TP53* tumor suppressor gene, results in malignant transformation of the cells (*222*).

GFI-1B and LSD1 are transcriptional regulators that also have been implicated in erythroid leukemia (*223, 224*). In particular, GFI-1B expression was increased in patients with erythroid and megakaryoblastic leukemia, but not in other AML types (*223*). GFI-1B regulates erythroid progenitor cell growth and differentiation, and an overexpression of this gene results in an increased proliferative capacity of erythroid progenitors. In fact, GFI-1B was shown to interact directly with LSD1, which represses GFI-1B target genes and hinders cells to differentiate towards erythrocytes. Inhibition of LSD1 resulted in disruption of GFI-1B/LSD1 interaction and cells differentiate, suggesting a specific mechanism against AEL (*223, 224*). Most recent studies indicated that LSD1-mediated repression of a GFI super-enhancer may play an important role in maintaining a transformed state in erythroleukemia (*225*).

FLI-1 belongs to the ETS transcription factor family, is expressed in multiple hematopoietic lineages and has been associated with erythroleukemia (*226*). The expression of FLI-1 was found to be induced by Friend Murine Leukemia Virus (F-MuLV) integration, which leads to a rearrangement of the *Fli-1* locus (*222, 227*). Specifically, 75% of the erythroleukemia cell clones targeted by *F-MuLV* present with rearranged *Fli-1* sites. Activation of FLI-1 seems to disrupt erythroid signaling pathways, such as EPO and SCF, thereby impairing erythroid differentiation. Furthermore, FLI-1 was shown to alter the expression of several genes that also control erythropoiesis such as BCL2, GATA1 and SHIP-1 (SH2 Domain-Containing Inositol Phosphatase 1) (*228-231*). A recent study even showed that exogenous expression of FLI-1 induces de-differentiation of erythroid progenitors towards more immature phenotypes (*232*). The same study also revealed that FLI-1 overexpression induces changes in the expression of important hematopoietic transcription factors, such as SCL/TAL1, GATA1, RUNX1, inducing self-renewal and favoring the transition to a more immature erythroid stage (*230, 232*). These data suggest that FLI-1 plays a critical role in erythroid differentiation and self-renewal and that aberrant expression may have important implications in the pathogenesis of erythroleukemia.

ERG, another ETS transcription factor was shown to play a crucial role in the regulation of normal HSCs, but also to harbor oncogenic potential when aberrantly expressed. It has been demonstrated that ERG is essential for HSC self-renewal and for definitive hematopoiesis (*233, 234*). Notably, overexpression of ERG has been associated with poor prognosis in AML and T-ALL (*235*). In mice, retroviral overexpression of ERG resulted in an erythroleukemia-like

disease *in vivo*, whilst *in vitro*, the leukemic cells showed an erythroid and megakaryocytic phenotype (*236*). These results suggest that ERG, like other ETS transcription factors related to erythroleukemia (PU.1, FLI-1), plays a crucial role in disease development.

1.5.4.2. Molecular lesions in primary patient-derived erythroleukemia cells

Targeted and genome-wide next generation sequencing allowed to define the genomic landscape of AEL patients (*237-242*) (Fagnan *et al,* submitted). These studies suggested that AEL patients can be grouped according to their mutations. Despite a significant heterogeneity between patients and cohorts, mutations mostly involved the TP53 tumor suppressor, epigenetic regulators (such as DNMT3A, IDH1/2 Isocitrate dehydrogenase 1/2, and TET2 Tet Methylcytosine Dioxygenase 2), transcription factors, splicing/RNA processing, tyrosine kinases, DNA repair genes, cohesions, NPM1 (Nucleophosmin 1) and others. Particularly, TP53 mutations were not only the most prevalent lesion in AEL (affecting about 30% of the patients) but present in almost all PEL patients (*237, 238, 242*). Most of the TP53 mutations were missense affecting the DNA binding domain (*242*). TP53 mutations are much less prevalent in MDS or non-erythroid AML cases often associated with secondary or therapy-related cases (*243*).

Cytogenetically, AEL is often associated with complex karyotypes, with frequent abnormalities such as monosomy of chromosome 7 and deletion of chromosome arm 5q (*244-247*). In rare cases of pediatric AEL, Iacobucci and colleagues also found fusion genes previously found in other AML forms in children (e.g. NUP98-NSD1, NUP98-JARID1A) in rare cases of pediatric AEL (*242*). In addition, another rare fusion consisting of the transcription factors GATA1 and MYB (V-Myb Avian Myeloblastosis Viral Oncogene Homolog), already described before in pediatric acute basophilic leukemia, was identified (*248*).

Chromosomal translocations in erythroid regulators, epigenetic regulators and signaling pathways, such as in GATA1, KMT2A, EPOR and STAT3 (Signal Transducer And Activator Of Transcription 3), were also found in AEL patients. However, they were very rare. Nevertheless, *in vitro* studies revealed that many of those fusion genes were not sufficient to develop erythroleukemia *in vivo,* suggesting the need of a "2nd hit" to drive leukemogenesis (*242*). Therefore, human AEL, like other non-erythroid AMLs, is most likely driven by at least two functionally cooperating mutations. Mutations of epigenetic regulators or transcription factors may block erythroid differentiation and a second hit most probably in signal transduction pathway genes, including TP53, would provide aberrant self-renewal leading to leukemic transformation (*201*) (**Fig.17**).

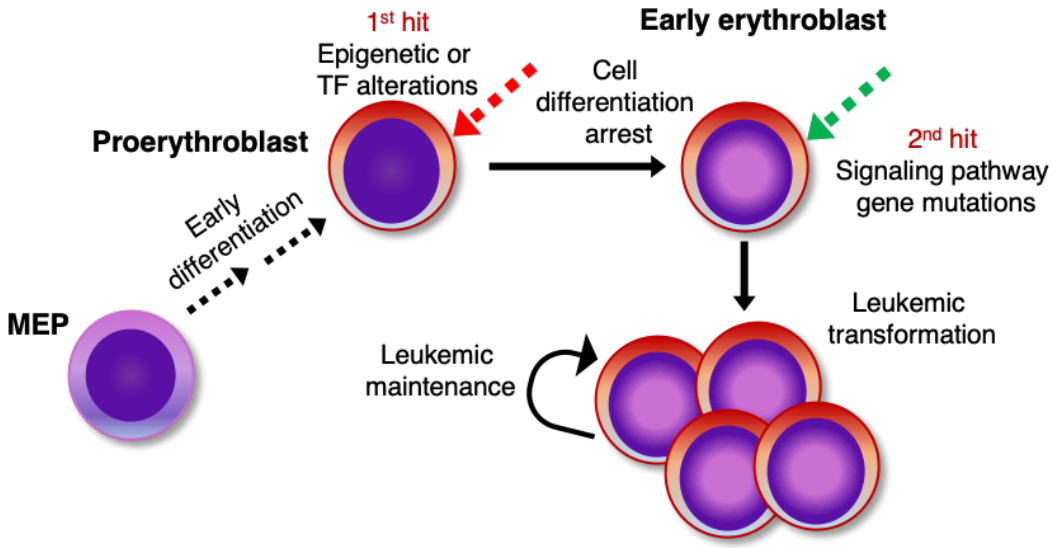


Figure 17. A genetic two-hit model of AEL. Proerythroblasts acquire an epigenetic or transcription factor (TF) alteration (1st hit), leading to cell differentiation arrest. Hyper-proliferative early erythroblasts acquire a second hit or mutations in a signaling pathway gene such as TP53, driving leukemic transformation and maintenance. Adapted from (*201*),

Two recurrent chromosomal translocations, never seen in other non-erythroid AML forms, were found in rare pediatric PEL cases: t(1;16)(p31;q24) and t(11;20)(p11;q11), resulting in the expression of NFIA-ETO2 (aka NFIA-CBFA2T3) (*249-251*) and ZMYND8-RELA (*252*) fusion genes respectively. Their early (mostly affecting infants) and exclusive association with erythroleukemia strongly suggested that they may act as specific inducers of the disease. Therefore, the main objective of my doctoral thesis was to study the role of NFIA-ETO2 and ZMYND8-RELA fusion genes in the pathogenesis in PEL.

1.5.4.3. NFIA-ETO2

The chromosomal translocation t(1:16)(p31;q24) was identified in four independent cases of young children diagnosed with PEL (*249, 250, 253, 254*). This translocation fuses exon 6 of the *nuclear factor IA (NFIA)* gene to exon 3 of the *ETO2* gene (**Fig.18**). The NFIA-ETO2 fusion gene leads to an open reading frame encoding for a chimeric protein of 811 amino-acid residues, containing 208 residues of NFIA and 603 residues of ETO2. The NFIA part of the fusion contains two annotated domains: the DNA binding domain (DNA-BD) and the MAD (Mother against Decapentaplegic) homology I Dwarfin-type (MH1). The ETO2 part contains all annotated functional domains of the wildtype protein: the four regions of sequence homology conserved from the *Drosophila melanogaster gene Nervy*, called “Nervy homology regions” (NHRs), termed NHR1-NHR4.

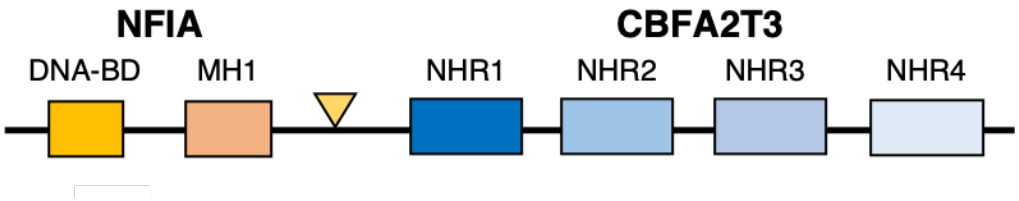


Figure 18. Schematic of the NFIA-ETO2 fusion. The NFIA-ETO2 contains the DNA-binding domain (DBD) and the Mad-homology domain (MH1) of NFIA (in yellow) fused to the Nervy-homology domains (NHR1-4) of ETO2 (in blue). Adapted from (*249, 250*).

1.5.4.3.1. NFIA

Nuclear Factor I-A (NFIA) belongs to a family of transcription factors consisting of four members (NFIA, NFIB, NFIC, NFIX). These genes encode for proteins that have a highly conserved N-terminal DNA-binding/dimerization domain, but differ on the C-terminal transactivation/repression domain, giving unique functions to each NFI family member (*255-257*). NFI proteins bind as hetero- or homodimers to the double-stranded palindromic DNA recognition sequence *TTGGC(N5)GCCAA*, thereby activating or repressing gene transcription (*257*).

NFIA has been shown to be part of a negative-feedback loop of granulocytic differentiation. It directly binds to the promoter of a myeloid regulatory miR-223 and maintains its expression at low levels and thereby suppresses granulocytic differentiation. On the other hand, miR-223 represses NFIA translation to start differentiation (*258*). Another study revealed that, in hematopoietic progenitor cells PU.1 activates the expression of miR-424 and its upregulation stimulates monocyte differentiation by repressing *NFIA* expression. Repression of *NFIA* led to the activation of differentiation-specific genes such as *M-CSFR* (macrophage colony-stimulating factor receptor) (*259*).

NFIA has also been shown to have a critical role in controlling erythroid/granulocyte fate (*255*). In normal human hematopoiesis, NFIA exhibited a marked lineage-specific expression pattern with upregulation in erythroid progenitors and suppression in granulocytes. *In vitro*, NFIA overexpression or knockdown accelerated or blocked erythropoiesis, respectively. Mechanistically, NFIA bound to the β-globin proximal promoter and activated transcription, allowing erythroid differentiation. On the other hand, NFIA negatively regulated expression of the granulocyte colony-stimulating factor receptor (G-CSFR) gene necessary for production and function of granulocytes. Interestingly, Starnes *et al.* also found that NFIA is capable to not only activate erythroid differentiation, but also to sustain it. In erythroid progenitor cells cultured in medium with low EPO stimulus, which is necessary for erythroid differentiation, NFIA overexpression restored the erythroid maturation. This result suggests that NFIA might directly target key molecules involved in the erythroid gene program (*255*). Further gene

expression studies revealed that NFIA promotes erythroid differentiation of normal human CD34+ HSPCs and cells from chronic myeloid leukemia (CML) patients (*260*). Taken together, these studies suggested that NFIA is an important regulator of myeloid-erythroid fate decisions in hematopoiesis.

1.5.4.3.2. ETO2

The ETO2 gene contains 15-exons encoding for two transcript variants: that differ in their 5' UTR and first exon, producing two different protein products (*261*). The longer isoform, ETO2a, codes for a 653-amino acid protein and contains a unique nucleolar localization sequence. Therefore, ETO2a is found in the nucleolus and nucleoplasm (*262*) while ETO2b, the shorter isoform coding for a 567-amino acid protein, is localized only in the nucleoplasm (**Fig.19**) (*263*). ETO2 contains four defined domains (Nervy homology domains, NHR1-NHR4), whose function is defined by interactions with DNA-binding transcription factors as well as with transcriptional corepressor complexes (*264*).

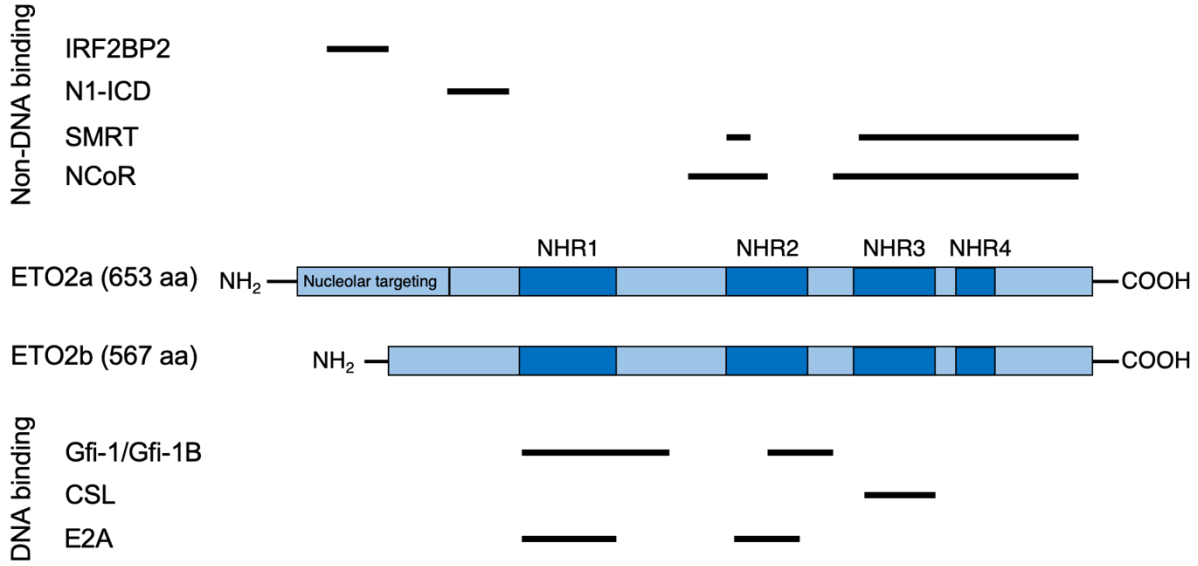


Figure 19. Schematic representation of ETO2 proteins and their domain structures. The black lines represent the binding sites between the ETO2 proteins and other transcriptional regulators. Adapted from (*127*).

Specifically, NHR1 is the region that interacts with E-proteins, such as E2A, which belongs to the SCL/TAL1 complex, and represses the activity of their target genes (*265*). Another important feature of ETO2 is the ability to form homo- and heterooligomers via the NHR2 oligomerization domain (*266*). The NHR2 domain contains a hydrophobic heptad repeat (HHR), which is critical for the oligomerization (*267*). Importantly, it has been suggested that quaternary structures of ETO2 organized by NHR2 confer the repression function to ETO2. For instance, NHR2 dimerization provides a stronger binding to E-proteins and allows NHR3/4

domains to bind in multimers to the nuclear corepressor NCoR/SMRT complexes, increasing the repression effect (*268*). The NHR3 domain is the least well-known domain of ETO2. This domain seems to mediate the interaction between ETO proteins binding to CBF1, Suppressor of Hairless, Lag-1 (CSL), a nuclear receptor that becomes an activator upon activation of upstream Notch signals (*269*). Last but not least, the NHR4 domain contains two noncanonical zinc finger motifs, which regulate the interaction with the corepressor NCoR/SMRT (*270*). Essentially, different combinations of these oligomer complexes might regulate varying subsets of target genes or produce different repression activities. As already shown in section 1.2.4, ETO2 is known to interact with multiple hematopoietic transcription factors to regulate HSPC proliferation and erythroid differentiation.

Besides its role as a regulator of erythroid differentiation, ETO2 is involved in modulation of other cell signaling pathways such as WNT and NOTCH. These two pathways are important for stem cell self-renewal and multicellular development implying that ETO2 has a central function as a regulator of self-renewal and differentiation of HSPCs (*271, 272*).

Knock out studies showed that ETO2 is indispensable for the self-renewal capacity of HSCs due to inappropriate entry of stem cells into the cell cycle (*119*). It has been proposed that ETO2 is required for the control of a cell-cycle regulator circuit, which, when misregulated, drives quiescent stem cells into DNA synthesis or prevents cycling stem cells from entering G_o (*273*). Chyla *et al.* revealed that ETO2$^{-/-}$ mice also suffer from anemia due to a failure to expand erythropoiesis in the spleen. Furthermore, BM presented a reduction in the total number of MEPs and abnormal CD34^{hi}/FCγR^{low} myeloid progenitor cell population (*120*). They also showed that multipotent progenitor cells and short-term stem cells suffer from a proliferation defect. Additionally, these findings were confirmed *in vitro*, where inactivation of ETO2 also impaired the proliferation of erythroid cells (*125*).

ETO2 has been found in several AML-associated chromosomal translocations. t(16;21)(q24;q22) leads to a fusion between RUNX1 and ETO2 (*274*). In the majority of the reported cases, RUNX1-ETO2 occurred in therapy-related AML (*275*). ETO2 was also reported to be fused to the DNA-binding domain of the GLIS2 transcription factor resulting in an ETO2-GLIS2 fusion found in 30% of pediatric non-Down syndrome acute megakaryoblastic leukemia (AMKL) (*276, 277*). Functional studies with RUNX1-ETO2 and ETO2-GLIS2 indicated that the transforming activity of these fusions is dependent on oligomerization by the NHR2 domain. Notably, inhibition of NHR2 oligomerization by competing peptides promotes differentiation and apoptosis of leukemic blasts (*277, 278*).

Considering that both NFIA and ETO2 fusion gene partners are regulators of transcription factors, whose functions are known to be crucial for normal erythropoiesis, and that

erythroleukemia in infants normally occurs during fetal hematopoiesis with limited time period to accumulate a large number of lesions, we hypothesized that this fusion gene might be the driver of aberrant erythropoiesis leading to PEL in children.

1.5.4.4. ZMYND8-RELA

The translocation t(11;20)(p11;q11) was found in very young children diagnosed with PEL (*242, 252*). This translocation leads to fusion of exon 21 of the ZMYND8 gene to exon 2 of RELA, forming a chimeric protein containing the Zinc-PHD finger domain, a bromodomain, a PWWP domain and a MYND type of zinc finger of ZMYND8 and the entire RELA protein (**Fig.20**). Both fusion partners are transcription factors involved in many biological processes such as differentiation, cell growth, tumorigenesis and apoptosis. Therefore, similarly to NFIA-ETO2 fusion gene, we hypothesized that the ZMYND8-RELA fusion gene might be the driver of PEL.

Figure 20. Schematic of the ZMYND8-RELA fusion. The N-terminus of ZMYND8 containing PHD, Bromo, PWWP and DUF3553 domains (in green) is fused to almost the entire ORF of RELA containing a RHD-n and IPT-NFka domain. The yellow triangle indicates the breakpoint. Adapted from (*252*).

1.5.4.4.1. ZMYND8

The Zinc finger MYD-type containing 8 (ZMYND8, aka RACK7) protein contains a Zinc-PHD finger domain, a bromodomain (Bromo), a PWWP, a MYND type of the zinc finger domain and a DUF3554 domain. These domains are needed for ZMYND8 function as a reader of acetylated chromatin (through its bromodomain) and recruitment of nucleosome remodeling and histone deacetylation complexes. In particular, ZMYND8 was shown to recognize H3K4me1 and H3K14ac, histone marks usually associated with gene activation (*279*). The same study also revealed that knockdown of ZMYND8 upregulated multiple co-target genes of ZMYND8, meaning that ZMYND8 acts as a transcriptional corepressor. However, ZMYND8 seems also to be involved in positive regulation of transcription through physical interaction with the P-TEFb elongation complex (*280*).

ZMYND8 has also been identified as a receptor for activated C-kinase protein that binds to activated protein kinase C beta I (PKCβI) (*281*). Additionally, it is also known for its function

as a cutaneous T cell lymphoma-associated antigen (*282*). ZMYND8 was also characterized as a critical DNA damage response factor involved in regulating transcriptional responses and DNA repair activities at DNA double-strand breaks (*283*).

In addition to erythroleukemia, ZMYND8 may also play a role in solid cancers. In a cohort of 91 colorectal cancer patients, some frameshift mutations of ZMYND8 were found in 17 patients (19%) (*202*) and increased ZMYND8 copy numbers have been reported in breast cancer cells (*284*). Recent studies showed that in breast cancer, ZMYND8 mediates tumor progression and metastasis in mice (*285*). According to these studies, including the ZMYND8-RELA fusion found in pediatric erythroleukemia (*242, 252*), aberrant ZMYND8 activity seems to be associated with cancer progression. However, the molecular mechanisms underlying its pro-tumoral activities and how ZMYND8 is related to malignant hematopoiesis remain completely unknown.

1.5.4.4.2. RELA

The *RELA* gene encodes for a member of the nuclear factor NF-Kappa B (NF-κB) family of transcription factors, known to regulate both differentiation and function of lineage-committed hematopoietic cells. RELA has been identified as a regulator of genes involved in HSC homeostasis and function. Stein *et al.* generated *Rela*$^{-/-}$ mice and found that HSCs lacking RELA had a defect in repopulation ability, with a reduction of lineage-negative, c-Kit^{+}, Sca-1^{+} (LSK) CD150^{+}CD48^{-} population and an increased number of more mature LSCK CD49^{+} and CD34^{+} cells in the BM (*286*). Transcriptome analysis revealed up-regulation of differentiation-associated factors and downregulation of factors controlling HSC quiescence. Moreover, ablation of *Rela* also disrupted the normal activity of multipotent progenitors (MPPs) and animals suffered from neutrophilia and reactive splenomegaly. Surprisingly, erythropoiesis appeared normal. Stein *et al.* concluded that loss of *Rela* increased the cycling of HSCs, inducing their differentiation to a progenitor-like stage and also decreasing their self-renewal capacity. Furthermore, genes expressed in lineage-restricted cells were up-regulated, leading to an abnormal distribution of hematopoietic cells(*286*). This data suggested that RELA has a crucial role in regulating genes involved in the fate of HSCs.

RELA was found to be higher expressed in immature precursors compared to more differentiated erythroid cells (*287*). However, ablation of *Rela* (by another group) was not associated with abnormal erythropoiesis (*288*). Grossmann *et al.* showed that mice lacking two NF-kB regulators, *Rela* and *c-Rel* were embryonic lethal (E13.5) (*289*). Transplantation of E12 FL cells into irradiated mice resulted in severe anemia, suggesting defective erythropoiesis. E12 fetuses also presented nucleated embryonic red blood cells in the peripheral blood,

however, the number of erythroid progenitors was normal, indicating a delay in the switch to definitive erythropoiesis (*289*). These data suggest that although loss of *Rela* and *c-Rel* did not affect the frequency of committed erythroid progenitors, both transcription factors seemed necessary for normal fetal erythropoiesis. Another study revealed that RELA represses the transcription activity of human ζ and α globin promoters through the NF-κB pathway (*290*).

Further studies found that about 50% of AML cases presented with aberrant activation of the NF-κB transcription factor complex containing RELA (*291*). RELA expressing leukemic blast cells were highly proliferating and protected from apoptotic cell death. Estrov *et al.* found that arsenic treatment of AML cell lines prevented the activation of NF-κB and induced apoptotic cell death, suggesting that NF-κB could be a therapeutic target in a subset of AML patients (*292, 293*). Overall, these studies suggested RELA could play a role in normal and malignant erythropoiesis, however, the underlying molecular mechanisms remain poorly understood.

Chapter 2 : Cellular and molecular transformation mechanisms of erythroleukemia-associated fusion genes

2.1. Working hypothesis

Recent next generation sequencing studies of tumor cells from acute erythroleukemia patients found recurrent mutations of genes such as *TP53*, *DNMT3A*, or *TET2* that are however not unique for the erythroid phenotype (*242*) (Fagnan *et al.* submitted). However, in rare cases of pediatric erythroleukemia patients, two chromosomal translocations t(1:16)(p31;q24) and t(11;20)(p11;q11) (*250, 252-254*) leading to expression of NFIA-ETO2 and ZMYND8-RELA have been described that so far have never been found in other AML forms. Considering that the fusion partners are transcriptional regulators, whose functions have been linked to normal erythropoiesis, makes them suitable candidates to be erythroid oncogenes. This idea is furthermore underlined by the fact that affected infants or young children have generally a very low mutational burden, making it more likely that these lesions are critical events for the disease. We therefore hypothesize that the NFIA-ETO2 and/or the ZMYND8-RELA fusions may be driver lesions of acute erythroleukemia.

2.1. Cellular and molecular transformation mechanisms of the NFIA-ETO2 fusion associated with acute erythroleukemia

Manuscript *in preparation*.

I present the most important data from the manuscript draft, and additional data that will not be integrated in the first submission of the manuscript.
Supplementary data: https://doi.org/10.5281/zenodo.4748892

Cellular and molecular transformation mechanisms of the NFIA-ETO2 fusion associated with acute erythroleukemia

Maria-Riera Piqué-Borràs[1,2], Frederik Otzen Bagger[1,2], Matheus Filgueira Bezerra[1,2], Amber Louwagie[1,2], Sabine Juge[1,2], Ioannis Nellas[1,2], Robert Ivanek[2], Alexandar Tzankov[3], Ute Moll[4,5], Ramona Schulz-Heddergott[4], Thomas Mercher[6], Juerg Schwaller[1,2]

[1] University Children's Hospital Basel & [2] Department of Biomedicine, University of Basel, 4031 Basel, Switzerland; [3] Institute for Pathology, University Hospital Basel, 4031 Basel, Switzerland; [4] Institute of Molecular Oncology, University of Göttingen, Göttingen, Germany; [5] Department of Pathology, Stony Brook University, Stony Brook, NY, USA; [6] INSERM U1170, Equipe Labellisée Ligue Contre le Cancer, Gustave Roussy Institute, Université Paris Diderot, Université Paris-Sud, Villejuif 94800, France

Running title: The NFIA-ETO2 fusion in erythroleukemia

Key words: NFIA-ETO2 - leukemia – erythroid differentiation - mouse model - gene expression

Correspondence:

Juerg Schwaller MD
University Children's Hospital Basel (UKBB)
Department of Biomedicine, University of Basel
ZLF, Lab 202
Hebelstrasse 20
CH-4031 Basel
Switzerland

Phone: ++41 61 265 3504/3517
Fax: ++41 61 265 2350
Email: J.Schwaller@unibas.ch

Abstract

The molecular drivers of acute erythroleukemia are poorly understood. Here we functionally characterized the transforming activities of a gene fusion between the nuclear factor I A (NFIA) and the transcriptional co-regulator ETO2 previously identified in pediatric erythroleukemia patients. We found that retroviral expression of the NFIA-ETO2 fusion significantly increased proliferation and blocked differentiation of primary murine erythroblasts dependent on the NFIA DNA-binding domain, and the ETO2 NHR2 or NHR4 transcriptional repression domains. As transplantation of NFIA-ETO2-expressing hematopoietic stem and progenitor cells (HSPC) did not induce any disease in mice, we explored functional cooperation with mutated TP53, the most prevalent genetic lesion in acute erythroleukemia. Interestingly, the presence of an *TP53*^{R248Q} mutation did not affect normal erythroid differentiation, but significantly increased serial propagation of colonies formed by NFIA-ETO2 expressing erythroblasts in methylcellulose. In addition, transplantation of NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ erythroblasts induced a transplantable fully penetrant erythroleukemia in mice. Transcriptome analysis found associations of NFIA-ETO2-mediated blocked erythroid differentiation with increased expression of proto-oncogenes (*Myb, Notch2*) and regulators of the erythroid master transcription factor GATA1 (*Trim35*), and downregulation of many NFIA- and GATA1 target genes related to erythropoiesis. The presence of *TP53*^{R248Q} resulted in mostly downregulation of few but well-known tumor suppressive TP53 targets including *Edar2* or *Phlda3*. Collectively, our work suggests that the NFIA-ETO2 fusion blocks erythroid differentiation by repression of genes that are critical for erythroid maturation and that mutated TP53 provides aberrant self-renewal necessary to induce erythroleukemia.

Introduction

Acute erythroleukemia (AEL) is a rare but aggressive human cancer characterized by uncontrolled accumulation of erythroid progenitor cells. Different subtypes have been described including pure erythroleukemia (PEL, AML-M6b, a.k.a. "Di Giuglielmo disease") with blasts committed exclusively to the erythroid lineage, and AML-M6a characterized by the presence of erythroid precursors together with myeloid blasts which turns the diagnosis of AEL challenging (*201, 246*). Due to the low reproducibility of non-erythroid blast counts and the potentially close biological relationship, the revised 2016 WHO classification integrated AML-M6a into the category of myelodysplastic syndrome (MDS) (*134*). Cytogenetic studies revealed that tumor cells of AEL patients often carry complex karyotypes of which alterations of chromosomes 5 and 7 were among the most prevalent (*203, 245*).

Our current understanding of AEL is heavily influenced by pioneer studies of Charlotte Friend and others establishing and studying virus-induced erythroleukemia models in chicken and mice (*294, 295*). Their work indicated that *Friend virus complex* induced erythroleukemia in mice is driven by aberrant expression of the *Spi1* gene encoding for PU.1 that functionally interferes with erythroid master transcription regulator GATA1 (*295, 296*). However, so far, viral integration events have never been reported to be causative events of human AEL (*297*).

More recent targeted or genome-wide sequencing studies found a rather heterogeneous genomic landscape of erythroleukemia in which mutations of the tumor suppressor *TP53* gene were among the most prevalent genetic lesions (*237, 239, 242, 298, 299*) (Fagnan *et al. submitted*). Notably, tumor cells of most PEL patients were reported to carry mono- or bi-allelic TP53 mutations mostly affecting its DNA binding domain (*241*). Cytogenetic analysis and RNA-sequencing of leukemic cells from rare pediatric erythroleukemia patients revealed the presence of a chromosomal translocations t(1;16)(p13;q24) leading to expression of a fusion between the nuclear factor IA (NFIA) to ETO2 (a.k.a. Core Binding Factor, runt domain, alpha subunit 2, translocated to, 3 (CBFA2T3) or Myeloid Translocation Gene on chromosome 16 (MTG16 (*249, 250*). The transcription factor NFIA has previously been shown to control erythroid fate of hematopoietic progenitors, while ETO2 was characterized as transcriptional repressor controlling hematopoietic stem cells (HSC) and differentiation of erythroid progenitor cells (*127, 255, 260*). The exclusive association of with erythroleukemia let us to hypothesis that the NFIA-ETO2 fusion plays a critical role for the initiation of the disease. We cloned a NFIA-ETO2 ORF and explored its cellular activity. We found that expression of NFIA-ETO2 significantly impaired induced *in vitro* terminal differentiation of murine erythroblasts. In presence of one of the most prevalent erythroleukemia-associated $TP53^{R248Q}$ mutation, NFIA-ETO2 expression induced aberrant colony formation in methylcellulose, and a PEL-like disease upon transplantation into irradiated mice. Molecular studies suggest that NFIA-ETO2

blocks differentiation by repression of NFIA erythroid target genes and that presence of the $TP53^{R248Q}$ mutation results in lower expression of tumor suppressive TP53 targets.

Results

NFIA-ETO2 blocks *in vitro* terminal erythroid differentiation

To address the potential to interfere with erythroid differentiation we first expressed NFIA-ETO2 in the murine erythroleukemia cell (MEL) line that undergo partial erythroid maturation after exposure to polar compounds such as dimethylsulfoxide (DMSO), reflected by reduced CD71 and increased Ter119 surface expression associated with augmented hemoglobin production leading to reddish cell pellets and benzidine-staining positive cells (*300*) (**Fig.1A**). NFIA-ETO2 expression did not significantly alter proliferation but impaired DMSO-induced erythroid differentiation as shown by reduced benzidine staining, lower Ter119 expression and lower hemoglobin (*Hbα1, Hbβ1)* mRNA levels (**Fig.1B-H**).

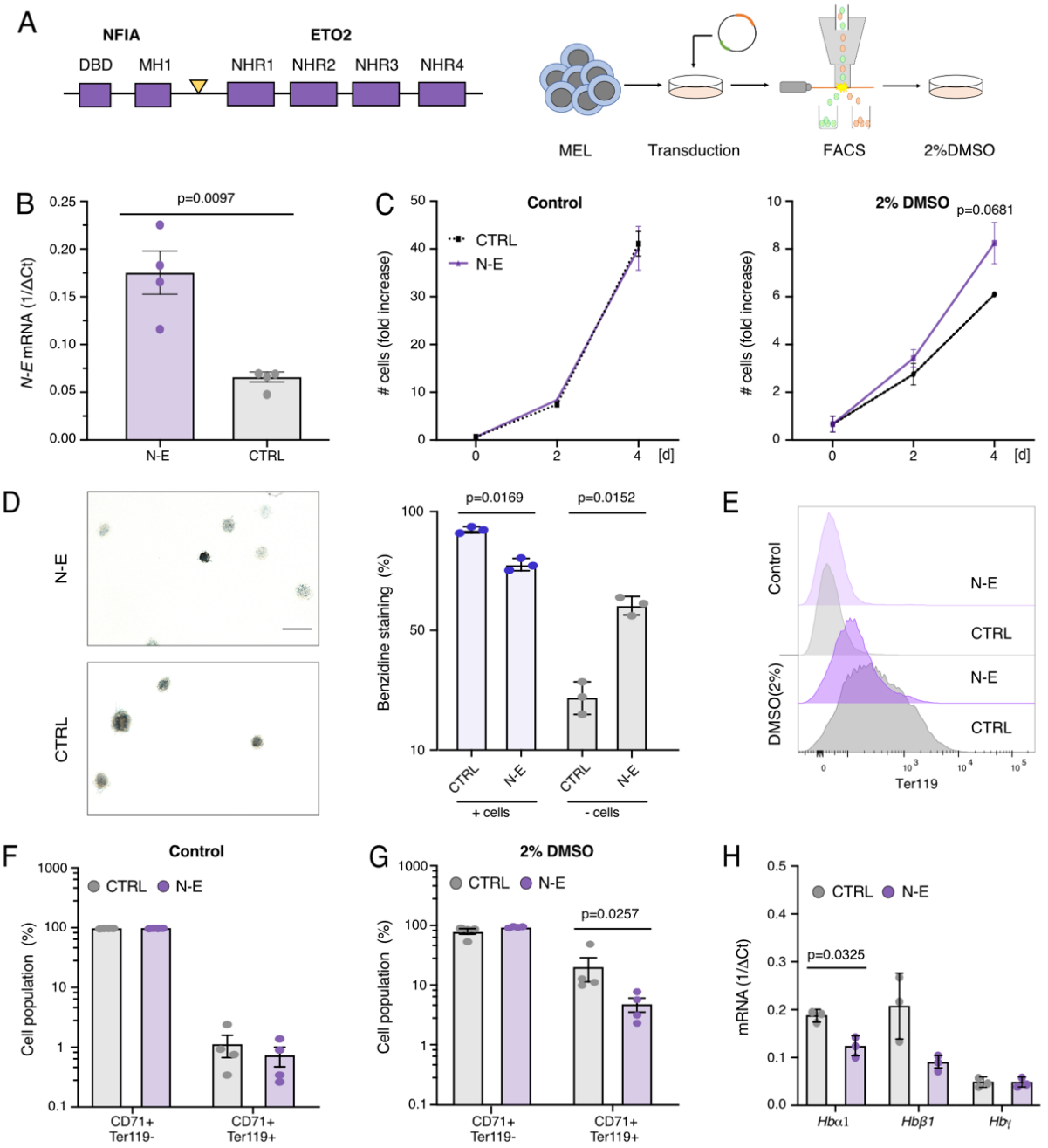


Figure 1. NFIA-ETO2 expression blocks erythroid differentiation of MEL cells. **(A)** Schematic representation of the NFIA-ETO2 (N-E) fusion and overview of experimental procedure. MEL cells were cultured in DMEM and transduced with NFIA-ETO2 or vector-control (CTRL), sorted and cultured with 2% DMSO to induce erythroid differentiation. **(B)** *NFIA-ETO2* mRNA expression levels in transduced MEL cells measured by RT-qPCR (p=0.0097). **(C)** Growth of MEL cells transduced with NFIA-ETO2 or vector control (CTRL) counted over 4 days cultured without DMSO (left graph) or with 2% DMSO (right graph) (n=4). **(D)** Benzidine staining of MEL cells transduced with NFIA-ETO2 or vector control (CTRL) after 4 days in 2% DMSO (left panel). Images were recorded with a 60x objective using a Nikon-TI. Scale-bar = 50 μM. Quantification of benzidine staining (n=3) (right panel). **(E)** Ter119 surface expression on NFIA-ETO2 expressing or vector-transduced control MEL cells before and after 4 days in 2% DMSO. **(F)** Ter119 and CD71 surface expression (%) on NFIA-ETO2 expressing MEL cells compared to vector-transduced control cells (CTRL) before DMSO induction (n=4). **(G)** Ter119 and CD71 surface expression (%) on NFIA-ETO2 expressing MEL cells compared to vector-transduced control cells (CTRL) after 4 days in 2% DMSO (n=4). **(H)** *Hbα1, Hbβ1* and *Hbγ* mRNA expression levels of MEL cells expressing NFIA-ETO2 compared to vector-transduced control cells (CTRL) before and after 4 days in 2% DMSO assessed by quantitative RT-PCR analysis (n=3). Values are presented as individual points, bar graphs represent the mean value of technical replicates, error bars as standard error of the mean. Statistical significances in A, B, C, D, F, G and H was tested with paired two-tailed t- test.

To corroborate this finding in primary cells we established erythroblasts from E14.5 fetal livers (FL) as well as from lineage marker-depleted bone marrow (BM) of 6-10-week old mice. The cells were expanded for 8-10 days in maintenance medium ("MM", containing stem cell factor (SCF), dexamethasone, insulin-like growth factor 1 (IGF1), cholesterol and erythropoietin (EPO)) before viral transduction (**Fig.2A**) (*301*). We observed that stable NFIA-ETO2 expression (confirmed on mRNA and protein level) significantly increased erythroblast proliferation in MM (**Fig.2B-D**). Independent of the cellular origin (FL or adult BM), vector-transduced cells almost completely differentiated towards mature erythrocytes within 4-6 days in differentiation medium ("DM", containing EPO and mSCF), while NFIA-ETO2 expressing cells continued proliferating (**Fig.2E**) and neither showed any major morphologic signs of differentiation nor did they form any red cell pellets (**Fig.2F**). Impaired erythroid differentiation of NFIA-ETO2-expressing erythroblasts was also reflected by increased c-Kit and reduced Ter119 surface expression (**Fig.2G-H, Supplementary Fig.2A-B**) associated with significantly reduced mRNA expression of several erythroid marker genes including *Hbα1, Hbβ1, Hbγ*, *EpoR*, *Gpa* and *Gata1* (**Fig.2I-K, Supplementary Fig.2C-E**). In contrast to vector control transduced cells, NFIA-ETO2 expressing erythroblasts could be propagated to the second plate and form very few and small colonies in the third plate in EPO-containing methylcellulose (M3434) (**Fig.2L, Supplementary Fig.2F**). Taken together, we found that stable expression of NFIA-ETO2 impaired induced terminal erythroid differentiation of MEL cells as well as of primary mouse erythroblasts.

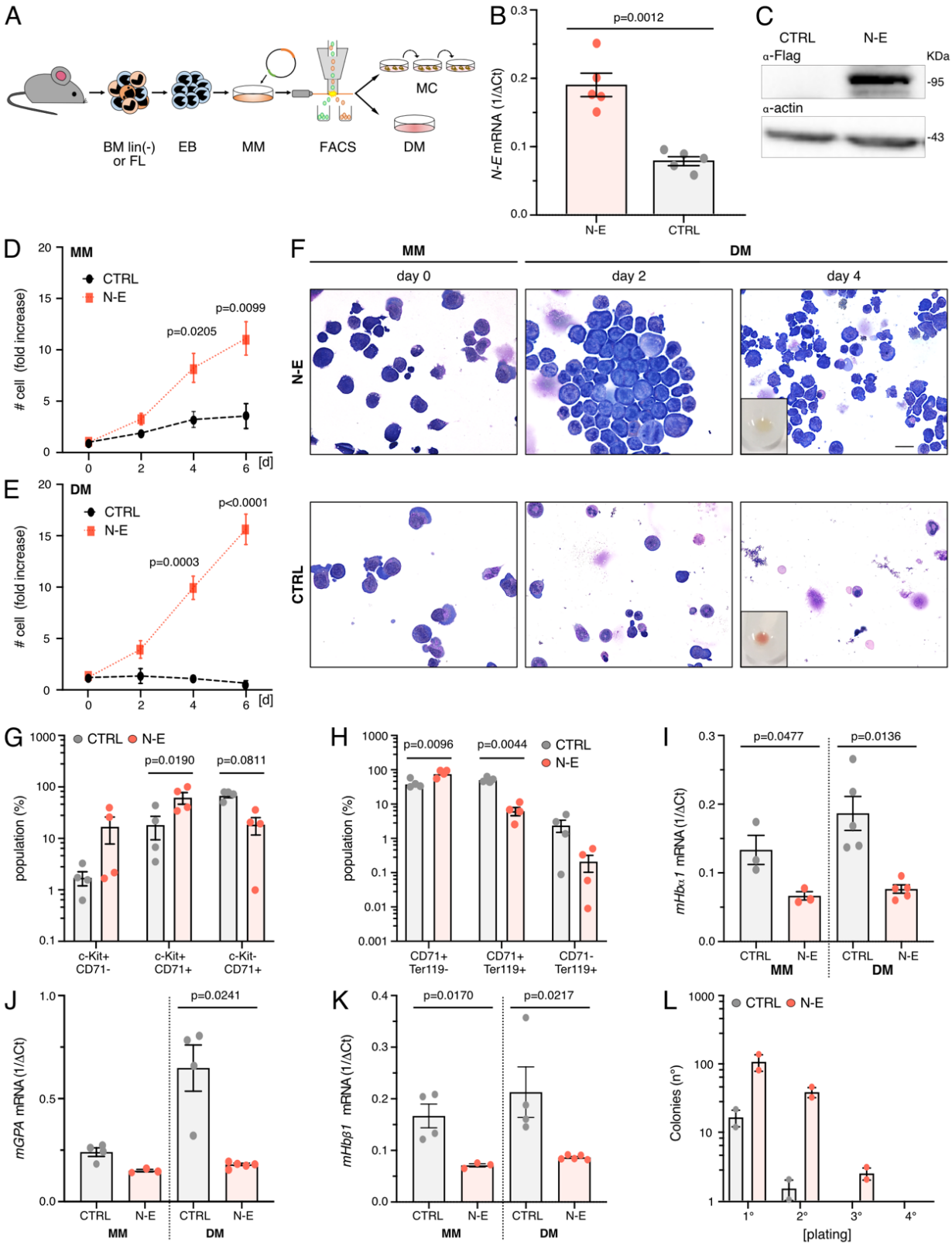


Figure 2. NFIA-ETO2 blocks erythroid differentiation of primary mouse BM erythroblasts. (A) Schematic overview of experimental procedure. BM or FL-derived erythroblasts were cultured in MM and transduced with NFIA-ETO2 or the empty vector (CTRL). Transduced cells were enriched by sorting (GFP+) and cultured in DM or MC (M3434). **(B)** *NFIA-ETO2* mRNA expression after transduction of BM erythroblasts measured by quantitative RT-PCR (p=0.0012, n=5). **(C)** NFIA-ETO2 protein expression in transduced BM erythroblasts compared to vector-transduced control cells (CTRL) using an α-flag antibody. Blotting with α-actin was used as loading control for nuclear proteins. **(D)** Growth of NFIA-ETO2 expressing BM-derived erythroblasts compared to vector-transduced

control cells (CTRL) over 6 days in MM. (day 4: p=0.0205, day 6: p=0.0099, n=4). **(E)** Growth of NFIA-ETO2-expressing BM-derived erythroblasts compared to vector-transduced control cells (CTRL) over 6 days in DM (day 4: p=0.0003, day 6: p=0.0001, n=4). **(F)** Representative images of Wright Giemsa-stained cytospin preparations from primary BM erythroblasts expressing NFIA-ETO2 compared to vector-transduced control cells (CTRL). Left panels: erythroblasts in MM (day 0). Middle panel: erythroblasts after 2 days in DM. Right panel: erythroblasts after 4 days in DM. The small insert shows a reddish cell pellets in CTRL cells, whereas the pellet of NFIA-ETO2-expressing cells appeared white. Images were recorded with a 60x objective using a Nikon-TI. Scale-bar = 50 μM. **(G)** c-Kit and CD71 surface expression (%) on NFIA-ETO2- and vector-transduced control (CTRL) BM-derived erythroblasts after 6 days in DM (n=4). **(H)** CD71 and Ter119 surface expression (%) on NFIA-ETO2- and vector-transduced control (CTRL) BM-derived erythroblasts after 6 days in DM (n=4). **(I)** *Hbα1* mRNA levels in NFIA-ETO2- and vector-transduced control (CTRL) BM-derived erythroblasts grown in MM (n=3) and 24h in DM (n=5) assessed by quantitative RT-PCR. **(J)** *mGpa* mRNA levels in NFIA-ETO2- and vector-transduced control (CTRL) BM-derived erythroblasts grown in MM (n=3) or 24h in DM (n=5). **(K)** *Hbβ1* mRNA levels in NFIA-ETO2- and vector-transduced control (CTRL) BM-derived erythroblasts grown in MM (n=3) and 24h in DM (n=4) assessed by quantitative RT-PCR. **(L)** Colony formation in MC (M3434) by BM-derived erythroblasts expressing NFIA-ETO2 compared to vector-transduced control cells (CTRL). Shown are absolute numbers of colonies after four consecutive platings. Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in B, D, E, G, H, I, J and K was tested with paired two-tailed t- test.

Functional cooperation of NFIA-ETO2 with TP53^{R248Q} *in vitro*

TP53 DNA-binding domain mutations are molecular hallmarks of acute erythroleukemia and PEL in particular (*241, 242, 302*). Although information about the TP53 status of the tumor cells from the patient in which the NFIA-ETO2 fusion was initially identified could not be obtained (*249, 250*), based on the phenotypic similarity of the disease, we wondered whether NFIA-ETO2 might functionally cooperate with TP53^{R248Q} that is one of the most prevalent lesions in PEL (*241*). Hereby, we took advantage of a mouse strain in which the murine *Tp53* exons 4-9 are replaced with the *floxed* human sequence containing the R248Q mutation in exon 7 (*303*) (**Supplementary Fig.3A,B**). We first checked the impact of the *Tp53* genotype in cultures of primary erythroblasts. Notably, we did not observe any significant difference in growth and expression of erythroid surface markers (c-Kit, CD71, Ter119) between *Tp53$^{+/-}$*, *Tp53$^{-/-}$*, *TP53$^{R248Q/+}$* or *TP53$^{R248Q/-}$* erythroblasts expanded in MM (**Supplementary Fig.3C-E**) or upon induced differentiation in DM (**Supplementary Fig.3.1F-H**). Retroviral NFIA-ETO2 expression increased the proliferation of erythroblasts in MM independent of the TP53 status and expressed similar populations of CD71, Ter119 and c-Kit cell markers (**Supplementary Fig.3I-K**). Expression of NFIA-ETO2 also blocked induced terminal erythroid maturation of *TP53$^{R248Q/-}$*, *TP53$^{R248Q/+}$*, *Tp53$^{+/-}$* and *Tp53$^{-/-}$* erythroblasts in DM as shown by cellular morphology, increased proliferation and reduced TER119 expression, while maintaining high c-Kit and CD71 surface expression (**Fig.3A-D, Supplementary Fig.3L)**.

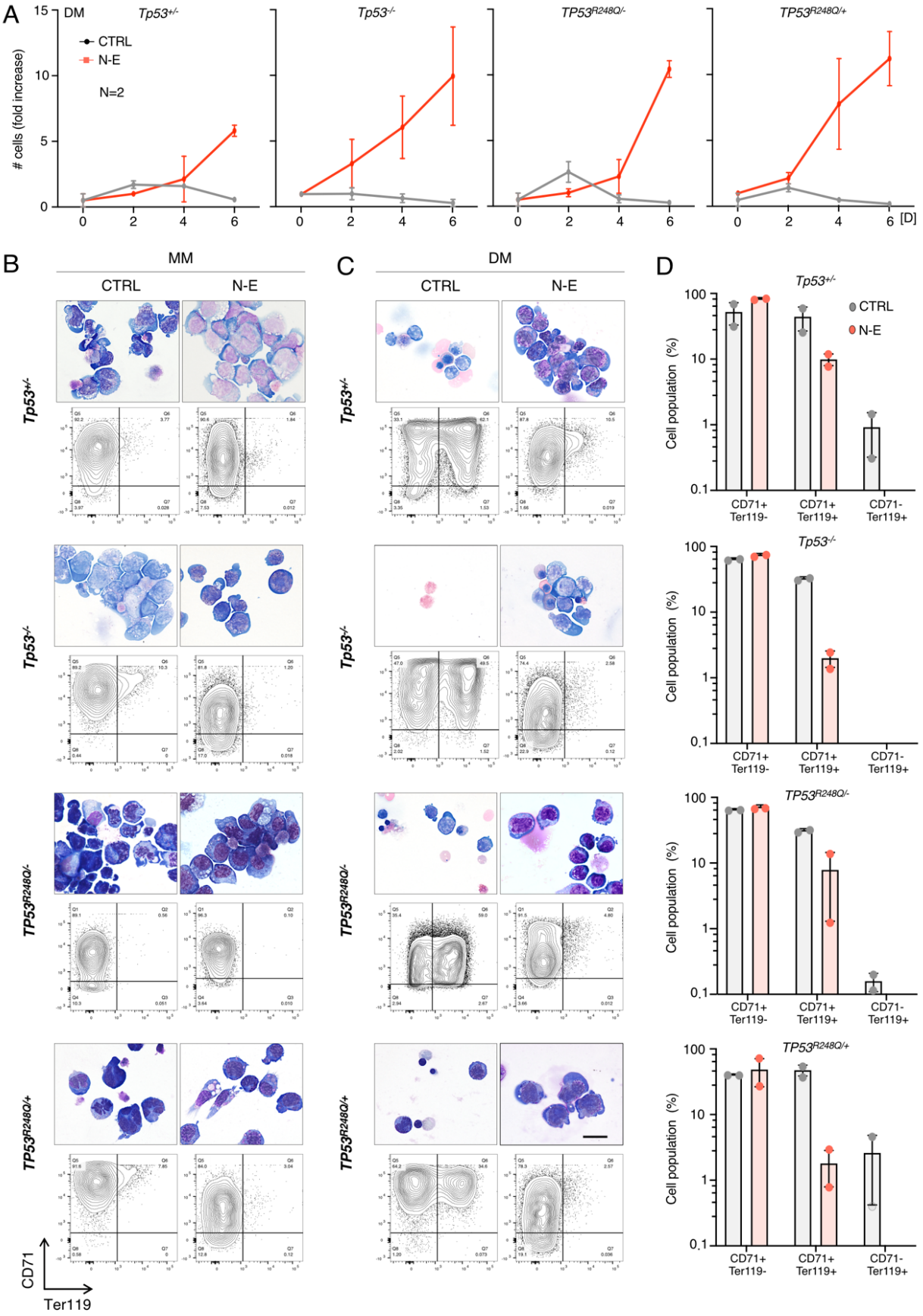


Figure 3. The *Tp53* genotype does not affect the NFIA-ETO2-mediated block in erythroid differentiation. (A) Growth of BM-derived erythroblasts with different *Tp53* genotypes expressing NFIA-ETO2 (orange line) compared to vector-transduced control cells (CTRL, gray line) over 6 days in DM (n=2). **(B)** Representative images of Wright Giemsa-stained cytospin preparations (top row) and flow cytometry panels showing CD71/Ter119 expression

(bottom row) of BM-derived erythroblasts with different *Tp53* genotypes (*Tp53*$^{+/-}$, *Tp53*$^{-/-}$, *TP53*$^{R248Q/-}$, *TP53*$^{R248Q/+}$) expressing NFIA-ETO2 compared to vector-transduced control cells (CTRL) grown in MM. Data represents 1 out of 2 independent experiments. Images were recorded with a 60x objective using a Nikon-TI. Scale-bars = 50 μM. **(C)** Representative images of Wright Giemsa-stained cytospin preparations (top row) and flow cytometry panels showing CD71/Ter119 expression (bottom row) of BM-derived erythroblasts with different *Tp53* genotypes (*Tp53*$^{+/-}$, *Tp53*$^{-/-}$, *TP53*$^{R248Q/-}$, *TP53*$^{R248Q/+}$) expressing NFIA-ETO2 compared to vector-transduced control cells (CTRL) grown in 6 days in DM. Data represents 1 out of 2 independent experiments. Size-bars = 50 μM. **(D)** CD71 and Ter119 surface expression (%) on BM-derived erythroblasts with different *Tp53* genotypes expressing NFIA-ETO2 compared to vector-transduced control cells (CTRL) grown for 6 days in DM (n=2). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean.

In contrast to liquid cultures, we found that the *Tp53* genotype significantly affected colony formation in MC. Compared to wildtype (**Fig.2L, Supplementary Fig.2F**) or *Tp53*$^{+/-}$ cells, we observed significantly increased serial propagation of NFIA-ETO2-expressing cells carrying *TP53*^{R248Q} in presence or absence or the wildtype allele, as well as of cells lacking both copies of the TP53 DNA binding domain (*Tp53*$^{-/-}$) (**Fig.4A-B**). Notably, NFIA-ETO2 expression provided efficient replating activity to *TP53*$^{R248Q/+}$, *TP53*$^{R248Q/-}$ or *Tp53*$^{-/-}$ cells and the cells maintained erythroblastic characteristics as shown by CD71 and c-kit expression (**Fig.4C-D**). Collectively these data show that in presence of *TP53*^{R248Q} mutation NFIA-ETO2 expression results in a significantly increased clonogenic activity of primary mouse erythroblasts.

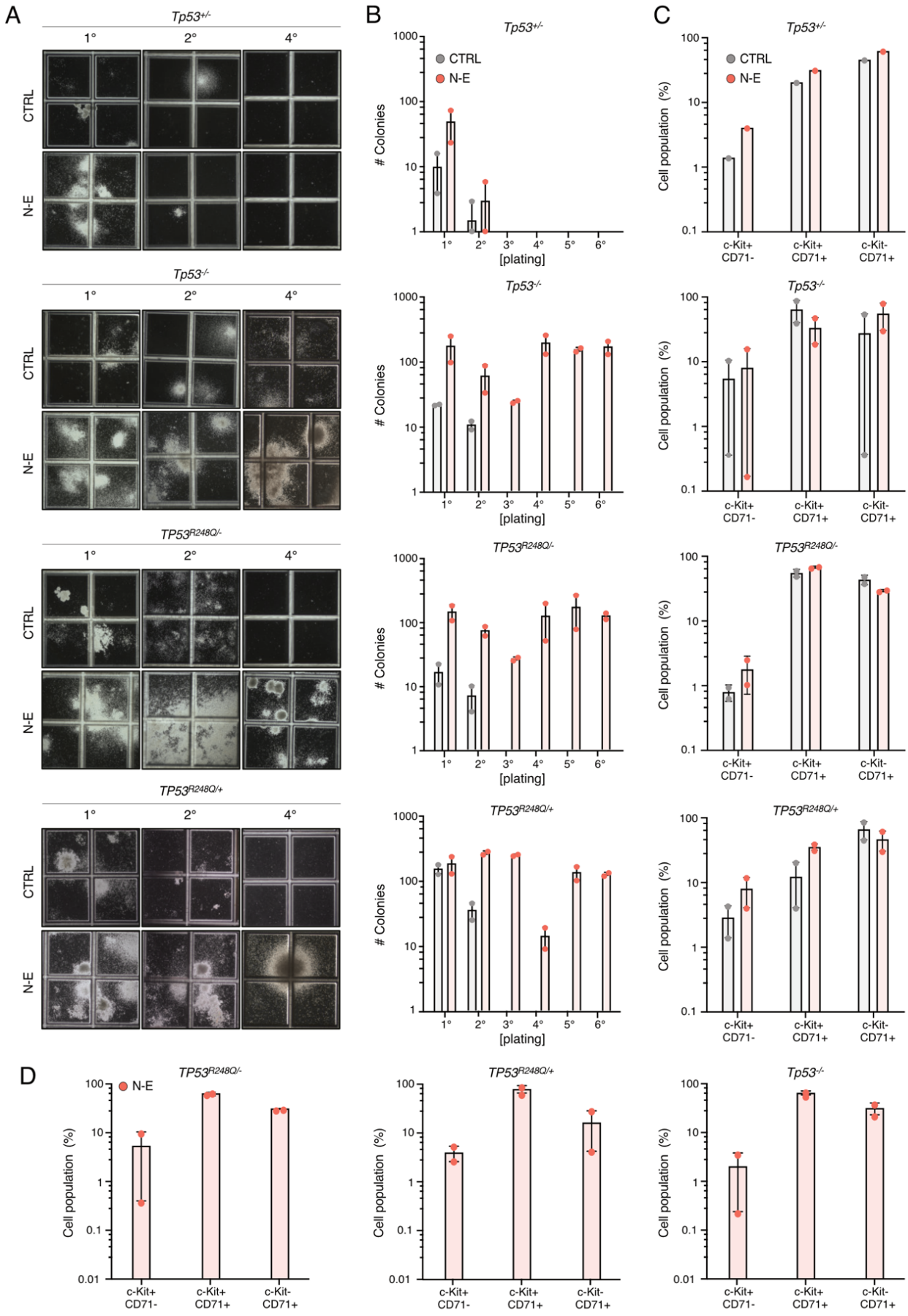

A
Tp53+/-
1°
2°
4°
CTRL
N-E
Tp53-/-
TP53R248Q/-
TP53R248Q/+
B
Tp53+/-
CTRL
N-E
Colonies
1000
100
10
1
1°
2°
3°
4°
5°
6°
[plating]
Tp53-/-
TP53R248Q/-
TP53R248Q/+
C
Tp53+/-
CTRL
N-E
Cell population (%)
100
10
1
0.1
c-Kit+
CD71-
c-Kit+
CD71+
c-Kit-
CD71+
Tp53-/-
TP53R248Q/-
TP53R248Q/+
D
TP53R248Q/-
N-E
Cell population (%)
100
10
1
0.1
0.01
c-Kit+
CD71-
c-Kit+
CD71+
c-Kit-
CD71+
TP53R248Q/+
Tp53-/-

Figure 4. NFIA-ETO2 expression increases the clonogenic potential of TP53 mutant erythroblasts. (A) Representative images (x4) of colonies formed in MC (M3434) cultures by BM-derived erythroblasts with different *Tp53* genotypes expressing NFIA-ETO2 compared to vector-transduced control cells (CTRL) from the first, second and fourth plating. **(B)** Colony formation in MC (M3434) by BM-derived erythroblasts with different *Tp53* genotypes expressing NFIA-ETO2 compared to vector-transduced control cells (CTRL). Shown are absolute numbers of colonies for 6 consecutive platings (n=2). **(C)** c-Kit and CD71 surface expression (%) on BM-derived erythroblasts with different *Tp53* genotypes expressing NFIA-ETO2 compared to vector-transduced control cells (CTRL) harvested from the second plating in MC (n=2). **(D)** c-Kit and CD71 surface expression (%) on BM primary erythroblasts with three different *Tp53* genotypes expressing NFIA-ETO2 harvested from the fourth plating in MC (n=2). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean.

Transplantation of NFIA-ETO2-expressing *TP53$^{R248Q/+}$* BM cells induces erythroleukemia in mice

To address the NFIA-ETO2 leukemia-inducing potential *in vivo* we transduced fetal (E14.5) liver-derived erythroblasts from *TP53$^{R248Q/+}$* or *WT* (B6) mice and intrafemorally (i.f) transplanted them into lethally-irradiated *WT* recipients (**Fig.5A**). None of the mice receiving NFIA-ETO2-expressing WT cells ever developed any disease during an observation time of >1 year (**Fig.5B**). In contrast, all mice transplanted with NFIA-ETO2 transduced *TP53$^{R248Q/+}$* erythroblasts developed signs of disease after a median latency of 3 months (2-6.5 months, n=8) (**Fig.5B**). Diseased mice presented with dysplastic erythroid progenitors on peripheral blood smears and hepatosplenomegaly with tumor cell infiltration in liver, spleen, and BM (**Fig.5C-E, Supplementary Fig.5A**). Symptomatic mice had significantly elevated white blood cells (WBC) and reduced red blood cells (RBC) and platelet (PLT) counts, whereas mice transplanted with WT erythroblasts expressing NFIA-ETO2 presented normal blood counts and only few BM cells remained GFP+ at the end of the experiment (**Fig.5F, Supplementary Fig.5B,C**). The BM of diseased mice was infiltrated by GFP$^+$ NFIA-ETO2-expressing cells (45.4, 8.3-82.5%, n=8) mostly expressing high levels of CD71$^+$ and c-Kit$^+$ and low levels of Ter119$^+$ (**Fig. 4G**, **Supplementary Fig.5D,E**). Importantly, total BM cells of symptomatic mice expressed significantly higher number of c-Kit$^+$/CD71$^+$ and CD71$^+$/Ter119$^+$ cell populations than mice transplanted with vector-transduced BM (CTRL) (p=0.0018, p=0.0002, respectively) (**Fig.5H**). NFIA-ETO2 expression in the BM of diseased mice was further confirmed by quantitative RT-PCR analysis (**Fig.5I**). Recent studies proposed that in tumors carrying missense TP53 mutations, the loss of the remaining WT TP53, also called loss of heterozygosity (LOH), is a prerequisite for stabilization of the mutant TP53 *in vivo* (*304, 305*). However, symptomatic mice transplanted with NFIA-ETO2 expressing *TP53$^{R248Q/+}$* cells showed similar copy numbers of the WT allele, whereas control cells carrying *TP53$^{R248Q/-}$* presented with reduced Tp53 alleles as expected (**Supplementary Fig.5F**). GFP$^+$ BM cells from diseased mice formed dense and round colonies that could be serially plated in MC

formed by mostly c-Kit$^+$ and CD71$^+$ cells (**Fig.5J-K**). Transplantation of 2x10^6 total BM cells from diseased mice into irradiated syngeneic secondary recipients propagated the disease after a short median latency of 25 days (20-29 days, n=4) (**Fig.5B**). Diseased mice presented again with anemia, thrombocytopenia and splenomegaly (**Supplementary Fig.5G,H**). Like primary transplanted diseased mice, most NFIA-ETO2-expressing GFP$^+$ BM cells (55.25, 17.6-92.9%, n=4) were c-Kit$^+$ and CD71$^+$ (**Supplementary Fig.5.2I-K**).

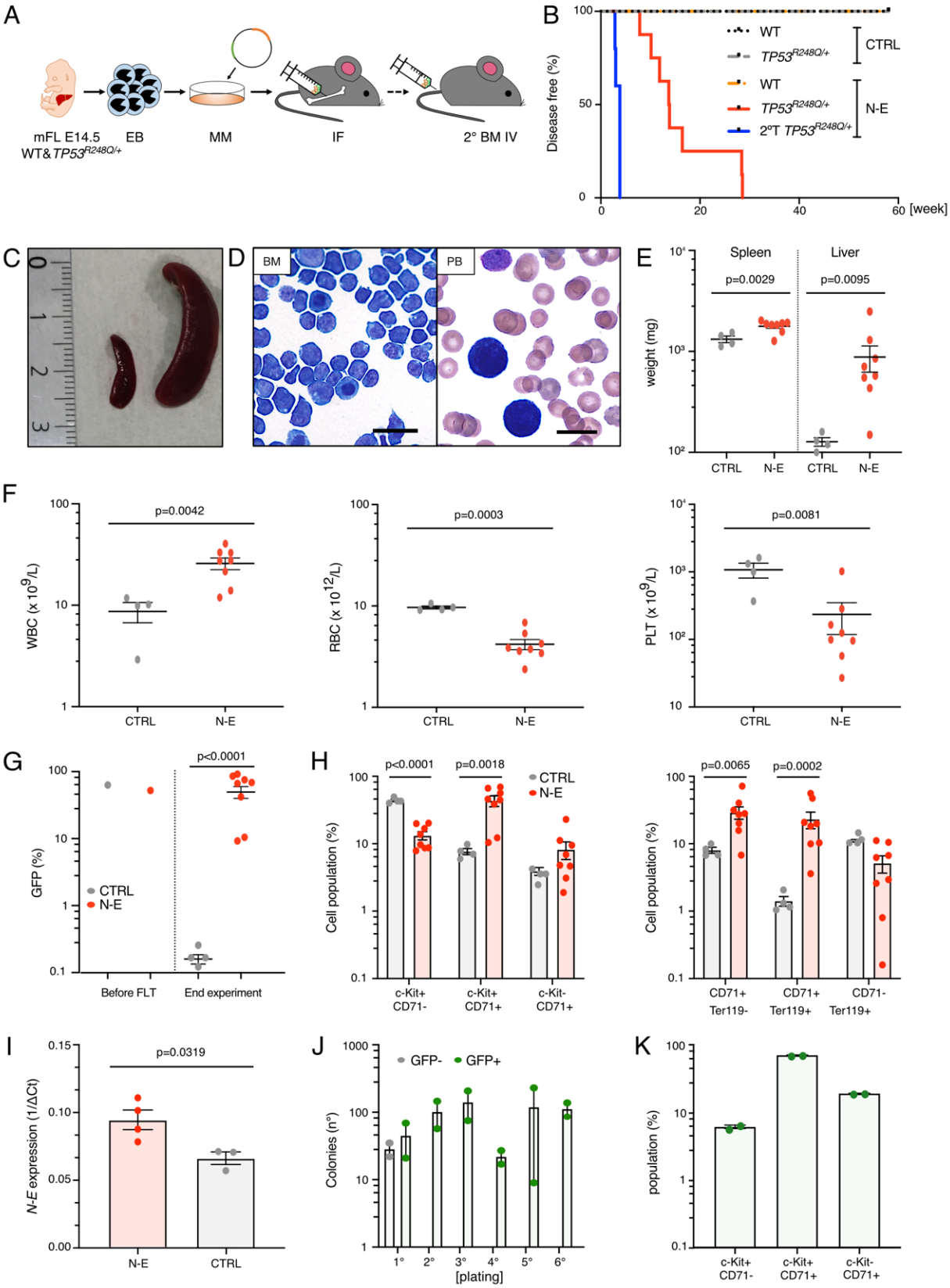


Figure 5: Transplantation of NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ erythroblasts induces a PEL-like disease mice. (A) Experimental setup: WT or *TP53*$^{R248Q/+}$ FL-derived erythroblasts were cultured in MM and transduced with NFIA-ETO2 or empty vector (CTRL) and i.f. transplanted into lethally irradiated recipients together with rescue BM. BM from diseased mice was transplanted i.v. into lethally irradiated secondary recipients. **(B)** Kaplan Meier plot of mice transplanted with NFIA-ETO2-expressing or vector-transduced control WT cells (CTRL) (NFIA-ETO2, n=9, orange line; CTRL, n=4, black dotted line) or *TP53*$^{R248Q/+}$ cells (NFIA-ETO2, n=8, red line; CTRL, n=4, grey line). Median survival of mice receiving NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ cells was 3 months, no mice of the other

groups did develop any disease within 1 year. **(C-D)** Representative images of spleens of control (left) and symptomatic mice transplanted with NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ cells (right); and images of Wright Giemsa-stained cytospin preparations of total BM cells (left panel, 60x, scale-bar = 50μM) and peripheral blood smears (100x, scale-bar = 30μM) showing two erythroblasts (right panel). Images were recorded using a Nikon-TI with x40 and x100 objectives respectively. **(E)** Spleen and liver weight (mg) of symptomatic mice transplanted NFIA-ETO2 expressing (orange dots, n=8) or vector-transduced (grey dots, n=4) *TP53*$^{R248Q/+}$ control cells (CTRL). **(F)** Peripheral WBC counts (x10^{9}/L) (left panel), RBC counts (x10^{12}/L) (middle panel) and PLT counts (x10^{12}/L) (right panel) in symptomatic mice transplanted with NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ (orange dots, n=8) compared to control mice (CTRL) transplanted with vector-transduced cells (grey dots, n=4). **(G)** Transduced GFP^{+} cells (%) before transplantation and at the end of the experiment in NFIA-ETO2- (orange dots, n=8) or vector-transduced (grey dots, n=4) *TP53*$^{R248Q/+}$ erythroblasts. Upon developing symptoms, the majority of their BM cells were GFP^{+} NFIA-ETO2 expressing cells ($p<0.0001$, t-test). **(H)** c-Kit and CD71 surface expression (%) (left plot) and CD71 and Ter119 surface expression (right plot) on total BM cells from symptomatic mice transplanted with NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ cells (n=8) or vector-transduced control (CTRL) cells (n=4). **(I)** *NFIA-ETO2* mRNA expression in total BM cells from diseased mice transplanted with NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ mice compared to vector-transduced control cell transplanted mice (CTRL). **(J)** Colon formation by GFP^{+} and GFP^{-} erythroblasts harvested from symptomatic mice transplanted with NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ cells. Shown are absolute numbers of colonies from first to the sixth consecutive plating in MC (M3434). **(K)** c-Kit and CD71 surface expression (%) on cells from the 2nd plating of MC cultures established from GFP^{+} cells (%) of symptomatic mice transplanted with NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ cells. Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in E, F, G, H and I was tested with unpaired two-tailed t- test.

In a parallel experiment we explored whether a more classical approach by intravenously (i.v.) transplantation of NFIA-ETO2-transduced FL-derived hematopoietic stem and progenitor cells (HSPCs) into lethally irradiation recipients would also induce erythroleukemia (**Fig.6A**). We observed that 6 out of 7 mice transplanted with NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ HSPCs developed symptoms of disease after a median latency of 8.5 months (4-12 months, n=6) (**Fig.6B**). Most diseased mice presented with mild splenomegaly and tumor cell infiltration in the BM, spleen and liver (**Fig.6C-E**). Notably, unlike mice i.f. transplanted with NFIA-ETO2-expressing erythroblasts, the majority of the symptomatic i.v. HSPC-transplanted mice presented with normal peripheral WBC and RBC counts, but they all had significantly reduced PLT counts (**Fig.6F**). Mice transplanted with WT HSPCs expressing NFIA-ETO2 presented normal blood counts and low numbers of GFP^{+} cells (**Supplementary Fig.6A,B**). Symptomatic mice showed 13.6 (2.77-24.5%) of NFIA-ETO2 expressing GFP^{+} cells in the BM that were mostly CD71^{+} and c-Kit^{+} (**Fig.6G-I, Supplementary Fig.6C**), however, BM cells also expressed some myeloid markers (data not shown). Transplantation of 2x10^{6} BM from diseased mice into irradiated secondary recipients induced a leukemic disease after a short median latency of 20.5 days (19-21days, n=4) characterized by increased WBC counts, reduced RBC counts and slightly reduced platelet counts with high infiltration of mostly CD71^{+}/c-KIT^{+} expressing GFP^{+} 46.4 (36.8-56.4%, n=4) in the BM (**Supplementary Fig.6D,E**). Collectively this data demonstrates that transplantation of NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ FL-derived HSPCs or erythroblasts is able to induce a lethal disease when transplanted in lethally irradiated mice.

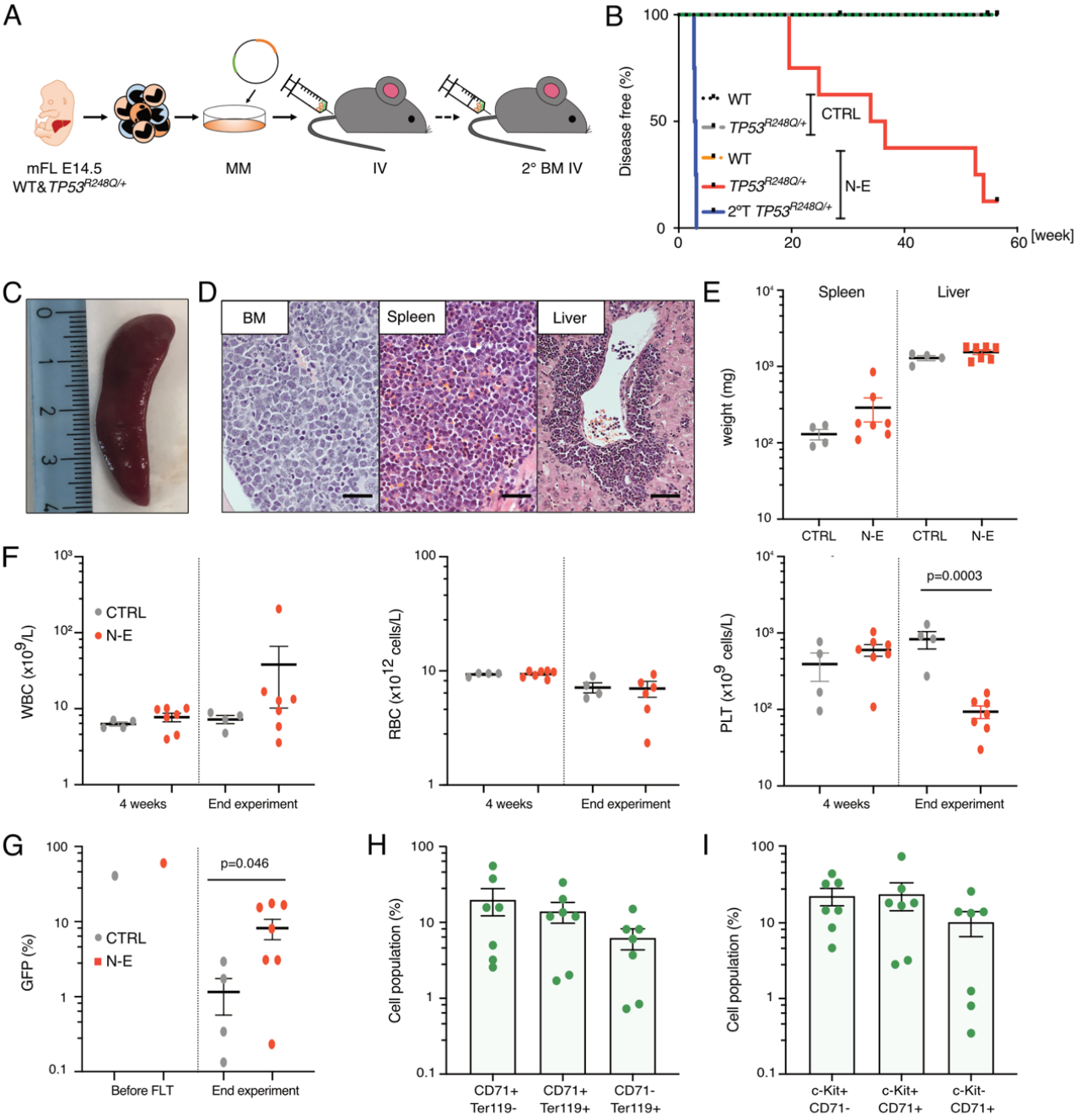


Figure 6. Transplantation of NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ HSPCs induces an erythroleukemia-like disease in mice(A) Experimental setup: WT or *TP53*$^{R248Q/+}$ FL-derived HSPCs were cultured in MM overnight, transduced with NFIA-ETO2 or empty vector (CTRL) and transplanted i.v. into lethal irradiated recipients. Total BM from diseased mice was transplanted into lethal irradiated secondary recipients. **(B)** Kaplan-Meier plot of mice transplanted with NFIA-ETO2 expressing or vector-transduced WT control (CTRL) FL-derived HSPCs (NFIA-ETO2, n=7, orange line; CTRL, n=4, black dotted line) or *TP53*$^{R248Q/+}$ cells (NFIA-ETO2, n=7, red line; CTRL, n=4, grey line). Median time to develop symptoms of disease of mice transplanted with NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ cells was 8.5 months; none of the other groups did develop any disease. Secondary transplanted mice (blue line, n=4) develop disease after 3 weeks. **(C-D)** Representative images of spleen and HE-stained sections of BM (left panel), spleen (middle panel) and liver (right panel) of symptomatic mice transplanted with NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ cells. (40x magnification, size-bar = 75 μM). Images were recorded using a Nikon-TI. **(E)** Spleen and liver weight (mg) of transplanted mice upon development of symptoms (NFIA-ETO2 *TP53*$^{R248Q/+}$, n=4, orange dots) or at terminal workup of CTRL mice (n=4, grey dots) transplanted with vector-transduced control *TP53*$^{R248Q/+}$ cells. **(F)** Peripheral WBC counts (x10⁹/L) (left panel), RBC counts (x10¹²/L) (middle panel) and PLT counts (x10¹²/L) (right panel) after 4 weeks and at the end of the experiment in mice transplanted with *TP53*$^{R248Q/+}$ cells transduced with NFIA-ETO2- (orange dots, n=7) or empty vector (CTRL) (grey dots, n=4). **(G)** Amount of GFP⁺ cells (%) before

transplantation and at the end of the experiment in BM of mice transplanted with NFIA-ETO2- (orange squares, n=7) or vector-transduced (CTRL) (grey squares, n=4) *TP53*$^{R248Q/+}$ HSPCs. **(H)** CD71 and Ter119 surface expression (%) on GFP^{+} cells from the BM of symptomatic mice transplanted with NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ cells (n=7). **(I)** c-Kit and CD71 surface expression (%) on GFP^{+} cells from the BM of symptomatic mice transplanted with NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ cells (n=7). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in F and G was tested with unpaired two-tailed t- test.

NFIA-ETO2 represses erythroid NFIA and GATA1 targets and increases the expression of known leukemia-associated proto-oncogenes

Based on its structural features we hypothesized that the NFIA-ETO2 fusion functions as a chimeric transcription factor in which the NFIA moiety provides contact to DNA whereas the fused ETO2 nervy-homology (NHR) domains mediate contact to transcriptional co-regulators (*127*). By expressing a series of deletion mutants in MEL cells and primary erythroblasts we found that both, the NFIA-DNA binding domain (DBD), as well as the ETO2-NHR2-4 domains appeared to be essential to impair induced erythroid maturation (**Supplementary Fig.7A-D**).

To better understand how NFIA-ETO2 interferes with terminal erythroid differentiation, we compared the gene expression profiles of FL-derived erythroblasts expressing full-length or the inactive NFIA-ETO2$^{\Delta NHR4}$ (ΔNHR4) mutant grown in MM and after 24 hours induced differentiation in DM (**Fig.7A**). As expected, in contrast to NFIA-ETO2$^{\Delta NHR4}$, NFIA-ETO2-transduced cells did not show any differentiation-associated changes of CD71, c-Kit or Ter119 surface expression after 24h in DM (**Supplementary Fig.7E)**. Principal Component Analysis (PCA) revealed a clear separation of the expression profiles according to nature of the transgene (NFIA-ETO2 vs. NFIA-ETO2$^{\Delta NHR4}$, PC1=54.7%) and the culture conditions (MM vs. DM, PC2=9.07%) (**Fig.7B**). Unclassified clustering resulted in a heatmap that clearly separated NFIA-ETO2 from NFIA-ETO2$^{\Delta NHR4}$ cells, whereas the culture conditions (DM vs. MM) did not clearly separate them within each class (**Fig.7C**).

When looking at differentially expressed genes (DEG) in NFIA-ETO2$^{\Delta NHR4}$-expressing cells before and after 24 induced differentiation, we found up and downregulation of 353 vs. 900 genes respectively (FDR>0.05, logFC>1.5) (**Fig.7D, Supplementary Table 1**). As expected from the morphological signs of erythroid maturation we found significant upregulation of erythroid marker genes including the erythroid differentiation-associated gene *Edag* (*306*) and hemoglobin (*Hb-βh1, Hb-βh2, Hb-β2, Hb-β1, Hb-γ*) genes, whereas genes controlling stemness including c-Kit were downregulated. Gene set enrichment analysis (GSEA) revealed positive correlations with signatures related to GATA1 target genes or genes regulating the hemoglobin complex, and negative correlation with genes characterizing HSCs (**Fig.7E**). These data indicate that in contrast to NFIA-ETO2, expression of the NFIA-ETO2$^{\Delta NHR4}$ mutant does not interfere with induced erythroid differentiation in DM.

We next compared differentially expressed genes between cells expressing NFIA-ETO2 or the NFIA-ETO2$^{\Delta NHR4}$ mutant. Under maintenance conditions, we found 1675 more highly and 1089 more lowly expressed genes (FDR<0.05, logFC>1.5) (**Supplementary Fig.7F, Supplementary Table 2**). Notably, one the most significantly higher expressed genes was the ETS transcription factor *Fli1* (FDR=3.71e-30) a well-known regulator of Friend virus-induced erythroleukemia (*307*). Interestingly, GSEA analysis showed a positive enrichment for genes reported to be up-regulated in pediatric AML with KMT2A rearrangements and negative correlation to heme metabolism (**Supplementary Fig.7G**).

24h induced differentiation of NFIA-ETO2 and NFIA-ETO2$^{\Delta NHR4}$-expressing cells was associated with significantly increased expression of 2313 genes and reduced expression of 1440 genes. 1344 genes were more highly expressed under both conditions, expansion in maintenance medium as well as 24 hours induced differentiation (**Supplementary Fig.7H, Supplementary Table 3**). Interestingly, GSEA revealed negative correlations with genes associated with development of the erythroid lineage and target genes of the erythroid master transcription factor GATA1, as well as positive enrichment for NFIA targets (**Fig.7F**). Surprisingly, genes reported to be up-regulated in BM cells from ETO2 knock-out mice were positive enriched in NFIA-ETO2 expressing erythroblasts (**Supplementary Fig.7I**).

Among the upregulated annotated NFIA targets, we found increased expression *Fli1* (FDR=1.26e-49), *Stat5a* (Signal transducer and activator of transcription 5A, FDR=4.24e-31), known proto-oncogenes found involved in other hematopoietic malignancies (*308-310*), and *Elk1*, a member of ETS oncogene family previously shown to regulate the over-expression of a very well-known gene involved in AML, the EVI1 gene (*311*). Four genes out of 234 down-regulated genes were also defined as GATA1 targets. For instance *Alas2* (5'-Aminolevulinate Synthase 2), which is known to be involved in the synthesis of heme (*312*), was significantly down-regulated in NFIA-ETO2 expressing cells (FDR=9.95e-28), as well as *Dmtn* (Dematin Actin Binding Protein, FDR=1.77e-71), which is involved in erythrocyte structure (*313*). *Ank1* (Ankyrin 1) was also downregulated (FDR=1.05e-76) and it is known to be down-regulated in other AML patients carrying the FLT3-ITD mutation (*314*). Moreover, 9 out of 234 down-regulated NFIA targets were also involved in the erythroid lineage development, such as *Ifrd2* (interferon-related developmental regulator 2, FDR=7.89e-19) and *Zranb3* (Zinc Finger RANBP2-Type Containing 3, FDR=1.84e-14) (**Supplementary Table 4**).

Intriguingly, we also observed reduced expression of genes encoding for known GATA1-interacting factors, such as *Tal1* (FDR=3.69e-36) and *Klf1* (Erythroid Krueppel-Like Transcription Factor, FDR=4.22e-33), as well as genes regulated by TAL1/SCL/GATA1 complexes, such as *Gypa* (FDR=1.37e-52) or *Cdkn1b* (FDR=1.05e-06) (**Supplementary Table 3**).

Interestingly, NFIA-ETO2 vs. NFIA-ETO2$^{\Delta NHR4}$ cell DEGs positively correlated to genes that contain motifs matching annotation for binding sites for ETS family transcription factors (such as ELF1 or ETS2) in their 3' end, which have been linked to leukemia of the erythro-megakaryocytic lineage (**Fig.7G**) (*315*). Notably, the ETS genes *Erg* (FDR=1.42e-22) and *Fli-1* (FDR=1.26e-49), both previously functionally related to erythroleukemia, were more highly expressed in NFIA-ETO2 compared to NFIA-ETO2$^{\Delta NHR4}$ expressing cells (**Fig.7H**) (*224, 232*). In facts, overexpression of ETS transcription factors has been shown to be associated with reduced of several erythroid lineage genes such as *Klf1* (FDR=4.26e-33), *Hbβ1* (a.k.a. Hbb.b1, FDR=4.32e-68), *Gypa* (FDR=1.38e-52) or *Ank1* (FDR=1.064-76) (*315*) that were indeed significantly lower expressed in NFIA-ETO2 cells (**Supplementary Fig.7J**).

To identify genes that are directly linked to erythroid differentiation we compared differentially expressed genes between NFIA-ETO2 and NFIA-ETO2 $^{\Delta NHR4}$-expressing cells before and after induced differentiation. We found 984 significantly more highly and 436 more lowly expressed genes (**Fig.7I, Supplementary Table 5**). Among them we found significant upregulation of *c-Myb* (FDR=1.87e-26, logFC=2.28), encoding for leukemogenic transcription factor known to maintain myeloid progenitor cells in an immature stage (*316*). We also observed significantly higher expression of the signaling mediator *Notch 2* (FDR=1.26e-21, logFC=1.63) and *Presenilin2* (*Psen2*) encoding for a regulator of gamma-secretase that mediates NOTCH processing previously shown to be differentially regulated during erythroid differentiation (*128*) (FDR=2.3e-31). Interestingly, one of the most significantly higher expressed genes was encoding for TRIM35 (tripartitie motif containing 35, aka hematopoietic lineage switch protein 5, HLS5) previously shown to suppress GATA1-mediated transcription (*317*) (FDR=9.23e-26). GSEA revealed positive correlation to genes downregulated in leukemic stem cells and negative correlations to cell cycle pathways (**Fig.7J**). Collectively, our data suggests that the NFIA-ETO2 impairs erythroid differentiation by repression of erythroid regulators that are normally controlled by NFIA and/or the erythroid master regulator GATA1. In addition, impaired differentiation is associated with increased expression of several leukemia-associated proto-oncogenes.

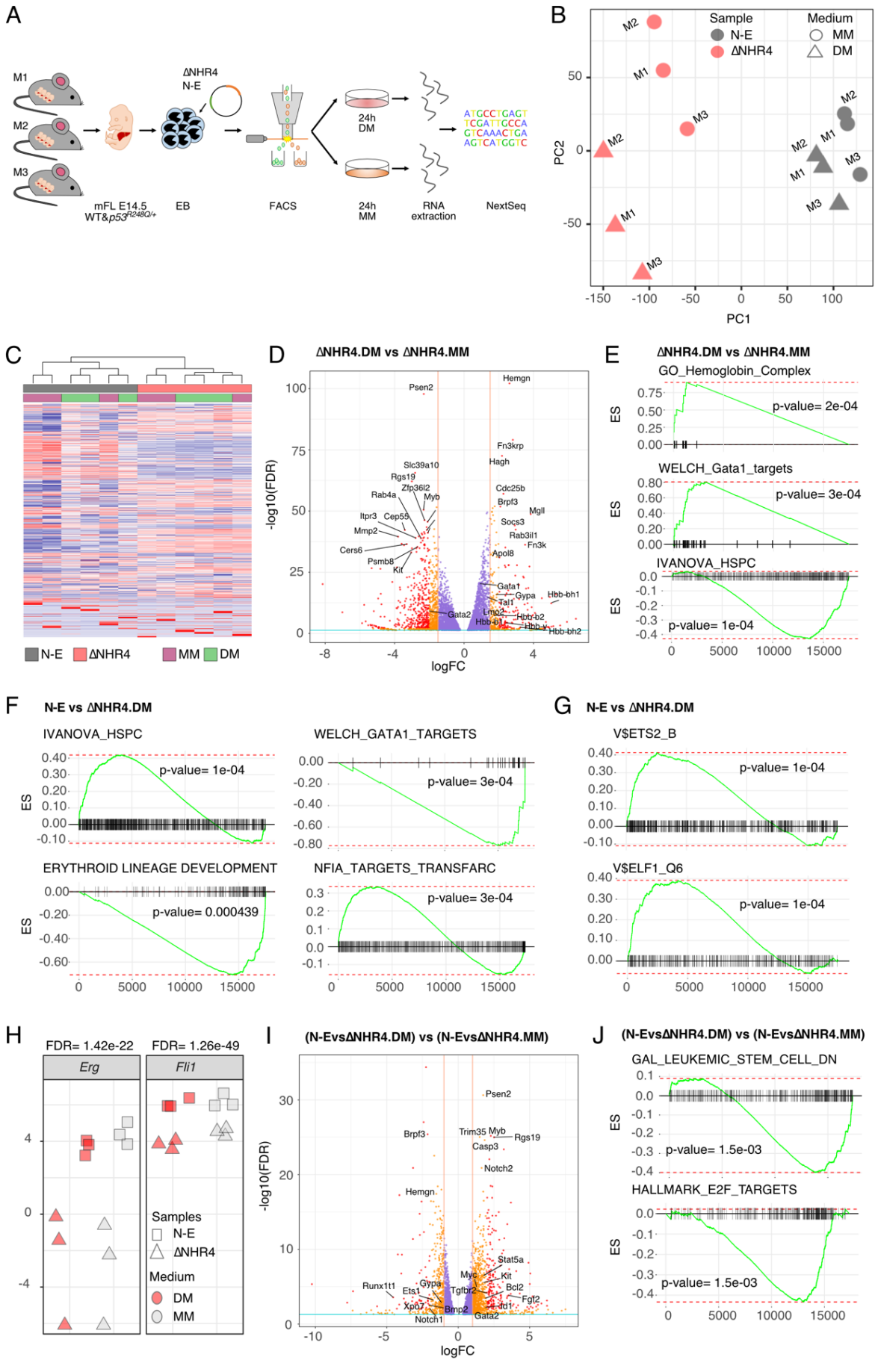
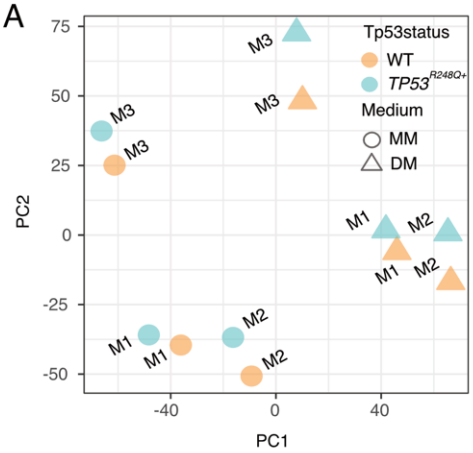

A
M1
M2
M3
ΔNHR4
N-E
mFL E14.5
WT&p53R248Q/+
EB
FACS
24h
DM
24h
MM
RNA
extraction
NextSeq
B
Sample
N-E
ΔNHR4
Medium
MM
DM
PC1
PC2
C
N-E
ΔNHR4
MM
DM
D
ΔNHR4.DM vs ΔNHR4.MM
-log10(FDR)
logFC
E
ΔNHR4.DM vs ΔNHR4.MM
GO_Hemoglobin_Complex
p-value= 2e-04
WELCH_Gata1_targets
p-value= 3e-04
IVANOVA_HSPC
p-value= 1e-04
F
N-E vs ΔNHR4.DM
IVANOVA_HSPC
p-value= 1e-04
WELCH_GATA1_TARGETS
p-value= 3e-04
ERYTHROID LINEAGE DEVELOPMENT
p-value= 0.000439
NFIA_TARGETS_TRANSFARC
p-value= 3e-04
G
N-E vs ΔNHR4.DM
V$ETS2_B
p-value= 1e-04
V$ELF1_Q6
p-value= 1e-04
H
FDR= 1.42e-22
FDR= 1.26e-49
Erg
Fli1
Samples
N-E
ΔNHR4
Medium
DM
MM
I
(N-EvsΔNHR4.DM) vs (N-EvsΔNHR4.MM)
-log10(FDR)
logFC
J
(N-EvsΔNHR4.DM) vs (N-EvsΔNHR4.MM)
GAL_LEUKEMIC_STEM_CELL_DN
p-value= 1.5e-03
HALLMARK_E2F_TARGETS
p-value= 1.5e-03

Figure 7. Gene expression signatures of NFIA-ETO2-expressing erythroblasts. (A) Experimental setup: WT and *TP53*$^{R248Q/+}$ FL-derived erythroblasts were isolated from three different mother mice and transduced with NFIA-ETO2 or NFIA-ETO2-ΔNHR4. After sorting for GFP^{+} erythroblasts were cultured in MM or DM, and after 24h, RNA was isolated and sequenced. **(B)** PCA of WT FL erythroblasts expressing NFIA-ETO2 or NFIA-ETO2-ΔNHR4, in MM or 24h in DM, based on gene expression. Each point represents one sample, colored and shaped according to NFIA-ETO2 (grey) or NFIA-ETO2-ΔNHR4 (orange) expression and medium where cells were cultured (MM, dot or DM, triangle). **(C)** Heatmap of gene expression of WT erythroblasts before (MM purple) and after 24h in DM (green) and expressing NFIA-ETO2 (grey) or NFIA-ETO2-ΔNHR4 NFIA-ETO2 (orange). **(D)** Volcano plot of DEG between NFIA-ETO2-ΔNHR4-expressing erythroblasts before and after 24h in DM (FDR <0.05, logFC>±1.5 showed by orange and blue lines). **(E)** GSEA of differential expression between NFIA-ETO2-ΔNHR4 expressing erythroblasts before and after 24h in DM (padj <0.05). **(F)** GSEA of gene expression between WT erythroblast expressing either NFIA-ETO2 or NFIA-ETO2-ΔNHR4 24h in DM (padj < 0.05). **(G)** GSEA of gene expression between WT erythroblast expressing either NFIA-ETO2 or NFIA-ETO2-ΔNHR4 24h in DM (padj < 0.05). **(H)** *Erg* (FDR=1.42e-22) and *Fli1* (FDR=1.26e-49) expression in NFIA-ETO2 (squares) and NFIA-ETO2-ΔNHR4 (triangles) cells in MM (grey) and 24h in DM (orange). **(I)** Volcano plot of DEG of (NFIA-ETO2 vs NFIA-ETO2-ΔNHR4 24h in DM) vs (NFIA-ETO2 vs NFIA-ETO2-ΔNHR4 24h in MM) comparison (FDR > 0.05, logFC>±1.5 black lines). **(J)** GSEA of gene expression of (NFIA-ETO2 vs NFIA-ETO2-ΔNHR4 24h in DM) vs (NFIA-ETO2 vs NFIA-ETO2-ΔNHR4 24h in MM) comparison.

The presence of TP53^{R248Q} results in downregulation of tumor suppressive TP53 target genes in NFIA-ETO2-expressing erythroblasts

To address the impact of the TP53^{R248Q} mutation we compared gene expression signatures between wildtype and *TP53*$^{R248Q/+}$ NFIA-ETO2-expressing erythroblasts grown in MM and after 24h induced differentiation in DM. Interestingly, PCA revealed only small changes between the growth conditions (MM vs. DM) (PC1=34.78) and the individual mice from which the cells were generated (PC2=25.76) (**Fig.8A**). Only 21 and 11 genes were significantly (FDR<0.05) differentially expressed in maintenance and upon induced differentiation, respectively. The majority 13/21 (in MM) (**Fig.8B,** left table down-regulated genes, right table up-regulated genes) and 7/11 (in DM) genes were downregulated (**Fig.8C, Supplementary Tables 6,7**).The most significantly down-regulated gene was *Eda2r* (a.k.a. Xedar) encoding for the ectodermal dysplasia receptor that belongs to the tumor necrosis factor receptor superfamily (a.k.a.TNFRSF27), previously characterized as direct TP53 target (*318*). In addition, we observed lower expression of genes encoding for the pleckstrin homology-like domain family A, member 3 (Phlda3), polo-like kinase 2 (Plk2) anoctamin 3 (Ano3) and the zinc finger protein 365 (Zfp365) all previously annotated TP53 targets (*319-322*). GSEA revealed positive correlations to stem cell proliferation, heme metabolic processes and metabolic pathways, and negative correlation to HSC differentiation and transcriptional TP53 pathway regulation (**Fig.8D**). Interestingly, up-regulated genes were significantly higher expressed (p=6.2e-06) and down-regulated were also lower expressed (p=0.055) in larger cohorts of acute erythroleukemia patients known to carry TP53 mutations (**Fig.8E**) (*242*). In addition, these genes were also observed significantly higher (p=0.0052) and lower expressed (p=3.7e-05) in patients with other AML subtypes carrying *TP53* mutations (**Fig.8F**) (TGCA, www.cbioportal.org).

A



B

TP53$^{R248Q/+}$ N-E vs WT N-E.MM

Gene Name	logFC	FD
Eda2r	-1.53	3.43E-11
Ano3	-2.64	5.52E-05
Phlda3	-0.85	0.0016
Plk2	-1.43	0.0016
Trp53	-0.65	0.0152
Sulf2	-1.82	0.0152
Cntn1	-0.67	0.0188
Ckmt1	-2.20	0.0188
Slc19a2	-0.54	0.0228
Zfp365	-1.93	0.0261
Zmat3	-0.67	0.0274
Fyn	-0.39	0.0442
Gm15446	-2.91	0.0464

TP53$^{R248Q/+}$ N-E vs WT N-E.MM

Gene Name	logFC	FD
Hip1r	0.60	0.0153
Chst13	1.55	0.0188
Ccl3	0.89	0.0188
Alox15	1.36	0.0188
Mboat2	3.19	0.0209
Acoxl	1.14	0.0274
Csf1	0.75	0.0352
Gpr34	1.13	0.0370

C

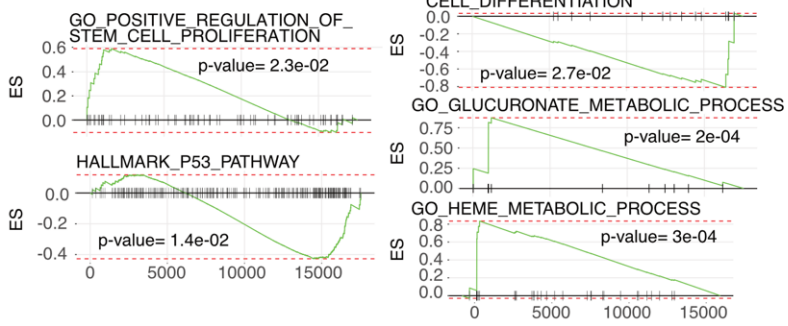
TP53$^{R248Q/+}$ N-E vs WT N-E.DM

Gene Name	logFC	FDR
Eda2r	-1.41	9.79E-08
Phlda3	-0.8987	0.0010
Zmat3	-0.8143	0.0030
Gm15446	-3.7377	0.0031
Psrc1	-0.7316	0.0234
Trp53	-0.6401	0.0234
Bax	-0.4688	0.0284

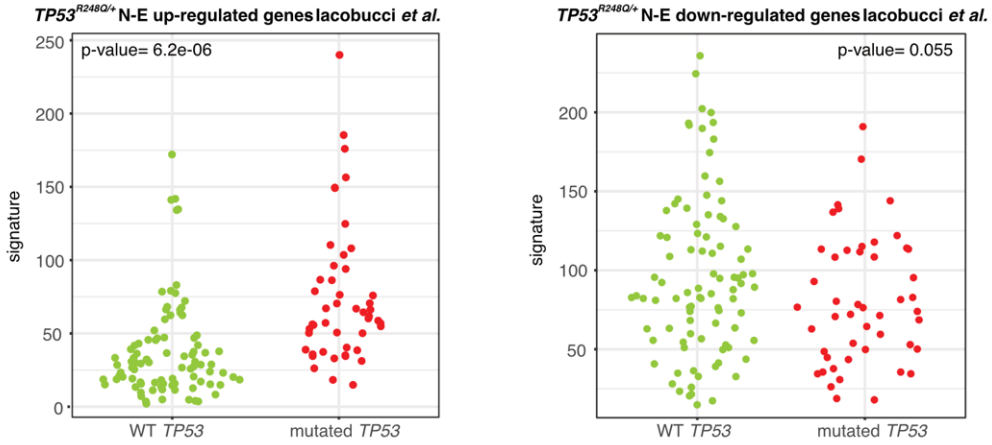
Gene Name	logFC	FDR
Milr1	1.1786	0.0284
Mboat1	0.9774	0.0354
Csf1	0.7702	0.0464
2210416O15Rik	3.1030	0.0284

D

TP53$^{R248Q/+}$ N-E vs WT N-E.MM



E



F

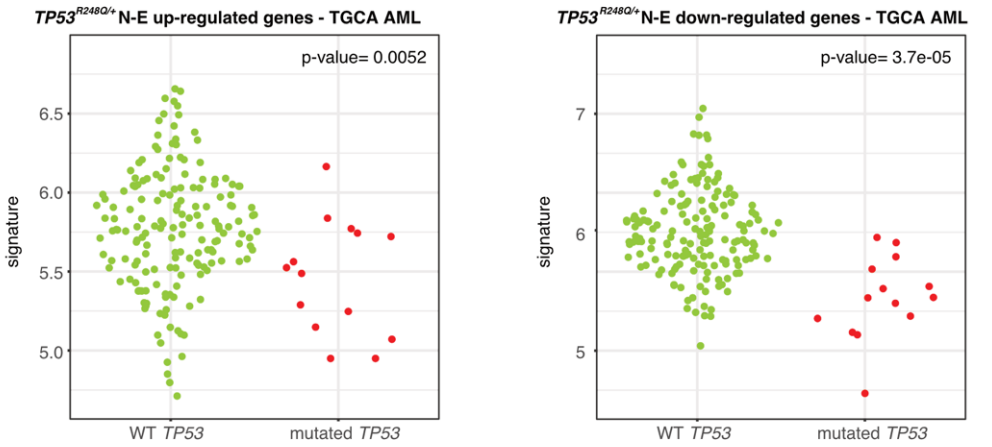
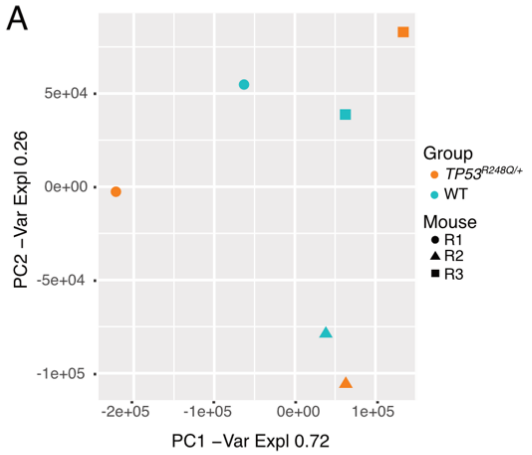


Figure 8. Gene expression profiles of *TP53*$^{R248Q/+}$ erythroblasts expressing NFIA-ETO2. (A) PCA of WT or *TP53*$^{R248Q/+}$ erythroblasts expressing NFIA-ETO2 (WT, orange, *TP53*$^{R248Q/+}$, blue) in MM (dot) or 24h in DM (triangle), based on gene expression. Each point represents one sample, colored and shaped according to the TP53 status (*WT* or *TP53*$^{R248Q/+}$) and the medium were cells were cultured (MM vs. DM), respectively. **(B)** Tables of differently expressed genes between NFIA-ETO2 WT and NFIA-ETO2 *TP53*$^{R248Q/+}$ in MM (FDR >0.05). 13 genes are downregulated (left table) and 8 genes are upregulated (right table) in NFIA-ETO2 *TP53*$^{R248Q/+}$ erythroblasts. **(C)** Table containing the differentially expressed genes between WT and *TP53*$^{R248Q/+}$ expressing NFIA-ETO2 cultured in DM for 24h (FDR >0.05). 7 genes are downregulated (upper table) and 4 upregulated (bottom table) in NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ erythroblasts. **(D)** GSEA of gene expression between NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ erythroblasts compared to NFIA-ETO2 expressing WT erythroblasts in MM (padj < 0.05). **(E)** Expression of the up-regulated (left panel) and down-regulated (right panel) genes found in NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ erythroblasts, expressed in AEL patients with (red dots) or without (green dots) *TP53* mutations. Shown are AEL samples from the Iacobucci *et al.* dataset (*242*). **(F)** Expression of the up-regulated (left panel) and down-regulated (right panel) genes found in NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ erythroblasts, expressed in AEL patients with (red dots) or without (green dots) TP53 mutations. Shown are AML samples from the TGCA project (www.cbioportal.org) (*167*).

Furthermore, Assay for Transposase-Accessible Chromatin (ATAC) sequencing revealed that the chromatin accessibility was similar between *TP53*$^{R248Q/+}$ and WT NFIA-ETO2-expressing erythroblasts cultured 24h in DM (PC1= 72%, PC2=26%) (**Fig.9A**). Interestingly, genes downregulated in *TP53*$^{R248Q/+}$ NFIA-ETO2-expressing cells cultured in both MM (**Fig.9B,** left table) and DM (**Fig.9B,** right table), were significantly highly enriched in their promoter region. Moreover, WT NFIA-ETO2-expressing cells showed similar enrichment, however, gene expression of those genes was higher (**Fig.9C, Fig.8B,C**). Prediction of functional factors, including transcription factors and chromatin regulators, showed that these down-regulated genes might be regulated by several subunits from the PCR2 complex (Polycomb repressive complex 2): JARID2 (Jumonji, AT rich interactive domain 2), EZH2 (Enhancer of zeste 2 polycomb repressive complex 2 subnuit) and SUZ12 (Polycomb repressive complex 2 subunit) (**Fig.9D**). Notably, recent work suggested that mutated TP53 interacts with EZH2 resulting in increased the levels of H3K27me3 in genes involved in HSC self-renewal and differentiation (*323*). Collectively, we found that the presence of *TP53*^{R248Q} is mostly associated with reduced expression of tumor suppressive TP53 target genes, eventually repressed by the PRC2 complex.



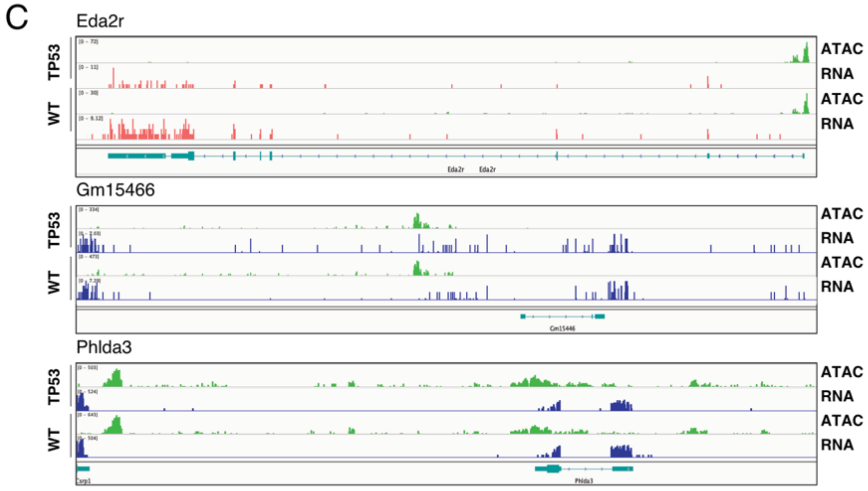
B

TP53$^{R248Q/+}$ N-E vs WT N-E.MM

Gene Name	Peak pval
Eda2r	9.11e-189
Ano3	6.06e-19
Phlda3	2.04e-68
Plk2	1.59e-58
Trp53	<<0.0001
Sulf2	<<0.0001
Cntn1	5.4e-132
Ckmt1	<<0.0001
Slc19a2	2.48e-253
Zfp365	<<0.0001
Zmat3	7.02e-35
Fyn	<<0.0001
Gm15446	<<0.0001

TP53$^{R248Q/+}$ N-E vs WT N-E.DM

Gene Name	Peak pval
Eda2r	9.11e-189
Phlda3	2.04e-68
Zmat3	7.01e-35
Gm15446	<<0.0001
Psrc1	3.77e-105
Trp53	<<0.0001
Bax	1.07e-72

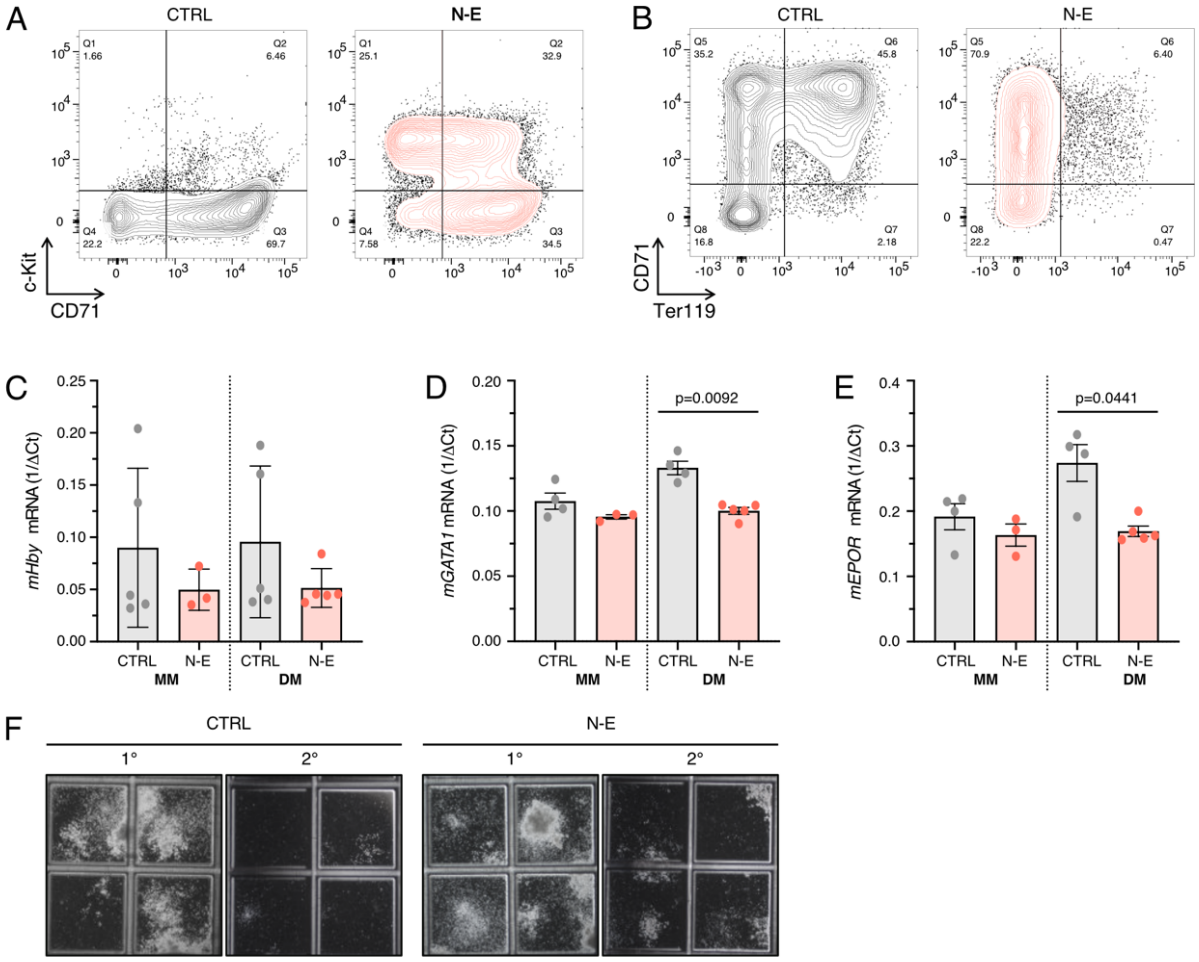


D

TF	p.value
JARID2	2.20e-06
EZH2	2.88e-18
SUZ12	1.72e-49
CTCF	2.94e-04
TP53	5.59e-06

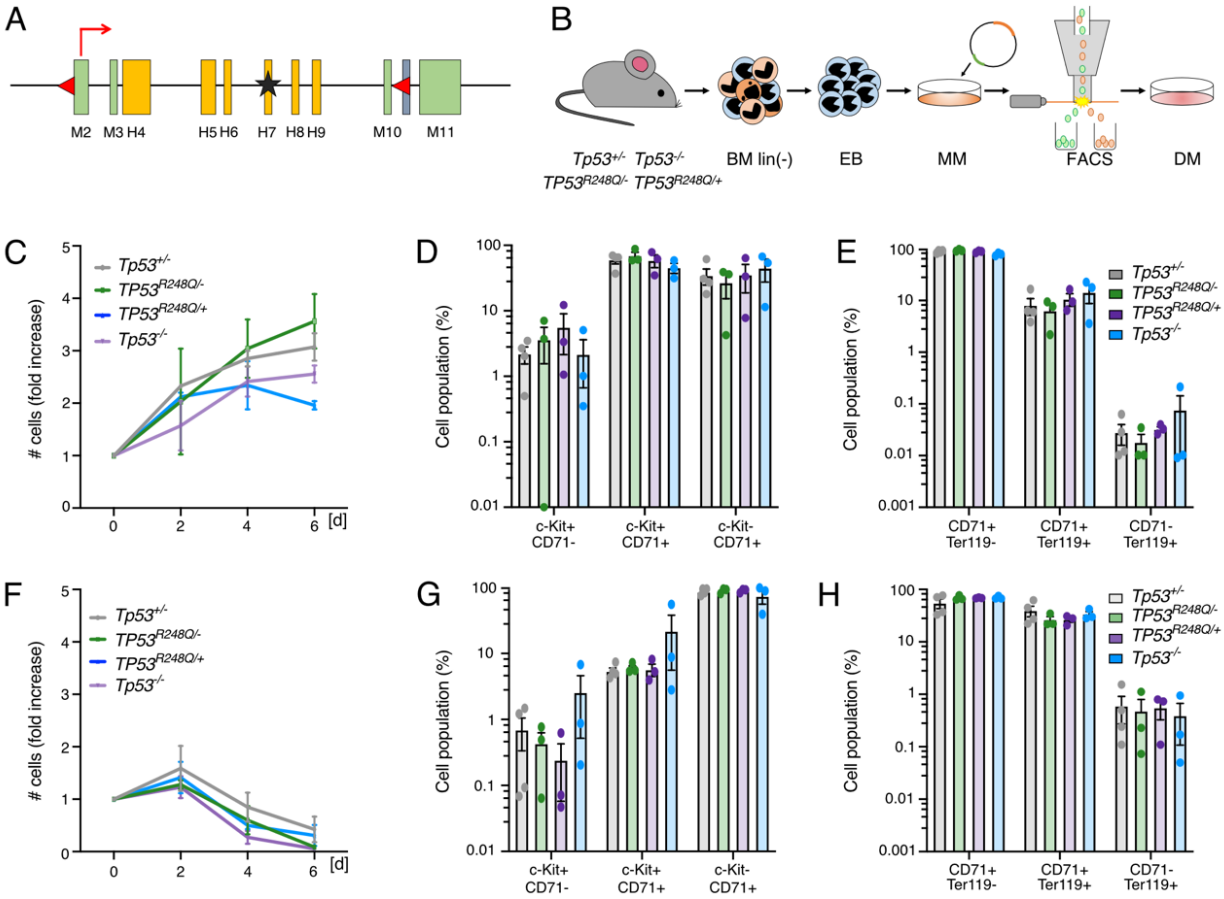
Figure 9: Chromatin analysis (ATAC-seq) of *TP53*$^{R248Q/+}$ and *WT* erythroblasts expressing NFIA-ETO2. (A) PCA of *WT* or *TP53*$^{R248Q/+}$ erythroblasts expressing NFIA-ETO2 (WT, orange, *TP53*$^{R248Q/+}$, blue) 24h in DM, based on chromatin accessibility. Each point represents one sample, colored and shaped according to TP53 status *(WT* or *TP53*$^{R248Q/+}$) and mouse replicate, respectively. **(B)** Tables of genes found down-regulated in *TP53*$^{R248Q/+}$ erythroblasts expressing NFIA-ETO2 (gene expression analysis) in MM (left table) and DM (right table). Compared to the background read coverage, promoters of these genes were significantly enriched in *TP53*$^{R248Q/+}$ erythroblasts expressing NFIA-ETO2 24h in DM (pval<0.05). **(C)** Representative image of integrative genomic viewer (IGV) showing ATAC peaks in the promoter and gene expression (RNA) of Eda2r (upper panel), Gm15466 (middle panel) and Phlda3 (bottom panel) in WT or *TP53*$^{R248Q/+}$ (TP53) erythroblasts expressing NFIA-ETO2 in 24h in DM. Promoters of these genes were enriched in both WT or *TP53*$^{R248Q/+}$ erythroblasts (ATAC), whereas gene expression was downregulated in *TP53*$^{R248Q/+}$ erythroblasts (RNA). **(D)** Predicted transcription factors and related p-values most likely regulating the down-regulated genes in *TP53*$^{R248Q/+}$ erythroblasts expressing NFIA-ETO2. BART software was used to predict functional transcription factors (*324*).

Supplementary Figures

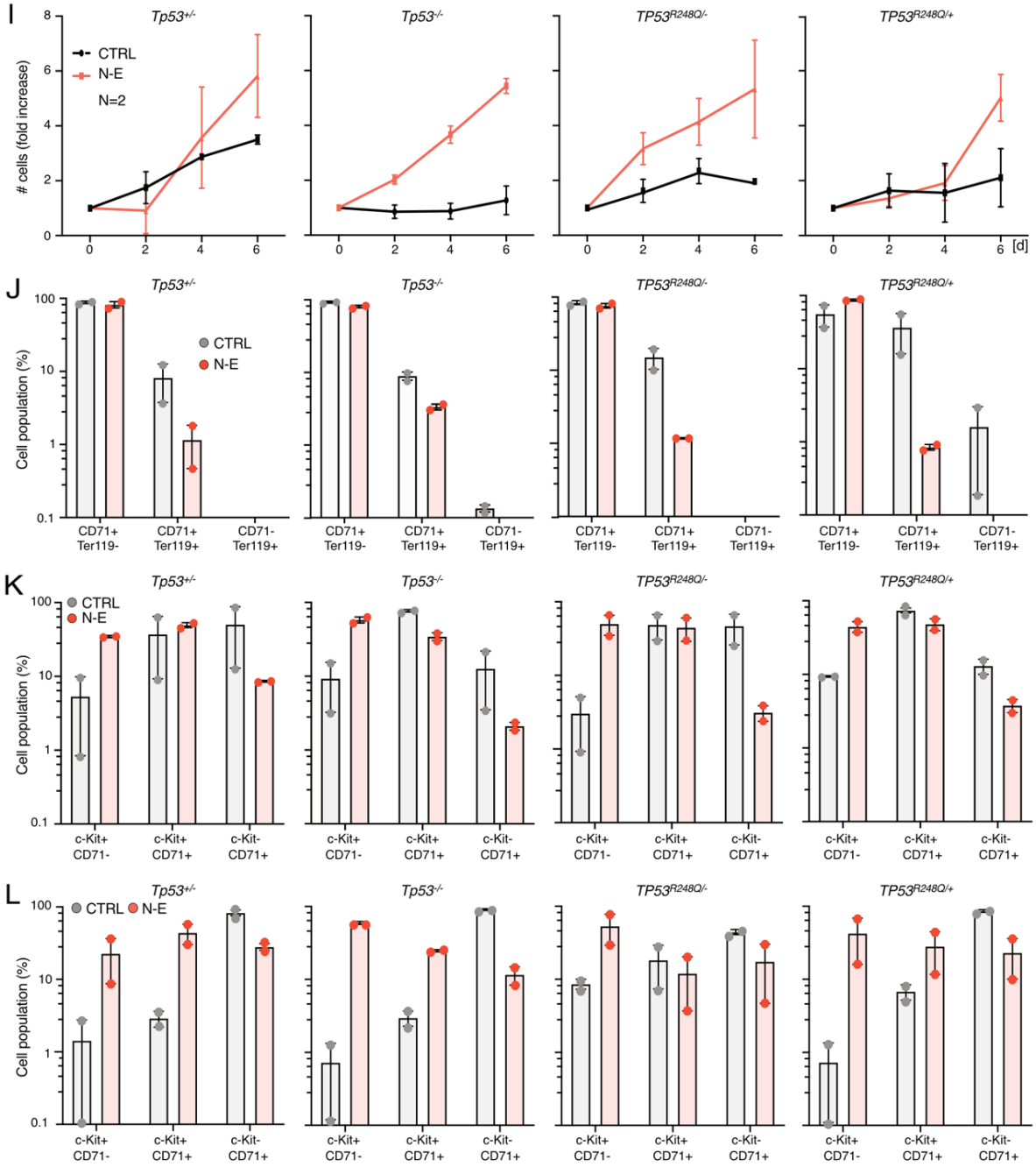


Supplementary Figure 2. NFIA-ETO2 blocks erythroid differentiation of mouse primary BM-derive erythroblasts.

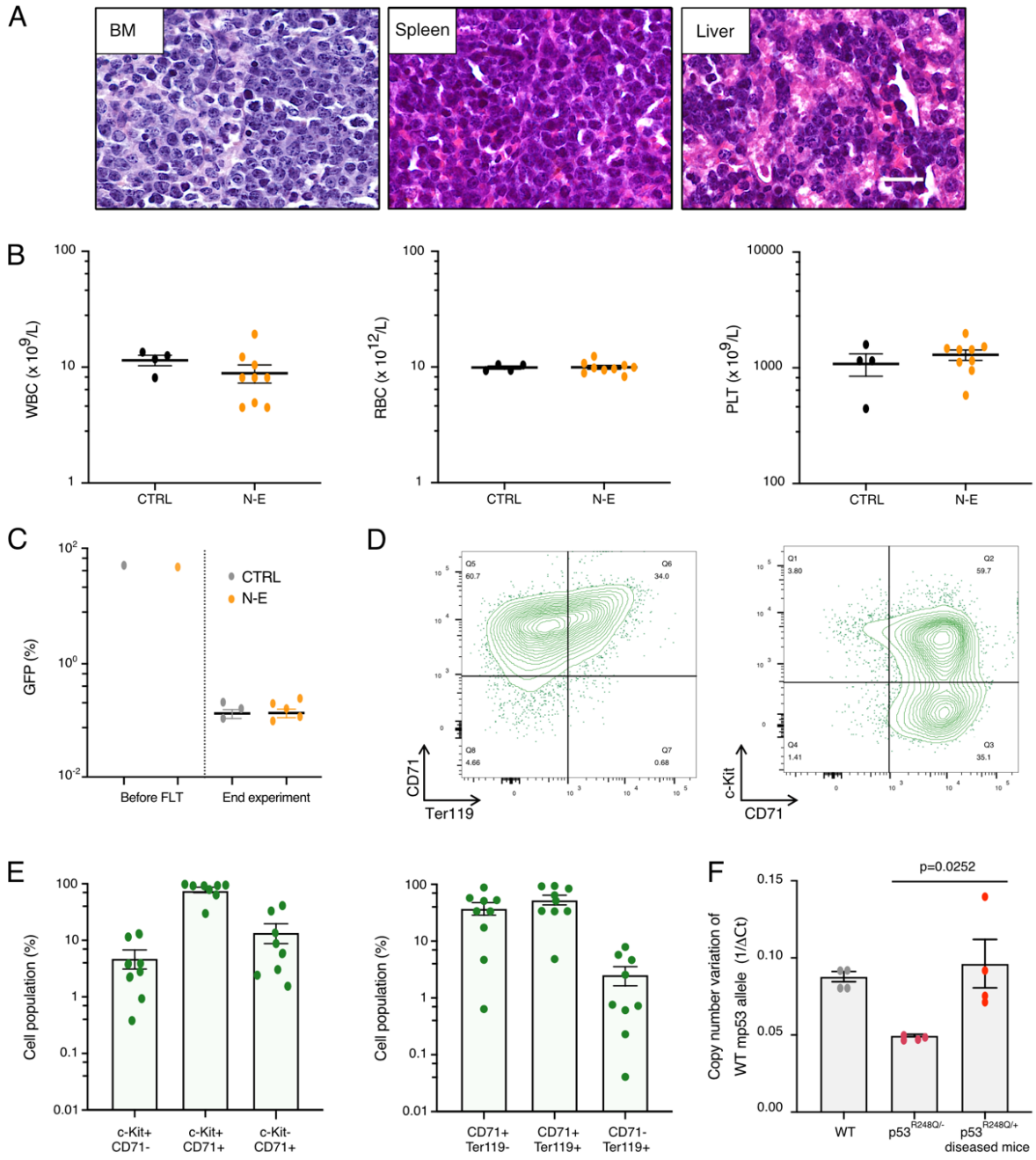
(A) c-Kit and CD71 surface expression (%) on NFIA-ETO2- and vector-transduced control (CTRL) BM-derived erythroblasts after 6 days in DM. **(B)** CD71 and Ter119 surface expression (%) on NFIA-ETO2- and CTRL-transduced BM-derived erythroblasts after 6 days in DM. **(C)** *Hbγ* mRNA levels in NFIA-ETO2- and CTRL-transduced cells grown in MM (CTRL: n=5, N-E: n=3) or 24h in DM (CTRL: n=5, N-E: n=4). **(D)** m*Gata1* mRNA levels in NFIA-ETO2- and CTRL-transduced cells grown in MM (CTRL: n=4, N-E: n=3) or 24h in DM (CTRL: n=4, N-E: n=5). **(E)** *Epor* mRNA levels in NFIA-ETO2- and CTRL-transduced cells grown in MM (CTRL: n=4, N-E: n=3) or 24h in DM (CTRL: n=4, N-E: n=5). **(F)** Images of MC (M3434) cultures (x4) taken seven days after the first and second plating of BM-derived erythroblasts expressing NFIA-ETO2 compared to CTRL-transduced cells. Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in D and E was tested with paired two-tailed t- test.



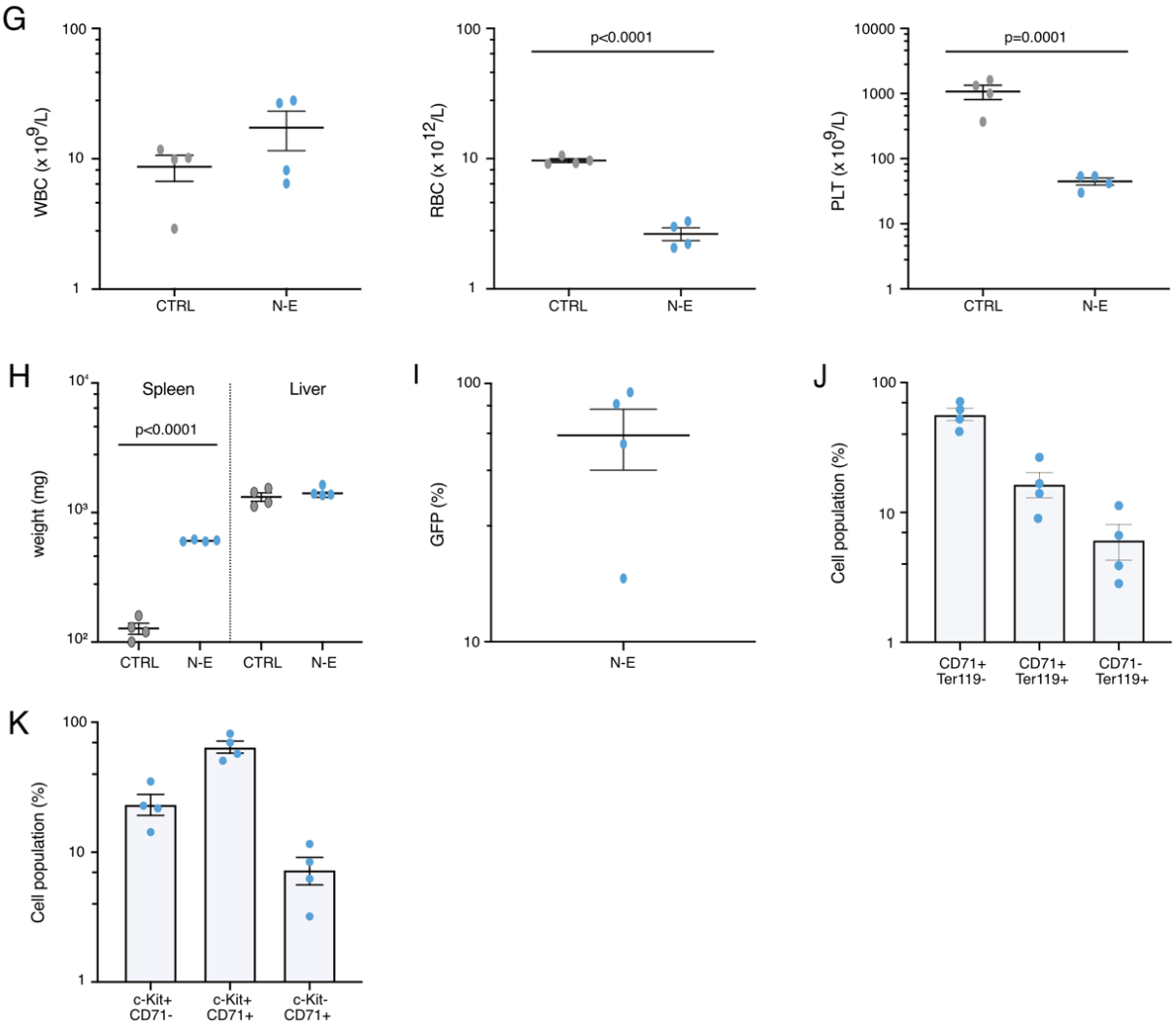
Supplementary Figure 3. The *Tp53* genotype does not affect NFIA-ETO2-mediated block in erythroid differentiation. (A) Schematic representation of the human *TP53*^{R248Q} knock-in (HUPKI) allele in mice. **(B)** Schematic overview of experimental procedure. **(C)** Growth of BM-derived erythroblasts with different *Tp53* genotypes (*Tp53*$^{+/-}$, gray line; *Tp53*$^{-/-}$, purple line; *TP53*$^{R248Q/-}$, green line; *TP53*$^{R248Q/+}$, blue line) over 6 days in MM (n=3). **(D)** c-Kit and CD71 surface expression (%) on BM-derived erythroblasts with four different *Tp53* genotypes grown for 6 days in MM. No significant differences between genotypes were observed (n=3). **(E)** CD71 and Ter119 surface expression (%) on BM-derived erythroblasts with four different *Tp53* genotypes grown for 6 days in MM. No significant differences between genotypes were observed (n=3). **(F)** Growth of BM-derived erythroblasts with different *Tp53* genotypes respectively (*Tp53*$^{+/-}$, gray line; *Tp53*$^{-/-}$, purple line; *TP53*$^{R248Q/-}$, green line; *TP53*$^{R248Q/+}$, blue line) grown in DM over 6 days. (G) c-Kit and CD71 surface expression (%) on BM-derived erythroblasts with four different *Tp53* genotypes grown for 6 days in DM. No significant differences between genotypes were observed (n=3). **(H)** CD71 and Ter119 surface expression (%) on BM-derived erythroblasts with different *Tp53* genotypes grown for 6 days in DM. No significant differences between genotypes were observed (n=3). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean.



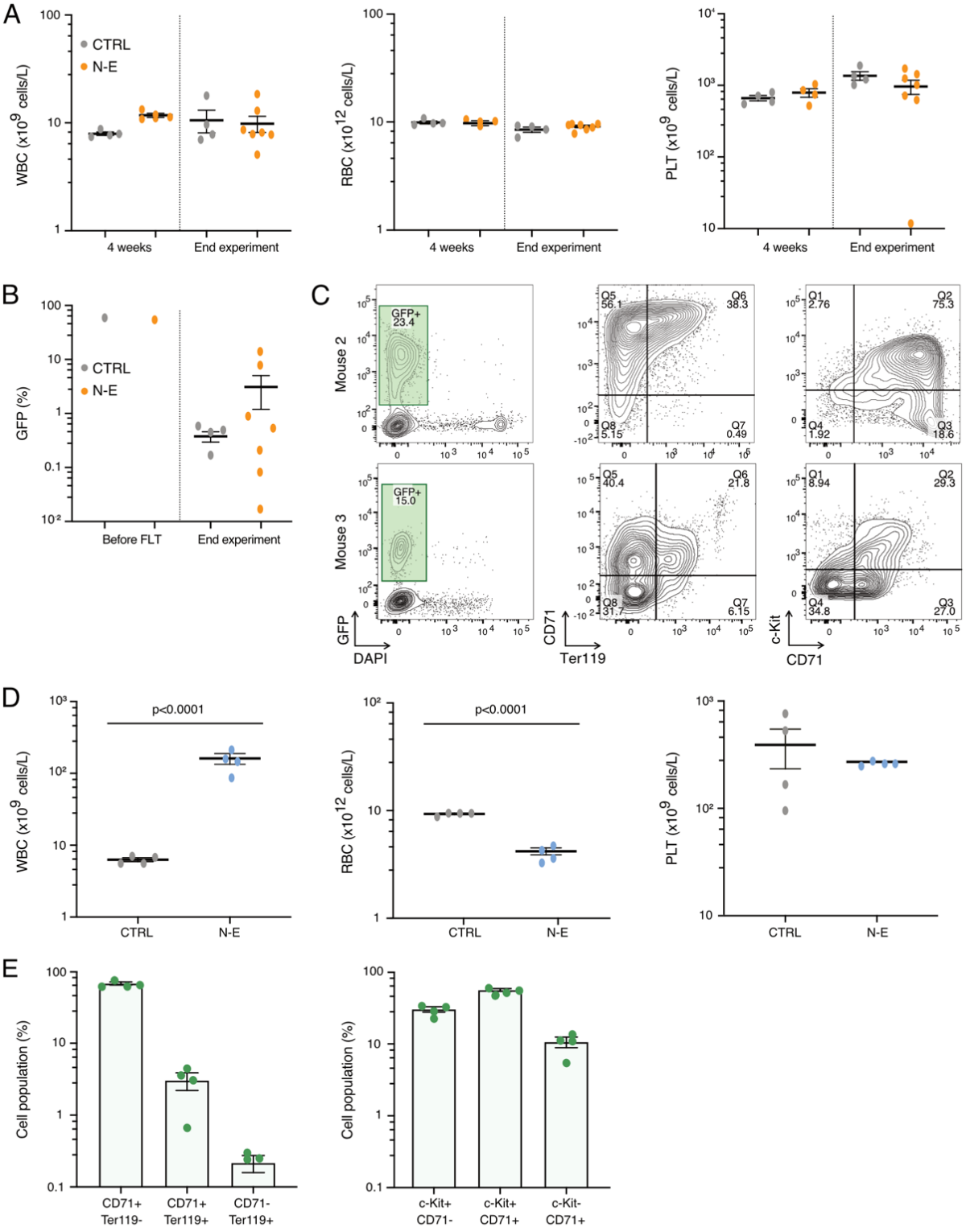
Supplementary Figure 3 (continuation).
(I) Growth of BM-derived erythroblasts with four different *Tp53* genotypes, expressing NFIA-ETO2 (orange line) compared to CTRL-transduced cells (black line) over 6 days in MM (n=2). **(J)** CD71 and Ter119 surface expression (%) on BM-derived erythroblast with different *Tp53* genotypes expressing NFIA-ETO2 compared to CTRL-transduced grown for 6 days in MM (n=2). **(K)** c-Kit and CD71 surface expression (%) on BM-derived erythroblast with different *Tp53* genotypes expressing NFIA-ETO2 compared to CTRL-transduced grown for 6 days in MM (n=2). **(L)** c-Kit and CD71 surface expression (%) on BM-derived erythroblasts with different *Tp53* genotypes expressing NFIA-ETO2 compared to CTRL-transduced grown for 6 days in DM (n=2). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean.



Supplementary Figure 5. Transplantation of NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ erythroblasts induces a PEL-like disease mice. (A) Representative images of HE-stained sections of BM (left panel), spleen (middle panel) and liver (right panel) of diseased mice transplanted with NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ erythroblasts. (60x, size bars = 75 μM). Images were recorded using a Nikon-TI. **(B)** Peripheral WBC counts (x10^{9}/L) (right panel), RBC counts (x10^{12}/L) and PLT counts (x10^{12}/L) in WT mice transplanted with NFIA-ETO2- (yellow dots) or vector-transduced control (CTRL) cells (black dots). **(C)** Amount of GFP^{+} cells (%) before transplantation and at the end of the experiment in the BM of WT mice transplanted with NFIA-ETO2 (yellow dots, n=5) or vector control (black dots, n=4-transduced erythroblasts). **(D)** Representative flow cytometry panels of a symptomatic mouse showing cells expressing CD71 and/or Ter119 (left panel) and c-Kit and/or CD71 (right panel). **(E)** c-Kit and CD71 surface expression (%) (left plot) and CD71 and Ter119 surface expression (right plot) on GFP^{+} cells from diseased mice (n=8) transplanted with NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ cells. **(F)** Copy number of *Tp53* alleles in BM cells from WT (grey dots), *TP53*$^{R248Q/-}$ (red dots) and from symptomatic mice transplanted with NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ cells (orange dots). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in F was tested with unpaired two-tailed t- test.



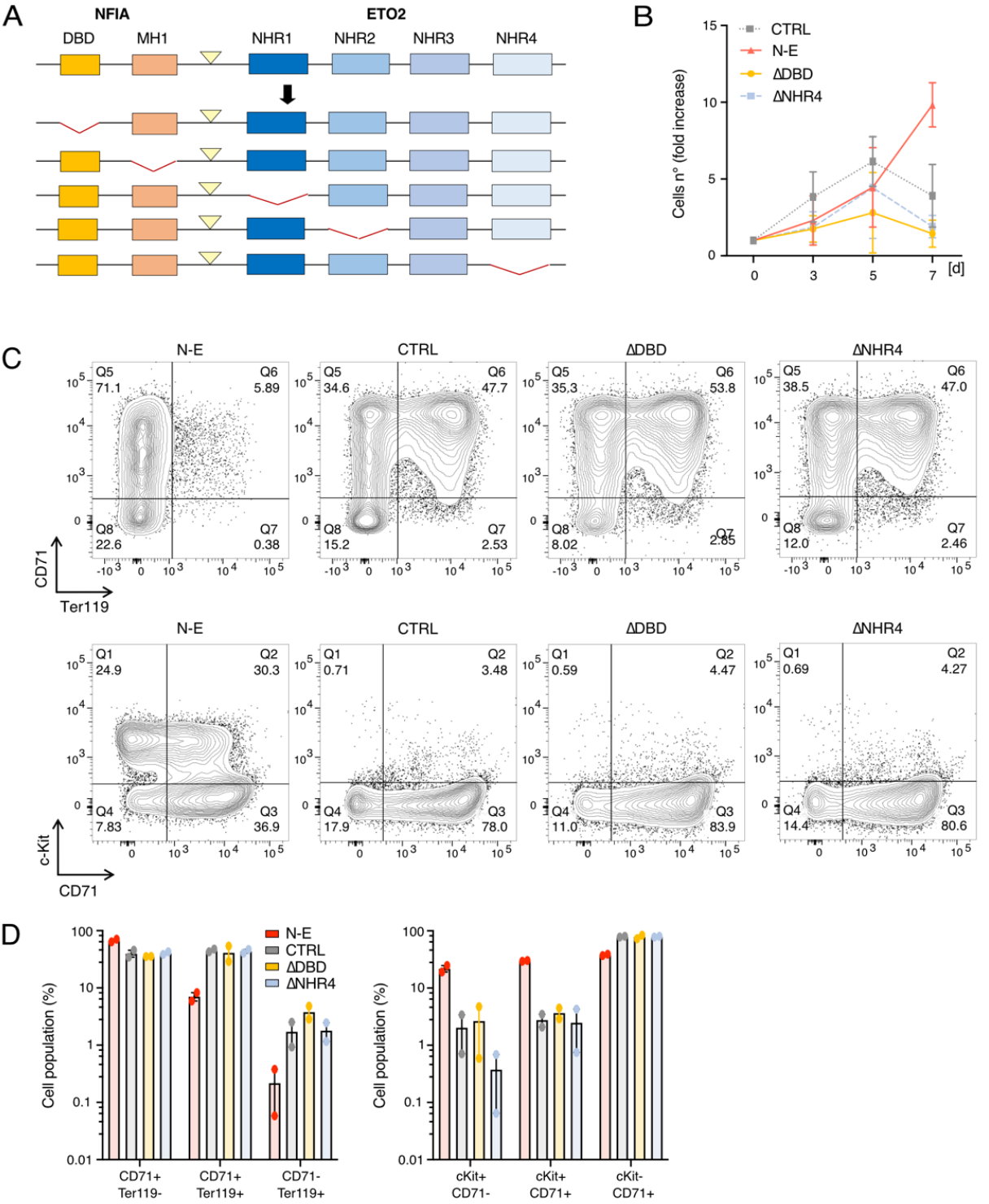
Supplementary Figure 5 (continuation)
(G) Peripheral WBC counts (x10^9/L; left panel), RBC counts (x10^{12}/L; middle panel) and PLT counts (x10^{12}/L; right panel) in symptomatic secondary transplanted mice (blue dots, n=4) compared to normal controls (grey dots, n=4). **(H)** Spleen and liver weight (mg) of symptomatic secondary transplanted mice with BM from diseased recipients of NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ cells (blue dots, n=4) compared to normal control mice (grey dots, n=4). **(I)** Amount of GFP$^+$ cells in the BM of symptomatic secondary transplanted mice (blue dots, n=4). **(J)** CD71 and Ter119 surface expression (%) on GFP$^+$ BM cells from symptomatic secondary transplanted mice (n=4). **(K)** c-Kit and CD71 surface expression (%) on GFP$^+$ BM cells from symptomatic secondary transplanted mice (n=4). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in G and H was tested with unpaired two-tailed t- test.



Supplementary Figure 6. Transplantation of NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ HSPCs induces an erythroleukemia-like disease in mice.

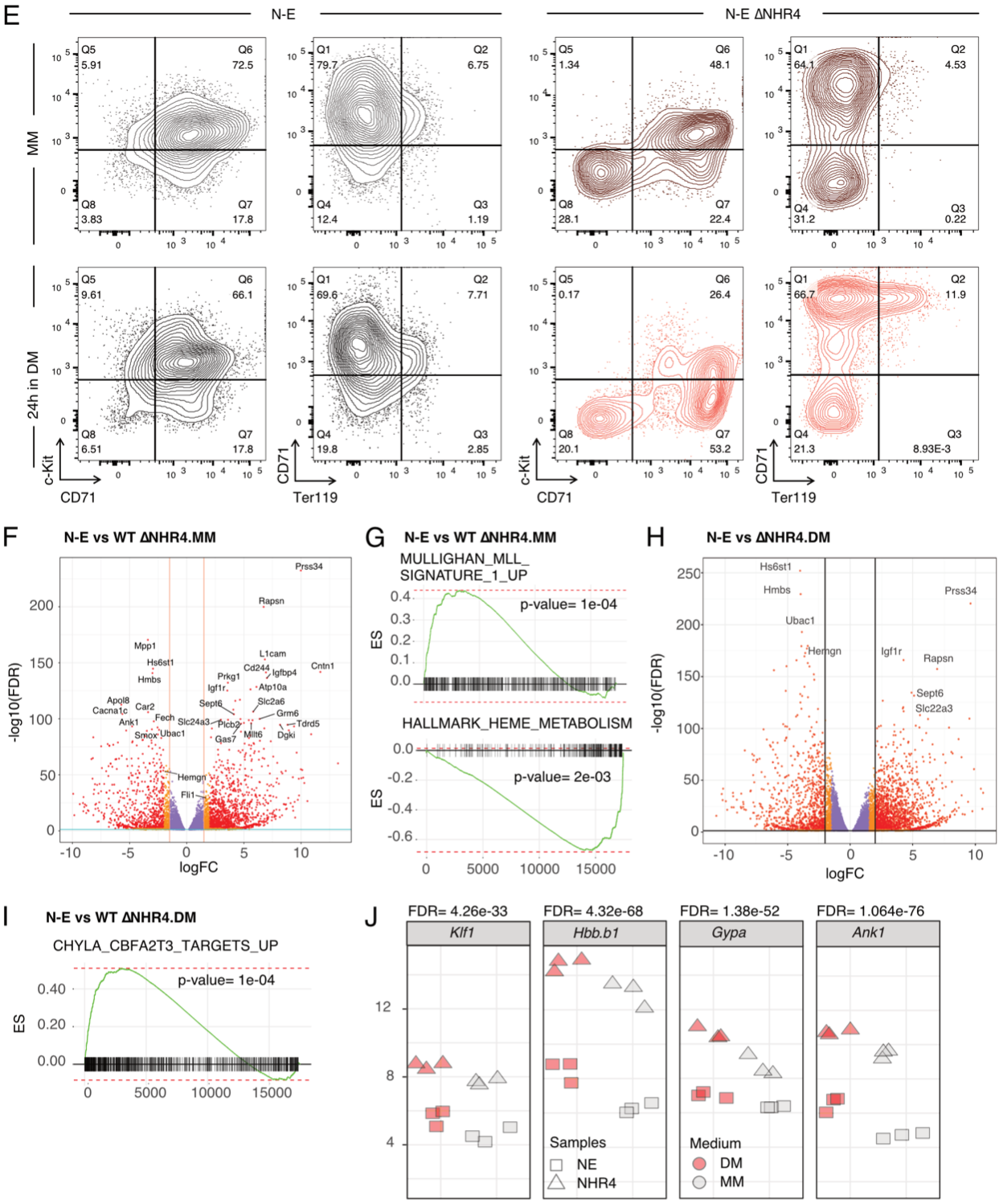
(A) Peripheral WBC counts (x 10^9/L) (left panel), RBC counts (x10^{12}/L) (middle panel) and PLT counts (x10^{12}/L) (right panel) of mice transplanted with NFIA-ETO2- (yellow squares, n=7) or vector-transduced control (CTRL) HSPCs (grey squares, n=4). **(B)** Amount of GFP$^+$ cells (%) before transplantation and at the end of the experiment in BM of NFIA-ETO2- or CTRL-transduced WT HSPCs. **(C)** Representative flow cytometry panels showing amount (%) of GFP$^+$ (left panels), CD71/Ter119 (middle panels) and c-Kit/CD71 (left panels) positive cells in the BM of two symptomatic mice transplanted with NFIA-ETO2-expressing WT erythroblasts. **(D)** Peripheral WBC counts (x10^9/L) (left panel), RBC counts (x10^{12}/L) (middle panel) and PLT counts (x10^{12}/L) (right panel) in secondary transplanted symptomatic mice receiving NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ BM cells (n=4) compared to control (CTRL) mice

transplanted with vector-transduced cells. **(E)** CD71 and Ter119 surface expression (%) (left plot) and c-Kit and CD71 surface expression (left plot) of GFP^+ BM cells from symptomatic secondary recipients of BM from diseased mice transplanted NFIA-ETO2-epressing *TP53*$^{R248Q+}$ cells (n=4). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in D was tested with unpaired two-tailed t- test



Supplementary Figure 7. NFIA-DBD and ETO2-ΔNHR4 deletion mutations both abrogate the potential of NFIA-ETO2 to block erythroid differentiation.

(A) Schematic overview of NFIA domains, DBD and MH1, and ETO2 domains, NHR1, NHR2, NHR3 and NHR4. Shown are single deletions of each NFIA-ETO2 domain that have been established. **(B)** Growth of BM primary erythroblasts expressing NFIA-ETO2 (red line) or single NFIA-ETO2 domain deletions (DBD, yellow line, NHR4, blue line) compared to CTRL (grey line) over 7 days in MM (n=2). **(C)** Representative flow cytometry panels of NFIA-ETO2- or single NFIA-ETO2 domain deletion mutant-transduced erythroblasts showing cells expressing CD71 and/or Ter119 (upper panels) and c-Kit and/or CD71 (right panel). **(D)** CD71 and Ter119 surface expression (left plot) (%) and c-Kit and CD71 surface expression (right plot) on BM-derived erythroblasts expressing NFIA-ETO2, CTRL or the three single NFIA-ETO2 domain deletion mutants. Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean.



Supplementary Figure 7 (continuation)**:**
(E) Representative flow cytometry panels of NFIA-ETO2- (left panels) and NFIA-ETO2-ΔNHR4-transduced (right panels) FL erythroblasts expressing CD71 and/or Ter119 and c-Kit and/or CD71 before (MM) and after 24h in DM. **(F)** Volcano plot of DEG between NFIA-ETO2 and NFIA-ETO2-ΔNHR4 in MM (FDR <0.05, logFC >±1.5 showed by orange and blue lines). **(G)** GSEA of gene expression between WT erythroblast expressing either NFIA-ETO2 or NFIA-ETO2-ΔNHR4 in MM (padj <0.05). (H) GSEA of gene expression between WT erythroblast expressing either NFIA-ETO2 or NFIA-ETO2-ΔNHR4 24h in DM (padj <0.05). **(H)** Volcano plot of DEG between NFIA-ETO2 and NFIA-ETO2-ΔNHR4 cells kept 24h in DM (FDR >0.05, logFC>±1.5 black lines). **(I)** GSEA of gene expression between WT erythroblast expressing either NFIA-ETO2 or NFIA-ETO2-ΔNHR4 24h in DM (padj <0.05). **(J)** *Klf1* (FDR=4.26e-33), Hbβ1 (a.k.a *Hbb.b1,* FDR=4.32e-68), *Gypa* (FDR=1.38e-53) and *Ank1* (FDR=1.064e-76) expression in NFIA-ETO2 (squares) and NFIA-ETO2-ΔNHR4 (triangles) cells in MM (grey) and 24h in DM (orange).

Chapter 3 : Additional experiments (not included in the manuscript)

3.1. Cellular Proliferation of NFIA-ETO2 expressing erythroblasts is EPO-independent

BFU-Es are the first erythroid cells that express EPOR, however, the cells become dependent on EPO for proliferation and survival only after the CFU-E stage (*76*). To study whether leukemic cells from diseased mice transplanted with *TP53*$^{R248Q/+}$ expressing NFIA-ETO2 erythroblasts depend on EPO, I cultured cells from three mice in maintenance medium (MM), in presence or absence of EPO. I used cells from mice with genetic inactivation of the nuclear set domain protein 1 (*Nsd1*$^{-/-}$) that develop an EPO-dependent erythroleukemia-like disease as controls (Leonards *et al.* submitted). Interestingly, in contrast to *Nsd1*$^{-/-}$ cells (**Fig.21B**), NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ erythroblasts were not significantly affected by the absence of EPO and continued proliferating (**Fig.21A**).

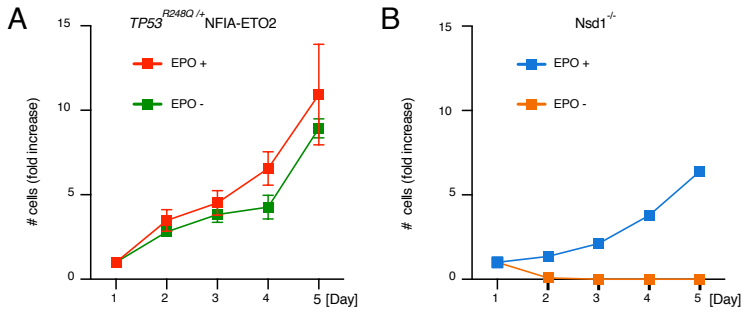


Figure 21. EPO-independent growth of NFIA-ETO2 expressing *TP53*$^{248Q/+}$ erythroblasts. (A) Growth of NFIA-ETO2-expressing *TP53*$^{248Q/+}$ erythroblasts in maintenance medium over 5 is not significantly affected by presence (red line) or absence (green line) of EPO (n=3). **(B)** Growth of *Nsd1*$^{-/-}$ erythroblasts in maintenance medium over 5 day is significantly reduced by absence (orange line) compared to presence (blue line) of EPO (n=1). Values are presented as individual points, error bars as standard error of the mean.

3.2. Expression of a small, ETO2-NHR2-derived peptide (NC128) induced differentiation of NFIA-ETO2-expressing erythroblasts

The oligomerization and co-factor recruitment by the NHR2 domain of ETO proteins was previously shown to be crucial for transforming activity of the AML-associated ETO2-GLIS2 and RUNX1-RUNX1T1 fusion proteins (*278, 325*). Expression of a competing 128aa peptide containing the entire NHR2 domain (known as "NC128") (**Fig.22A**) was shown to interfere with the capacity of ETO proteins to form high molecular weight complexes and to impair the leukemogenic activity of RUNX1-RUNX1T1 and ETO2-GLIS2 respectively (*277, 278*). To explore whether NC128 would interfere with the oncogenic activity of NFIA-ETO2, I virally expressed the peptide in WT or *TP53*$^{R248Q/+}$ NFIA-ETO2-expressing erythroblasts. Compared to vector controls, NC128 transduced WT NFIA-ETO2-expressing cells showed a significantly higher fraction of more mature erythroid cells characterized by expression of CD71^{+} and Ter119^{+} (p=0.0294) cells and reduced c-Kit^{+} expression (**Fig.22B-C**). Expression of NC128 in NFIA-ETO2 expressing *TP53*$^{R248Q/+}$ erythroblasts showed similar effects (p=0.0015) (**Fig.22D-E**). Taken together, this data suggests that the NC128 ETO-intercalating NHR2 peptide at least partly overcomes the NFIA-ETO-mediated terminal erythroid differentiation block in mouse cells in presence or absence of *TP53*$^{R248Q/+}$.

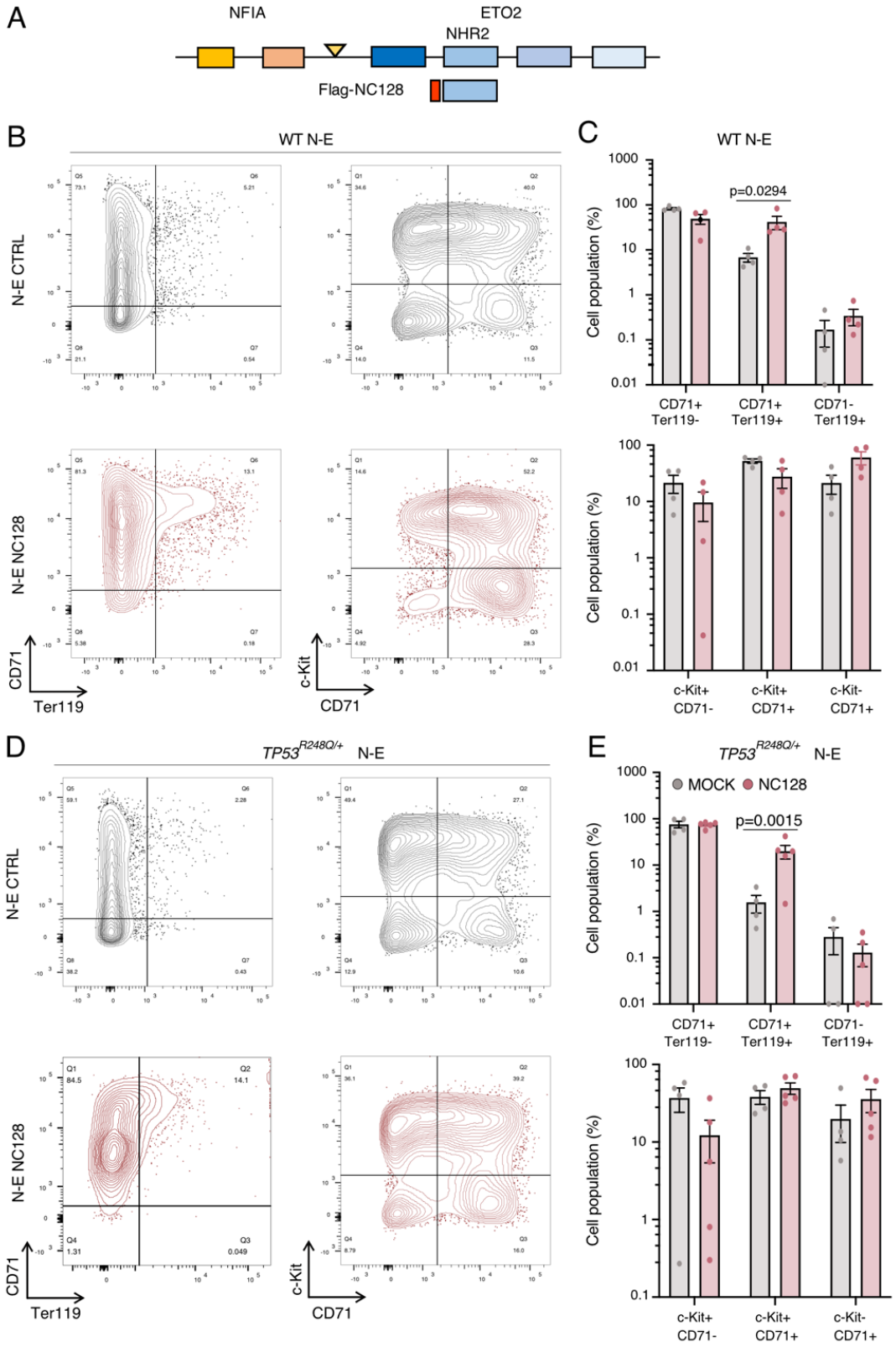


Figure 22. Expression of a competing NHR2-derived peptide (NC128) induces terminal differentiation of NFIA-ETO2-expressing erythroblasts. (A) Schematic overview of NFIA-ETO2 domains. Peptide NC128 corresponds to the NHR2 oligomerization domain (490-578 amino acids). **(B)** Representative flow cytometry panels of NFIA-ETO2-expressing WT BM-derived erythroblasts transduced with a retrovirus expressing NC128 (lower panels) or empty vector (CTRL) (upper panels), showing cells expressing CD71 and/or Ter119 (left panels) and c-Kit and/or CD71 (right panels), on cells grown in MM or 4 days in DM. **(C)** CD71 and Ter119 surface expression (%) (upper plot) and c-Kit and CD71 surface expression (lower plot). Expression of NC128 resulted in a significant increase of $CD71^{+}/Ter119^{+}$ cells not seen in CTRL cells (n=54, p=0.003). **(D)** Representative flow cytometry panels

of NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ BM-derived erythroblast transduced with NC128 (lower panels) or empty vector (CTRL) (upper panels), showing cells expressing CD71 and/or Ter119 (left panels), and c-Kit and/or CD71 (right panels), on cells grown in MM or 4 days in DM. **(E)** CD71 and Ter119 surface expression (%) (upper plot) and c-Kit and CD71 expression (lower plot). Expression of NC128 resulted in a significant increase of $CD71^+/Ter119^+$ cells not seen in vector control cells (n=5, p=0.0015). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in C and E was tested with paired two-tailed t- test.

3.3. Induction of NFIA-ETO2-induced disease by transplantation of BM-derived HSPCs

As interfemoral transplantation of NFIA-ETO2 expressing *TP53*^{R248Q} FL erythroblasts was able to induce an erythroleukemia-like disease (**Fig.5** of the manuscript), we wondered whether transplantation of NFIA-ETO2-transduced BM-derived HSPCs could also induce a PEL-like disease in mice. Before transduction, I cultured BM lineage marker-depleted HSPCs in "stem and progenitor medium" (SPM) composed of 50ng/ml hTPO (Peprotech, London, UK), 20ng/ml hIL-11 (Peprotech, London, UK), 20ng/ml hFLT3 ligand (Peprotech, London, UK) and 100 ng/ml mSCF (Peprotech, London, UK) for 24h. This medium was shown to support *ex vivo* growth of HSCs and early progenitors (*326*). I transduced BM-derived HSPCs with NFIA-ETO2-viral supernatants and before flow-sorting for the GFP-marker, I transplanted 5 x10^5 HSPCs expressing 25% GFP$^+$ cells directly intravenously into lethally irradiated mice. To study the colony formation capacity of these cells, I enriched transduced cells by GFP$^+$ marker and plated them into growth factor-containing methylcellulose (M3434) (**Fig.23A**). Hereby, I observed that NFIA-ETO2 transduced cells formed slightly increased numbers of colonies in MC, with colonies being similar to CFU-GM (granulocytes and monocytes) and CFU-GEMM (granulocytes, erythrocytes, monocytes and megakaryocytes) (data not shown). However, it was not possible to serially propagate them beyond the 3rd plate (**Fig.23B,C**). Although cells harvested from MC cultures expressed similar c-Kit/CD71 levels in the 1st plating (**Fig.23D**), I observed a shift to higher expression of myeloid markers Mac-1 and Gr-1 in NFIA-ETO2 expressing colonies (**Fig.23E**). However, none of the transplanted mice ever developed any symptoms during an observation period of >1 year (**Fig.23F**). Cells in the peripheral blood expressed very low numbers of GFP$^+$ cells after the first 4 weeks post-transplantation, further decreasing over the 32 weeks after transplantation (**Fig.23G**). In addition, NFIA-ETO2 transplanted mice did not present any aberrant blood counts (**Fig.23H**). Taken together, I found that retroviral transduction of BM-derived HSPCs by NFIA-ETO2 slightly increased the colony formation capacity up to 3° rounds, however, transplanted mice did not develop any disease.

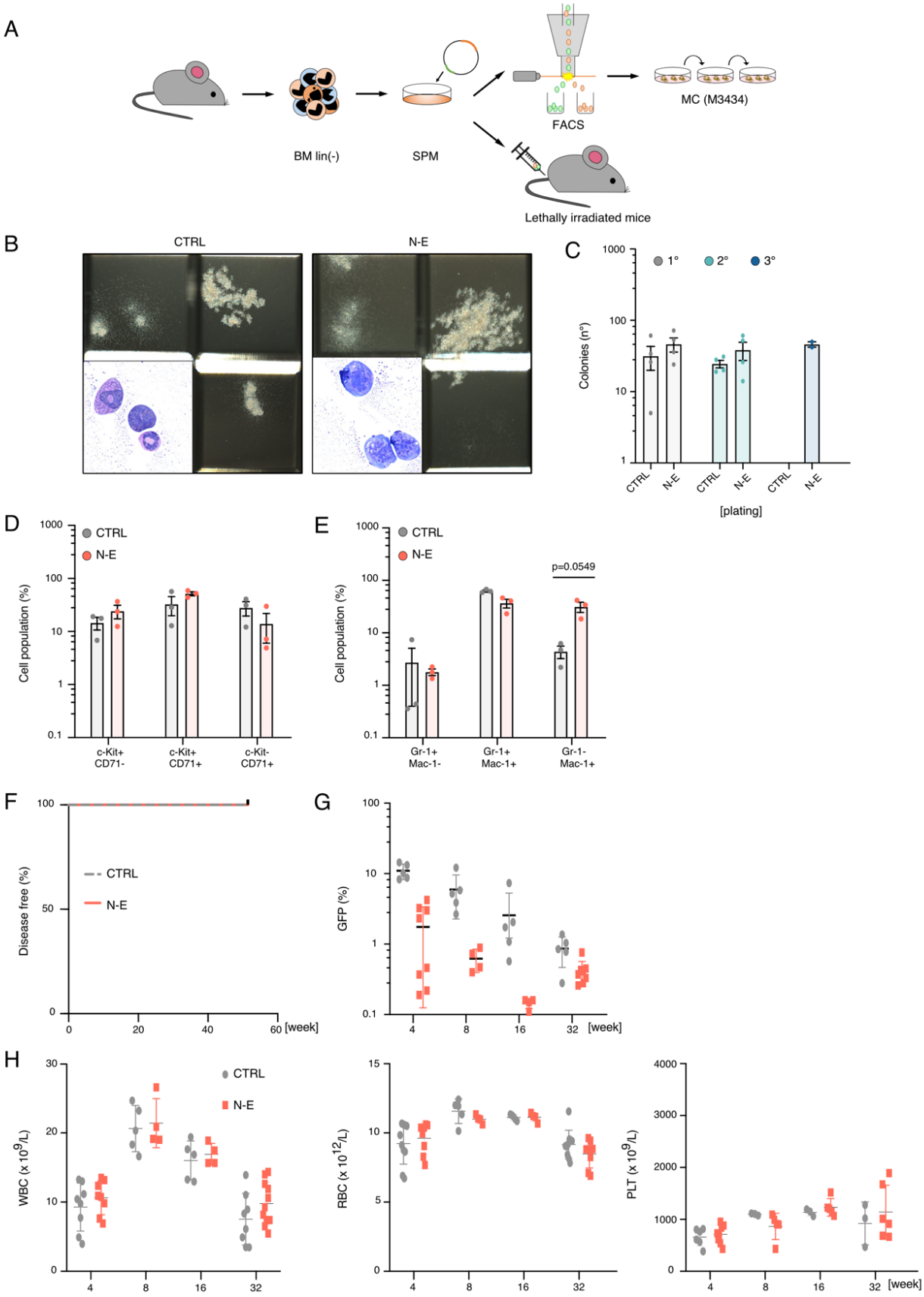


Figure 23. Expression of NFIA-ETO2 modestly increased the clonogenic activity of BM-derived HSPCs. (A) Schematic overview of experimental procedure. **(B)** Images of methylcellulose cultures (MC) (x4) recorded seven days after the first plating of HSPCs expressing NFIA-ETO2 plated in M3434 compared to vector-transduced control cells (CTRL). Lower left inserted are representative images of Wright Giemsa-stained cytospin preparations of cells harvested from the first plating. Images were recorded with a 60x objective using a Nikon-TI. These pictures are

representative for 3 independent experiments. **(C)** Quantification of transformation assay of HSPCs expressing NFIA-ETO2 compared to CTRL. Shown are absolute numbers of colonies after the first, second and third plating in MC. In contrast to vector controls, NFIA-ETO2 cells could be propagated to the third round but not beyond (2°plate: n=4, 3°plate: n=2). **(D)** c-Kit and CD71 surface expression (%) on cells harvested from the first MC plating of HSPCs expressing NFIA-ETO2 compared to CTRL did not reveal any difference (n=3). **(E)** Mac-1 and Gr-1 surface expression on cells harvested from the first plating of HSPCs expressing NFIA-ETO2 compared to CTRL revealed increased number of $Gr1^{-}/Mac1^{+}$ cells expressing NFIA-ETO than CTRL (n=3, p=0.005). **(F)** Kaplan Meier plot of mice transplanted with NFIA-ETO2 (orange line) or vector (grey line) transduced HSPCs (n=8, n=5, respectively). **(G)** Percentage of GFP^{+} NFIA-ETO2 (orange dots) or CTRL-transduced (grey dots) cells in the peripheral blood of transplanted mice after 4, 8, 16 and 32 weeks after transplantation. **(H)** Peripheral WBC counts ($x10^{9}$/L) (left panel), RBC counts ($x10^{12}$/L) (middle panel) and PLT counts ($x10^{12}$/L) (right panel) of mice transplanted with NFIA-ETO2 (orange dots) or CTRL (grey dots) transduced HSPCs. Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in E was tested with paired two-tailed t- test.

3.4. Structure-function analysis of NFIA-ETO2

As outlined before, to dissect critical domains of the NFIA-ETO2 fusion I generated FLAG-tagged, single domain deletion mutants of the DNA-binding domain (DBD) and the Mad homology domain (MH1) of the NFIA moiety and NHR1, NHR2 and NHR4 domain of the ETO2 part, respectively. I then confirmed the expression of these mutants in primary mouse erythroblasts at the mRNA level (**Supplementary Fig.7.1** of the manuscript). However, as none of the mutants possessed transforming potential of NFIA-ETO2, I was unable to grow sufficient primary erythroblasts to show expression on protein level by Western blot. Therefore, I also cloned the mutants into *pcDNA3.1*, a CMV-promoter-driven expression vector and transfected HeLa cells. I detected expression of WT and mutant NFIA-ETO2 by using a FLAG Ab (**Fig.24A**). However, despite the fact, that I boiled the samples (directly lysed in Laemmli sample buffer) to denature the proteins, we detected a second band in WT and most mutants except the ΔNHR1, which was previously reported to be responsible of high weight complex formation (*127*). We therefore hypothesized that this second band might be the result of oligomerization of NFIA-ETO2 with other ETO proteins. Remarkably, the lack of NHR1 may also affect the stability of the NFIA-ETO2 proteins, as this mutant was expressed at lower level than the others.

I next expressed the mutants in MEL cells and validated their expression at the mRNA level (**Fig.24B**). I particularly checked for protein expression of the NFIA-ETO2-ΔNHR4 mutant that I used for the functional experiments in primary cells and found similar levels as in HeLa cells (**Fig.24C**). After DMSO-induced erythroid differentiation, MEL cells expressing the mutants expressed higher mRNA levels of Hbα1 and Hbβ1 compared to WT NFIA-ETO2, as well as to non-induced control cells (**Fig.24D,E**).

All in all, I found that most deletion mutants lacking single NFIA-ETO2 domains could be expressed to similar levels (except ΔNHR1). As expression of the NFIA-ETO-ΔNHR4 mutant was unable to impair DMSO-induced erythroid differentiation of MEL cells, I used this mutant to compare it with the full-length NFIA-ETO fusion in primary cells.

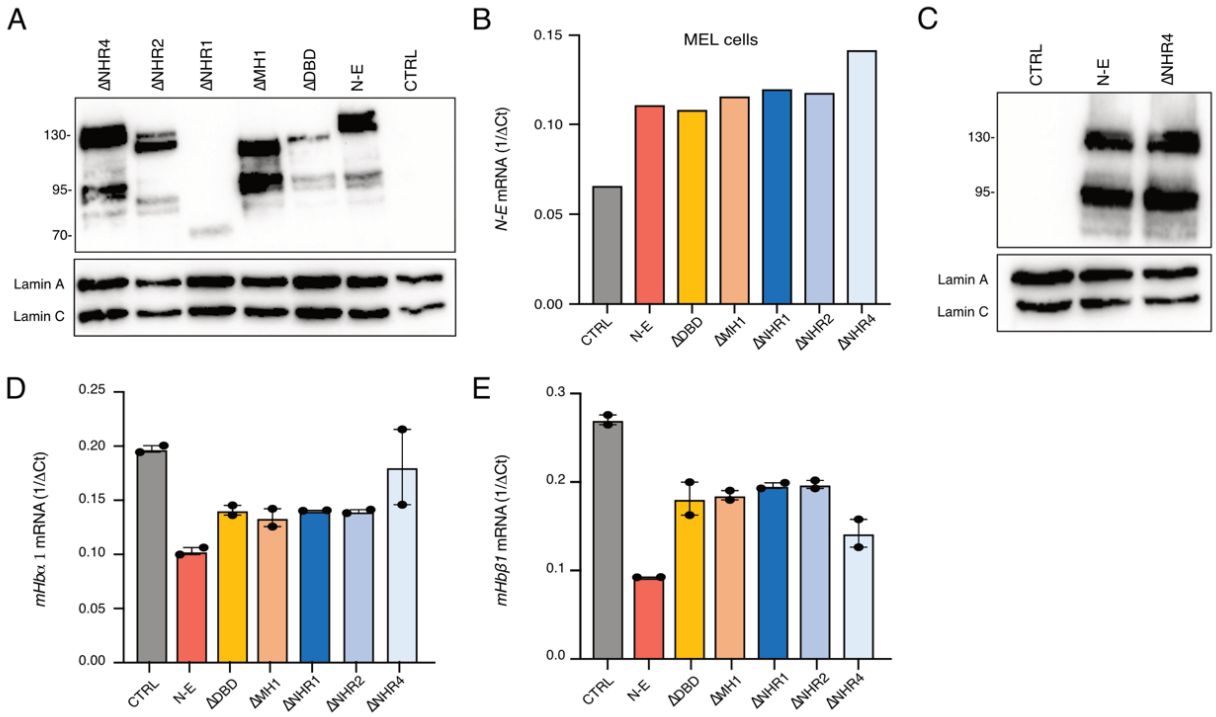


Figure 24. Expression of NFIA-ETO2 deletion mutants in HeLa and MEL cells. (A) Expression of transiently transfected full-length NFIA-ETO2 and derived deletion mutants (ΔNHR4, ΔNHR2, ΔNHR1, ΔMH1, ΔDBD) in HeLa cells assessed by immunoblotting of nuclear extracts. LAMIN-A/C was used as loading control. Vector-transduced cells served negative controls (CTRL). **(B)** Expression of full-length *NFIA-ETO2* and derived deletion mutants (ΔNHR4, ΔNHR2, ΔNHR1, ΔMH1, ΔDBD) in MEL cells at the mRNA level assessed by quantitative RT-PCR (n=2). **(C)** Expression of full-length NFIA-ETO2 and the ΔNHR4 deletion mutant in MEL cells assessed by immunoblotting of nuclear extracts. LAMIN-A/C was used as loading control. CTRL-transduced cells as negative control. **(D)** Expression of *mHbα1* at the mRNA level in MEL cells expressing full-length or NFIA-ETO2 deletion mutants after 4 days of induced differentiation by 2% DMSO measured by quantitative RT-PCR (n=2). **(E)** Expression of *mHbβ1* at the mRNA level in MEL cells expressing full-length or NFIA-ETO2 deletion mutants after 4 days of induced of differentiation by 2% DMSO measured by quantitative RT-PCR (n=2). Values are presented as individual points, bar graphs represent the mean value of technical replicates, error bars as standard error of the mean.

3.5. Transforming potential of the ZMYND8-RELA fusion

To study the impact of the ZMYND8-RELA fusion in terminal erythroid differentiation, I used the same models as for NFIA-ETO2: MEL cells (**Fig.25A**) and primary (adult or FL-derived) mouse erythroblasts. First, I transduced MEL cells and confirmed *ZMYND8-RELA* expression on mRNA level (**Fig.25B**). I found that expression of the fusion did not significantly alter MEL cell proliferation compared to vector-transduced control cells. ZMYND8-RELA transduction resulted in slightly increased proliferation of MEL cells upon induced differentiation by 2% DMSO (**Fig.25C**). ZMYND8-RELA-transduced MEL cells expressed slightly lower hemoglobin levels assessed by benzidine staining (**Fig.25D,E**). Flow cytometry analysis showed that ZMYND8-RELA expression resulted in a small reduction of $CD71^{+}/Ter119^{+}$ population compared to control cells after 4 days of induced differentiation (**Fig.25F**). Notably, mRNA levels of hemoglobin β and γ did not differ from control cells, however, even though the variance between samples was large, levels of hemoglobin α were significantly increased in ZMYND8-RELA (**Fig.25G**), suggesting that erythroid differentiation was partially blocked by the fusion gene.

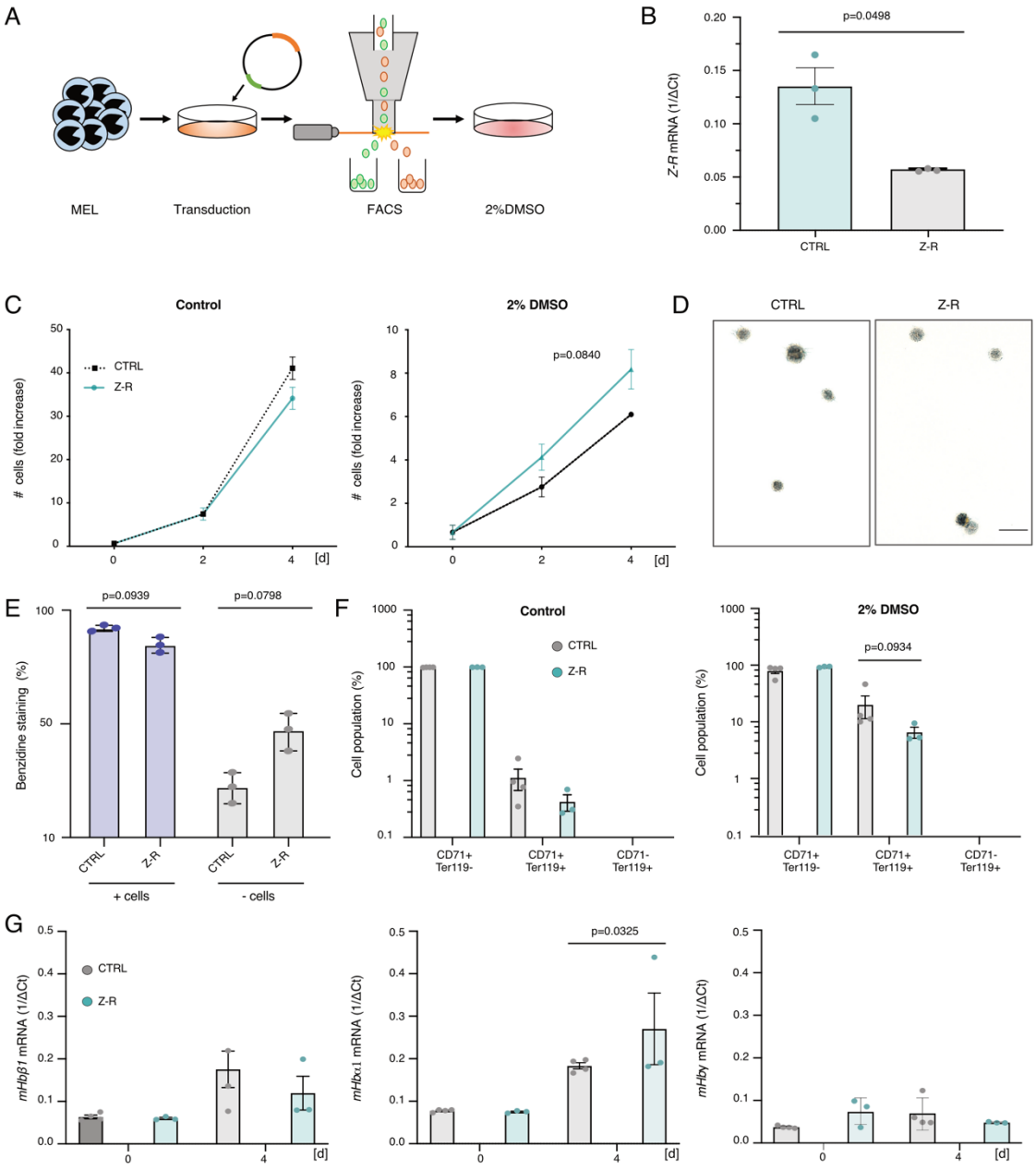


Figure 25. Expression of the ZMYND8-RELA (Z-R) fusion impaired DMSO-induced differentiation of MEL cells. **(A)** Schematic overview of experimental procedure. MEL cells were cultured in DMEM and transduced with ZMYND8-RELA or empty vector (CTRL), enriched by sorting for transduced cells (GFP+) and cultured in DMEM with 2% DMSO to induce erythroid differentiation. **(B)** *ZMYND8-RELA* mRNA expression in transduced MEL cells measured by RT-qPCR (n=3). **(C)** Growth of MEL cells transduced with ZMYND8-RELA or CTRL in DMEM over 4 days (n=3). Left graph represents proliferation without DMSO induction; right graph represents proliferation with 2% DMSO. **(D)** Benzidine staining of cytospins of MEL cells expressing CTRL or ZMYND8-RELA after 4 days of DMSO induction. Images were recorded with a 60x objective using a Nikon-TI. Scale-bars = 50 μm. **(E)** Quantification of benzidine staining (n=3) (right panel) (n=3). **(F)** CD71 and Ter119 surface expression (%) on ZMYND8-RELA- and CTRL-transduced MEL cells before (Control) and after 4 days of induced differentiation in 2% DMSO (CTRL: n=4, Z-R: n=3). **(G)** *mHbα1, mHbβ1* and *mHbγ* mRNA levels in MEL cells expressing ZMYND8-RELA compared to vector CTRL before and after 4 days of induced differentiation in 2% DMSO (n=3). Values are presented as individual points, bar graphs represent the mean value of technical replicates, error bars as standard error of the mean. Statistical significances in B, C, E, F and G was tested with paired two-tailed t- test.

I also transduced the ZMYND8-RELA fusion into primary adult and FL-derived mouse erythroblasts. In maintenance medium, ZMYND8-RELA transduced cells did not proliferate more than vector-transduced control cells and started to die after a few days (**Fig.26A**). Interestingly, upon induced differentiation, ZMYND8-RELA transduced cells seemed to proliferate without differentiation for the first two days, however, the cells rapidly lost GFP (reflecting mRNA expression of the transduced transgene) and showed a highly variable phenotype with some of them being apoptotic or differentiating towards erythrocytes, as visualized on cytospin preparations (**Fig.26B,C**). Upon plating in methylcellulose (M3434), ZMYND8-RELA transduced erythroblasts could not be propagated beyond the first plate (**Fig.26D,E**).

I also transduced lineage-marker depleted BM cells grown in SPM medium with ZMYND8-RELA. Upon plating in in methylcellulose (M3434) I was able to replate the cells, similar to NFIA-ETO2, for up to 3 rounds (**Fig.26F,G**). However, I observed that ZMYND8-RELA transduced cells from the first plate did not differ from the vector-transduced controls in colony morphology and c-Kit/CD71 and Gr-1/Mac-1 surface marker expression (**Fig.26H,I**). All in all, even though ZMYND8-RELA seemed to slightly block terminal maturation of MEL cells, I could not establish primary BM cells or erythroblasts that stably expressed ZMYND8-RELA.

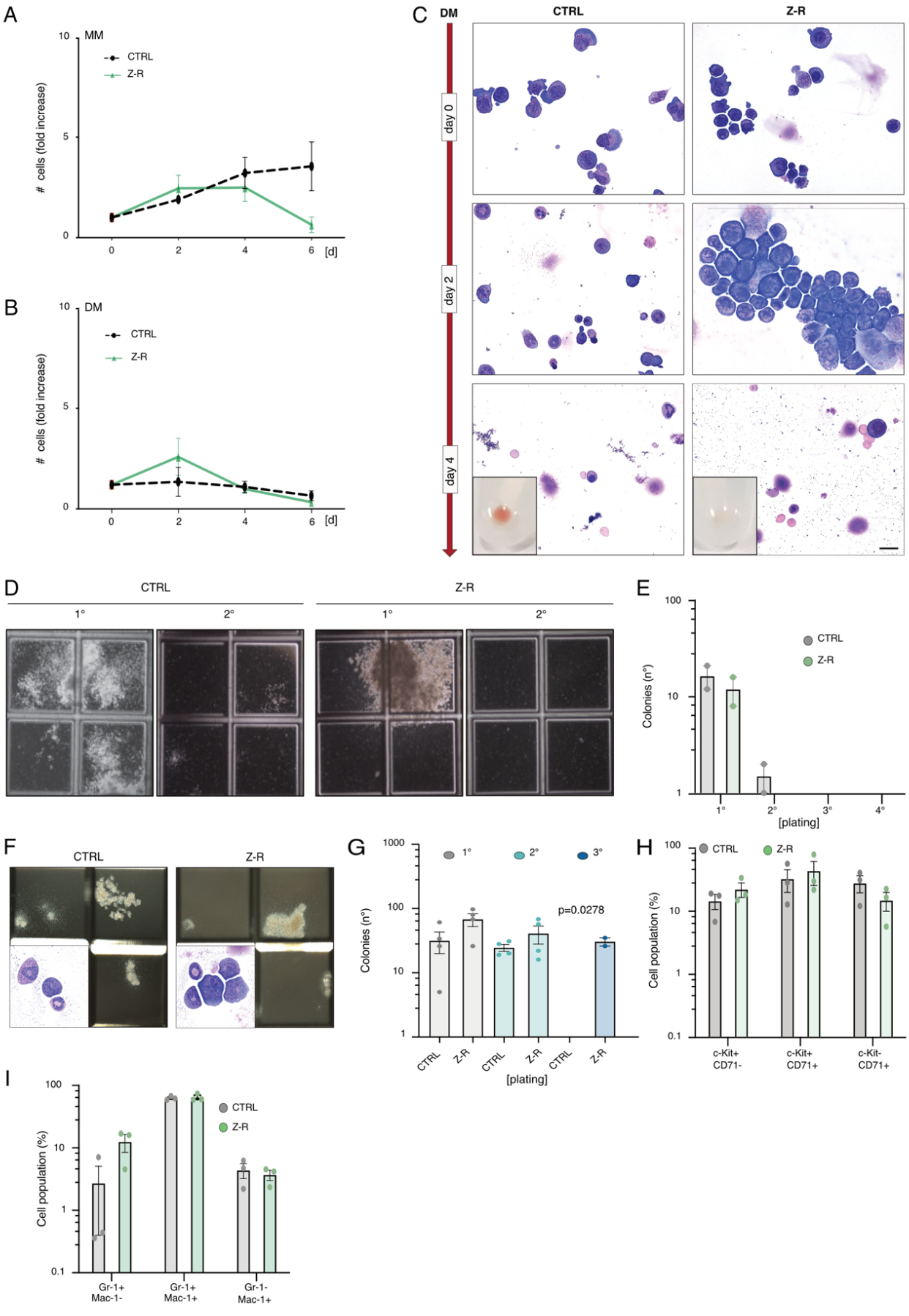

A
MM
CTRL
Z-R
cells (fold increase)
10
5
0
0
2
4
6
[d]
B
DM
CTRL
Z-R
cells (fold increase)
C
DM
CTRL
Z-R
day 0
day 2
day 4
D
CTRL
Z-R
1°
2°
1°
2°
E
Colonies (n°)
CTRL
Z-R
100
10
1
1°
2°
3°
4°
[plating]
F
CTRL
Z-R
G
Colonies (n°)
1°
2°
3°
p=0.0278
1000
100
10
1
CTRL
Z-R
CTRL
Z-R
CTRL
Z-R
[plating]
H
CTRL
Z-R
Cell population (%)
100
10
1
0.1
c-Kit+
CD71-
c-Kit+
CD71+
c-Kit-
CD71+
I
Cell population (%)
CTRL
Z-R
100
10
1
0.1
Gr-1+
Mac-1-
Gr-1+
Mac-1+
Gr-1-
Mac-1+

Figure 26. Effects of ZMYND8-RELA expression in BM-derived mouse primary erythroblasts. (A) Growth of BM-derived erythroblasts transduced with ZMYND8-RELA or empty vector (CTRL) over 6 days in MM (n=3). **(B)** Growth of BM-derived erythroblasts transduced with ZMYND8-RELA or empty vector (CTRL) over 6 days in DM (n=3). **(C)** Representative images (x600) of Wright Giemsa-stained cytospin preparations from primary erythroblasts transduced with ZMYND8-RELA compared to empty vector (CTRL). Upper panels: erythroblasts upon expansion in MM (day 0). Middle panel: erythroblasts after 2 days in DM. Lower panel: erythroblasts after 4 days in DM. Images were recorded with a 60x objective using a Nikon TI. **(D)** Representative images of MC cultures (x4) recorded seven days after the first plating of erythroblasts transduced with ZMYND8-RELA compared to empty vector (CTRL). **(E)** Numbers of colonies formed by erythroblasts transduced with ZMYND8-RELA compared to vector CTRL in MC (3434). Shown are absolute numbers after the first and second plating (n=2). **(F)** Representative images (x4) of MC cultures (M3434) recorded seven days after the first plating of HSPCs transduced with ZMYND8-RELA compared to CTRL transduced cells. Lower left inserts are representative images (x100) of Wright Giemsa-stained cytospin preparations from cells harvested seven days after the first plating. **(G)** Numbers of colonies formed by HSPCs transduced with ZMYND8-RELA compared to empty vector (CTRL) in MC (3434) after the first, second and third plating (1st and 2nd plates: n=4, 3rd plate: n=2). **(H)** c-Kit and CD71 surface expression (%) on cells harvested after the first MC plating of HSPCs transduced with ZMYND8-RELA (Z-R) compared to empty vector (CTRL) (n=3). **(I)** Mac-1 and Gr-1 surface expression (%) on cells harvested from the first MC plating of HSPCs transduced with ZMYND8-RELA compared to empty vector (CTRL) (n=3). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in G was tested with paired two-tailed t- test.

3.5.1. Cooperation of ZMYND8-RELA with TP53^{R248Q}

NFIA-ETO2 increased proliferation as well as the clonogenic activity of primary TP53^{R248Q} erythroblasts in methylcellulose (**Supplementary Fig.4** of the manuscript). However, as outlined above, I could not establish BM-derived erythroblasts from WT mice that stably expressed ZMYND8-RELA, either in MM (**Fig.27A**), nor DM (**Fig.27B**).

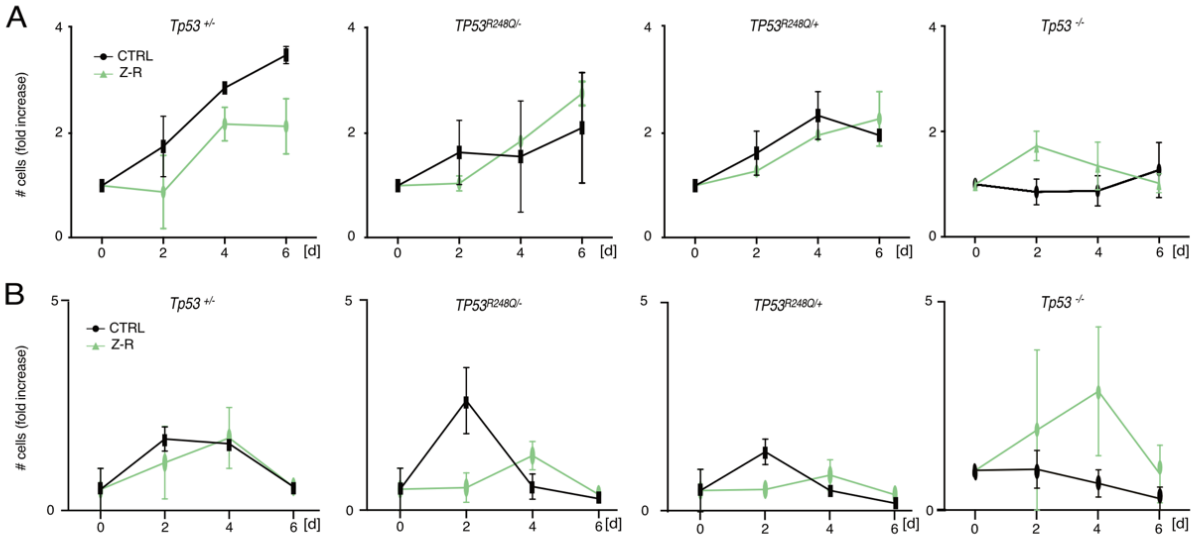


Figure 27. Growth of ZMYND8-RELA transduced BM-derived erythroblasts with different *Tp53* genotypes. (A) Total cell number (fold increase) of erythroblasts carrying different *Tp53* genotypes (*Tp53*$^{+/-}$, *TP53*$^{R248Q/-}$, *TP53*$^{R248Q/+}$, *Tp53*$^{-/-}$), transduced with ZMYND8-RELA compared to vector-transduced (CTRL) grown in MM for 6 days (n=2). **(B)** Total cell number (fold increase) of erythroblasts carrying different *Tp53* genotypes (*Tp53*$^{+/-}$, *TP53*$^{R248Q/-}$, *TP53*$^{R248Q/+}$, *Tp53*$^{-/-}$), transduced with ZMYND8-RELA compared to vector controls (CTRL) grown in DM for 6 days (n=2). Values are presented as individual points, error bars as standard error of the mean.

However, rather surprisingly, I observed that ZMYND8-RELA transduced *TP53*^{R248Q} erythroblasts (in presence or absence of a normal *Tp53* allele) showed a significantly increased clonogenic activity and formed very round and dense colonies in MC (**Fig.28A**). In contrast to WT cells, ZMYND8-RELA transduced *TP53*$^{R248Q/+}$ and *TP53*$^{R248Q/-}$ erythroblasts could be propagated up to 5 rounds. *Tp53*$^{+/-}$ cells expressing the fusion, on the other hand, formed only few colonies in the first plating and were not able to replate, whereas *Tp53*$^{-/-}$ cells could be propagated up to 4 rounds (**Fig.28B**). Interestingly, in the 2nd plate, ZMYND8-RELA-transduced cells formed colonies composed of cells, which were similar to the non-colony forming cells in vector-transduced control cultures. However, they expressed higher levels of the stemness marker c-Kit^{+} (**Fig.28C**). Additionally, in contrast to controls, Mac-1^{+} and Gr-1^{+} populations seemed to be reduced in *TP53*$^{R248Q/-}$ ZMYND8-RELA cells (**Fig.28D**). After 4 platings, ZMYND8-RELA-transduced cells carrying a *TP53*^{R248Q} allele kept a stem-like phenotype expressing high c-Kit levels, whereas *Tp53*$^{-/-}$ cells lost the c-Kit^{+} population (**Fig.28E**).

Collectively, I found that TP53 mutations did not affect the growth of ZMYND8-RELA transduced erythroblasts in liquid medium, however expression of *TP53*^{R248Q} or loss of both WT *Tp53* alleles increased their clonogenic activity suggesting functional cooperation between the fusion and reduced *TP53* activity.

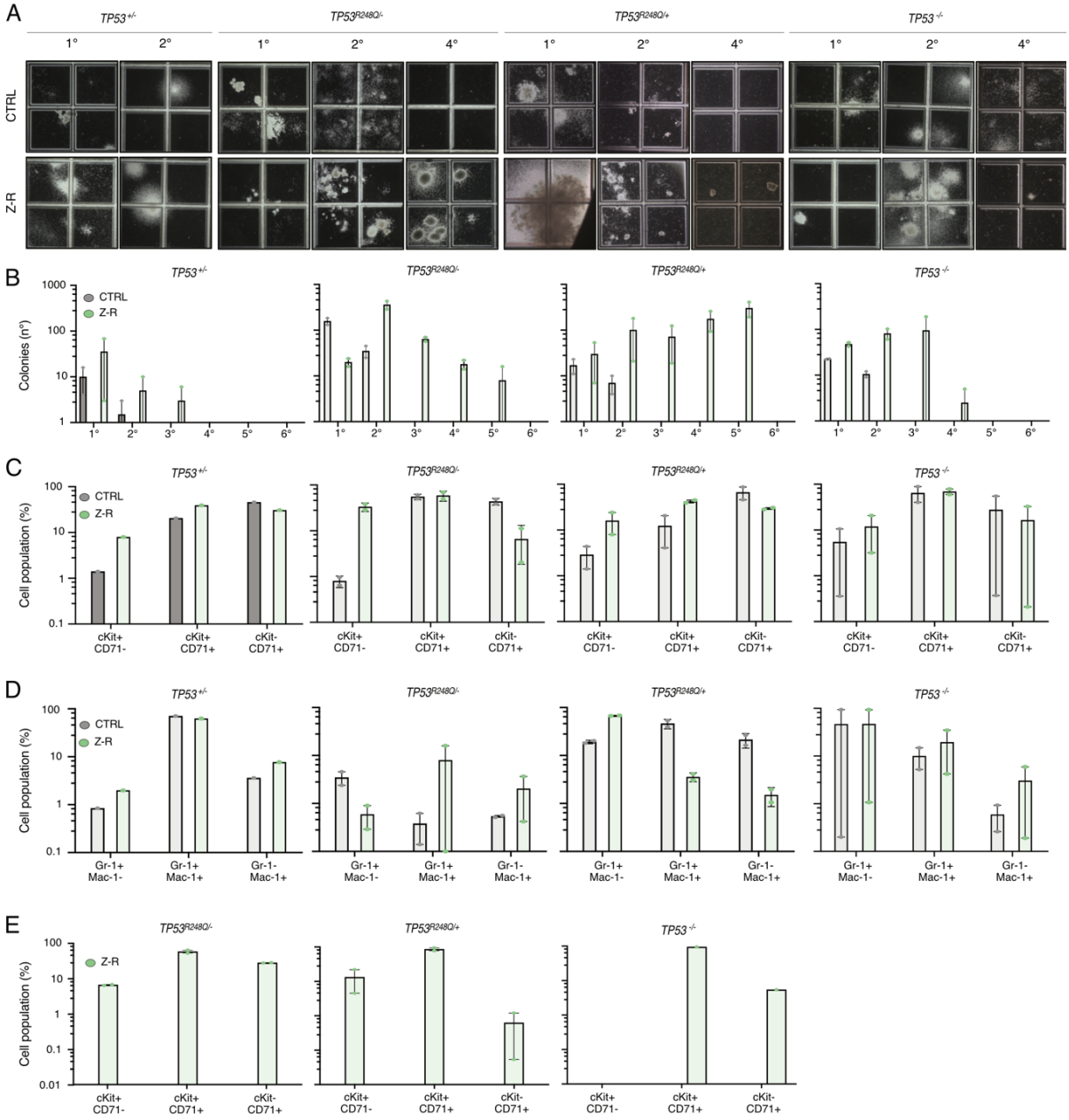


Figure 28. Clonogenic activity of ZMYND8-RELA transduced BM-derived erythroblasts with different *Tp53* genotypes. (A) Representative images (x4) of MC cultures (3434) of BM-derived erythroblasts with different *Tp53* genotypes (*Tp53*$^{+/-}$, *TP53*$^{R248Q/-}$, *TP53*$^{R248Q/+}$, *TP53*$^{-/-}$) transduced with ZMYND8-RELA compared to vector control (CTRL). Shown are the first, second and fourth platings. **(B)** Number of colonies in MC formed by BM-derived erythroblast with different *Tp53* genotypes (*Tp53*$^{+/-}$, *TP53*$^{R248Q/-}$, *TP53*$^{R248Q/+}$, *Tp53*$^{-/-}$) transduced with ZMYND8-RELA compared to CTRL. Shown are absolute numbers from the first to sixth plating (n=2). **(C)** c-Kit and CD71 surface expression (%) on BM-derived erythroblasts with different *Tp53* genotypes (*Tp53*$^{+/-}$, *TP53*$^{R248Q/-}$, *TP53*$^{R248Q/+}$, *Tp53*$^{-/-}$) transduced with ZMYND8-RELA compared to CTRL harvested from the second plating in MC (n=2). **(D)** Gr-1 and Mac-1 surface expression (%) on BM-derived erythroblasts with different *Tp53* genotypes (*Tp53*$^{+/-}$, *TP53*$^{R248Q/-}$, *TP53*$^{R248Q/+}$, *Tp53*$^{-/-}$) transduced with ZMYND8-RELA compared to CTRL harvested from the second plating in MC (n=2). **(E)** c-Kit and CD71 surface expression (%) on BM-derived erythroblasts with different *Tp53* genotypes (*Tp53*$^{+/-}$, *TP53*$^{R248Q/-}$, *TP53*$^{R248Q/+}$, *Tp53*$^{-/-}$) transduced with ZMYND8-RELA compared to CTRL harvested from the fourth plating in MC. Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean.

3.5.2. Transplantation of ZMYND8-RELA-transduced *TP53*$^{R248Q/+}$ cells did not result in a leukemic phenotype

Since ZMYND8-RELA, similar to NFIA-ETO2, increased the clonogenic activity of *TP53*^{R248Q} erythroblasts, I hypothesized that these cells might be also be able to induce erythroleukemia-like disease in mice. Therefore, I intravenously transplanted *TP53*$^{R248Q/+}$ FL-derived HSPCs expressing ZMYND8-RELA or vector control into lethally irradiated mice (**Fig.29A**). However, transplanted mice did not develop any symptoms over one year of observation (**Fig.29B**) and the blood counts never differed from control mice (**Fig.29C**). Although I transplanted cells expressing high levels of GFP, indicating high ZMYND8-RELA expression, I observed a drastic reduction of GFP$^+$ cells in peripheral blood already 4 weeks post-transplantation, and I could barely detect any GFP$^+$ cells in the BM at the end of the experiment (**Fig.29D**).

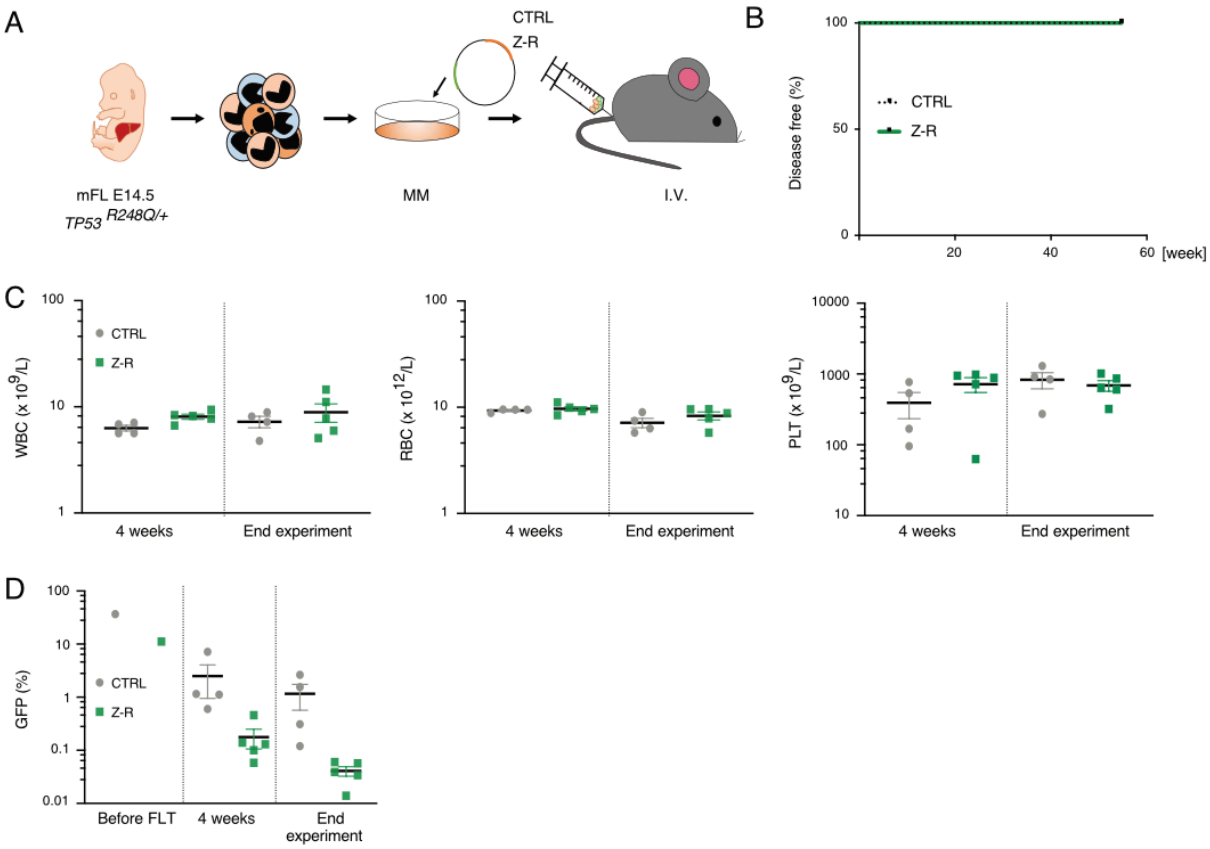


Figure 29. Transplantation of FL-derived *TP53*$^{R248Q/+}$ ZMYND8-RELA transduced HSPCs into lethally irradiated mice did not result in any disease. (A) Schematic overview of the experiment: FL-derived *TP53*$^{R248Q/+}$ HSPCs transduced with ZMYND8-RELA or CTRL were transplanted intravenously (I.V.) (with normal BM cells) into lethally irradiated mice. **(B)** Kaplan Meier plot of mice transplanted with ZMYND8-RELA (green line, n=5) or CTRL (black dotted line, n=4) HSPCs. None of the mice did develop any symptoms of disease. **(C)** Peripheral WBC counts (x10^9/L) (left panel), RBC counts (x10^{12}/L) (middle panel) and PLT counts (x10^{12}/L) (right panel) of mice transplanted with ZMYND8-RELA (green dots, n=5) or CTRL (grey dots, n=4) transduced *TP53*$^{R248Q/+}$ HSPCs. **(D)** Percentage of GFP$^+$ cells of transduced cells before transplantation, 4 weeks after transplantation in peripheral blood and at the end of the experiment (BM). ZMYND8-RELA transduced cells are given in green, CTRL in grey dots. Values are presented as individual points, error bars as standard error of the mean.

Chapter 4 : Discussion

In this project, I studied the transforming mechanisms of the erythroleukemia-associated NFIA-ETO2 and ZMYND8-RELA fusion genes. All fusion partners are transcription factors or transcriptional corepressors previously shown to be regulators of normal hematopoiesis and erythropoiesis. The fact that both fusion proteins were first identified in erythroleukemia cells from young infants, in an age when not many additional genetic defects are normally found, suggests that they might be drivers of the disease (*169*). Therefore, we hypothesized that NFIA-ETO2 and ZMYND8-RELA fusion genes interfere with normal erythropoiesis leading to pediatric erythroleukemia, which in most cases is characterized as pure erythroleukemia (PEL).

My experiments revealed that expression of NFIA-ETO2 impairs terminal erythroid differentiation of murine erythroblasts and cooperates with the $TP53^{R248Q}$ mutation to induce a PEL-like disease in mice. Molecular studies showed that upon induced erythroid differentiation, NFIA-ETO2 represses genes that are normally controlled by NFIA and the erythroid master regulators GATA1 and TAL1. Furthermore, proto-oncogenes, previously shown to be involved in malignant hematopoiesis, were found highly upregulated. Of note, expressing NFIA-ETO2 in the presence of the $TP53^{R248Q}$ mutation resulted in downregulation of several known tumor suppressive *TP53* target genes.

In contrast to NFIA-ETO2, I could not stably express the ZMYND8-RELA fusion in primary erythroblasts cultured in liquid medium. However, I observed that ZMYND8-RELA expression allowed serial replating of $TP53^{R248Q}$ erythroblasts in methylcellulose, but, unlike NFIA-ETO2, transplantation of ZMYND8-RELA transduced $TP53^{R248Q/+}$ BM-derived HSPCs did not induced any hematopoietic disease in lethally irradiated recipients.

NFIA-ETO2 expression impairs erythroid differentiation but is not sufficient to induce leukemia

I found that retroviral expression of NFIA-ETO2 not only blocked terminal erythroid differentiation of murine erythroleukemia (MEL) cells but also primary BM- or FL-derived erythroblasts. Although a wide range of fusion genes had been shown to play a role in AML, such as RUNX1-RUNXT1, PML-RARA, NUP98-HOXA9 or various MLL-fusions (*170, 171, 327*), to date no fusion has been mechanistically linked to the erythroleukemia subtype.

NFIA-ETO2 expression only slightly increased colony formation of erythroblasts in MC enriched for myeloid and erythroid lineage cytokines (**Fig.2**). Similarly, NFIA-ETO2 expression also increased colony formation of HSPCs enabling replating for maximally three rounds, forming mainly CFU-GM (granulocytes and monocytes) and CFU-GEMM (granulocytes,

erythrocytes, monocytes and megakaryocytes)-like colonies (**Fig.23**). This shift towards a myeloid fate might be due to the presence of myeloid specific cytokines in the medium, which allow cells that are blocked for erythroid differentiation to follow the alternative myeloid differentiation routes.

Generally, the limited increase in self-renewal capacity of NFIA-ETO2 expressing cells in MC cultures was rather unexpected, as expression of other leukemia-associated ETO containing fusions (RUNX1-RUNX1T1, ETO2-GLIS2) were shown to significantly increase the serial propagation of HSPCs-derived colonies in MC (*276, 277, 328*).

To investigate whether expression of the NFIA-ETO2 fusion would be able to induce leukemia *in vivo*, I intravenously transplanted BM-derived lineage-marker depleted HSPCs expressing NFIA-ETO2 into irradiated mice. Four weeks post-transplantation, I found very low numbers (0.5-5%) of GFP^+ cells in peripheral blood of transplanted mice. After 1 year of observation, none of the mice ever developed symptoms and showed barely any GFP^+ cells in the peripheral blood. These observations suggest that I either did not transplant enough HSPCs expressing NFIA-ETO2 (that efficiently engraft) due to the low transduction efficiency, or that NFIA-ETO-expressing HSPCs are not able to induce a disease, suggesting the requirement for cooperative mutations to induce leukemogenesis. The latter being supported by the observation of only marginally increased clonogenic activity of HSPCs expressing NFIA-ETO2 in MC. Additionally, it could also be that retroviral transduction did not reach the most susceptible cells: it is well-known that HSC proliferation is associated with differentiation during the short culture period that is necessary to bring cells into cycle (*329*).

Notably, these findings are supported by an earlier study in which BM cells expressing the RUNX1-RUNX1T1 fusion were transplanted into lethally irradiated recipients, but additional treatment with a mutagenic dose of DNA alkylating agent, N-ethyl-N-nitrosoure (ENU) was necessary to development acute myeloid leukemia/granulocytic sarcoma (*330*). Another previous work described a significantly increased colony formation *in vitro* of HSPCs retrovirally overexpressing the ETO2-GLIS2. Preliminary data showed that transplantation of transduced cells into syngeneic mice failed to induce a leukemia phenotype, suggesting the necessity of a cooperative mutation to develop leukemia (*331*). However, recently published study showed that transgenic mice expressing ETO2-GLIS2 only induced rapid megakaryoblastic leukemia when the fusion was expressed in fetal hematopoietic stem cells, suggesting that aggressiveness and phenotype depend on the cell-of-origin (*332*).

What are the critical domains of NFIA-ETO2 to block erythroid differentiation?

NFIA-ETO2 contains the DNA-binding domain (DBD) of NFIA fused to almost the entire ORF of ETO2 containing the Nervy homology domains (NHR) 1-4. The NHR1 domain is the

region that interacts with E-proteins, such as E2A which belongs to the SCL/TAL1 complex and represses target gene activity (*265*). The NHR2 domain contains a hydrophobic heptad repeat which is crucial for oligomerization and allows ETO2 proteins to form homodimers with other ETO2 proteins or heterodimers with other ETO family proteins (*267*). In fact, it has been proposed that quaternary structures of ETO2 organized by the NHR2 domain confer the repression function to ETO2, providing, for instance, stronger binding to E-proteins via NHR1 and to the nuclear corepressor NCoR/SMRT via NHR3/NHR4 (*268*). The NHR3 domain seems to mediate the interaction between ETO proteins binding to CSL (CBF1, Suppressor of Hairless, Lag-1), a nuclear receptor that becomes an activator upon induction of upstream Notch signals (*269*). Lastly, the NHR4 domain contains two noncanonical zinc finger motifs which regulate the interaction with the nuclear corepressor NCoR/SMRT complex (*268*).

To dissect the role of the different domains for the cellular activity of NFIA-ETO2, I cloned a series of deletion mutants of NFIA-ETO2 carrying an N-terminal 3xFLAG tag. I used HeLa cells to check for the protein expression of the NFIA-ETO2 deletion mutants. HeLa cells are adherent human cervix carcinoma-derived cells that can be efficiently transfected by expression plasmids containing a CMV promoter (cytomegalovirus), which has been shown to direct a high level of transiently expressed proteins (*333, 334*). I checked protein expression by Western blot using an anti-FLAG antibody (**Fig.24** Chapter 3). Surprisingly, even though the samples were denatured by boiling in sample buffer, two bands were detected in the WT NFIA-ETO2 (≈93KDa) expressing cells as well as ΔDBD (≈88KDa), ΔMH1 (≈81KDa), ΔNHR2 (≈85KDa) and ΔNHR4 (≈89KDa), but not in ΔNHR1 (≈83KDa). The nature of this upper higher molecular weight band (≈120-135KDa) remains for now unclear. I analyzed whether there is a second ORF that would allow expression of an NFIA-ETO2 fusion lacking the triple FLAG tag. However, I did not find any alternative protein that could be the one expressed.

Alternatively, one may also consider that the upper molecular weight band might be a result of post-translational modifications (such as SUMOylation or glycosylation). However, we have not found any hints from published ETO data for such modifications and normally, the differences in molecular weight are not as prominent as seen in our immunoblots. In fact, it has been suggested that fusions carrying ETO, such as RUNX1-RUNX1T1, are not SUMOylated since co-localization assays revealed that RUNX1-RUNX1T1 could not be co-localized with PML (Promyelocytic Leukemia) or SUMO-1 (Small Ubiquitin Like Modifier 1) (*335*).

The relatively high molecular weight, appearing as almost double the size of the expected size suggests indeed formation of a dimer, however its resistance to disruption in boiling sample buffer seems rather unusual. Nevertheless, previous studies have described protein complexes that remain intact even after boiling of the nuclear extracts in Lämmli sample buffer

for >20 minutes (*336, 337*). Surprisingly, previous studies showing immunoblots of RUNX1-RUNX1T1 or ETO2-GLIS2 never showed such high molecular weight bands. However, most studies show single bands cut out from their immunoblots (*277, 338, 339*).

Previous studies showed that the NHR1 domains are necessary for the stabilization of high molecular weight complex formation (*339*), such as binding to E-proteins like E2A/HEB and establishing pentameric complexes with SCL/TAL (*104, 340, 341*). Therefore, deletion of the entire NHR1 domain may disrupt formation of a highly stable complex resulting in the absence of the second band. Indeed, previous immunoprecipitation analysis showed that the lack of NHR1 also destabilizes the formation of high molecular weight complexes (*338*). They found that deletion of NHR1 domain but not NHR2 abrogated the binding of the RUNX1-RUNX1T1 fusion to p300, a transcriptional coactivator that acetylates RUNX1-RUNX1T1 necessary to mediate self-renewal and promote leukemogenicity in mouse models. Western blot assays showed a reduction in the molecular weight of complex formed by RUNX1-RUNX1T1 lacking the NHR1 domain in human Kasumi-1 leukemic cells. However, deletion of exon 9 of RUNX1-RUNX1T1, containing both NHR3 and NHR4, also abrogated the binding to p300 and showed similar effects (*338*).

I found that MEL cells expressing the NFIA-ETO2 deletion mutants (ΔDBD, ΔMH1, ΔNHR1, ΔNHR2 or ΔNHR4) underwent terminal differentiation after 4 days in 2% DMSO induction (**Fig.24**). Upon expression in primary erythroblasts, I found that cells expressing the NFIA-ETO2 ΔDBD, ΔNHR2 and ΔNHR4 mutants still also underwent terminal erythroid differentiation (**Supplementary Fig.7**). Altogether, these experiments suggested that erythroid differentiation blockage properties of NFIA-ETO2 needs the DNA-binding domains of NFIA, as well as one of the NHR domains. However, we cannot rule out that ablation of any of these domains will result in inappropriate folding of the protein and therefore indirectly modulate its activity. In future experiments we will therefore also establish point mutants that inactivate the NFIA-DBD and the NHR2 and NHR4 of ETO2. Similar observations were made by others in the study of the ETO2-GLIS2 fusion (*277*). A point mutation in the DNA-binding domain of GLIS2 or the lack of the NHR2 domain of ETO2 indicated that both are required to promote self-renewal and to block megakaryocyte differentiation by this fusion. Also, mutations in the DBD of RUNX1-RUNX1T1 quickly led to exhaustion of the proliferation ability and failure to increase colony formation capacity. However, only the deletion of the NHR2 domain of ETO affected the oligomerization and *in vitro* transformation capacity of RUNX1-RUNX1T1 (*328, 342*). Earlier structural studies showed that deletion of the NHR2 domain reduced the capacity of RUNX1-RUNX1T1 to block granulocyte differentiation, and that deletion of both NHR2 and NHR4 domains completely eliminated the activity of the fusion protein and inhibited cell proliferation, suggesting a cooperative activity of NHR2 and NHR4 domains (*343*). Although

this last study suggested that single deletion of NHR1 and NHR4 domains did not affect the transformation capacity of RUNX1-RUNX1T1 as it did in NFIA-ETO2, the function of ETO2 domains in erythropoiesis could differ from the ones of ETO domains in granulopoiesis. Complex formation of E2A and HEB witch SCL/TAL1 for instance, is only known to be mediated by ETO2, but not ETO (*125*). Furthermore, early studies proposed a model in which ETO can recruit at least two different corepressors complexes: NCoR and mSin3a (*344, 345*). This finding further suggests that ETO2, which is not known to bind mSin3a, may be a component of corepressor complexes that are different from ETO-containing complexes, hence, deletion of their domains could result into different phenotypes.

These data suggest that NHR2 has a critical oligomerization function for the transforming potential of the ETO protein fusions. Thus, NHR2 could be a potential target to abrogate the leukemogenesis potential of these fusions. Wichmann and colleagues showed that overexpression of small NHR2-derived peptides impeded oligomerization in a competitive manner and reduced transformation by RUNX1-RUNX1T1 (*278*). More recently, expression of a small 128aa NHR2 peptide (NC128) was shown to impair transformation by the ETO2-GLIS2 fusion (*277*). We therefore also explored the effect of the competing NC128 peptide in NFIA-ETO2 expressing cells. Expression of NC128 resulted in erythroid differentiation of NFIA-ETO2 expressing erythroblasts. Interestingly, we observed similar effects in both *WT* and *TP53*$^{R248Q/+}$ NFIA-ETO2-expressing erythroblasts (**Fig.22**). This observation suggests that small molecules targeting the NHR2 domain could represent a therapeutic opportunity for human cancers carrying ETO fusions, including erythroleukemia expressing NFIA-ETO2. Although smaller TAT-NHR2 fusion peptides seemed to be able to induce differentiation of RUNX1-RUNX1T1 in AML cells, based on the structure, potential targeting of the NHR2/NHR2 interface with small molecules seems challenging (*346, 347*).

Collectively, these findings let us propose a model in which NFIA-ETO2 blocks erythroid differentiation by binding the DNA through NFIA, most probably binding to NFIA erythroid targets and other erythroid genes, and by recruiting corepressor complexes and forming homo- and heterodimers, thereby modifying transcription of genes that are essential for erythroid differentiation (**Fig.30**).

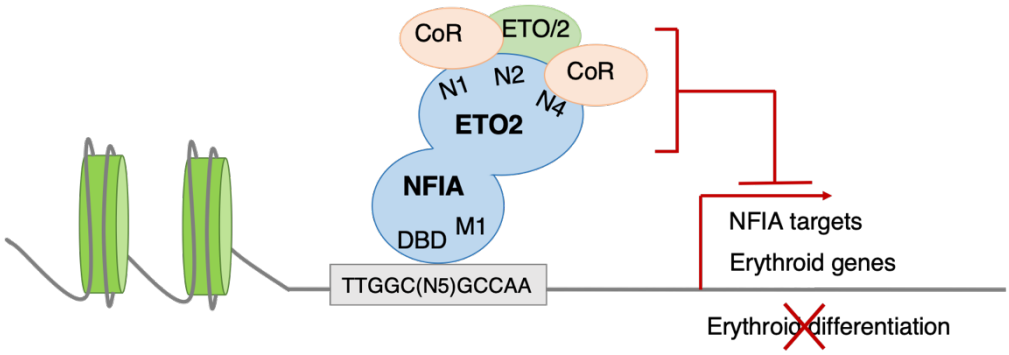


Figure 30. Hypothetical mechanism of NFIA-ETO2 to block erythroid differentiation. Schematic representation of NFIA-ETO2 (blue) and its domains (DBD: DNA-binding domain, M1: MH1, N1-4: NHR1-4). NFIA binds to DNA through the double-stranded palindromic DNA recognition sequence (*256*) and ETO2 recruits corepressors (CoR) complexes and forms homodimers (ETO2) or heterodimers with other members of the ETO family proteins, repressing the transcription of NFIA targets and erythroid genes leading to blockage in erythroid differentiation.

NFIA-ETO2 expression increased the clonogenic activity of *TP53^{R248Q}* erythroblasts

Recent studies have shown that almost all PEL patients carry mostly missense mutations in the DNA-binding domain of TP53 (*237, 238, 241, 242*). Unfortunately, due to patients' data protection, we did not obtain any information about the mutational landscape of the pediatric PEL case in which NFIA-ETO2 was initially identified (*249, 250*). However, since TP53 mutations are the most prevalent genetic lesions in PEL, we aimed to investigate any functional cooperation between NFIA-ETO2 and one of the most predominant TP53 mutation, TP53^{R248Q} *(241)*. We took advantage of a human TP53 knock-in (HUPKI) mouse model carrying the DNA binding domain R248Q mutation, analysed it in the presence or absence of the mouse WT *Tp53* allele (*303*) and compared it to the *Tp53$^{+/-}$* and *Tp53$^{-/-}$* genotypes. Interestingly, I observed that *in vitro* induced erythroid differentiation of *TP53$^{R248Q/+}$*, *TP53$^{R248Q/-}$*, *Tp53$^{+/-}$* and *Tp53$^{-/-}$* BM-derived erythroblasts developed normally and did not differ between the genotypes (**Supplementary Fig.3**). Also, transduction of NFIA-ETO2 into the four different erythroblast TP53-mutants also caused a similar impairment in erythroid differentiation as seen before with the WT *Tp53$^{+/+}$* cells. However, *Tp53$^{-/-}$*, *TP53$^{R248Q/+}$ and TP53$^{R248Q/-}$* cells seemed to proliferate faster in erythroid differentiation-inducing medium. These findings suggest that the NFIA-ETO2 differentiation blockage properties *in vitro* are not affected by the *Tp53* status, but that the proliferation rate increases upon TP53 mutation (**Fig.3**). Accelerated cell proliferation caused by mutated TP53 has been also shown earlier in other cancer types (*348-350*).

In MC cultures, *TP53^{R248Q}* erythroblasts expressing NFIA-ETO2 with or without a wildtype allele, were able to form dense and round colonies that could be serially propagated for at least six rounds, composed by cells that maintained an erythroid phenotype (c-Kit^{+}/CD71^{+}) (**Fig.4**). Importantly, we did not observe any serial plating of erythroblasts carrying only *TP53^{R248Q}* or those that lost both *Tp53* alleles without expression of NFIA-ETO2.

A recent study found that *TP53*$^{R248Q/+}$ leukemic cells are more impaired in TP53-regulated G_1 cell cycle arrest and apoptosis than *Tp53*$^{+/-}$ cells, suggesting that the TP53^{R248Q} mutant acts in a dominant-negative (DNE) manner on the functionality of the wildtype protein (*351*). These results support the idea that, also in our model the TP53^{R248Q} mutant impairs the activity of the WT protein, which is underlined by the finding that NFIA-ETO2 expression increased the self-renewal capacity of *TP53*$^{R248Q/+}$, *TP53*$^{R248Q/-}$ and *Tp53*$^{-/-}$ erythroblasts in a similar manner, but not *Tp53*$^{+/-}$ cells (**Fig.4** Chapter 2). Notably, supporting the idea of DNE activity of the *TP53*^{R248Q} allele, Boettcher and colleagues showed that *TP53*$^{R248Q/+}$, *TP53*$^{R248Q/-}$ and *Tp53*$^{-/-}$ cells were functionally equivalent, as they showed equal degree of resistance to nutlin-3a, an inhibitor of MDM2-mediated proteasomal degradation of TP53. This indicates that the missense variant inhibits the transcriptional activity of the WT TP53 protein. Furthermore, both *TP53*$^{R248Q/+}$, *TP53*$^{R248Q/-}$ rapidly outcompeted *Tp53*$^{+/+}$ cells in coculture assays in the presence of the inhibitor (*351*).

Studies comparing tumor formation in mice found earlier onset and shorter survival of *TP53*$^{R248Q/-}$ compared to *Tp53*$^{-/-}$ mice, indicating a "gain-of-function" activity (*304, 352*). However, similar to Boettcher *et al.,* we observed that the TP53^{R248Q} missense mutation and null alleles have equal oncogenic properties *in vitro*.

Cooperation of NFIA-ETO2 and *TP53*$^{R248Q/+}$ for disease induction

The largest study investigating the genetic landscape in an erythroleukemia cohort found some potentially leukemogenic fusion genes including NUP98-KDMA5A, MYB-GATA1 and NPM1-MLF1 (*242*). Functional validation studies revealed that only the NUP98-KDM5A fusion previously found in other AML subtypes was sufficient to promote self-renewal *in vitro* and to develop a serially transplantable myeloid but not erythroid leukemia in mice. The myeloid rather than erythroid phenotype could speak for the need of an additional mutation in target cells of the erythroid lineage, or that a different cell of origin is responsible for the erythroleukemia phenotype, or even that the erythroleukemia patients initially studied suffered from AML with some erythroid predominant clones. Proper diagnosis of erythroleukemia often appears to be difficult, reflected by constant changing of classification and diagnostic parameters. In particular, the presence of both aberrant erythroid and myeloid cells can indeed cause difficulties to distinguish MDS from AML with some minor erythroid clones and from erythroleukemia (*134, 246, 353, 354*). The other fusion genes found in the study by Iacobucci and colleagues were unable to develop disease upon transplantation of retrovirally transduced BM cells. The fact that they all occurred together with other genetic alterations in patients also suggests that additional genetic hits are necessary to induce the erythroid phenotype (*242*). Studies in other AML forms actually indicate that in the majority of the cases, the initiating

lesion, like a particular fusion oncogene, is not sufficient to phenocopy the disease when expressed in the hematopoietic system in mice. Several transgenic mice have shown that, if ever, they developed leukemias after long latency, suggesting that AML-associated fusion proteins might induce a preleukemic state, but other genetic lesions are necessary for leukemia development (*330, 355*).

These facts led me to explore whether NFIA-ETO2 would cooperate with the *TP53*^{R248Q} mutation *in vivo.* I used the classical transplantation assay with NFIA-ETO2 expressing E14.5 FL-derived HSPCs or erythroblasts We decided to work with E14-5 FL-derived HSPCs or erythroblasts for three reasons: firstly, the NFIA-ETO2 fusion gene was found in very young children, and most likely originates during fetal hematopoiesis. Secondly, FL HSCs are known to have a high cycling activity, not found in adult HSCs (*356*), which may result in increased retroviral transduction. Lastly, FL HSCs transplanted into irradiated mice are characterized by a higher level of regeneration indicating increased self-renewal capacity compared to adult BM (*357*). Since the majority of TP53 mutations in clonal hematopoiesis and AML are mono-allelic missense mutations (*358-361*), I utilized heterozygous *TP53*$^{R248Q/+}$ mutant mice for transplantation assays.

Transplantation of FL-derived *TP53*$^{R248Q/+}$ HSPCs expressing NFIA-ETO2 induced a transplantable leukemia phenotype in 6 out of 7 mice after a latency of 20 to 60 weeks (**Fig.6**). However, there were some interindividual differences in the disease phenotype. Some, but not all, of the mice had characteristics of PEL with extensive BM infiltration by erythroblasts, splenomegaly and erythroblasts in the peripheral blood. The other mice showed a decrease of platelets numbers in the peripheral blood, but normal white blood cell counts and did not present splenomegaly. However, all symptomatic mice had similar elevated numbers of myeloid cells compared to controls. These findings suggest that these mice did not suffer from PEL, but most likely from AML with erythroid predominance or MDS (beforehand classified as AML-M6a.) These differences in disease phenotype might be due to the heterogeneity of the transduced HSPC population, as the composition of the transduced cell pool transplanted into individual mice was most likely highly variable. It has been shown that activation of the same oncogenic pathway in different cellular compartments strongly influences malignant potential. For instance, transplantation of BM cells expressing AML-associated fusion genes such as MOZ-TIF2, MLL-AF9 or MLL-ENL, expressed either in HSCs or GMPs, also initiate AML in recipient mice, however, disease latency, aggressiveness and treatment resistance was different depending on the cell of origin (*139-141*).

I.F transplantation of FL-derived *TP53*$^{R248Q/+}$ erythroblasts expressing NFIA-ETO2 into irradiated mice induced a more severe disease after a latency of 12 to 30 weeks with a PEL phenotype in all recipients (**Fig.5**). I decided to transplant transduced erythroblasts

intrafemorally as erythroblasts were reported to have lower engraftment capacity than HSCs (*362*). Symptomatic mice presented with fully infiltrated BM with c-Kit, CD71 and Ter119 positive cells and almost no remaining myeloid cells as well as infiltration in spleen and liver with the presence of erythroblasts in the peripheral blood. The observed differences in disease latency and aggressiveness between mice transplanted with NFIA-ETO2-expressing erythroblasts or HSPCs could again be due to different cells of origin. Erythroblasts are highly committed to erythropoiesis, hence NFIA-ETO2 expression in these cells blocks their natural differentiation and, together with the *TP53^{R248Q}* mutation, induces a PEL-like phenotype. On the other hand, the HSPC population is very heterogenous and transduction of NFIA-ETO2 might have affected cells of different lineages, resulting in differences in disease severity and lineage contribution. Our observation further suggests that in pediatric PEL carrying NFIA-ETO2, the transformation process may occur in committed erythroid progenitor cells. However, to ultimately answer the question, of which mutation is acquired first in which stage of hematopoietic differentiation, further experiments are required.

Recent studies proposed that in tumors carrying missense TP53 mutations, the loss of the remaining WT Tp53, also called loss of heterozygosity (LOH), is a prerequisite for stabilization and gain-of-function activity of the mutant TP53 *in vivo* (*304, 305*). LOH is recurrent in human cancers with missense TP53 mutations and tumors in mice carrying heterozygous TP53 mutations also often undergo LOH (*304*). The same group reported earlier that not all tumors in a *TP53$^{R248Q/+}$;*Neu transgenic breast cancer model, overexpressing the ErbB2 gene, underwent LOH suggesting that the *Neu* oncogene might eliminate the pressure for LOH and mutated TP53 might have a DNE activity over the WT allele. We observed that diseased mice transplanted with NFIA-ETO2-expressing *TP53^{R248Q}* cells did not undergo LOH of the wildtype allele, suggesting that, similar to the *TP53$^{R248Q/+}$;*Neu model, the R248Q allele in our NFIA-ETO2 *TP53$^{R248Q/+}$* erythroleukemia has DNE effects and is sufficient to repress the activity of the wildtype protein (**Supplementary Fig.5**).

Previously described erythroleukemia models were induced by two functionally cooperating mutations, such as MEL cells with deregulation of PU.1 expression due to integration of the Friend's virus and Tp53 mutation (*211*), expression of a constitutively active *JAK2^{V617F}* mutation in *Tp53$^{-/-}$* cells (*214, 215*), or expression of constitutively active NTRKA in cells carrying a *Tp53^{R173H}* mutation (*242*). These observations strongly suggest that erythroleukemia is driven by at least two functionally cooperating mutations.

Is NFIA-ETO2 a simple erythroid transcriptional repressor?

Based on NFIA-ETO2 structure, we could hypothesize that this fusion protein functions as a chimeric transcription factor, in which NFIA provides contact to erythroid gene loci (*255, 260*) and the fused ETO2 nervy-homology (NHR) domains mediate oligomerization and recruitment of transcriptional co-regulators (*127*). RNA-seq analysis identified a transcriptional signature of erythroid regulatory genes that were down-regulated in erythroblasts expressing NFIA-ETO2 and enriched in cells expressing the inactive NFIA-ETO2-ΔNHR4 mutant. Interestingly, NFIA-ETO2 expression also negatively correlated with previously proposed erythroid NFIA target genes supporting the hypothesis that the NFIA part targets ETO2 and other co-repressors to suppress their transcription (**Fig.7&30**). In addition, target genes of the erythroid master transcriptional regulator GATA1 were negatively enriched in NFIA-ETO2 expressing erythroblasts. These initial findings would go along with the idea that the ETO2 part of the fusion (known to primarily act as transcriptional co-repressor) might impair transcription of genes necessary for erythroid terminal differentiation (*118, 122, 363, 364*).

Unexpectedly, expression of NFIA-ETO2 compared to NFIA-ETO2-ΔNHR4 was associated with upregulation of genes previously found to be over-expressed in lineage marker-depleted BM HSPCs from mice lacking *Eto2* (*120*). This finding suggests that aberrant overexpression of NFIA-ETO2 may also impair ETO2 repression activity in genes that regulate HSPCs. However, we may have to take-into-account that ETO2 could also have completely different functions and mechanisms for maintenance of HSCs than during erythroid maturation (*127*). In contrast to our model, Iacobucci *et al.* showed that gene expression signatures of lineage marker-depleted BM HSPCs of mice that developed erythroleukemia after transplantation of *TP53* mutant cells retrovirally expressing a constitutively active NTRKA receptor tyrosine kinase (which is mutated in rare cases of erythroleukemia) was enriched for genes down-regulated in *Eto2*$^{-/-}$ lineage marker-depleted BM HSPCs, indicating aberrantly high ETO2 activity. They hypothesized that this finding most likely reflected the differences in the relative frequency of HSPCs within the heterogeneous populations that were compared (diseased vs. not diseased BM), rather than being directly linked to molecular mechanisms related to transformation by NTRKA and mutated *TP53* (*242*). These data support the idea that our correlations may more be based on the differentiation state of the cell populations that were compared rather than the fact that NFIA-ETO2 expression downregulates ETO2 activity.

Intriguingly, NFIA-ETO2 activity during induced erythroid differentiation was associated with high expression levels of *Notch2* and *Psen2*, genes related to the NOTCH signaling pathway, shown to be deregulated in some human cancers (*365*). *Notch2* was shown to be highly expressed during early erythropoiesis, reaching the highest point at the proerythroblast stage, but starts decreasing at the basophilic erythroblast stage (*128*). These results potentially

indicate that NFIA-ETO2 expression could maintain NOTCH2 at high levels for prolonged periods and thereby prevents cells from undergoing terminal erythroid differentiation. Moreover, the fact that we found NOTCH2 significantly upregulated, whilst its canonical activator in the hematopoietic system, the Kit ligand or stem cell factor (SCF), was drastically downregulated in NFIA-ETO2 expressing erythroblasts upon differentiation, implies that NOTCH2 levels were kept high by an alternative pathway. Interestingly, it has been shown that ETO2, together with the NCoR/SMRT complex, binds to the CSL (CBF1/Su[H/Lag-1) complex and represses Notch1 transcription (*269*).

We also found that upon induction of erythroid differentiation, NFIA-ETO2 expression was associated with the expression of several proto-oncogenes such as *c-Myb*, which encodes for a leukemogenic transcription factor known to maintain myeloid progenitor cells in an immature stage (*316*) (**Fig.7**). Caspase 3, another proto-oncogene highly up-regulated, is known to promote genetic instability and carcinogenesis, as well as cleavage of GATA1 leading to blockage in terminal maturation (*114, 115*). Interestingly, one of the top positively differentially expressed genes was encoding for TRIM35 (tripartitie motif containing 35, aka hematopoietic lineage switch protein 5, HLS5) previously shown to suppress GATA1-mediated transcription (*317*). In addition, other protooncogenes like S*tat5a*, *c-Kit*, or *Bcl2* were found over-expressed in NFIA-ETO2 expressing cells blocked in erythroid maturation. These proto-oncogenes are well-known for their roles in several hematological malignancies (*366-369*). These findings suggest that NFIA-ETO2 not only represses transcription of genes related to erythropoiesis, it might also repress other genes involved in repressing oncoproteins. This fact would explain the overexpression of some of the beforehand mentioned proto-oncogenes. Surprisingly, however, according to the TRANSFAC database, used to search for potential transcription factor binding sites, NFIA is known to directly bind to Stat5a. Considering the fact, that NFIA-ETO2 harbors the DNA binding domain of NFIA, it is probable, that the fusion directly binds STAT5a as well. This is conflictive with the hypothesis that repressed repressors lead to the upregulation of Stat5a and other direct NFIA targets. A potential mechanism explaining these findings might be, that NFIA-ETO2 might bind the DNA binding sites of repressors of NFIA targets and thereby prevents repression of these genes.

Preliminary investigations via ATAC-seq show little or no difference in accessible chromatin (data not shown) between NFIA-ETO2 and WT erythroblasts, suggesting a competitive regulatory effect of NFIA-ETO2 on already accessible chromatin. In order to better understand the regulatory mechanisms of NFIA-ETO2, we think that chromatin immunoprecipitation sequencing (ChiP-seq) analysis to obtain information about binding of the fusion to chromatin is necessary. However, my first attempts to perform a ChiP-seq so far failed (data not shown). This is most likely due to several factors, such as: excess of cross-

linking, primary cells are more sensitive than cell lines, resulting in reduction of antigen availability in chromatin or loss of material, shearing conditions were not optimal due to the settings of the sonicator apparatus, too harsh DNA extraction buffer, or simply that the epitope is embedded within the protein complex and thus cannot be recognized by the antibody. In order to optimize this assay, I could implement several adjustments to our protocol, such as: use of commercial buffers, increase antibody concentration or use a cross-linking protocol specific for primary cells. However, to avoid the cross-linking and sheering step, which are very critical, I will also try a recently developed novel technique: CUT&RUN. This technique is performed *in situ* and promises quantitative high-resolution chromatin mapping and probing of the local chromatin environment (*370*).

Taken together, upon terminal erythroid maturation, NFIA-ETO2 expression leads to down-regulation of GATA1 and NFIA target genes and erythroid genes necessary for the induction of erythropoiesis. Furthermore, the fusion triggers the up-regulation of proto-oncogenes related to blockage in differentiation, genetic instability and carcinogenesis. The NFIA part of NFIA-ETO2 might bind to NFIA targets or adjacent to GATA1 targets resulting in enriched ETO2 corepressor activity blocking their transcription (**Fig.31B**). However, we cannot exclude that the ETO2 fusion partner might bind directly to corepressor complexes, therefore, further analysis, such as ChiP-seq or CUT&RUN, could give as more insights about the molecular mechanism of NFIA-ETO2.

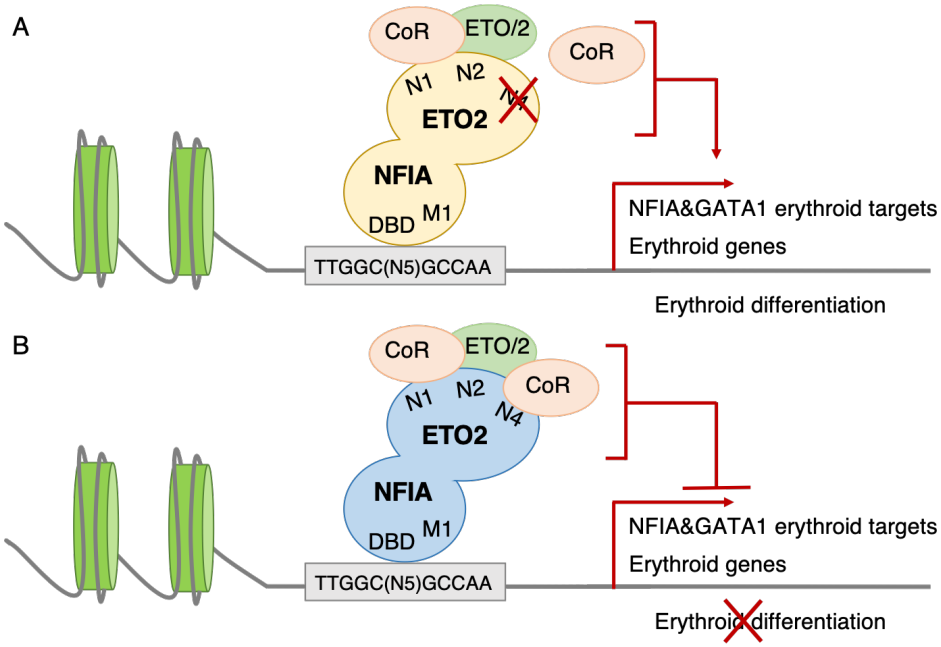


Figure 31. Hypothetical mechanism of NFIA-ETO2 to block erythroid differentiation and activate proto-oncogenes expression. (A) Schematic representation of NFIA-ETO2-ΔNHR4 (yellow) and its domains (DBD: DNA-binding domain, M1: MH1, N1-4: NHR1-4) lacking the NHR4 domain (N4). NFIA binds to DNA through the double-stranded palindromic DNA recognition sequence (Chen *et al.* 2017), however, ETO2 lost the capacity to recruit corepressors (CoR), lacking the capacity to repress the transcription of NFIA and GATA1 targets and erythroid genes, leading to erythroid differentiation. **(B)** Schematic representation of NFIA-ETO2 (blue) and its

domains, binding to DNA through NFIA, while ETO2 forms oligomers and recruits corepressors, leading to repression of NFIA and GATA1 targets, as well as erythroid genes.

TP53^{R248Q} is associated with downregulation of tumor-suppressive TP53 target genes in NFIA-ETO2-expressing erythroblasts

Although we found significant biological differences (increased clonogenic activity *in vitro* and leukemia induction *in vivo*) we observed only small changes in the transcriptional signatures of NFIA-ETO2-expressing erythroblasts with or without the TP53^{R248Q} mutation. The majority of the significantly differentially expressed genes was downregulated and previously characterized as direct TP53 targets. *Eda2r* (a.k.a. *Xedar*) encoding for the ectodermal dysplasia receptor, that belongs to the tumor necrosis factor receptor superfamily (TNFRSF27), was the most significantly downregulated gene. EDA2R and its ligand EDA-A2 have been shown to be controlled by *TP53* signaling, and its inactivation resulted in resistance to *TP53* induced apoptosis (*318*). Additionally, the pleckstrin homology-like domain family A member 3 (*Phlda3*), polo-like kinase 2 (*Plk2*), anoctamin 3 (*Ano3*) and the gene encoding for the zinc finger protein 365 (*Zfp365*) were significantly downregulated as well and also previously known as *TP53* targets (*319-322*). PHLDA3 was shown to inhibit activation of the AKT serine/threonine kinase, a central cellular signaling hub controlling growth. Reduced PHLDA3 levels enhanced AKT activity and decreased *TP53*-dependent apoptosis (*319*). These results support previous studies showing that mutant TP53 interferes with the WT TP53 protein and impairs the normal transcriptional activity of TP53 (*371*) and somehow also endorses a dominant-negative mode of action in hematopoietic cells (*305, 351*). Predictive analysis of transcription factor activity (*324*) revealed correlations between downregulation of TP53 targets and the activity of the Polycomb repressive complex 2 (PRC2), containing the enhancer of Zeste-2 (EZH2) methyltransferase, the Jumonji AT rich interactive domain (JARID2) and the PRC 2 subunit (SUZ12), which have been implicated in stem cell maintenance and cancer development (*372-374*). PRC2 maintains stable repression of differentiation-associated genes by activating the trimethylation of histone 3 at lysine 27 (H3K27me3) (*375*). Recent RNA-seq studies showed that an EZH2 target gene signature was negatively enriched in HSPCs carrying leukemia-associated TP53 mutations (*376, 377*). Compared to WT and *TP53$^{-/-}$,* ectopic expression of mutated TP53 HSPCs presented significantly higher levels of H3K27me3 at the transcription start sites (TSS) of target genes (*323*). ChiP-seq assays revealed that the majority of H3K27me3 peaks in mutated TP53 HSPCs were overlapping with the WT HSPCs, suggesting that mutated TP53 enhances the association of EZH2 with the chromatin, increasing the levels of H3K27me3 in genes involved in HSC self-renewal and differentiation (*323*). Furthermore, it has been shown that knockdown

of *EZH2* noticeably inhibited cell invasion augmented by mutated TP53 *in vivo* (*376*) and brought the aberrantly increased replating potential of *Tp53* mutant mouse BM cells back to normal levels (*323*). These studies suggested that the epigenetic regulator EZH2 could be a novel therapeutic target for hematological malignancies with TP53 mutations. Notably, our chromatin analysis also revealed enrichment in the promoters of the downregulated genes in *TP53*$^{R248Q+/-}$ NFIA-ETO2 erythroblasts, suggesting that the mutant TP53 could bind together with the PCR2 complex to the promoter of these genes and repress their transcription.

I also found a positive enrichment in HSC proliferation genes and metabolic pathways, as well as negative correlation to WT TP53 signaling pathway in *TP53*$^{R248Q/+}$ erythroblasts expressing NFIA-ETO2. These results support other studies which also found that the dominant-negative activity of the majority of TP53 mutants had a remarkable target gene selectivity for the wildtype TP53 protein, resulting in deregulation of several processes involved in cell proliferation, metabolism and escaping the tumor suppressive activities of *TP53* (*305*). Importantly, we also found some negative enrichment of gene expression signatures of AEL patients carrying *TP53* mutations for genes that were downregulated in NFIA-ETO2-expressing *TP53*$^{R248Q/+}$ erythroblasts, suggesting that in these tumor cells, TP53 targets may also be downregulated by PRC2 complex (**Fig.9**). In fact, GSEA analysis of AEL patients with TP53 mutations (*302*) revealed a negative enrichment for PCR2 and SUZ12 target genes supporting this possibility (see Chapter 5).

Taken together, these published and our own data support the idea of cooperation between TP53^{R248Q} and PRC2 most likely by EZH2-mediated increased H3K27me3 levels at gene promoters, repressing genes related to WT TP53 activity (**Fig.32**). I therefore plan to explore the effects of small EZH2 inhibitors (*378*) as well as *EZH2* knockdown on *TP53*$^{R248Q/+}$ erythroblasts expressing NFIA-ETO2 and particularly to study the impact on clonogenic activity *in vitro* and leukemia induction *in vivo*.

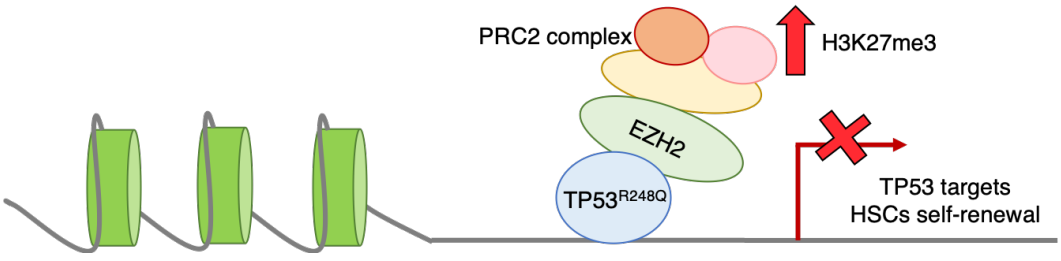


Figure 32. Hypothetical mechanism of TP53^{R248Q} repressed transcription. Mutated TP53 cooperates with EZH2, enhancing its association with the chromatin and increasing the levels of H3K27me3 in genes involved in HSCs self-renewal and TP53 targets, leading to transcription repression.

ZMYND8-RELA induces self-renewal of *TP53^{R248Q}* erythroblasts but seems unable to induce erythroleukemia *in vivo*

Similar to NFIA-ETO2, transduction of ZMYND8-RELA somehow reduced DMSO-induced erythroid maturation of MEL cells. However, even though the cells appeared slightly blocked in erythroid differentiation, hemoglobin mRNA expression levels remained similar to control cells (**Fig. 25** Chapter 3). These results suggest that ZMYND8-RELA may less efficiently block terminal maturation than NFIA-ETO2. In fact, although RELA is known to be involved in cell fate decisions of HSCs and regulates transcription of genes that maintain HSC quiescence (*286*), its function in erythropoiesis is not well established. Early studies revealed that RELA KO in adult mice did not show any effect in erythropoiesis (*288*), however, an older study found that loss of RELA activity together with c-REL was embryonic lethal and mice presented an erythroid differentiation deficiency (*289*). These data suggest that RELA could be involved in erythroid differentiation, however, further studies are needed to better understand the mechanism. On the other hand, ZMYND8 was shown to have oncogenic potential in solid cancers, enhancing progression and breast cancer metastasis in mice (*285*).

I also transduced ZMYND8-RELA into primary mouse erythroblasts (**Fig.26**), which turned out to be very inefficient. Although this could have been related to the relatively large size of the viral expression plasmid resulting in low titer and poor transduction, in few experiments, I managed to grow a larger number of cells to induce erythroid differentiation. The first day in differentiation-inducing medium the cells seemed to be morphologically blocked in differentiation when compared to vector-transduced control cells. Unfortunately, the low number of cells did not allow us to analyze the phenotype of the cells by flow cytometry. However, the cells stopped proliferating and underwent terminal maturation or cell death. The decreased viability upon ZMYND8-RELA expression could be due to ZMYND8, as overexpression has been shown to reduce proliferation and viability of HeLa cells (*379*).

Surprisingly, upon plating ZMYND8-RELA transduced HSPCs, I observed slightly increased colony formation that even allowed replating of the cells for three rounds. The cells mainly formed CFU-M (macrophages) and CFU-GM (granulocytes and monocytes). Similar to NFIA-ETO2 colonies, these results could suggest that the cytokines from the medium induce a shift towards granulocyte-monocyte progenitors due to the fact, that the cells are blocked in erythroid differentiation. Additionally, ZMYND8-RELA-transduced erythroblasts formed very few colonies in the first plating, and could not be replated further, supporting the hypothesis, that ZMYND8-RELA affects more immature cells like HSPCs rather than erythroblasts. This is in accordance with the endogenous function of RELA itself, which is involved in HSC maintenance and fate decisions (*286*).

I also studied the impact of ZMYND8-RELA in erythroblasts with different *Tp53* genotypes (*TP53*$^{+/-}$, *TP53*$^{-/-}$, *TP53*$^{R248Q/-}$ and *TP53*$^{R248Q/+}$). In liquid medium, cells expressing ZMYND8-RELA did not behave differently than WT and rapidly underwent maturation and died (**Fig. 27**). However, rather surprisingly, *TP53*^{R248Q} erythroblasts expressing ZMYND8-RELA formed dense and round colonies in methylcellulose that could be replated up to five rounds (**Fig. 28**). The cells maintained an immature phenotype (c-Kit^{+}/CD71^{+}). However, colonies in the 5th plate were very small and could not be propagated further. *TP53*$^{-/-}$ expressing ZMYND8-RELA also showed an increase in colony formation, cells, however, lost c-Kit^{+} cell marker and could only be replated until the 4th plating. ZMYND8-RELA did not affect colony-formation capacity of heterozygous *TP53*$^{+/-}$ erythroblasts. These data suggest that ZMYND8-RELA expressing erythroblasts acquire self-renewal capacity only upon reduced WT TP53 activity. Additionally, the phenotype is enhanced when the cells carry the missense TP53^{R248Q} mutation. Similar to NFIA-ETO2, ZMYND8-RELA seems also to profit from a dominant-negative activity of the *TP53*^{R248Q} allele. The fact that we could not stably express ZMYND8-RELA in normal cells could be based on the induction of cellular stress leading to activation of TP53, growth arrest and cell death. However, so far, we have not found any reports linking ZMYND8 or RELA to activation of TP53. Lastly, I also investigated the leukemogenesis activity of ZMYND8-RELA *in vivo*. I transplanted *TP53*$^{R248Q/+}$ FL-derived HSPCs expressing ZMYND8-RELA directly after transduction into lethally irradiated mice. However, in contrast to NFIA-ETO2, mice did not develop any symptoms over >1-year and, even though the cells expressed high levels of GFP before transplant, mice presented very low numbers of GFP^{+} cells already 4 weeks post-transplantation (**Fig.29** Chapter 3). These findings suggest that although ZMYND8-RELA increases the clonogenic activity of *TP53*$^{R248Q/+}$ erythroblasts *in vitro*, additional genetic hits might be necessary to induce erythroleukemia. Similar to NFIA-ETO2, we could not get more information about the coexisting cytogenetic lesions in the PEL patients carrying ZMYND8-RELA (*252*). Interestingly, Iacobucci *et al.* also identified the ZMYND8-RELA fusion in a single pediatric case, suggesting that it is also recurrent in pediatric AEL. However, they also mention that the patient had a very poor cytogenetic risk suggesting the presences of other mutations (*242*).

Taken together, I functionally characterized the potential oncogenic fusions NFIA-ETO2 and ZMYND8-RELA in the context of erythroleukemia. I found that NFIA-ETO2 expression is blocking erythroid differentiation in erythroblasts and that the fusion cooperates with TP53^{R248Q} mutation to induce fully penetrant erythroleukemia *in vivo*. Mechanistically, RNA-seq analysis profoundly supports the hypothesis, that NFIA-ETO2 binds to endogenous NFIA targets related to erythroid development and causes their repression, thereby interfering with erythroid

differentiation. Furthermore, RNA-seq and ATAC-seq analysis propose that $TP53^{R248Q}$ down-regulates WT TP53 targets, most likely in cooperation with PRC2 complex.

ZMYND8-RELA on the other hand was shown to only partially block differentiation of erythroblasts. Furthermore, transplantation of ZMYND8-RELA expressing HSPCs into lethally irradiated mice, even in cooperation with TP53 mutations, did not result in any leukemic phenotype.

All in all, the herein presented data does not only confirm NFIA-ETO2 and, to some part, ZMYND8-RELA as drivers of PEL, but also provides important novel insights into the mode of action of fusion proteins and directly adds to the general understanding of the mechanisms leading to erythroleukemia.

Chapter 5 :Human erythroleukemia genetics and transcriptomes

5.1. Introduction

Erythroleukemia (AEL or AML-M6) is a rare disease accounting for less than 5% of all AML cases. Generally, AEL is associated with a poor outcome and the molecular mechanism of the disease is poorly understood (*201, 246, 380*). Earlier reports described that leukemic cells from AEL patients often carry complex karyotypes. Targeted sequencing studies revealed mutations in several genes known to be altered in other AML forms, however, neither were any erythroleukemia-specific mutations found, nor were they functionally validated (*237, 239, 240, 298*). Recent studies indicated that the by far most frequently mutated gene in patients with pure erythroid leukemia (PEL) is TP53, detected in 11 (92%) out of 12 patients (*241*).

To better understand the molecular mechanisms of erythroleukemia, we collected primary tumor cell samples from AEL patients worldwide and exome- and RNA-sequencing was performed by our collaborator Thomas Mercher (Paris). Three genetic subgroups were identified: patients carrying TP53 mutations, patients with various combinations of mutations already found before in AML and MDS such as DNMT3A, TET2 or IDH2, and those patients with none of these recurrent alterations. I was involved in analyzing the transcriptomes of the erythroleukemia patients from our cohort. Comparisons with previously reported gene expression signatures allowed me to divide our AEL patients into 4 different groups based on the degree of erythroid differentiation of the tumor cells. Additionally, we found that leukemic cells from several patients expressed aberrantly high levels of transcription factors and co-regulators related to the GATA1 master transcription factor complex, such as SKI, ERG and ETO2. Finally, *in vivo* experiments (performed by Alexandre Fagan in the Mercher lab) revealed, that depending on the cellular hierarchy and cooperating mutations, overexpression of some of these genes in HSPCs induced lethal erythroid or mixed erythroid/myeloid disease in mice, phenocopying several hallmarks of the human disease. The related paper is provided in the section Appendices.

5.2. Correlation of AEL transcriptomes with erythroid differentiation

Overall, we were able to collect samples from 58 AEL patients. According to the 2008 WHO classification (*117*), 33 patients were diagnosed with *de novo* AEL, of which 29 were AML-M6a and 4 AML-M6b (=PEL), 20 patients were diagnosed with AML-M6a secondary to other hematological malignancies, 1 patient was diagnosed with AML-M6b secondary to plexus choroid carcinoma and 4 patients were not diagnosed precisely. RNA was successfully isolated from 29 of these patients and gene expression profiles were compared to investigate erythroid identity. However, some of the primary patient samples did not have enough material and they had to be expanded before RNA extraction. Hereby patient blood or BM cells were xenotransplanted into NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ (NGS) mice before being sequenced. Previous studies have shown that patient-derived xenotransplants in immune deficient mice are a suitable tool to expand primary leukemia cells without significantly affecting their genotypes (*381*). Samples were enriched according to the expression of CD71, CD36, CD33, CD34, CD45 and c-Kit cell surface markers. The processed patient samples were classified in four groups: “immature progenitors” (CD36^{-}, CD34^{+}, c-Kit $^{+}$), “CFU-BFU-like” (CD36^{+}, CD34^{+}, c-Kit$^{+/-}$), “erythroid-progenitors” (CD36^{+}, CD34^{-}, c-Kit^{-}) and unknown.

In a first step of the analysis, I investigated whether the gene expression signatures of the 29 AEL samples were affected by the xenotransplantation. By performing principle component analysis (PCA) I did not observe any significant correlation between the gene expression and the origin of the samples (**Fig.33A**). In addition, I also checked, whether there was a correlation between the gene expression and the immunophenotype assigned by the clinical laboratory. However, again, PCA did not show any significant correlations between the gene expression signatures and the immunophenotypes (**Fig.33B**). Notably, the percentage of immature cells and erythroblasts in the BM did also not significantly affect the gene expression and the samples did not cluster together according to cell populations (**Fig.33C,D**).

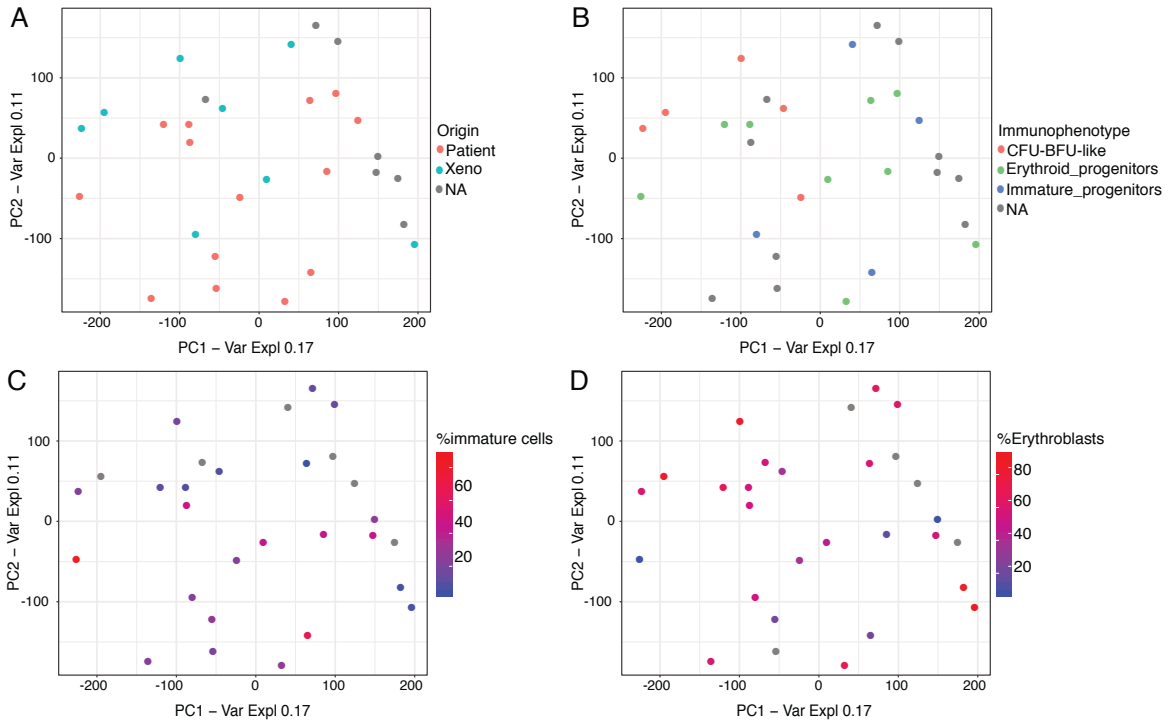


Figure 33. Correlation of AEL gene expression signatures with sample origin, immunophenotype, and number of immature cells and erythroblasts in the sample. PCA of AEL patient sample expression signatures in relation to **(A)** sample origin (directly patient derived, red dots; expanded by xenotransplantation, blue dots; and unknown, grey dots), **(B)** immunophenotype (CFU-BFU-E-like, ($CD34^+$ c-$Kit^{+/-}$ $CD36^+$), red dots; erythroid progenitor like, (CD34- c-Kit- CD36+), green dots; immature progenitors, (CD34+ c-Kit+ CD36-), blue dots; not known, grey dots), **(C)** percentage of immature cells (defined as $CD34^+$, c-$Kit^{+/-}$, $CD36^-$, $CD14^-$, $CD15^-$, $CD235^-$) (ranging from <20%, blue dots to >60% shown in red dots) and **(D)** percentage of erythroblasts (defined as $CD71^+$, $CD36^+$) (ranging from <20%, blue dots to >80% shown in red dots) in the patient's BM at diagnosis.

I next analyzed whether the AEL gene expression signatures may reflect the degree of maturation of the erythroid leukemic cells. Hereby, I compared the transcriptomes from our patients with those obtained experimentally by Merryweather-Clarke *et al.* after *in vitro* differentiation of human peripheral blood mononuclear cells into four defined erythroid progenitor populations: colony forming unit erythroid (CFU-E, "Day 2", $CD71^+/CD235^-$), pro-erythroblast (Pro-E, "Day 4", $CD71^+/CD235^{low}$), intermediate (Int-E, "Day 7", $CD71^+/CD235^{high}$) and late erythroblasts (Late-E, "Day 10", $CD71^{low}/CD235^{high}$) (**Fig.34**) (*382*).

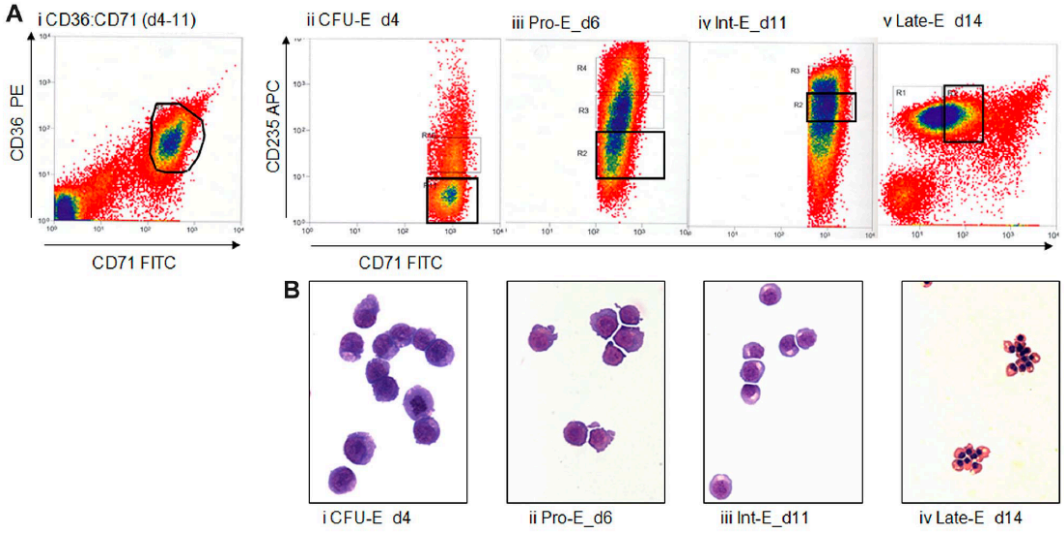


Figure 34. Erythroid maturation stages upon *in vitro* maturation of human peripheral mononuclear cells. Definition of *in vitro* generated human "CFU-E", "Pro-E", "Int-E" and "Late-E" based on flow cytometry **(A)** and morphology **(B)** extracted from Merryweather-Clarke *et al.* (*382*). **(A)** CD36, CD71 and CD235a (corresponding antigen in mice is Ter119) expression determined by flow cytometry of peripheral blood mononuclear cell-derived erythroblasts. CD235a expression increases until the intermediate erythroblast. stage, when it remains high and CD71 expression begins to decrease. **(B)** Cytospin preparations of cells stained with May-Grünwald-Giemsa illustrating the morphologic changes as cultured erythroblast mature of the four sorted populations.

Unsupervised analysis revealed that our AEL samples clustered according to their maturation stages and samples separated into three groups: "Int-E & Late-E", "CFU-E & Pro-E" and "Pro-E" (**Fig.35A,B**). The bioinformatician associated to our group (Frederik Otzen Bagger) analyzed the gene expression signatures of another, recently published, independent larger cohort of AEL samples (*242*), which clustered similarly to our samples (**Fig.35C,D**). Differential gene expression analysis of this larger independent cohort revealed the 50 most significant genes to classify AEL samples according to the four Merryweather-Clarke's groups. Importantly, most of these genes are related to erythroid differentiation such as EPOR, KLF1, but also glycoprotein-A (GYPA) and ankyrin 1(ANK1), both erythroid maturation markers up-regulated mainly in CFU-E & Pro-E-like patients and down-regulated in more differentiated samples grouped as Int-E & Late-E respectively. Collectively our data suggests that AEL-associated transcriptomes reflect the state of erythroid maturation.

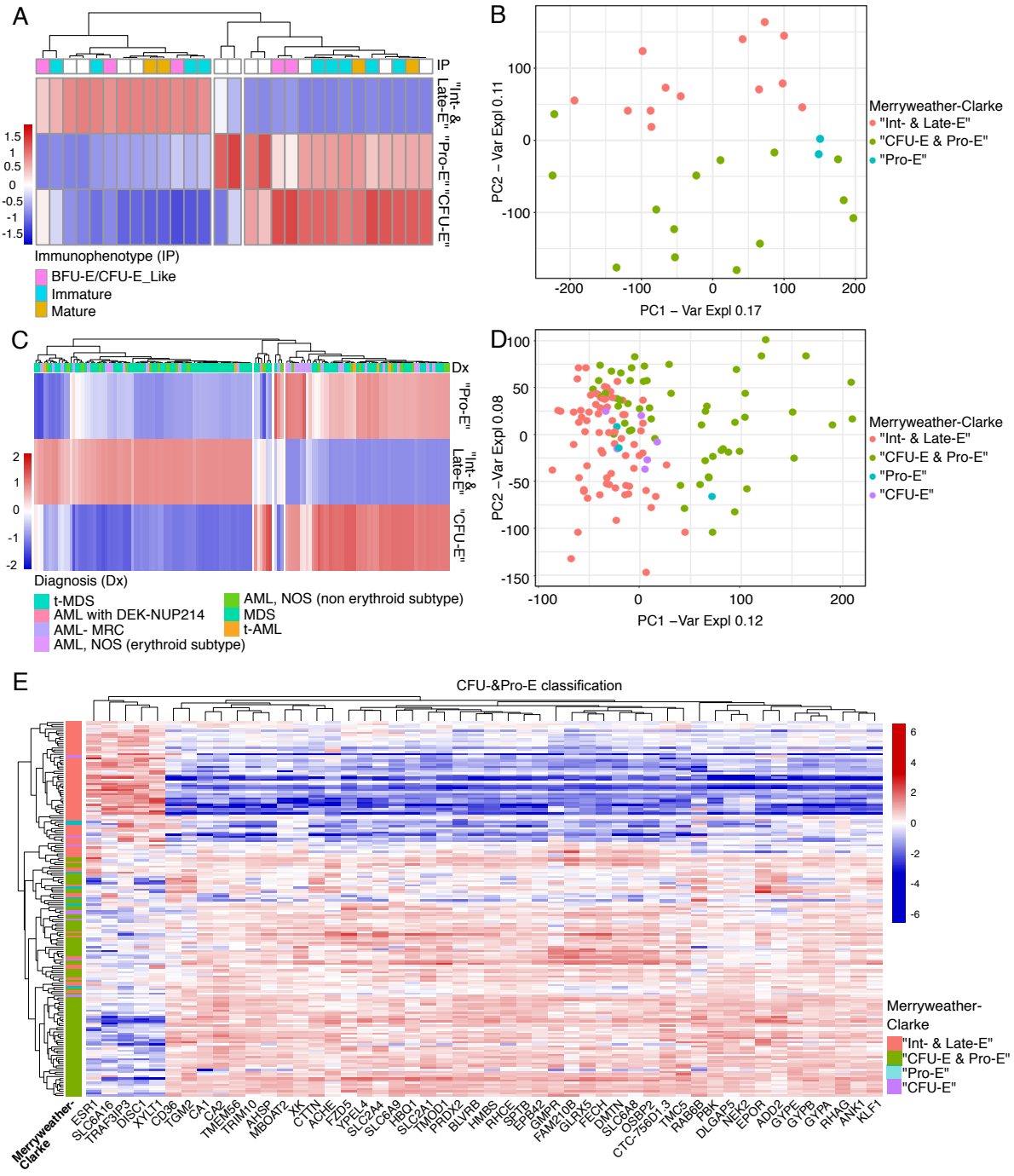


Figure 35. Correlation of AEL gene expression signatures with erythroid differentiation. **(A)** Heatmap showing the correlation between AEL patient samples and the maturation stages. Merryweather-Clarke *et al.* described the expression profile of enriched human colony-formation unit-erythroid (CFU-E), pro-erythroblasts (PRO-E), intermediate (Int-E), and late (Late-E) erythroblast based on surface cell marker expression (*382*). Patient-derived M6 cell expression signatures clustered in an unsupervised manner into four groups according to the four differentiation stages. **(B)** PCA of AEL patient samples, colored according to the Merryweather-Clarke differentiation stages. **(C)** Heatmap showing the correlation between AEL samples from Iacobucci *et al.* (*242*) and the Merryweather-Clarke *et al.* erythroid differentiation stages. Samples clustered in an unsupervised manner into the four erythroid differentiation stages. **(D)** PCA of AEL samples from Iacobucci *et al.* colored according to differentiation stages by Merryweather-Clarke *et al.* **(E)** Heatmap of 50 most significantly differentially expressed genes that characterized each differentiation stage group. Significantly expressed genes were derived from a random forest model. This model predicts which genes are defining the groups obtained comparing our samples with the Merryweather-Clarke dataset.

5.3. Correlation of AEL transcriptomes according to the mutational state

By combining exome- and RNA-seq data, Alexandre Fagan (PhD student in the Mercher lab) identified variants with predicted functional consequences in 62 genes. The genetic data, including high variant allele frequency (**Fig.36A**), supports the classification of AEL patients into three molecular groups (**Fig.36B**). Group 1 presented TP53 mutations (n=12), also associated with higher cytogenetic risk and poorer outcome. Group 2 characterized by mostly TET2 nonsense mutations (n=8) and DNMT3A mutations (n=5), including two patients with both TET2 and DNMT3A mutations. Finally, group 3 (n=10) contained samples without TP53, epigenetic or recurrent known variants. Overall, we found similar mutations already described in other AEL cohorts analyzed by targeted or exome-wide sequencing (*240, 242, 383*).

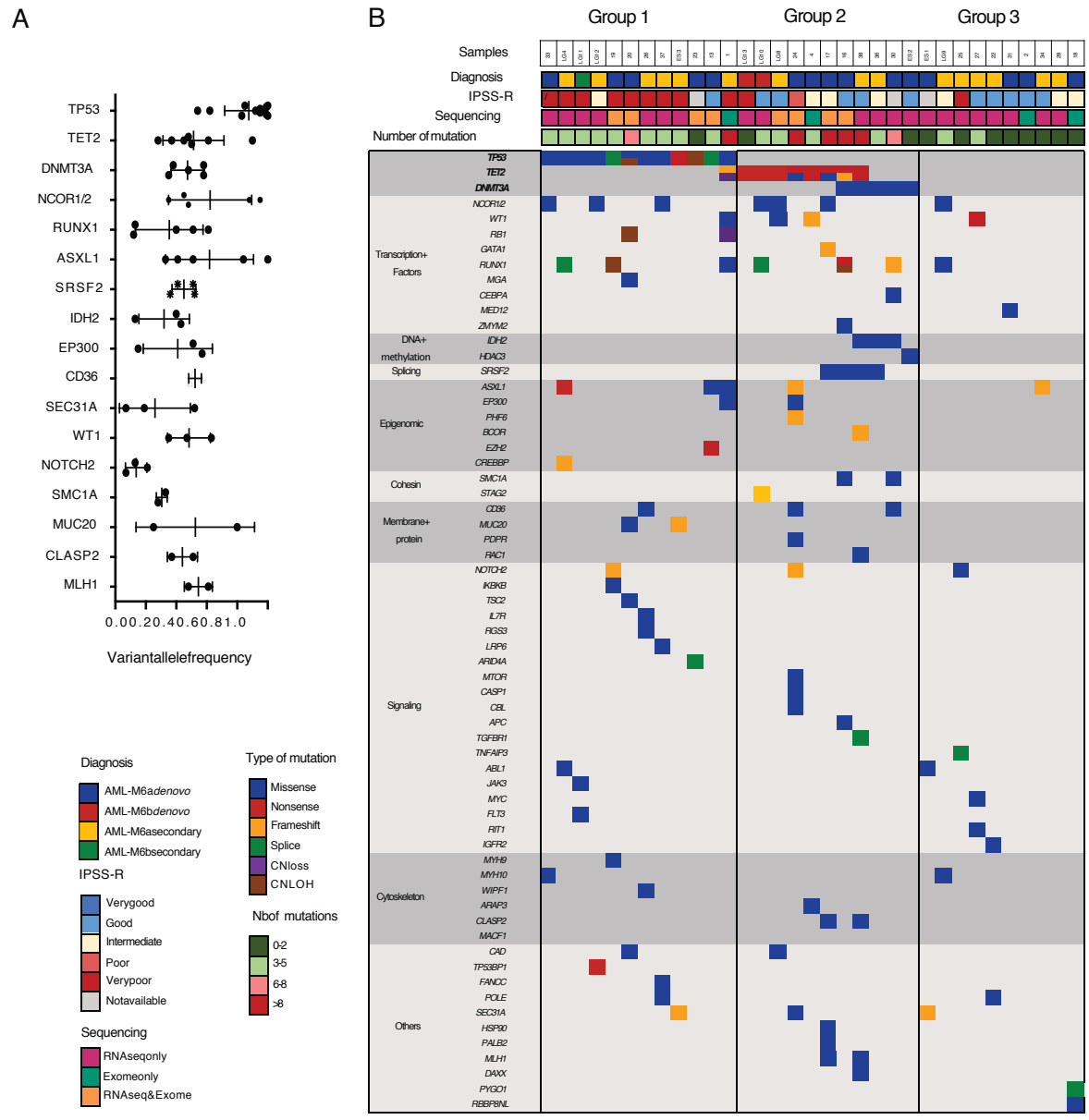


Figure 36. Genomic landscape of our cohort of AEL patients. **(A)** Variant allele frequency of recurrent mutated genes in AEL patients. **(B)** Heatmap representing 3 genomic AEL groups: group 1 characterized by TP53 mutations, group 2 presenting epigenetic modifier-mutations, and group 3 with no recurrent mutations, according to diagnosis,

cytogenetic risk group, methods of sequencing, number of predicted driver mutations and type of predicted driver mutations. (Analysis and plots done by Alexandre Fagnan, Mercher lab, Paris).

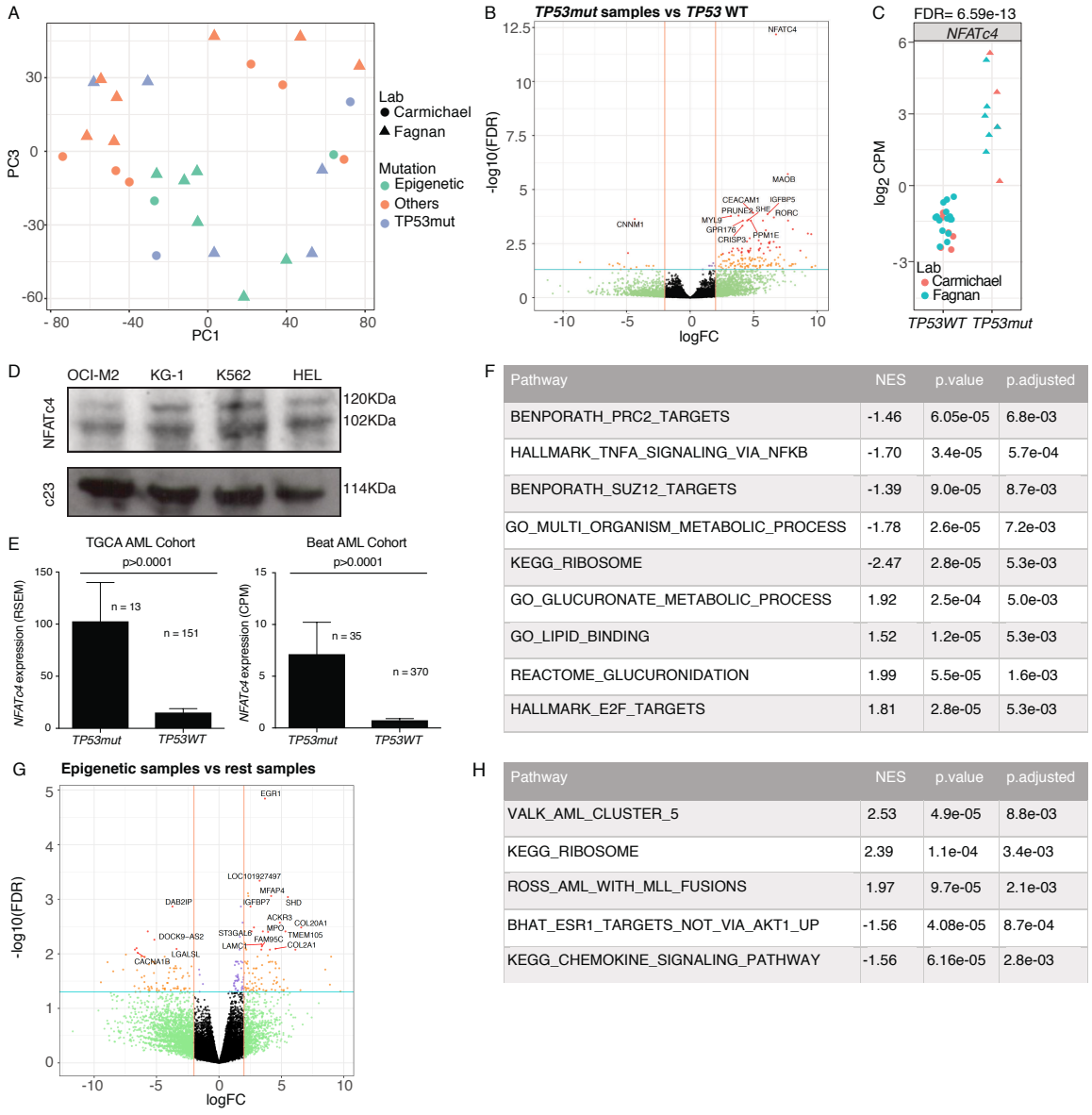
We next asked whether gene expression signatures of the AEL patient-derived samples were affected by the mutational landscape. I performed this analysis with 21 patients and combined them with an AEL expression signature dataset provided by Catherine Carmichael (Melbourne, Australia) from additional 10 patients. Unexpectedly, PCA revealed no significant correlation between the gene expression signature and the three genetic subgroups (**Fig.37A**). These results suggest that the AEL gene expression program seems more affected by the stage of erythroid differentiation than by the presence of particular genetic lesions, suggesting that the erythroid identity in human AEL might depend on the activity of regulators driving erythroid differentiation.

I also analyzed differently expressed genes between the three genetic AEL groups. Comparing TP53 mutated samples (Group 1) with those carrying WT TP53, we found highly significant upregulation of transcripts encoding for the nuclear factor of activated T cells 4 (NFATc4) (FDR=6.59e-13) in TP53 mutated samples (**Fig.37B,C**). We validated this finding in erythroleukemic cell lines known to carry TP53 mutations and interestingly, K562 cells, that carry an inactivating mutation of TP53, expressed higher levels of NFATc4 protein than the other AEL cell lines (OCI-M2, KG-1 and HEL) (**Fig.37D**) (*384*). Furthermore, transcriptomes of AML (n=164) patients from the cancer genome atlas program (TGCA) dataset and transcriptomes of AML (n=405) from the BEAT AML program (www.cbioportal.org), showed highly significant over-expression of NFATc4 ($p<0.0001$) in patients carrying TP53 mutations (TGCA: n=12, BEAT: n=35) compared to those with WT TP53 (TGCA: n=152, BEAT: n=370) (**Fig.37E**). Interestingly, gene set enrichment analysis (GSEA, $p<<0.001$) also revealed that target genes of PRC2 and SUZ12, a member of the PRC2 complex, were negatively enriched in patients carrying TP53 mutations. Furthermore, genes regulated by nuclear factor kappa-B (NF-κB) in response to tumor necrosis factor α (TNFα) were also negatively enriched in patients with TP53 mutations.

I also found that genes related to ribosome pathways were negatively enriched in patients carrying TP53 mutations. Furthermore, also metabolic pathways, such as glucose and lipid metabolism, were dysregulated in TP53 mutated patients (**Fig.37F**). Genes were enriched for glucoronate metabolism and E2F targets, a group of transcription factors required for the regulation of DNA replication and cell cycle progression (*385*).

When comparing the transcriptome of the AEL samples carrying epigenetic mutations (Group 2) with the rest of the samples without epigenetic mutations, I identified early growth response 1 (EGR1) gene significantly up-regulated in epigenetic mutated samples (FDR=6.5e-3) (**Fig.37G**). EGR1 is known to be a transcription factor that regulates proliferation and

localization of hematopoietic stem cells (*386*). GSEA analysis identified genes that are otherwise enriched in AML samples classified as M4 or M5 subtypes, as well as genes enriched in pediatric AML subtypes with chimeric MLL-fusions (**Fig.37H**). In contrast to patients carrying TP53 mutations, genes related to ribosomal pathways were positively enriched in patients carrying epigenetic mutations. Moreover, genes were negatively correlated to chemokine signaling pathway, which regulates biological processes of HSCs, such as cellular activation, differentiation and survival (*387*).



F

Pathway	NES	p.value	p.adjusted
BENPORATH_PRC2_TARGETS	-1.46	6.05e-05	6.8e-03
HALLMARK_TNFA_SIGNALING_VIA_NFKB	-1.70	3.4e-05	5.7e-04
BENPORATH_SUZ12_TARGETS	-1.39	9.0e-05	8.7e-03
GO_MULTI_ORGANISM_METABOLIC_PROCESS	-1.78	2.6e-05	7.2e-03
KEGG_RIBOSOME	-2.47	2.8e-05	5.3e-03
GO_GLUCURONATE_METABOLIC_PROCESS	1.92	2.5e-04	5.0e-03
GO_LIPID_BINDING	1.52	1.2e-05	5.3e-03
REACTOME_GLUCURONIDATION	1.99	5.5e-05	1.6e-03
HALLMARK_E2F_TARGETS	1.81	2.8e-05	5.3e-03

H

Pathway	NES	p.value	p.adjusted
VALK_AML_CLUSTER_5	2.53	4.9e-05	8.8e-03
KEGG_RIBOSOME	2.39	1.1e-04	3.4e-03
ROSS_AML_WITH_MLL_FUSIONS	1.97	9.7e-05	2.1e-03
BHAT_ESR1_TARGETS_NOT_VIA_AKT1_UP	-1.56	4.08e-05	8.7e-04
KEGG_CHEMOKINE_SIGNALING_PATHWAY	-1.56	6.16e-05	2.8e-03

Figure 37. Correlation of AEL expression signatures with genetic mutations. **(A)** PCA of AEL patient samples expression signatures, either from Fagnan *et al.* (triangle) (*302*) or provided by Catherine Carmichael (Melbourne) (dots), with TP53 mutations (blue), mutations in epigenetic regulators (TET2, DNMT3A, IDH1/2, green) and other mutations (Group 3, orange). **(B)** Differentially expressed genes (DEG) between TP53 mutated samples (Group 1) and TP53 WT samples (FDR<0.05, logFC>2.5). **(C)** *NFATc4* mRNA expression (count per million, CPM) in TP53 mutated (triangles) and *TP53* WT (dots) AEL samples, colored according to their source, either from Fagnan *et al.* (green dots/triangles) or Catherine Carmichael (red dots/triangles). **(D)** Western Blot analysis of NFATc4 protein in nuclear lysates of OCI-M2, KG-1, K562 and HEL human AML cell lines. **(E)** *NFATc4* gene expression in AML

patients from the TGCA dataset with or without TP53 mutations normalized by RSEM (RNA-Seq by Expectation Maximization) (n=12, n=152 respectively, t-test, p<0.0001) and in AML patients from the Beat AML Cohort with or without TP53 mutations normalized by CPM (count per million) (n=35, n=152 respectively, t-test, p<0.0001). **(F)** GSEA of gene expression between Group 1 carrying TP53 mutation and rest of samples (padj < 0.05). **(G)** Differentially expressed genes between Group 2 (with mutations in epigenetic regulators) samples and rest of the samples (FDR<0.05, logFC>2.5). **(H)** GSEA of gene expression between Group 2 and rest of samples (padj < 0.05). Values are presented as individual points, bar graphs represent the mean value of samples, error bars as standard error of the mean. Statistical significances in E was tested with Mann Whitney t- test.

5.4. Potential cooperation of EPO overexpression and *TP53*$^{R248Q/+}$

Among our AEL cohort, we identified a patient with cells carrying a TP53 mutation and an *out of frame* fusion between the promoter/enhancer of YWHAE (Tyrosine 3-monooxygenase/tryptophan 5- monooxygenase activation protein epsilon) and EPO (**Fig.38A,B**), resulting in an abnormal expression of *EPO* mRNA (**Fig.38C**).

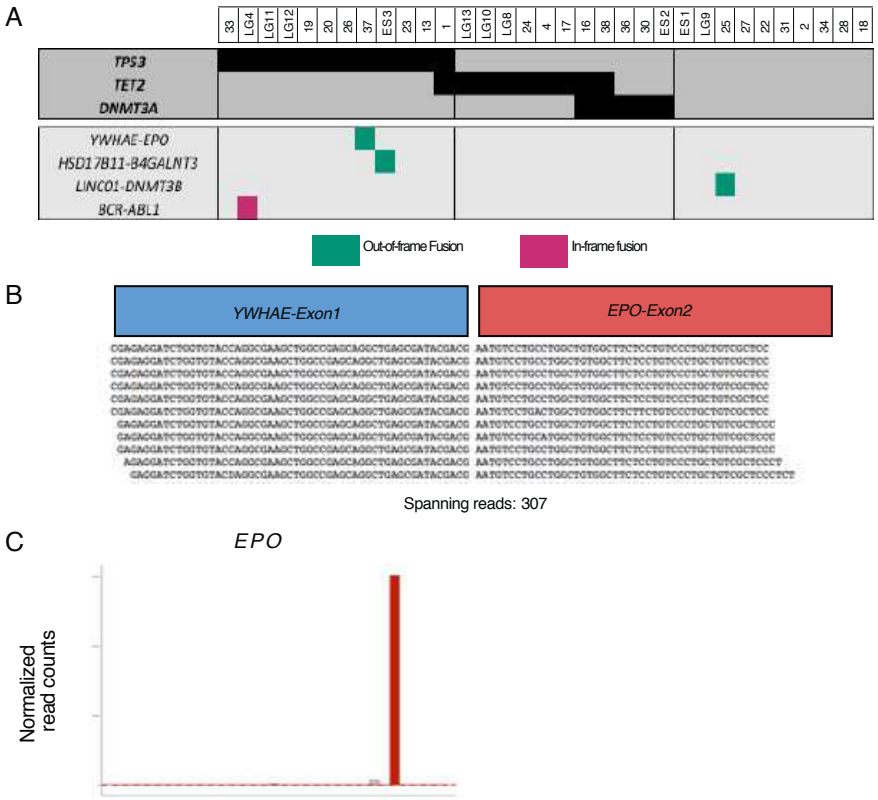


Figure 38. Fusion transcript in human AEL. **(A)** Heatmap representation of patients carrying out-of-frame (green) or in-frame (red) fusion transcripts according to their molecular subgroup. **(B)** Sequences of the breakpoint *YWHAE-EPO* fusion transcript. The number of reads supporting the fusion are indicated. **(C)** Paired histogram representation of *EPO* gene expression in AEL patients, normal human BFU-E and CFU-E. Patient with red carries the fusion transcript. (Analysis and plots done by Alexandre Fagnan, Mercher's lab, Paris).

Alterations of cellular signaling pathways including EPO/EPOR were previously reported in up to 48% of human AEL samples (*388*). Therefore, we wondered whether experimental EPO overexpression would functionally cooperate with the TP53^{R248Q} mutation. To address this possibility, I first retrovirally transduced *TP53*$^{R248Q/+}$ FL-derived HSPCs with an EPO-expression viral vector (*pMSCV-IRES-GFP*) and transplanted flow-sorted transduced (GFP^{+}) cells together with rescue FL-derived HSPCs into lethally irradiated syngeneic recipient mice (**Fig.39A**). Interestingly, after 10 weeks I realized that the skin of the mice that received EPO-expressing cells was more reddish than animals obtaining vector-transduced control cells. The first mouse developed symptoms of disease after less than 10 weeks post-transplantation and

showed a very severe phenotype. We found some infiltration of dark blue cells (most likely erythroblasts) in BM, spleen, lungs, liver and kidney (**Fig.39B**). Although difficult to judge, we speculated that the mouse developed symptoms due to rheostatic problems as it was almost impossible to properly prepare smears; the blood appeared almost glue-like. However, not all symptomatic mice showed the same degree of organ infiltration by cells, that morphologically looked like erythroblasts. The other three symptomatic mice presented erythroblast infiltrations in the lung, less in the liver or/and kidney, variable in the spleen, but not really prominent in other extramedullary tissues (**Fig.39C**) and showed disease symptoms much later after 1-year post-transplantation (**Fig.39D**). Histology suggested that the mice developed a systemic erythroblastosis presenting polychromasia, abnormal high numbers of immature red blood cells (**Fig.39B,C**, red arrows) and few Howell-Jolly body inclusions (clusters of DNA in circulating erythrocytes, **Fig.39B,C,** black arrows) in the peripheral blood, which are a marker of elevated red cell production by the BM and markedly decreased splenic function. However, mice did not suffer from a true blockage in erythroid differentiation that characterizes acute erythroleukemia.

All transplanted mice kept GFP-expressing cells in the peripheral blood over several weeks after transplantation (**Fig.39E**). Importantly, unlike erythroleukemic mice, the spleen and liver weight remained in the normal range (**Fig.39F**). Symptomatic mice presented significantly increased WBCs after 2 weeks post-transplantation ($p=0.0005$), however, at the end of the experiment WBC counts were more variable between the diseased mice. RBC counts were significantly higher already after 4 weeks post-transplantation ($p=0.0004$) and mice presented similar values at the end of the experiment ($p=0.023$). In contrast, PLT counts did not significantly differ from control mice (**Fig.39G**). The immunophenotype of the BM of symptomatic mice was quite variable between diseased mice. Some of them barely expressed single $CD71^+$ and c-Kit^+ cells, whilst others presented significant numbers of single $CD71^+$ as well as c-Kit^+ cells (**Fig.39H,I**). Taken together, transplantation of *TP53*$^{R248Q/+}$ BM cells retrovirally overexpressing EPO resulted in a symptomatic disease that presented (although with variable severity) more as a systemic chronic erythroblastosis rather than acute erythroleukemia.

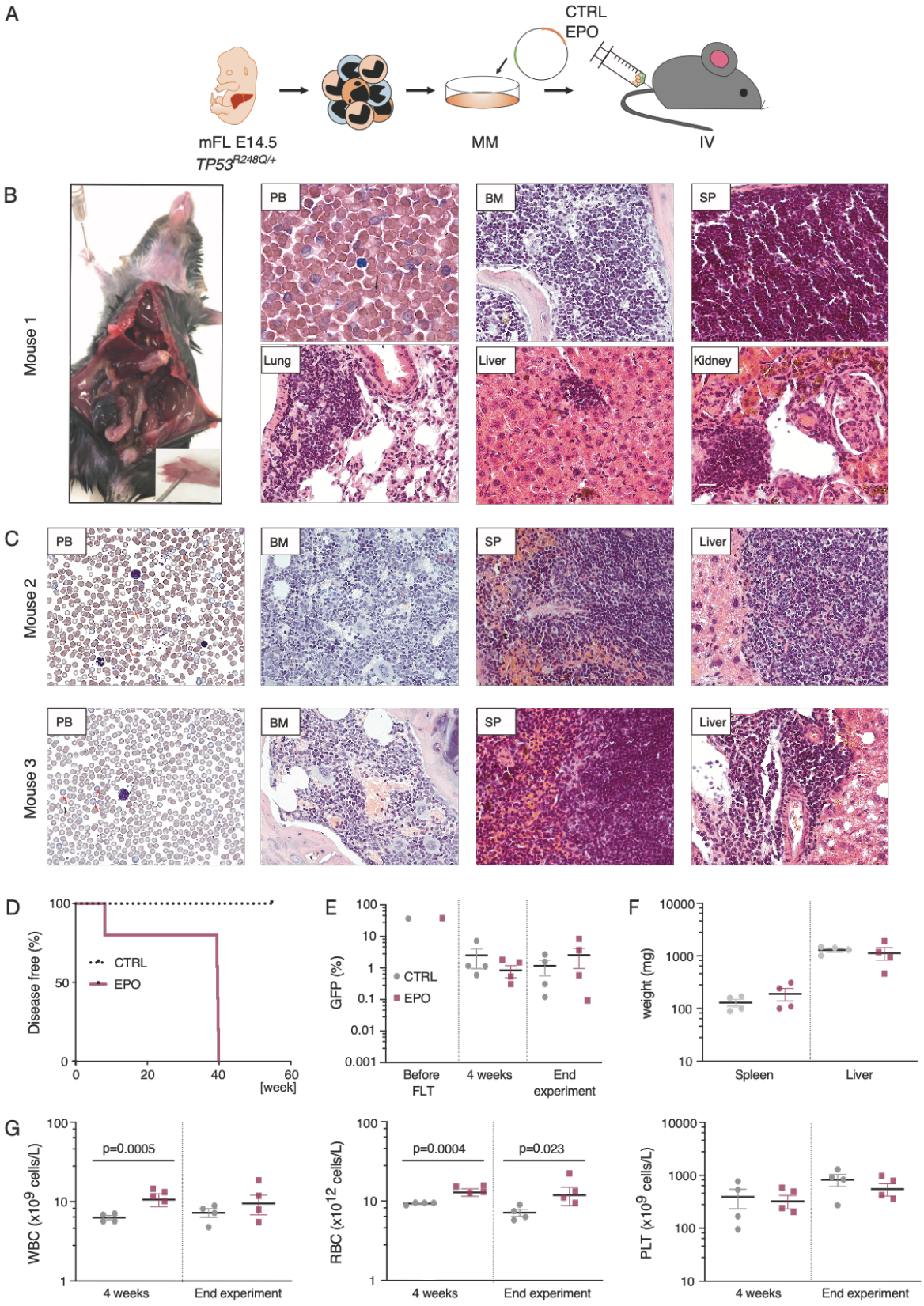


Figure 39. Exploring functional cooperation of EPO overexpression and mutated TP53. (A) Experimental setup. *TP53*$^{R248Q/+}$ FL-derived HSPCs were retrovirally transduced with a human EPO ORF, flow-sorted, and transplanted intravenously into lethally irradiated mice. **(B)** First panel shows a picture of a symptomatic transplanted mouse, followed by a peripheral blood smear (PB), and sections through BM, spleen (SP), lung, liver and kidney. Images were recorded with a 40x objective using a Nikon-TI. **(C)** First panels show pictures of peripheral blood smears (PB), followed by sections through BM, spleen (SP), lung, liver and kidney of symptomatic mice 2

and 3. Images were recorded with a 40x objective using a Nikon-TI. **(D)** Kaplan-Meier plot of transplanted mice with *TP53*$^{R248Q/+}$ FL-derived HSPCs retrovirally expressing EPO (EPO, purple line, n=4) or vector-transduced (CTRL, black dotted line, n=4). **(D)** Amount of transduced GFP^{+} cells (%) before transplantation, 4 weeks post-transplantation in the PB, and at the end of the experiment in the BM. **(E)** Spleen and liver weight (mg) of CTRL (n=4, purple dots) and symptomatic EPO (n=4, grey dots) transplanted mice. **(F)** Peripheral WBC counts (x10^{9}/L) (left panel), RBC counts (x10^{12}/L) (middle panel) and PLT counts (x10^{12}/L) (right panel) in mice transplanted with EPO (n=4, purple dots) or CTRL (n=4, grey dots) *TP53*$^{R248Q/+}$ HSPCs. **(G)** Flow cytometry of two symptomatic mice. Shown are GFP^{+} cells (left panels), CD71/Ter119 (middle panels) and c-Kit/CD71 (right panels) populations. Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in G was tested with unpaired two-tailed t- test.

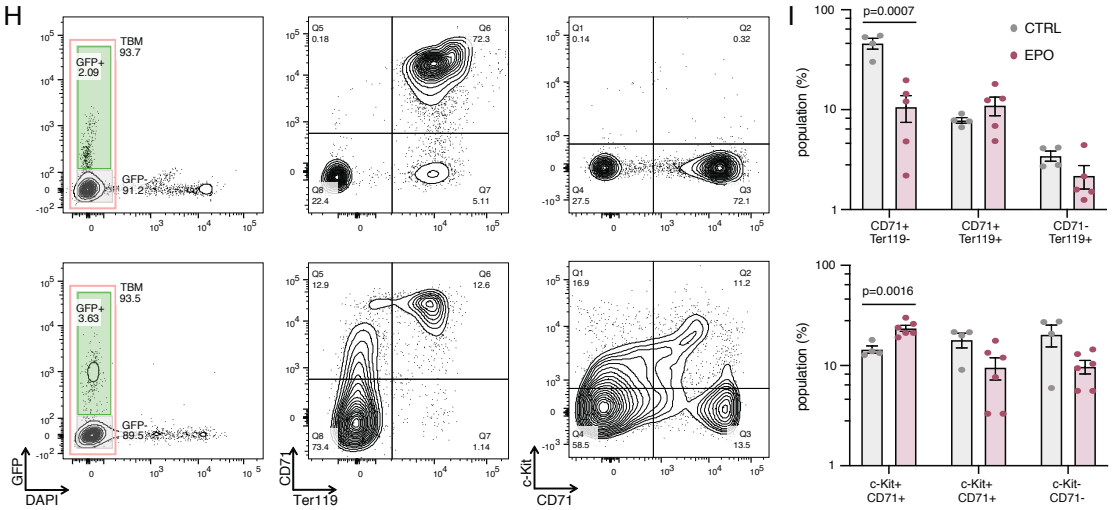


Figure 39 (continuation).
(H) Flow cytometric panels of total BM of mice transplanted with control and *TP53*$^{R248Q/+}$ HSPC transduced with EPO (purple dots) or CTRL (grey dots). Upper plot shows CD71/Ter119, lower plot shows c-Kit/CD71 expressing cell populations. **(I)** CD71 and/or Ter119 surface expression (upper plot) (%) and c-Kit and/or CD71 surface expression on total BM from symptomatic EPO mice (n=4) compared to CTRL mice (n=4). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in I was tested with unpaired two-tailed t- test.

5.5. Discussion

AEL is a very heterogenous and aggressive disease, often difficult to diagnose and classify due to its resemblance with other neoplasms, such as MDS or other AML subtypes.

Comparative transcriptome analysis identified different AEL groups according to their erythroid maturation stage: immature samples clustering together with human CFU-E & Pro-E or only Pro-E signatures, or more mature samples clustering together with human Int-E & Late-E (**Fig.35**). Similar results were obtained using an independent dataset of AEL patients (*242*). These data suggest that the identity of human AEL could reflect the activity of differentiation regulators. However, further analysis, such as the recently published human erythroid lineage cell RNA-seq dataset from Sankaran and colleagues (*389*), would be necessary to validate our findings.

Interestingly, classical immunophenotyping of the BM cells of the patients did not correlate well with the gene expression signature of the leukemic cells (**Fig.35**). This finding suggests that the erythroid phenotype at the time of diagnosis might be initiated by alterations that primarily interfere with erythroid differentiation, or by mutations in multipotent progenitors with a subsequent epigenetic drift towards the erythroid lineage. In fact, Cervera *et al.* recently revealed that similar potential driver mutations were found in the red cell lineage in AEL and other AML types, suggesting that in AEL, mutations occur in a progenitor that has a potency or fate different from a non-AEL progenitor (*299*).

Exome sequencing classified our AEL patients into three different genetic groups. These three groups were characterized by either the presence of TP53 mutations (36.3% of all cases), the presence of genetic alterations in epigenetic regulators already reported in MDS such as TET2, IDH1/2 and DNMT3A mutations (33.3%) and finally, another group characterized by the presence of no recurrent alterations (30.4%) (**Fig.36**). Similar groups have been proposed in targeted sequencing studies (*237, 239, 240, 298, 383*). While our study was ongoing, the Dr. Iacobucci from the Mullighan group from St. Jude (Memphis) reported the genetic landscape of 159 AEL patients. They suggested the presence of 6 genetic groups, of which, like in our study, those patients carrying TP53 mutations made the largest group. (*242*). They also found *NUP98* fusion genes and mutations of the *KMT2A* (MLL1) gene that we did not observe. In contrast to us, they did not report any out-of-frame fusions like *YWAHE-EPO* leading to *EPO* overexpression. Interestingly, YWHAE has been previously reported to be involved in a recurrent t(10;17)(q23;p13) translocation leading to a fusion with the NUTM2A (NUT family member 2A, aka FAM22) in rare endometrial or kidney stromal sarcomas (*390, 391*). However, as far as we know, YWHAE alterations have never been reported in AEL before.

Interestingly, we identified *NFATc4* mRNA (nuclear factor of activated T cells 4) as significantly over-expressed in AEL samples carrying TP53 mutations (**Fig.37**). NFATc4 belongs to the NFAT transcription factor family which is mostly expressed in immune-system cells and plays a critical role in the transcription of cytokine genes and other genes related to immune cell activation. NFATs are hyperphosphorylated in the cytoplasm, where cell stimulation and calcium release activate them leading dephosphorylation and translocation to the nucleus. In the nucleus, NFATs then cooperate with other factors to promote transcription (*392*). The ratio between phosphorylated and fully dephosphorylated forms of NFAT defines the index to assess the calcineurin activity (*393*). Several studies have found that different NFAT isoforms are overexpressed in human solid tumors and hematologic malignancies and have an important role in invasive migration, differentiation and survival of cells in the tumor and its microenvironment (*394, 395*). Medyouf and colleagues found that primary T-cell leukemia and lymphoma exhibited fully dephosphorylated NFAT with increased in calcineurin activity. These observations suggest that calcineurin is activated in lymphoid malignancies (*395*). Notably, the treatment of leukemic mice with FK506 treatment (inhibitor of calcineurin) increased the survival, associated with an inhibition of calcineurin activation and nearly complete NFAT re-phosphorylation (*395*). Furthermore, another recent study showed that diffuse large B-cell lymphoma cells, frequently affected by mutations in TP53, presented constitutive NFAT activation and could be treated by FK506, suggesting inhibition of NFAT-calcineurin axis as a novel strategy for the treatment of tumor cells affected by high expression of NFAT (*396-398*). However, these researchers did not correlate NFAT expression with the presence of mutations or LOH of TP53.

Interestingly, Ding *et al.* suggested that TP53 and NFAT proteins form a reciprocal regulatory complex to control cell growth and differentiation. They found that arsenic trioxide (ATO), an FDA approved drug for acute promyelocytic leukemia (APL), inhibited NFATc4 via TP53 activation. Notably, this group described that *TP53* knockdown or *TP53* inhibition dramatically increased the basal level of NFATc4, whereas *TP53* overexpression repressed it (*399, 400*). Taken together, these data suggest that in our AEL patients with TP53 mutations, reduced TP53 activity may lead to increase in NFATc4 expression, and most probably an increase of calcineurin activity. As outlined before, most leukemia-associated TP53 mutations were shown to impair activity of TP53 in a dominant negative manner (*351*).

GSEA analysis showed that target genes of PRC2 complex and SUZ12 were negatively enriched in patients carrying TP53 mutations (**Fig.37**). Recent studies have shown that HSPCs carrying TP53 mutations express lower levels of genes modified by EZH2 with significantly higher levels of H3K27me3 at TSS compared to WT HSPCs (*323, 376, 377*). I also found similar results in the NFIA-ETO2-expressing $TP53^{R248Q/+,}$ which presented down-regulation of

TP53 targets, most probably regulated by the mutated TP53 together with the PRC2 complex (see Chapter 2). These observations may explain our findings, that gene expression signature of AEL patients carrying TP53 mutations is negatively enriched for PRC2 complex and SUZ12 target signatures.

In AEL samples carrying mutations in epigenetic regulators (TET2, DNMTA3A, IDH1/2), I found higher expression of EGR1(Early growth response 1), a gene that binds to DNA acting as either an activator or repressor of transcription (**Fig.37**) (*401*). EGR1 was shown to be expressed in HSCs and drastically reduced when HSCs were stimulated to proliferate and migrate (*386*). EGR1 is also expressed in both, lymphoid and myeloid cells (*402, 403*) and its overexpression leads to differentiation towards macrophage lineage at the expense of developing along either the granulocyte or erythroid lineage (*404*). Interestingly, since high expression of *Egr1* resulted in differentiation towards macrophage progenitors, mice transplanted with hematopoietic progenitor cells ectopically expressing *Egr1* failed to repopulate the BM and spleen (*404*). This finding could suggest that the overexpression of EGR1 in AEL patients might impair the erythroid progenitors' development, thus resulting in a blockage of erythroid differentiation.

I also tried to model the consequence of aberrantly high EPO expression in erythroblasts carrying a TP53 mutation (**Fig.39**). Mice transplanted with FL-derived HSPCs carrying *TP53*$^{R248Q/+}$ retrovirally overexpressing EPO developed a symptomatic disease that presented as systematic erythroblastosis rather than erythroleukemia or excessive erythrocytosis, with some similarities to previously described mouse models overexpressing EPO (*405-407*). Symptomatic mice did not present a dramatically enlarged spleen, suggesting that there was no extramedullary erythropoiesis, as found before in mice overexpressing EPO (*405*). Vogel *et al.* also reported increased blood viscosity. In our mouse model on the other hand, although we did not check for viscosity, the peripheral blood of only one symptomatic mouse appeared very thick and we had more difficulties to prepare blood smears, and morphologically we observed very large and dark red erythrocytes.

Lastly, since the 2016 WHO classification assigned most cases previously diagnosed as AEL to MDS or AML, Dr. Frederik Otzen-Bagger from our group computed a transcriptome-based space to directly compare AEL, MDS and other AML forms. Hereby he found that our samples clustered between the erythroid and myeloid trajectories (**Fig.40**).

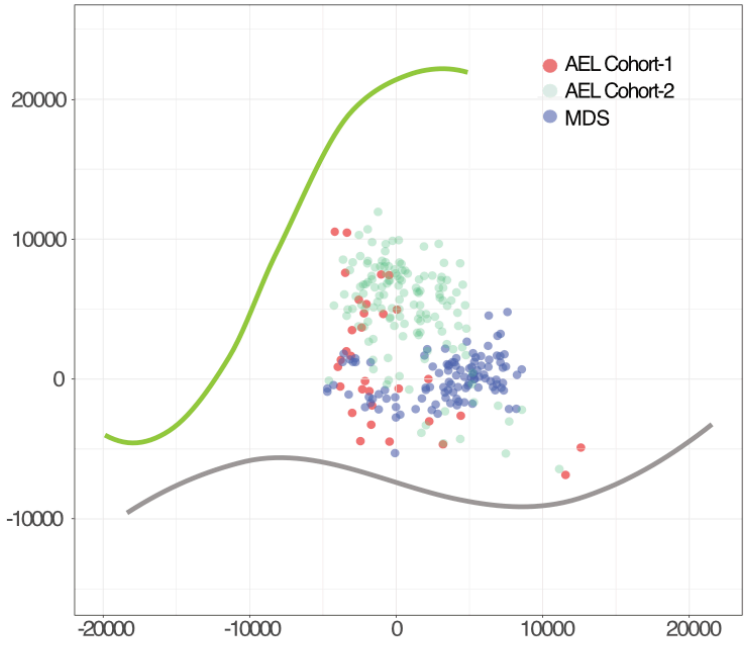


Figure 40. Partial overlap in AEL and MDS transcriptomes in an erythro-myeloid expression space. Principal component analysis (PCA) of data from Differentiation Map (DMAP) (*408*) with regression to cell types in an erythroid and myeloid compartment. The regression line fits to erythroid (green) and myeloid (grey) cells in the PCA space of genes significantly (FDR>0.05, LogFC>2). Shown are projections of AEL patient samples (Cohort 1 Fagnan, red dots, Cohort 2, Iacobucci *et al.,* green dots) and MDS samples (blue dots) (*409, 410*).

Specifically, 17 cases mapped closer to the erythroid axis and 6 to the myeloid axis. These last 6 samples appeared very similar to MDS-derived signatures and one sample showed high expression of *SPI1*, generally associated with other AML subtypes. Interestingly, the two samples that clustered very close to the myeloid expression axis were recently re-diagnosed as AML rather than AEL. These data suggest that, even though we saw some overlaps, the transcriptome of the majority of AEL cases differed from those of MDS, however, some AEL cases could also be reclassified as other AML subtypes. Overall, our findings demonstrate that AEL classification might be more complex than the latest 2016 WHO classification, which simply divided them into PEL and MDS. Our data proposes that AEL is heterogeneously spread along differentiation, with some patients mapping closer to progenitors retaining myeloid features, and other patients mapping closer to more mature erythroid cells. These findings indicate that the myeloid and erythroid composition of the samples at the time of diagnosis could be based on the type of progenitor targeted by the mutations.

Material and methods

Cloning NFIA-ETO and ZMYND8-RELA fusion cDNAs into *MSCV* retroviral expression vectors

Full-length cDNA for murine *Nfia* was acquired from Addgene (#112698) and full-length cDNA for human *ETO2* was obtained from Thomas Mercher (Paris). ORFs fusing *Nfia* to *ETO2* were first generated. That involved *Nfia* from the beginning until the site of the breaking point (BP, 509aa) and *ETO2* from the site of the BP (50aa) until the end (653) (**Fig.41**, right scheme). Full-length cDNA for human ZMYND8 and full-length cDNA for mouse RelA were acquired from Addgene (#65401, #23255). A fusion ORF was cloned composed of ZMYND8 from the beginning until the site of BP before Znf_MYND domain fused the entire *RelA* sequence (lacking the ATG) (**Fig.41B**, left scheme). The fusion genes were ligated into *pMSCV* and *pcDNA3.1* expression vector containing selection markers such as green fluorescent protein (GFP) or neomycin respectively.

To clone NFIA-ETO2 single mutant deletions, Q5 Site-Direct Mutagenesis Kit (New England BioLabs) was used according to the manufacturer's protocol. Primers are listed in **table 7**.

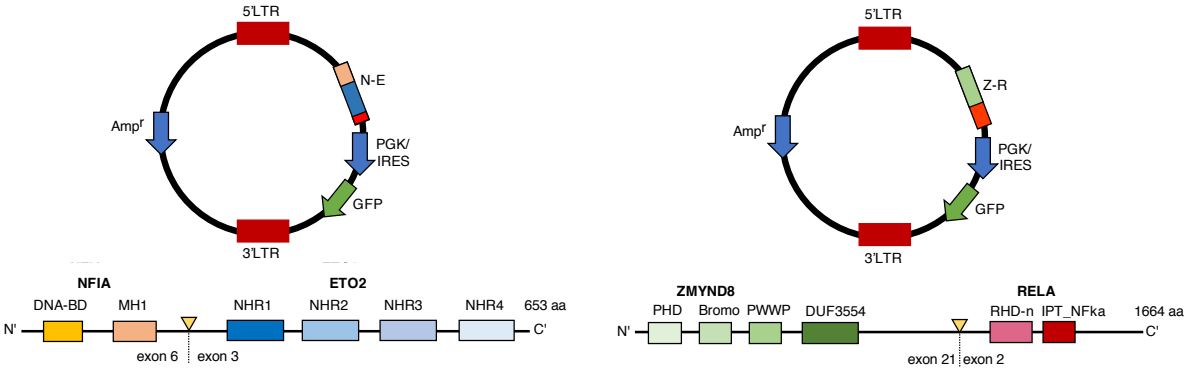


Figure 41. Schematic representation of NFIA-ETO2 and ZMYND8-RELA constructs and their domains.

MEL cells

Murine erythroleukemia (MEL) cells were cultured in suspension in DMEM (Dulbecco's Modified Eagle Medium) (Gibco, Lubio, Thermo Fisher Scientific, Reinach, Switzerland), 5% FCS (Fetal Calf Serum), 1% P/S (Penicillin-Streptomycin) (Reddy and Shen 1992). To induce differentiation, 10^5 cells were cultured in DMEM with 2%DMSO (sigma). After 2 days, cells were washed and counted, and cultured again in right volume of newly prepared medium. Differentiation was continued for 4 days in presence of DMSO. Differentiation was shown by flow cytometry ($CD71^+/TER119^+$), benzidine staining and hemoglobin expression by qPCR (Real-time polymerase chain reaction).

Benzidine staining

Benzidine staining is used to score erythroid differentiation and determines the production of hemoglobin. For the staining, cells were incubated with 0.3% hydrogen peroxide and 0.2% di-hydrochloride benzidine (Sigma Aldrich, Buchs, Switzerland) in a ration 2:5, in 0.5M acetic acid/1x PBS for 5 min at 37°C. Cells were washed and prepare for cytospin.

Isolation of FL- and BM-derived hematopoietic cells

To isolate adult mouse BM, long bones and spine were crushed and filtered through a 70μm cell strainer (Cat. 352350, BD, New Jersey, USA) and the red blood cells were lysed with ammonium-chloride potassium (ACK) lysis buffer (150mM NH_4Cl, 10mM $KJCO_3$, and 0.1 mM EDTA, pH 8.0) for 10min on ice. Mouse hematopoietic lineage depletion kit (Cat. 130-090-858, Milteny Biotech, Bergisch Gladbach, Germany) was used to enrich for HSPCs, according to manufacturer's protocol. To obtain FL cells, one male mouse and one or two females were places together overnight. 14.5 days later, pregnant females were sacrificed by CO_2 asphyxia, and FL cells from individual fetuses were isolated, chopped and passed through 50μm cell strainer.

Cytospin preparation

Cytospin preparation of 10^5 cells were made in non-coated cytoslides (Cat. 5991051, Thermo Fisher Scientific, Reinach, Switzerland) by centrifugation of 3min at 200rpm using a Shandon Cytospin 3 centrifuge using cytofunnel disposable sample chambers (Cat. 5991040, Thermo Fisher Scientific, Reinach, Switzerland). Cytospins were stained with Wright-Giemsa solution.

In vitro erythroid differentiation assay

In vitro erythroid differentiation of primary erythroblasts was studied using a previously published protocol (*301*). Adult mice erythroblasts derived from BM lineage depleted cells lin(-) were cultured for more than one week and split every two days in "maintenance medium (MM)" composed of StemSpan SFEM (Stem Cell Technologies, Vancouver, Canada), supplemented with 1%Pen/Strep, 0.4%cholesterol (Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 2U/ml hEpo (Eprex 4000, 9096976, Pharmacy of University Hospital Basel), 100ng/ml mScf (Peprotech, London, UK), 10^{-6}M dexamethasone (Calbiochem, Sigma Aldrich, Buchs, Switzerland) and 50ng/ml hIGF-1 (Peprotech, London, UK). The presence of erythroblasts was confirmed by flow cytometry ($DAPI^-$/c-Kit^+/$CD71^+$/$Ter119^-$). Erythroid maturation of erythroblasts was induced in "differentiation medium (DM)" composed of IMDM

(Iscove's Modified Dulbecco's Media) (Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 5% hPDS (0.45μM filtered, Blood donation center, University Hospital Basel), monothioglycerol (Sigma Aldrich, Buchs, Switzerland), 100ng/ml mSCF and 2U/ml hEPO. Cells were also split every two days, and cultured 500.000 cells/ml.

Colony formation assay

For colony formation analysis of erythroblasts ($CD71^+/Ter119^-$), approximately $10x10^3$ cells were plated in methylcellulose M3434 (Methocult, Stem Cell Technologies, Vancouver, Canada). After 7 to 10 days, colonies were counted and pictures were taken on Olympus IX50 microscope with 2x and 4x magnification. For replating and cell counting, cells were washed with warm PBS 2x (8min,1800rpm) and cells were counted with trypan blue and if there were enough cells, same number of cells were replated into fresh methylcellulose. For colony formation of lin(-) HSPCs analysis, approximately $4x10^4$ cells were plated in also M3434. Cells were transduced twice after isolation and cultured in Stem and progenitor medium (SPM) composed of 50ng/ml hTPO (Peprotech, London, UK), 20ng/ml hIL-11 (Peprotech, London, UK), 20ng/ml hFLT3 ligand (Peprotech, London, UK) and 100ng/ml mScf (Peprotech, London, UK). HSPCs were plated into methylcellulose directly after sorting.

Flow cytometry

Cells were washed with FACS buffer (0.5% BSA, 1mM EDTA in PBS, filtered) and incubated for 15 min at RT with specified antibodies in 100μl FACS buffer. After staining, cells were again washed and stained with 1μg/ml DAPI (Life Techologies, Paisley, UK) in FACS buffer. LSR Fortessa (BD, New Jersey, USA) was used to analyze the stained cells. The data was analyzed with FlowJo software (Tree Star). For cell sorting, cells were washed with PBS and resuspended with 300μl of FACS buffer with 1μg/ml DAPI. Cells were sorted on BD FASC ARIA III (BD, New Jersey, USA) in 1.5mL Eppendorf with 300μl of medium. All antibodies used in this analysis are listed in **Table 8**.

Preparation of retroviral supernatants and transduction

For viral transfection, HEK293T-LX were cultured in DMEM (Gibco, Lubio, Thermo Fisher Scientific, Reinach, Switzerland) with 10%FSC and 1%P/S. Transfection was performed by co-transfection of packaging vector (*pIPAK6*) and respective plasmids using Turbofect or Jetprime transfection reagent (Life Technologies, Paisley, UK) in HEK293T LX. Supernatant containing retrovirus was harvested 48 and 72 hours after transfection, using 10x Vivaspin 20 (Sartorius, Göttingen, Germany) concentrated at 3500rpm for 2h at 4°C and snap frozen in liquid nitrogen ad stored at -80°C until usage. All plasmid used in this study are listed in **Table**

9. HeLa cells were transfected with *pcDNA3* plasmid encoding for NFIA-ETO2 or the single deletion mutants, using the Jetprime transfection reagent.

MEL cells, BM lin(-) HSPC and BM-derived erythroblasts were transduced in DMEM (5%FCS, 1%P/S), in SPM or in MM respectively. Cells were transduced in presence of 5ug/ml polybrene (Sigma Aldrich, Buchs, Switzerland) with virus for 90min, 1800 rpm at 30°C. Cells were incubated overnight with the cells, and washed with PBS and plated in their respective mediums. MEL and BM lin(-) cells were spin infected twice before sorting. Two days later the cells were $EGFP^+$ sorted as previously described.

BM transplantation

Transplantation were performed using E14.5 FL-derived erythroblasts expressing NFIA-ETO2 or vector-control, together with rescue total BM from 6-9 weeks old mice into lethal irradiated recipient mice (BC57BL/6NCrl) (2x 600cGy). Around $1.5x10^6$ erythroblast cells were transplanted per mouse. Around $4x10^6$ rescue total BM cells were also transplanted per mouse. E14.5 FL-derived or BM-derived HSPCs from 6-9 weeks old mice, expressing or not the fusions or vector control, were transplanted into lethal irradiated recipient mice. Around $5x10^5$ HSPCs were transplanted per mouse. For secondary transplantation, $2x10^6$ total BM cells of symptomatic mice were transplanted into each lethal irradiated recipient mice. For $p53^{R248Q}$ genotyping, list of primers is found in **Table 10.**

Analysis of symptomatic mice

Symptomatic mice were sacrificed by CO_2 asphyxia, and one long bone and organs (lungs, heart, liver, spleen, intestine, stomach, kidney) were removed and fixed in buffered 4% formalin solution. Liver and spleen were weighted to check for disease symptoms. Blood collection of live mice was performed in the tail vein and by *vena cava inferior* puncture when it was the end of experiment. Blood counts were determined using an Advia120 Hematology Analyzer using Multispecies Version 5.9.0-MS software (Bayer, Leverkusen, Germany). Hematoxylin and eosin (H&E) staining was used for paraffin-embedded tissue sections and Wright-Giemsa staining (Hematology, University Hospital Basel) was used to stain blood smears. Sections and smears were analyzed on Nikon TI (Tokyo, Japan), and pictures were taken 20x,40x and 60x.

Quantitative RT-PCR

RNA Plus extraction kit (Macherey-Nagel, Düren, Germany) was used to extract total RNA according to the manufacturer's protocol. Nucleospin RNA Plus kit (Ref 740990.250, Macherey-Nagel, Düren, Germany) was used when cell number was lower than $50x10^4$ cells,

also according to the manufacturer's protocol. The high capacity cDNA reverse transcription kit (Cat. 4368814, Applied Biosystems, Foster City, USA) was used to perform the cDNA synthesis. Quantitative PCR was performed using SYBR Green reagent (applied Biosystems, Foster City, USA) or TaqMan (Thermo Fisher Scientific) and analysis was performed in an ABI prism 7500 sequence detection system. Gapdh expression was used to normalized the Ct values and relative expression was quantified using 1/dCt. Primers are listed in a **Table 11**.

Western Blotting

For primary cells, nuclear protein extraction was performed. All used buffers were supplemented with Complete Mini protease inhibitors (Cat. 11836153001, Roche). Nuclear protein lysates were prepared by resuspending the cell pellets in hypotonic lysis buffer (10mM HEPES pH=7.9, 10mM KCl, 0.1mM EDTA, 0.1mM EGTA, 1mM DTT) for 15 min on ice and then treated with 0.1% NP-40 with 15sec vortexing. Nuclei were then spun down at 14.000 rpm for 2min at 4°C and the supernatant which contained the cytoplasmic fraction was kept for further analysis. Pellets were well resuspended in nuclear lysis buffer (20mM HEPES, pH=7.9, 0.4M NaCl, 1mM EDTA, 1mM EGTA, 1mM DTT). Finally, pellets were sonicated for 5 cycles (30sec sonication, 30sec pause) on a Bioruptor Pico Sonicator (Diagenode, Seraing, Belgium) and kept on ice for 20min. After incubation, nuclear extraction was spin down at 14.000rpm for 15min at 4°C and the pellet which contained histones was discarded and supernatant kept for nuclear protein analysis. To quantify the proteins, Bradford assay (Biorad, München, Germany) was used. Before loading, the samples were prepared in 4x Laemmli buffer (Biorad, München, Germany) and boiled for 10min at 95°C. For whole lysate extraction of MEL and HELA cells, 1 million cells were boiled in Laemmli buffer 10min at 100°C and centrifuge 15 min at 4°C at maximum speed. Supernatant was kept for whole lysate analysis. Nuclear extraction and whole lysate samples were loaded on a handmade 7% polyacrylamide gels. For nuclear extraction of NFIA-ETO2, 50ug of nuclear extract was loaded. Wet transfer was done for 100min on ice in 5%Methanol/0.1%SDS/Tris-Base-Bicine buffer on 0.45μM nitrocellulose membranes. Membranes were blocked in 5% non-fatty milk (NFM) in PBS-1% Tween for 2 hours at room temperature and probed overnight with the selected antibody at 4°C in PBS-1%Tween. Blots were washed three times for 15min in PBS-1% Tween and probed with a secondary antibody in PBS-1%, and again washed three times for 15min in PBS-1%. Blots were then probed with Supersignal West Femto Max substrate (Thermo Scientific, Reinach, Switzerland) and develop using Carestream Biomax Kodak films (Sigma, New York, USA). List of antibodies and respective dilutions can be found in **Table 12**.

Quantitative LOH analysis

Genomic DNA was isolated from total BM from diseased *TP53*$^{R248Q/+}$ mice, and used as control *Tp53*$^{+/+}$ and *TP53*$^{R248Q/-}$ total BM cells, using the Puregene Core Kit A (Qiagen). Quantitative real-time PCR was performed in duplicates with SYBR Green reagent (applied Biosystems, Foster City, USA) and analysis was performed in an ABI prims 7500 sequence detection system, using 8ng genomic DNA and the following mouse wtp53 allele-specific primers: 5′-ACAGCGTGGTGGTACCTTAT-3′ (forward) and 5′-TATACTCAGAGCCGGCCT-3′ (reverse). These primers anneal to mouse exon 5 and 6 of for the mouse *Tp53*, but do not recognize the humanized *TP53*^{R248Q} allele. The WT *Tp53* signal was normalized to the *Rosa26* signal using the following primers: 5′-AAAGTCGCTCTGAGTTGTTAT-3′ (forward) and 5′-GGAGCGGGAGAAATGGATATG-3′ (reverse) (*304*).

RNA sequencing

WT and *TP53*$^{R248Q/+}$ FL erythroblasts were isolated from fetuses from three pregnant mice. Erythroblasts were transduced either with NFIA-ETO2 or NFIA-ETO2 lacking NHR4 domain. Total RNA was isolated using Nucleospin RNA Plus XS kit (Ref 740990.250, Macherey-Nagel). RNA was quality-checked on the Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA) using the RNA 6000 Nano Chip (Agilent, Cat# 5067-1511) and quantified by Fluorometry using the QuantiFluor RNA System (Cat# E3310, Promega, Madison, WI, USA). Library preparation was performed, starting from 200ng total RNA, using the TruSeq Stranded mRNA Library Kit (Cat# 20020595, Illumina, San Diego, CA, USA) and the TruSeq RNA UD Indexes (Cat# 20022371, Illumina, San Diego, CA, USA) / and the TruSeq RNA CD Index Plate (Cat# 20019792, Illumina, San Diego, CA, USA) /. 15 cycles of PCR were performed. Libraries were quality-checked on the Fragment Analyzer (Advanced Analytical, Ames, IA, USA) using the Standard Sensitivity NGS Fragment Analysis Kit (Cat# DNF-473, Advanced Analytical) revealing excellent quality of libraries (average concentration was 103±20 nmol/L and average library size was 381±5 base pairs). Samples were pooled to equal molarity. The pool was quantified by Fluorometry using the QuantiFluor ONE dsDNA System (Cat# E4871, Promega, Madison, WI, USA). Libraries were pooled and sequenced SR76 with the Illumina NextSeq 500 system using the High Output Kit v2.5 (75 cycles). Primary data analysis was performed with the Illumina RTA version 2.4.11, bcl2fastq v2.20.0.422.

RNA-sequencing data analysis

To quantified the reads, QuasR was used. Count per million (CPM) normalization was used to normalize the counts and plot the log2 normalized data in gene expression plots as well as principal analysis complement (PCA) (*411*). Differentially gene expression analysis was

performed with the EdgeR package. The design to analyze the data was: ~ 0 + group + mother, where groups where classified according to fusion gene expression and medium were cells were cultured. fGSEA (*412*) was used to perform gene set enrichment analysis (GSEA). Parameters that were used were: nperm=100000, minSize=8, maxSize=5000. Several databases were used for the gene signatures analysis: MsigDB [https://doi.org/10.1016/j.cels.2015.12.004], Gene Ontology (*413*), HALLMARK (*414*), KEGG (*415*) databases and erythronDB (*416*). Using the functionality "build gene lists" comparing proerythroblasts and basophilic erythroblast cells, the "ERYTHROID LINAGE DEVELOPMENT" pathway was obtained. The list can be found in **Supplementary Table 8**. NFIA targets were obtained using the TRANSFAC, a database on transcription factors and their DNA binding sites (*417*). The list can be found in **Supplementary Table 9**.

Gene expression of external cohorts of AML patients was obtained from Iaccobucci et. al (*242*) and TCGA (*167*), using age groups and mutational state as defined in respective studies. Iaccobucci mRNA expression from RNA-seq was kindly provided processed as count table format from the authors, and TCGA data was obtained raw .cel files from TCGA (project LAML HG-U133_Plus_2, Level 1120) and rma normalised with ReadAffy (*418*).

ATAC sequencing

ATAC sequencing was performed using a previously published protocol optimized for blood cells (*419*). Digitonin (Sigma Aldrich, Buchs, Switzerland), a very gentle detergent, was used. Same experimental setup was used as the RNA sequencing experiment. WT and $p53^{R248Q/+}$ FL erythroblasts from fetuses from three different pregnant mice were transduced with NFIA-ETO2. After sorting, cells were cultured in maintenance and differentiation medium for 24h. $75x10^5$ cells were washed two times with cold PBS and spun down at 500g for 5 min at 4°C. Supernatant was removed using two pipetting steps. Transposase mixture was prepared (25 µl 2x TD buffer, 2.5 µl TDE1, 22 µl H2O, 0.5 µl Digitonin) (FC-121-1030, Illumina; G9441, Promega) and added directly to the pellet and transferred the pellet to a new PCR tube were pellet was disrupted by pipetting. Samples were incubated at 37°C for 30 min. Next step was to purified the transposed DNA using the QIAGEN MinElute Reaction cleanup kit (28204) following the manufacturer's protocol. Purified DNA was eluted in 12 µl of elution buffer (10 mM Tris-HCl, pH=8). Transposed fragments were amplified using 2X KAPA HiFi HOTSTART Ready mix (Roche) with modified primers (*420*). PCR purification was performed using Qiagen PCR purification kit (Cat no. 28104) according to the manufacturer's protocol.

Libraries were quality-checked on the Fragment Analyzer (Advanced Analytical, Ames, IA, USA) using the High Sensitivity NGS Fragment Analysis Kit (Cat# DNF-474, Advanced Analytical). AMPure XP beads (Beckman Coulter) Double Size Selection (1.2-0.8x) was

performed to enrich for 150-400 bp library fragments before mixing library pools for sequencing. Before sequencing with the NovaSeq, the pool of libraries was treated with Free Adapter Blocking Reagent (Illumina) according to manufacturer's specifications.

Library pools were sequenced PE 2x38 and PE 2x51 with the Illumina NextSeq 500 system using the High Output Kit v2.5 (75 cycles) and the Illumina NovaSeq 6000 system using the S1 Reagent Kit (100 cycles), respectively. NextSeq primary data analysis was performed with the Illumina RTA version 2.4.11, bcl2fastq v2.20.0.422, and NovaSeq primary data analysis with pipeline version v3.4.4.

ATAC-sequencing: alignment and quality control

Reads were aligned with bowtie2 to the mouse genome (UCSC version mm10). The output was sorted and indexed with samtools. Duplicated reads were marked with picard. Coverage tracks per sample were generated by tiling the genome in 20bp windows and counting 5'end of reads per window using the function bamCount from the bioconductor package bamsignals. These window counts were exported in bigWig format using the bioconductor package rtracklayer.

ATAC-sequencing: peak calling

For each group of biological replicates, regions of accessible chromation where called with macs2 using the option '-f BAM -g 2652783500 --nomodel --shift -100 --extsize 200 --broad --keep-dup all --qvalue 0.05'. The resulting peak lists were cleaned from peaks called in mitochondria and ENCODE blacklist regions. Additionally, we applied a log-fold-change > 1.5 and FDR<0.05 filter. The filtered peak lists were converted to bigBed format using the UCSC command line tool bedToBigBed. Additionally, peaks were annotated using the ensembl transcript annotation (version ensembldb_Mus_musculus_GRCm38_ensembl_96.sqlite). For the filtered peak set, the fraction of peaks overlapping with promoters, genes (intro/exons) and intergenic regions and their respective enrichment assuming a random distribution was calculated. For transcription factor prediction, a novel computational method and software package was used: Binding Analysis for Regulation of Transcription (BART)(*324*).

Data presentation and statistical analysis

Bar plots represent the mean value of biological replicates. Error bars are plotted as standard error of the mean (Mean±SEM). For mouse statistics, unpaired two-tailed t-test was used. For viral transduction experiments, paired two-tailed t-test was used. In both cases, equal variance was assumed. Statistical test was performed in $\log_{10}$ space, or $\log_2$ space for RT-PCR.

Table 7: Primers for cloning NFIA-ETO2 mutants

Target	Forward primer (5'-3')	Reverse primer (5'-3')
DBD	CGCATGTCGAAAGAAGAG	ACAGAGCGGAGAATACATG
MH1	TCTGTAAAGGAACTCGATTTATATTTG	CTCGGGCTTCTCGCTTAG
NHR1	TTGGCCCAGCATGAGCAGCTCCTGC	CTGCCGGGCCCCGCAGGC
NHR2	AAGGGCCCCGCTCCCGCC	CCGGGACCCAGGCACCACAAG
NHR4	GGCCAGAGCCTGCAGGGC	GCTCTCGCTGGAGTCCTCCTG

Table 8: Antibodies for Flow cytometry analysis.

Antibody	Clone	Fluorochrome	Concentration	Supplier
CD71	C2	PE	1:100	BD Biosciences
Ter119	TER-119	APC	1:100	BD Biosciences
c-Kit	2B8	PE-Cy7	1:100	BD Biosciences
Mac-1	M1/70	APC	1:100	BD Biosciences
Gr-1	RB6-8C5	APC-Cy7	1:100	BD Biosciences

Table 9: Plasmids

Name	Description/Source
NFIA-ETO2 (#535)	*pMSCV-Nfia-ETO2-pgk-IRES-GFP*
NFIA-ETO2-Flag (#535.2)	*pMSCV-Nfia-ETO2-Flag-pgk-IRES-GFP*
ZMYND8-RELA (#575)	*pMSCV-ZMYND8-RELA-pgk-IRES-GFP*
GFP-Ctrl (#14)	*pMSCV-pgk-IRES-GFP*
NFIA-ETO2-ΔDNA.BD	*pMSCV-Nfia-ETO2-ΔDNA.BD-Flag-pgk-IRES-GFP*
NFIA-ETO2-ΔMH1	*pMSCV-Nfia-ETO2-ΔMH1-Flag-pgk-IRES-GFP*
NFIA-ETO2-ΔNHR1	*pMSCV-Nfia-ETO2-ΔNHR1-Flag-pgk-IRES-GFP*
NFIA-ETO2-ΔNHR2	*pMSCV-Nfia-ETO2-ΔNHR2-Flag-pgk-IRES-GFP*
NFIA-ETO2-ΔNHR4	*pMSCV-Nfia-ETO2-ΔNHR4-Flag-pgk-IRES-GFP*
pcDNA3	Invitrogen, CMV promoter base

Table 10: Genotyping primers

Target	Forward primer (5'-3')	Reverse primer (5'-3')
R248Q	CCCTCCAGCTCAGCCTTTGTAG	CTTGATCAAGGCTTGGAAGGC
Cre	CGATGCAACGAGTGATGAGG	GCATTGCTGTCACTTGGTCGT

Table 11: RT-qPCR primers.

Target	Forward primer (5'-3')	Reverse primer (5'-3')
mGapdh	ATGACATCAAGAAGGTGGTG	CATACCAGGAAATGAGCTTG
mGata1	GTGTCCTCACCATCAGATTCCAC	TCCCTCCATACTGTTGAGCAGTG
mHb-β	GTCTCTTGCCTGTGGGGAAA	CAACCAGCAGCCTGCCC
mHb-α	TGATGTAAGCCACGGCTCTG	CAGTGGCTCAGGAGCTTGAA
mHb-γ	CCTCATCAATGGCCTGTGGA	CACAACAAGAAGCCTTCCCAAG
mGypa	TGAAGTGTCTGCTGCGTT	CCGATAATCCCTGCCATCA
mEPOR	GGGCTGCATCATGGACAAAC	AGGGCCGCTTTGCTCTC
NFIA-ETO2	TTGGCATACTTTGTACATGCAGC	ATGTGTGTGTGGCGTGAAGGAGGGGGGGC
ZMYND8-RELA	AGAGTGGCTCGACCCTTGACCTTTCTGGC	AGCGGAATCGCATGCCCCGTTGCTTCGGC

Table 12: Antibodies used for Western blot analysis

Target protein	Clone	Dilution	Supplier
GATA1	D52H6 XP	1:1000	Cell signaling
ETO2	C-20/sc-9739	1:500	Santa Cruz
FLAG	D6W5B	1:1000	Cell signaling
LAMIN-A/C	E-1/ sc-376248	1:5000	Santa Cruz
ACTIN	C-11	1:6000	Santa Cruz
P53	1C12	1:1000	Cell signaling

References

1. A. Silva, A. R. Anderson, R. Gatenby, A multiscale model of the bone marrow and hematopoiesis. *Math Biosci Eng* **8**, 643-658 (2011).
2. F. Notta *et al.*, Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016).
3. C. Nombela-Arrieta, M. G. Manz, Quantification and three-dimensional microanatomical organization of the bone marrow. *Blood Adv* **1**, 407-416 (2017).
4. J. C. Boisset, C. Robin, On the origin of hematopoietic stem cells: progress and controversy. *Stem Cell Res* **8**, 1-13 (2012).
5. A. J. Becker, C. E. Mc, J. E. Till, Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452-454 (1963).
6. E. D. Thomas, H. L. Lochte, Jr., W. C. Lu, J. W. Ferrebee, Intravenous infusion of bone marrow in patients receiving radiation and chemotherapy. *N Engl J Med* **257**, 491-496 (1957).
7. M. Jagannathan-Bogdan, L. I. Zon, Hematopoiesis. *Development* **140**, 2463-2467 (2013).
8. K. E. McGrath, A. D. Koniski, J. Malik, J. Palis, Circulation is established in a stepwise pattern in the mammalian embryo. *Blood* **101**, 1669-1676 (2003).
9. L. Silver, J. Palis, Initiation of murine embryonic erythropoiesis: a spatial analysis. *Blood* **89**, 1154-1164 (1997).
10. J. Palis, M. C. Yoder, Yolk-sac hematopoiesis: the first blood cells of mouse and man. *Exp Hematol* **29**, 927-936 (2001).
11. K. Choi, M. Kennedy, A. Kazarov, J. C. Papadimitriou, G. Keller, A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725-732 (1998).
12. T. L. Huber, V. Kouskoff, H. J. Fehling, J. Palis, G. Keller, Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432**, 625-630 (2004).
13. M. Kennedy *et al.*, A common precursor for primitive erythropoiesis and definitive haematopoiesis. *Nature* **386**, 488-493 (1997).
14. J. Tober *et al.*, The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood* **109**, 1433-1441 (2007).
15. K. K. Hirschi, Hemogenic endothelium during development and beyond. *Blood* **119**, 4823-4827 (2012).
16. J. Y. Bertrand *et al.*, Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. *Development* **134**, 4147-4156 (2007).
17. K. E. McGrath *et al.*, A transient definitive erythroid lineage with unique regulation of the beta-globin locus in the mammalian embryo. *Blood* **117**, 4600-4608 (2011).
18. J. M. Frame, K. E. McGrath, J. Palis, Erythro-myeloid progenitors: "definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells. *Blood Cells Mol Dis* **51**, 220-225 (2013).
19. I. Pietilä, S. Vainio, The embryonic aorta-gonad-mesonephros region as a generator of haematopoietic stem cells. *APMIS* **113**, 804-812 (2005).
20. K. Kissa, P. Herbomel, Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112-115 (2010).
21. J. Y. Bertrand *et al.*, Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108-111 (2010).
22. M. Ogawa, Differentiation and proliferation of hematopoietic stem cells. *Blood* **81**, 2844-2853 (1993).
23. S. J. Morrison, J. Kimble, Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* **441**, 1068-1074 (2006).
24. L. I. Zon, Intrinsic and extrinsic control of haematopoietic stem-cell self-renewal. *Nature* **453**, 306-313 (2008).
25. L. Velten *et al.*, Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271-281 (2017).
26. K. Akashi, D. Traver, T. Miyamoto, I. L. Weissman, A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193-197 (2000).
27. R. Lu, A. Czechowicz, J. Seita, D. Jiang, I. L. Weissman, Clonal-level lineage commitment pathways of hematopoietic stem cells in vivo. *Proc Natl Acad Sci U S A* **116**, 1447-1456 (2019).

28. A. D. McLellan, E. Kampgen, Functions of myeloid and lymphoid dendritic cells. *Immunol Lett* **72**, 101-105 (2000).
29. S. Scala, A. Aiuti, In vivo dynamics of human hematopoietic stem cells: novel concepts and future directions. *Blood Adv* **3**, 1916-1924 (2019).
30. E. M. Pietras *et al.*, Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* **17**, 35-46 (2015).
31. F. Ye, W. Huang, G. Guo, Studying hematopoiesis using single-cell technologies. *J Hematol Oncol* **10**, 27 (2017).
32. B. J. Povinelli, A. Rodriguez-Meira, A. J. Mead, Single cell analysis of normal and leukemic hematopoiesis. *Mol Aspects Med* **59**, 85-94 (2018).
33. F. Paul *et al.*, Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163**, 1663-1677 (2015).
34. J. D. Buenrostro *et al.*, Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535-1548 e1516 (2018).
35. D. Karamitros *et al.*, Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol* **19**, 85-97 (2018).
36. B. Psaila *et al.*, Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biol* **17**, 83 (2016).
37. S. H. Orkin, Diversification of haematopoietic stem cells to specific lineages. *Nat Rev Genet* **1**, 57-64 (2000).
38. H. W. Detrich, 3rd *et al.*, Intraembryonic hematopoietic cell migration during vertebrate development. *Proc Natl Acad Sci U S A* **92**, 10713-10717 (1995).
39. F. Y. Tsai, S. H. Orkin, Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood* **89**, 3636-3643 (1997).
40. E. C. Liao *et al.*, SCL/Tal-1 transcription factor acts downstream of cloche to specify hematopoietic and vascular progenitors in zebrafish. *Genes Dev* **12**, 621-626 (1998).
41. A. J. Davidson, L. I. Zon, The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis. *Oncogene* **23**, 7233-7246 (2004).
42. E. J. Paik, L. I. Zon, Hematopoietic development in the zebrafish. *Int J Dev Biol* **54**, 1127-1137 (2010).
43. C. M. Bennett *et al.*, Myelopoiesis in the zebrafish, Danio rerio. *Blood* **98**, 643-651 (2001).
44. G. J. Lieschke *et al.*, Zebrafish SPI-1 (PU.1) marks a site of myeloid development independent of primitive erythropoiesis: implications for axial patterning. *Dev Biol* **246**, 274-295 (2002).
45. A. Kumar, S. S. D'Souza, A. S. Thakur, Understanding the Journey of Human Hematopoietic Stem Cell Development. *Stem Cells Int* **2019**, 2141475 (2019).
46. S. H. Orkin, L. I. Zon, Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644 (2008).
47. K. G. Guruharsha, M. W. Kankel, S. Artavanis-Tsakonas, The Notch signalling system: recent insights into the complexity of a conserved pathway. *Nat Rev Genet* **13**, 654-666 (2012).
48. W. Saelens, R. Cannoodt, Y. Saeys, A comprehensive evaluation of module detection methods for gene expression data. *Nat Commun* **9**, 1090 (2018).
49. E. Dzierzak, S. Philipsen, Erythropoiesis: development and differentiation. *Cold Spring Harb Perspect Med* **3**, a011601 (2013).
50. K. A. Lawson, J. J. Meneses, R. A. Pedersen, Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891-911 (1991).
51. S. J. Kinder *et al.*, The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development* **126**, 4691-4701 (1999).
52. M. Belaoussoff, S. M. Farrington, M. H. Baron, Hematopoietic induction and respecification of A-P identity by visceral endoderm signaling in the mouse embryo. *Development* **125**, 5009-5018 (1998).
53. P. D. F. Murray, J. T. Wilson, The development in vitro of the blood of the early chick embryo. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **111**, 497-521 (1932).

54. A. A. Maximow, Untersuchungen uber blut und bindegewebe 1. Die fruhesten entwicklungsstadien der blut- und binde- gewebszellan bein saugetierembryo, bis zum anfang der blutbilding unden leber. *Arch. Mikroskop. Anat.* **73**, 444-561 (1909).
55. C. S. Lin, S. K. Lim, V. D'Agati, F. Costantini, Differential effects of an erythropoietin receptor gene disruption on primitive and definitive erythropoiesis. *Genes Dev* **10**, 154-164 (1996).
56. J. Palis, S. Robertson, M. Kennedy, C. Wall, G. Keller, Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development* **126**, 5073-5084 (1999).
57. G. Migliaccio *et al.*, Human embryonic hemopoiesis. Kinetics of progenitors and precursors underlying the yolk sac----liver transition. *J Clin Invest* **78**, 51-60 (1986).
58. P. Kumaravelu *et al.*, Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. *Development* **129**, 4891-4899 (2002).
59. H. Ema, H. Nakauchi, Expansion of hematopoietic stem cells in the developing liver of a mouse embryo. *Blood* **95**, 2284-2288 (2000).
60. C. W. Kelemen E., Fliedner T. M. , Atlas of Human Hematopoietic Development. *Springer-Verlag*, (1979).
61. I. Blazsek, J. Chagraoui, B. Peault, Ontogenic emergence of the hematon, a morphogenetic stromal unit that supports multipotential hematopoietic progenitors in mouse bone marrow. *Blood* **96**, 3763-3771 (2000).
62. P. Carnot, Sur l'activite hematopoietique de differents organes au cours de la regeneration du sang. *Compt Rend Acad Sci* **143**, 432 (1906).
63. H. Wu, X. Liu, R. Jaenisch, H. F. Lodish, Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell* **83**, 59-67 (1995).
64. A. R. Migliaccio, S. Campisi, G. Migliaccio, Standardization of progenitor cell assay for cord blood banking. *Ann Ist Super Sanita* **37**, 595-600 (2001).
65. S. Elliott, Erythropoiesis-stimulating agents and other methods to enhance oxygen transport. *Br J Pharmacol* **154**, 529-541 (2008).
66. M. J. Koury, M. C. Bondurant, Erythropoietin retards DNA breakdown and prevents programmed death in erythroid progenitor cells. *Science* **248**, 378-381 (1990).
67. K. Sawada *et al.*, Purification of human erythroid colony-forming units and demonstration of specific binding of erythropoietin. *J Clin Invest* **80**, 357-366 (1987).
68. M. J. Koury, V. H. Haase, Anaemia in kidney disease: harnessing hypoxia responses for therapy. *Nat Rev Nephrol* **11**, 394-410 (2015).
69. L. Samsel, J. P. McCoy, Jr., Imaging flow cytometry for the study of erythroid cell biology and pathology. *J Immunol Methods* **423**, 52-59 (2015).
70. T. Kina *et al.*, The monoclonal antibody TER-119 recognizes a molecule associated with glycophorin A and specifically marks the late stages of murine erythroid lineage. *Br J Haematol* **109**, 280-287 (2000).
71. D. K. Marsee, G. S. Pinkus, H. Yu, CD71 (transferrin receptor): an effective marker for erythroid precursors in bone marrow biopsy specimens. *Am J Clin Pathol* **134**, 429-435 (2010).
72. J. Palis, Primitive and definitive erythropoiesis in mammals. *Front Physiol* **5**, 3 (2014).
73. P. A. Ney, Normal and disordered reticulocyte maturation. *Curr Opin Hematol* **18**, 152-157 (2011).
74. R. E. Waugh, A. Mantalaris, R. G. Bauserman, W. C. Hwang, J. H. Wu, Membrane instability in late-stage erythropoiesis. *Blood* **97**, 1869-1875 (2001).
75. J. A. Chasis, N. Mohandas, Erythroblastic islands: niches for erythropoiesis. *Blood* **112**, 470-478 (2008).
76. M. J. Koury, S. T. Sawyer, S. J. Brandt, New insights into erythropoiesis. *Curr Opin Hematol* **9**, 93-100 (2002).
77. J. Back, A. Dierich, C. Bronn, P. Kastner, S. Chan, PU.1 dtermines the self-renewal capacity of erythroid progenitor cells. *Blood* **103**, (2004).
78. E. Mancini *et al.*, FOG-1 and GATA-1 act sequentially to specify definitive megakaryocytic and erythroid progenitors. *EMBO J* **31**, 351-365 (2012).
79. M. Siatecka, J. J. Bieker, The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood* **118**, 2044-2054 (2011).

80. P. D. Kingsley, J. Malik, K. A. Fantauzzo, J. Palis, Yolk sac-derived primitive erythroblasts enucleate during mammalian embryogenesis. *Blood* **104**, 19-25 (2004).
81. H. Wu, U. Klingmuller, P. Besmer, H. F. Lodish, Interaction of the Erythropoietin and Stem-Cell-Factor Receptors. *Nature* **377**, 242-246 (1995).
82. A. Wickrema, J. D. Crispino, Erythroid and megakaryocytic transformation. *Oncogene* **26**, 6803-6815 (2007).
83. M. A. Kerenyi, S. H. Orkin, Networking erythropoiesis. *J Exp Med* **207**, 2537-2541 (2010).
84. T. A. Moreau-Gachelin F, Tambourin P. , Spi-1 is a putative oncogene in virally induced murine erythroleukaemias. *Nature* **331**, 277-280 (1988).
85. G. Rao, N. Rekhtman, G. Cheng, T. Krasikov, A. I. Skoultchi, Deregulated expression of the PU.1 transcription factor blocks murine erythroleukemia cell terminal differentiation. *Oncogene* **14**, 123-131 (1997).
86. N. Rekhtman, F. Radparvar, T. Evans, A. I. Skoultchi, Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev* **13**, 1398-1411 (1999).
87. C. Vicente, A. Conchillo, M. A. Garcia-Sanchez, M. D. Odero, The role of the GATA2 transcription factor in normal and malignant hematopoiesis. *Crit Rev Oncol Hematol* **82**, 1-17 (2012).
88. O. G. H. Fujiwara T, Keles S, Blahnik K, Linnemann AK, Kang YA, Choi K, Farnham PJ, Bresnick EH, Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell.* **36(4)**, 667-681. (2009).
89. J. A. Grass *et al.*, GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A* **100**, 8811-8816 (2003).
90. M. J. Weiss, C. Yu, S. H. Orkin, Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol* **17**, 1642-1651 (1997).
91. L. Pevny *et al.*, Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257-260 (1991).
92. Y. Fujiwara, C. P. Browne, K. Cunniff, S. C. Goff, S. H. Orkin, Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* **93**, 12355-12358 (1996).
93. L. Pevny *et al.*, Development of hematopoietic cells lacking transcription factor GATA-1. *Development* **121**, 163-172 (1995).
94. K. Kitajima, J. Zheng, H. Yen, D. Sugiyama, T. Nakano, Multipotential differentiation ability of GATA-1-null erythroid-committed cells. *Genes Dev* **20**, 654-659 (2006).
95. M. Suzuki, T. Moriguchi, K. Ohneda, M. Yamamoto, Differential contribution of the Gata1 gene hematopoietic enhancer to erythroid differentiation. *Mol Cell Biol* **29**, 1163-1175 (2009).
96. L. I. Zon, H. Youssoufian, C. Mather, H. F. Lodish, S. H. Orkin, Activation of the erythropoietin receptor promoter by transcription factor GATA-1. *Proc Natl Acad Sci U S A* **88**, 10638-10641 (1991).
97. R. M. Evans T, Felsenfeld G. , An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc Natl Acad Sci USA.* **85**, 5976–5980 (1988).
98. D. I. Martin, S. H. Orkin, Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev* **4**, 1886-1898 (1990).
99. M. Yu *et al.*, Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell* **36**, 682-695 (2009).
100. Y. Cheng *et al.*, Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res* **19**, 2172-2184 (2009).
101. H. K. Mikkola *et al.*, Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature* **421**, 547-551 (2003).
102. S. W. Porcher C, Rockwell K, Fujiwara Y, Alt FW, Orkin SH., The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86(1)**, 47-57 (1996).
103. L. H. Wang, N. E. Baker, E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease. *Dev Cell* **35**, 269-280 (2015).

104. A. B. Cantor, S. H. Orkin, Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* **21**, 3368-3376 (2002).
105. O. H. Wadman IA, Grütz GG, Agulnick AD, Westphal H, Forster A, Rabbitts TH, The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* **16(11)**, 3145-3157. (1997).
106. H. S. Lecuyer E, Saint-Denis M, Martin R, Begley CG, Porcher C, Orkin SH, Hoang T The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. *Blood* **100**, 2430–2440 (2002).
107. L. Q. Li *et al.*, Nuclear adaptor Ldb1 regulates a transcriptional program essential for the maintenance of hematopoietic stem cells. *Nature Immunology* **12**, 129-U138 (2011).
108. P. E. Love, C. Warzecha, L. Q. Li, Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends in Genetics* **30**, 1-9 (2014).
109. M. Crossley, A. P. Tsang, J. J. Bieker, S. H. Orkin, Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. *J Biol Chem* **269**, 15440-15444 (1994).
110. M. R. Tallack *et al.*, Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. *Genome Research* **22**, 2385-2398 (2012).
111. F. Bouilloux *et al.*, EKLF restricts megakaryocytic differentiation at the benefit of erythrocytic differentiation. *Blood* **112**, 576-584 (2008).
112. Miller IJ, B. JJ., A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Mol Cell Biol.* **13**, 2776-2786 (1993).
113. R. C. Taylor, S. P. Cullen, S. J. Martin, Apoptosis: controlled demolition at the cellular level. *Nat Rev Mol Cell Biol* **9**, 231-241 (2008).
114. R. De Maria *et al.*, Negative regulation of erythropoiesis by caspase-mediated cleavage of GATA-1. *Nature* **401**, 489-493 (1999).
115. J. A. Ribeil *et al.*, Hsp70 regulates erythropoiesis by preventing caspase-3-mediated cleavage of GATA-1. *Nature* **445**, 102-105 (2007).
116. H. Kaneko, R. Shimizu, M. Yamamoto, GATA factor switching during erythroid differentiation. *Curr Opin Hematol* **17**, 163-168 (2010).
117. P. Valent *et al.*, Normal and pathological erythropoiesis in adults: from gene regulation to targeted treatment concepts. *Haematologica* **103**, 1593-1603 (2018).
118. E. Soler *et al.*, The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* **24**, 277-289 (2010).
119. M.-M. I. Fischer MA, Hunt A, Chyla BJ, Hiebert SW. , Myeloid translocation gene 16 is required for maintenance of haematopoietic stem cell quiescence. *EMBO Journal* **31**, 1494–1505 (2012).
120. B. J. Chyla *et al.*, Deletion of Mtg16, a target of t(16;21), alters hematopoietic progenitor cell proliferation and lineage allocation. *Mol Cell Biol* **28**, 6234-6247 (2008).
121. T. Fujiwara *et al.*, Role of transcriptional corepressor ETO2 in erythroid cells. *Exp Hematol* **41**, 303-315 e301 (2013).
122. R. Stadhouders *et al.*, Control of developmentally primed erythroid genes by combinatorial co-repressor actions. *Nat Commun* **6**, 8893 (2015).
123. A. H. Schuh *et al.*, ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol Cell Biol* **25**, 10235-10250 (2005).
124. L. McGhee *et al.*, Gfi-1 attaches to the nuclear matrix, associates with ETO (MTG8) and histone deacetylase proteins, and represses transcription using a TSA-sensitive mechanism. *Journal of Cellular Biochemistry* **89**, 1005-1018 (2003).
125. N. Goardon *et al.*, ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *EMBO J* **25**, 357-366 (2006).
126. J. Lacombe *et al.*, Scl regulates the quiescence and the long-term competence of hematopoietic stem cells. *Blood* **115**, 792-803 (2010).
127. N. Steinauer, C. Guo, J. S. Zhang, Emerging Roles of MTG16 in Cell-Fate Control of Hematopoietic Stem Cells and Cancer. *Stem Cells International* **2017**, 12 pages (2017).
128. A. Zeuner *et al.*, The Notch2-Jagged1 interaction mediates stem cell factor signaling in erythropoiesis. *Cell Death Differ* **18**, 371-380 (2011).
129. L. Zhou *et al.*, Notch-dependent control of myelopoiesis is regulated by fucosylation. *Blood* **112**, 308-319 (2008).

130. W. L *et al.*, The Notch/Jagged Pathway Inhibits Proliferation of Human Hematopoietic Progenitors In Vitro. *Stem cells* **17**, 162-171 (2009).
131. H. T. Cheng X, Chen VC, Gadue P, Keller GM, Numb mediates the interaction between Wnt and Notch to modulate primitive erythropoietic specification from the hemangioblast. *Development* **135**, 3447-3458 (2008).
132. J. Enciso, L. Mendoza, R. Pelayo, Normal vs. Malignant hematopoiesis: the complexity of acute leukemia through systems biology. *Front Genet* **6**, 290 (2015).
133. J. R. Downing, K. M. Shannon, Acute leukemia: a pediatric perspective. *Cancer Cell* **2**, 437-445 (2002).
134. D. A. Arber *et al.*, The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391-2405 (2016).
135. A. S. Davis, A. J. Viera, M. D. Mead, Leukemia: an overview for primary care. *Am Fam Physician* **89**, 731-738 (2014).
136. E. Papaemmanuil *et al.*, Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* **374**, 2209-2221 (2016).
137. J. S. Welch, A. A. Petti, T. J. Ley, Decitabine in TP53-Mutated AML. *N Engl J Med* **376**, 797-798 (2017).
138. B. J. Huntly *et al.*, MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell* **6**, 587-596 (2004).
139. A. V. Krivtsov *et al.*, Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818-822 (2006).
140. A. Cozzio *et al.*, Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes Dev* **17**, 3029-3035 (2003).
141. V. Stavropoulou *et al.*, MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome. *Cancer Cell* **30**, 43-58 (2016).
142. J. E. Visvader, Cells of origin in cancer. *Nature* **469**, 314-322 (2011).
143. P. Riemke *et al.*, Myeloid leukemia with transdifferentiation plasticity developing from T-cell progenitors. *EMBO J* **35**, 2399-2416 (2016).
144. M. Chopra, S. K. Bohlander, The cell of origin and the leukemia stem cell in acute myeloid leukemia. *Genes Chromosomes Cancer* **58**, 850-858 (2019).
145. J. George *et al.*, Leukaemia cell of origin identified by chromatin landscape of bulk tumour cells. *Nat Commun* **7**, 12166 (2016).
146. P. B. Gupta, C. L. Chaffer, R. A. Weinberg, Cancer stem cells: mirage or reality? *Nat Med* **15**, 1010-1012 (2009).
147. X. Wang, S. Huang, J. L. Chen, Understanding of leukemic stem cells and their clinical implications. *Mol Cancer* **16**, 2 (2017).
148. A. Reinisch, S. M. Chan, D. Thomas, R. Majeti, Biology and Clinical Relevance of Acute Myeloid Leukemia Stem Cells. *Semin Hematol* **52**, 150-164 (2015).
149. D. Hanekamp, J. Cloos, G. J. Schuurhuis, Leukemic stem cells: identification and clinical application. *Int J Hematol* **105**, 549-557 (2017).
150. B. Deschler, M. Lubbert, Acute myeloid leukemia: epidemiology and etiology. *Cancer* **107**, 2099-2107 (2006).
151. A. VM, Acute myelogenous leukemia. *Curr Probl Pediatr Adolesc Health Care* **32**, 50–58. (2002).
152. W. L. C. Pui CG , Soheil Meshinchi , Robert J. Arceci, S. More, Biology, Risk Stratification, and Therapy of Pediatric Acute Leukemias: An Update. *J Clin Oncol* **29**, 551-565 (2011).
153. B. JM *et al.*, Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Co-operative Group. *Brisith Journal of Haematology* **33**, (1976).
154. I. Uribesalgo, L. Di Croce, Dynamics of epigenetic modifications in leukemia. *Brief Funct Genomics* **10**, 18-29 (2011).
155. J. W. Vardiman *et al.*, The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-951 (2009).
156. D. A. Arber, The 2016 WHO classification of acute myeloid leukemia: What the practicing clinician needs to know. *Semin Hematol* **56**, 90-95 (2019).
157. F. Mitelman, F. Mertens, B. Johansson, Prevalence estimates of recurrent balanced cytogenetic aberrations and gene fusions in unselected patients with neoplastic disorders. *Genes Chromosomes Cancer* **43**, 350-366 (2005).

158. F. Mitelman, B. Johansson, F. Mertens, The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* **7**, 233-245 (2007).
159. S. Frohling, C. Scholl, D. G. Gilliland, R. L. Levine, Genetics of myeloid malignancies: pathogenetic and clinical implications. *J Clin Oncol* **23**, 6285-6295 (2005).
160. T. S., Current findings for recurring mutations in acute myeloid leukemia. *J Hematol Oncol* **4**, (2011).
161. K. CC., Genetic abnormalities and challenges in the treatment of acute myeloid leukemia. *Genes Cancer* **2**, 95–107 (2011).
162. D. G. Gilliland, Molecular genetics of human leukemias: new insights into therapy. *Semin Hematol* **39**, 6-11 (2002).
163. E. Conway O'Brien, S. Prideaux, T. Chevassut, The epigenetic landscape of acute myeloid leukemia. *Adv Hematol* **2014**, 103175 (2014).
164. C. Riether, C. M. Schurch, A. F. Ochsenbein, Regulation of hematopoietic and leukemic stem cells by the immune system. *Cell Death Differ* **22**, 187-198 (2015).
165. K. Tomczak, P. Czerwinska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* **19**, A68-77 (2015).
166. S. C. Meyer, R. L. Levine, Translational implications of somatic genomics in acute myeloid leukaemia. *Lancet Oncol* **15**, e382-394 (2014).
167. N. Cancer Genome Atlas Research *et al.*, Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059-2074 (2013).
168. J. H. Martens, H. G. Stunnenberg, The molecular signature of oncofusion proteins in acute myeloid leukemia. *FEBS Lett* **584**, 2662-2669 (2010).
169. H. Bolouri *et al.*, The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* **24**, 103-112 (2018).
170. K. E. Elagib, A. N. Goldfarb, Oncogenic pathways of AML1-ETO in acute myeloid leukemia: multifaceted manipulation of marrow maturation. *Cancer Lett* **251**, 179-186 (2007).
171. A. V. Krivtsov, S. A. Armstrong, MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* **7**, 823-833 (2007).
172. T. Katsumoto, N. Yoshida, I. Kitabayashi, Roles of the histone acetyltransferase monocytic leukemia zinc finger protein in normal and malignant hematopoiesis. *Cancer Sci* **99**, 1523-1527 (2008).
173. I. Kitabayashi, Y. Aikawa, L. A. Nguyen, A. Yokoyama, M. Ohki, Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein. *EMBO J* **20**, 7184-7196 (2001).
174. M. E. Figueroa *et al.*, DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13-27 (2010).
175. R. Villa *et al.*, Epigenetic gene silencing in acute promyelocytic leukemia. *Biochem Pharmacol* **68**, 1247-1254 (2004).
176. L. Di Croce *et al.*, Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. *Science* **295**, 1079-1082 (2002).
177. V. Gelmetti *et al.*, Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. *Mol Cell Biol* **18**, 7185-7191 (1998).
178. A. T. Thiel *et al.*, MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele. *Cancer Cell* **17**, 148-159 (2010).
179. H. Dombret, C. Gardin, An update of current treatments for adult acute myeloid leukemia. *Blood* **127**, 53-61 (2016).
180. C. D. DiNardo, J. E. Cortes, Mutations in AML: prognostic and therapeutic implications. *Hematology Am Soc Hematol Educ Program* **2016**, 348-355 (2016).
181. N. J. Short, H. Kantarjian, F. Ravandi, N. Daver, Emerging treatment paradigms with FLT3 inhibitors in acute myeloid leukemia. *Ther Adv Hematol* **10**, 2040620719827310 (2019).
182. N. Daver, R. F. Schlenk, N. H. Russell, M. J. Levis, Targeting FLT3 mutations in AML: review of current knowledge and evidence. *Leukemia* **33**, 299-312 (2019).
183. M. Levis, Midostaurin approved for FLT3-mutated AML. *Blood* **129**, 3403-3406 (2017).
184. N. R. Mathew *et al.*, Sorafenib promotes graft-versus-leukemia activity in mice and humans through IL-15 production in FLT3-ITD-mutant leukemia cells. *Nat Med* **24**, 282-291 (2018).
185. A. Antar *et al.*, Inhibition of FLT3 in AML: a focus on sorafenib. *Bone Marrow Transplant* **52**, 344-351 (2017).

186. A. E. Perl *et al.*, Gilteritinib or Chemotherapy for Relapsed or Refractory FLT3-Mutated AML. *N Engl J Med* **381**, 1728-1740 (2019).
187. C. M. McMahon *et al.*, Clonal Selection with RAS Pathway Activation Mediates Secondary Clinical Resistance to Selective FLT3 Inhibition in Acute Myeloid Leukemia. *Cancer Discov* **9**, 1050-1063 (2019).
188. M. A. Sanz *et al.*, A modified AIDA protocol with anthracycline-based consolidation results in high antileukemic efficacy and reduced toxicity in newly diagnosed PML/RARalpha-positive acute promyelocytic leukemia. PETHEMA group. *Blood* **94**, 3015-3021 (1999).
189. F. Lo-Coco *et al.*, Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *N Engl J Med* **369**, 111-121 (2013).
190. A. K. Burnett *et al.*, Arsenic trioxide and all-trans retinoic acid treatment for acute promyelocytic leukaemia in all risk groups (AML17): results of a randomised, controlled, phase 3 trial. *Lancet Oncol* **16**, 1295-1305 (2015).
191. H. de Thé, Lessons taught by acute promyelocytic leukemia cure. *Lancet* **386**, 247-248 (2015).
192. P. JT., Polycythemia vera and other primary polycythemias. *Curr Opin Hematol.* **12**, 112-116 (2005).
193. W. S. Wang W, Medeiros LJ, Khoury JD., Pure erythroid leukemia. *Am J Hematol.* **92**, 292-296 (2017).
194. D. A. Arber, R. P. Hasserjian, Reclassifying myelodysplastic syndromes: what's where in the new WHO and why. *Hematology Am Soc Hematol Educ Program* **2015**, 294-298 (2015).
195. W. Liu *et al.*, Pure erythroid leukemia: a reassessment of the entity using the 2008 World Health Organization classification. *Mod Pathol* **24**, 375-383 (2011).
196. S. O. Schwartz, J. Critchlow, Erythremic myelosis (Dl Guglielmo's disease); critical review with report of four cases, and comments on erythroleukemia. *Blood* **7**, 765-793 (1952).
197. B. J. Bain, Di Guglielmo and his syndromes. *Br J Haematol* **120**, 939-943 (2003).
198. J. L. Villeval *et al.*, Phenotype of early erythroblastic leukemias. *Blood* **68**, 1167-1174 (1986).
199. N. L. Harris *et al.*, The World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues. Report of the Clinical Advisory Committee meeting, Airlie House, Virginia, November, 1997. *Ann Oncol* **10**, 1419-1432 (1999).
200. S. Park *et al.*, Erythroleukemia: a comparison between the previous FAB approach and the WHO classification. *Leuk Res* **26**, 423-429 (2002).
201. P. Boddu *et al.*, Erythroleukemia-historical perspectives and recent advances in diagnosis and management. *Blood Rev* **32**, 96-105 (2018).
202. J. Park *et al.*, Mutation profiling of mismatch repair-deficient colorectal cncers using an in silico genome scan to identify coding microsatellites. *Cancer Res* **62**, 1284-1288 (2002).
203. M. Lessard *et al.*, Cytogenetic study of 75 erythroleukemias. *Cancer Genet Cytogenet* **163**, 113-122 (2005).
204. R. S. Ohgami, K. M. Chisholm, L. Ma, D. A. Arber, E-cadherin is a specific marker for erythroid differentiation and has utility, in combination with CD117 and CD34, for enumerating myeloblasts in hematopoietic neoplasms. *Am J Clin Pathol* **141**, 656-664 (2014).
205. J. M. Coffin *et al.*, Proposal for naming host cell-derived inserts in retrovirus genomes. *J Virol* **40**, 953-957 (1981).
206. B. Vennstrom, J. M. Bishop, Isolation and characterization of chicken DNA homologous to the two putative oncogenes of avian erythroblastosis virus. *Cell* **28**, 135-143 (1982).
207. M. von Lindern *et al.*, Leukemic transformation of normal murine erythroid progenitors: v- and c-ErbB act through signaling pathways activated by the EpoR and c-Kit in stress erythropoiesis. *Oncogene* **20**, 3651-3664 (2001).
208. J. Downward *et al.*, Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. *Nature* **307**, 521-527 (1984).
209. A. S. Tsiftsoglou, I. S. Vizirianakis, J. Strouboulis, Erythropoiesis: model systems, molecular regulators, and developmental programs. *IUBMB Life* **61**, 800-830 (2009).
210. C. Friend, W. Scher, J. G. Holland, T. Sato, Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethyl sulfoxide. *Proc Natl Acad Sci U S A* **68**, 378-382 (1971).
211. T. Yamada *et al.*, Overexpression of PU.1 induces growth and differentiation inhibition and apoptotic cell death in murine erythroleukemia cells. *Blood* **89**, 1383-1393 (1997).

212. O. Kosmider *et al.*, Kit-activating mutations cooperate with Spi-1/PU.1 overexpression to promote tumorigenic progression during erythroleukemia in mice. *Cancer Cell* **8**, 467-478 (2005).
213. H. G. Drexler, Y. Matsuo, R. A. MacLeod, Malignant hematopoietic cell lines: in vitro models for the study of erythroleukemia. *Leuk Res* **28**, 1243-1251 (2004).
214. R. Rampal *et al.*, Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. *Blood* **123**, e123-133 (2014).
215. T. Tsuruta-Kishino *et al.*, Loss of p53 induces leukemic transformation in a murine model of Jak2 V617F-driven polycythemia vera. *Oncogene* **36**, 3300-3311 (2017).
216. R. Shimizu *et al.*, Leukemogenesis caused by incapacitated GATA-1 function. *Mol Cell Biol* **24**, 10814-10825 (2004).
217. R. Shimizu, J. D. Engel, M. Yamamoto, GATA1-related leukaemias. *Nat Rev Cancer* **8**, 279-287 (2008).
218. S. S. Clark *et al.*, Expression of a distinctive BCR-ABL oncogene in Ph1-positive acute lymphocytic leukemia (ALL). *Science* **239**, 775-777 (1988).
219. K. Wagner *et al.*, Absence of the transcription factor CCAAT enhancer binding protein alpha results in loss of myeloid identity in bcr/abl-induced malignancy. *Proc Natl Acad Sci U S A* **103**, 6338-6343 (2006).
220. A. J. McGahon *et al.*, Downregulation of Bcr-Abl in K562 cells restores susceptibility to apoptosis: characterization of the apoptotic death. *Cell Death Differ* **4**, 95-104 (1997).
221. P. Burda *et al.*, GATA-1 Inhibits PU.1 Gene via DNA and Histone H3K9 Methylation of Its Distal Enhancer in Erythroleukemia. *PLoS One* **11**, e0152234 (2016).
222. Y. Ben-David, E. B. Giddens, K. Letwin, A. Bernstein, Erythroleukemia induction by Friend murine leukemia virus: insertional activation of a new member of the ets gene family, Fli-1, closely linked to c-ets-1. *Genes Dev* **5**, 908-918 (1991).
223. A. H. Elmaagacli *et al.*, Growth factor-independent 1B gene (GFI1B) is overexpressed in erythropoietic and megakaryocytic malignancies and increases their proliferation rate. *Br J Haematol* **136**, 212-219 (2007).
224. R. Yamamoto *et al.*, Selective dissociation between LSD1 and GFI1B by a LSD1 inhibitor NCD38 induces the activation of ERG super-enhancer in erythroleukemia cells. *Oncotarget* **9**, 21007-21021 (2018).
225. G. Tatsumi *et al.*, LSD1-mediated repression of GFI1 super-enhancer plays an essential role in erythroleukemia. *Leukemia*, (2019).
226. J. Dittmer, A. Nordheim, Ets transcription factors and human disease. *Biochim Biophys Acta* **1377**, F1-11 (1998).
227. J. W. Cui, L. M. Vecchiarelli-Federico, Y. J. Li, G. J. Wang, Y. Ben-David, Continuous Fli-1 expression plays an essential role in the proliferation and survival of F-MuLV-induced erythroleukemia and human erythroleukemia. *Leukemia* **23**, 1311-1319 (2009).
228. A. Tamir *et al.*, Fli-1, an Ets-related transcription factor, regulates erythropoietin-induced erythroid proliferation and differentiation: evidence for direct transcriptional repression of the Rb gene during differentiation. *Mol Cell Biol* **19**, 4452-4464 (1999).
229. I. Lesault, C. T. Quang, J. Frampton, J. Ghysdael, Direct regulation of BCL-2 by FLI-1 is involved in the survival of FLI-1-transformed erythroblasts. *EMBO J* **21**, 694-703 (2002).
230. M. Athanasiou, G. Mavrothalassitis, L. Sun-Hoffman, D. G. Blair, FLI-1 is a suppressor of erythroid differentiation in human hematopoietic cells. *Leukemia* **14**, 439-445 (2000).
231. G. K. Lakhanpal *et al.*, The inositol phosphatase SHIP-1 is negatively regulated by Fli-1 and its loss accelerates leukemogenesis. *Blood* **116**, 428-436 (2010).
232. L. M. Vecchiarelli-Federico *et al.*, Fli-1 overexpression in erythroleukemic cells promotes erythroid de-differentiation while Spi-1/PU.1 exerts the opposite effect. *Int J Oncol* **51**, 456-466 (2017).
233. A. P. Ng *et al.*, Erg is required for self-renewal of hematopoietic stem cells during stress hematopoiesis in mice. *Blood* **118**, 2454-2461 (2011).
234. S. J. Loughran *et al.*, The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat Immunol* **9**, 810-819 (2008).
235. G. Marcucci *et al.*, High expression levels of the ETS-related gene, ERG, predict adverse outcome and improve molecular risk-based classification of cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B Study. *J Clin Oncol* **25**, 3337-3343 (2007).

236. C. L. Carmichael *et al.*, Hematopoietic overexpression of the transcription factor Erg induces lymphoid and erythro-megakaryocytic leukemia. *Proc Natl Acad Sci U S A* **109**, 15437-15442 (2012).
237. V. Grossmann *et al.*, Acute erythroid leukemia (AEL) can be separated into distinct prognostic subsets based on cytogenetic and molecular genetic characteristics. *Leukemia* **27**, 1940-1943 (2013).
238. D. Rose *et al.*, Subtype-specific patterns of molecular mutations in acute myeloid leukemia. *Leukemia* **31**, 11-17 (2017).
239. N. Ping *et al.*, Exome sequencing identifies highly recurrent somatic GATA2 and CEBPA mutations in acute erythroid leukemia. *Leukemia* **31**, 195-202 (2017).
240. N. Cervera *et al.*, Revisiting gene mutations and prognosis of ex-M6a-acute erythroid leukemia with regard to the new WHO classification. *Blood Cancer J* **7**, e594 (2017).
241. G. Montalban-Bravo *et al.*, More than 1 TP53 abnormality is a dominant characteristic of pure erythroid leukemia. *Blood* **129**, 2584-2587 (2017).
242. I. Iacobucci *et al.*, Genomic subtyping and therapeutic targeting of acute erythroleukemia. *Nat Genet* **51**, 694-704 (2019).
243. M. Prokocimer, A. Molchadsky, V. Rotter, Dysfunctional diversity of p53 proteins in adult acute myeloid leukemia: projections on diagnostic workup and therapy. *Blood* **130**, 699-712 (2017).
244. J. C. Cigudosa *et al.*, De novo erythroleukemia chromosome features include multiple rearrangements, with special involvement of chromosomes 11 and 19. *Genes Chromosomes Cancer* **36**, 406-412 (2003).
245. F. P. Santos *et al.*, Adult acute erythroleukemia: an analysis of 91 patients treated at a single institution. *Leukemia* **23**, 2275-2280 (2009).
246. R. P. Hasserjian *et al.*, Acute erythroid leukemia: a reassessment using criteria refined in the 2008 WHO classification. *Blood* **115**, 1985-1992 (2010).
247. A. N. de Almeida *et al.*, Clinical Outcome, Tumor Recurrence, and Causes of Death: A Long-Term Follow-Up of Surgically Treated Meningiomas. *World Neurosurg* **102**, 139-143 (2017).
248. C. Quelen *et al.*, Identification of a transforming MYB-GATA1 fusion gene in acute basophilic leukemia: a new entity in male infants. *Blood* **117**, 5719-5722 (2011).
249. F. Micci *et al.*, Translocation t(1;16)(p31;q24) rearranging CBFA2T3 is specific for acute erythroid leukemia. *Leukemia* **25**, 1510-1512 (2011).
250. F. Micci *et al.*, High-throughput sequencing identifies an NFIA/CBFA2T3 fusion gene in acute erythroid leukemia with t(1;16)(p31;q24). *Leukemia* **27**, 980-982 (2013).
251. H. Liu *et al.*, De novo primary central nervous system pure erythroid leukemia/sarcoma with t(1;16)(p31;q24) NFIA/CBFA2T3 translocation. *Haematologica*, (2020).
252. I. Panagopoulos *et al.*, Fusion of ZMYND8 and RELA genes in acute erythroid leukemia. *PLoS One* **8**, e63663 (2013).
253. V. L. Castaneda, R. T. Parmley, V. A. Saldivar, M. S. Cheah, Childhood undifferentiated leukemia with early erythroid markers and c-myb duplication. *Leukemia* **5**, 142-149 (1991).
254. U. Koller *et al.*, Phenotypic and genotypic heterogeneity in infant acute leukemia. II. Acute nonlymphoblastic leukemia. *Leukemia* **3**, 708-714 (1989).
255. L. M. Starnes *et al.*, NFI-A directs the fate of hematopoietic progenitors to the erythroid or granulocytic lineage and controls beta-globin and G-CSF receptor expression. *Blood* **114**, 1753-1763 (2009).
256. K. S. Chen, J. W. C. Lim, L. J. Richards, J. Bunt, The convergent roles of the nuclear factor I transcription factors in development and cancer. *Cancer Lett* **410**, 124-138 (2017).
257. M. Fane, L. Harris, A. G. Smith, M. Piper, Nuclear factor one transcription factors as epigenetic regulators in cancer. *Int J Cancer* **140**, 2634-2641 (2017).
258. G. Zardo *et al.*, Polycombs and microRNA-223 regulate human granulopoiesis by transcriptional control of target gene expression. *Blood* **119**, 4034-4046 (2012).
259. A. Rosa *et al.*, The interplay between the master transcription factor PU.1 and miR-424 regulates human monocyte/macrophage differentiation. *Proc Natl Acad Sci U S A* **104**, 19849-19854 (2007).
260. L. M. Starnes *et al.*, A transcriptome-wide approach reveals the key contribution of NFI-A in promoting erythroid differentiation of human CD34(+) progenitors and CML cells. *Leukemia* **24**, 1220-1223 (2010).
261. N. Steinauer, C. Guo, J. Zhang, Emerging Roles of MTG16 in Cell-Fate Control of Hematopoietic Stem Cells and Cancer. *Stem Cells Int* **2017**, 6301385 (2017).

262. A. T. Hoogeveen *et al.*, The transcriptional corepressor MTG16a contains a novel nucleolar targeting sequence deranged in t (16; 21)-positive myeloid malignancies. *Oncogene* **21**, 6703-6712 (2002).
263. S. R. Lindberg, A. Olsson, A. M. Persson, I. Olsson, The Leukemia-associated ETO homologues are differently expressed during hematopoietic differentiation. *Experimental Hematology* **33**, 189-198 (2005).
264. J. N. Davis, L. McGhee, S. Meyers, The ETO (MTG8) gene family. *Gene* **303**, 1-10 (2003).
265. J. S. Zhang, M. Kalkum, S. Yamamura, B. T. Chait, R. G. Roeder, E protein silencing by the leukemogenic AML1-ETO fusion protein. *Science* **305**, 1286-1289 (2004).
266. I. Kitabayashi *et al.*, The AML1-MTG8 leukemic fusion protein forms a complex with a novel member of the MTG8(ETO/CDR) family, MTGR1. *Molecular and Cellular Biology* **18**, 846-858 (1998).
267. J. N. Davis, B. J. Williams, J. T. Herron, F. J. Galiano, S. Meyers, ETO-2, a new member of the ETO-family of nuclear proteins. *Oncogene* **18**, 1375-1383 (1999).
268. J. S. Zhang *et al.*, Oligomerization of ETO is obligatory for corepressor interaction. *Molecular and Cellular Biology* **21**, 156-163 (2001).
269. M. E. Engel, H. N. Nguyen, J. Mariotti, A. Hunt, S. W. Hiebert, Myeloid translocation gene 16 (MTG16) interacts with Notch transcription complex components to integrate Notch signaling in hematopoietic cell fate specification. *Mol Cell Biol* **30**, 1852-1863 (2010).
270. T. H. J. Wang, R. L. Redner, S. Kajigaya, and J. M. Liu, ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human NCoR/mSin3/HDAC1 complex. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10860–10865 (1998).
271. T. Reya, H. Clevers, Wnt signalling in stem cells and cancer. *Nature* **434**, 843-850 (2005).
272. B. T. MacDonald, K. Tamai, X. He, Wnt/beta-Catenin Signaling: Components, Mechanisms, and Diseases. *Developmental Cell* **17**, 9-26 (2009).
273. J. DeGregori, G. Leone, A. Miron, L. Jakoi, J. R. Nevins, Distinct roles for E2F proteins in cell growth control and apoptosis. *Proc Natl Acad Sci U S A* **94**, 7245-7250 (1997).
274. T. Gamou *et al.*, The partner gene of AML1 in t(16;21) myeloid malignancies is a novel member of the MTG8(ETO) family. *Blood* **91**, 4028-4037 (1998).
275. E. De Braekeleer *et al.*, RUNX1-MTG16 fusion gene in acute myeloblastic leukemia with t(16;21)(q24;q22): case report and review of the literature. *Cancer Genet Cytogenet* **185**, 47-50 (2008).
276. T. A. Gruber *et al.*, An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell* **22**, 683-697 (2012).
277. C. Thirant *et al.*, ETO2-GLIS2 Hijacks Transcriptional Complexes to Drive Cellular Identity and Self-Renewal in Pediatric Acute Megakaryoblastic Leukemia. *Cancer Cell* **31**, 452-465 (2017).
278. C. Wichmann *et al.*, Targeting the oligomerization domain of ETO interferes with RUNX1/ETO oncogenic activity in t(8;21)-positive leukemic cells. *Cancer Res* **67**, 2280-2289 (2007).
279. N. Li *et al.*, ZMYND8 Reads the Dual Histone Mark H3K4me1-H3K14ac to Antagonize the Expression of Metastasis-Linked Genes. *Mol Cell* **63**, 470-484 (2016).
280. K. Ghosh *et al.*, Positive Regulation of Transcription by Human ZMYND8 through Its Association with P-TEFb Complex. *Cell Rep* **24**, 2141-2154 e2146 (2018).
281. S. C. Fossey *et al.*, Identification and characterization of PRKCBP1, a candidate RACK-like protein. *Mamm Genome* **11**, 919-925 (2000).
282. S. Eichmuller *et al.*, Serological detection of cutaneous T-cell lymphoma-associated antigens. *Proc Natl Acad Sci U S A* **98**, 629-634 (2001).
283. F. Gong, T. Clouaire, M. Aguirrebengoa, G. Legube, K. M. Miller, Histone demethylase KDM5A regulates the ZMYND8-NuRD chromatin remodeler to promote DNA repair. *J Cell Biol* **216**, 1959-1974 (2017).
284. H. Yu *et al.*, Integrative genomic and transcriptomic analysis for pinpointing recurrent alterations of plant homeodomain genes and their clinical significance in breast cancer. *Oncotarget* **8**, 13099-13115 (2017).
285. Y. Chen *et al.*, ZMYND8 acetylation mediates HIF-dependent breast cancer progression and metastasis. *J Clin Invest* **128**, 1937-1955 (2018).
286. S. J. Stein, A. S. Baldwin, Deletion of the NF-kappaB subunit p65/RelA in the hematopoietic compartment leads to defects in hematopoietic stem cell function. *Blood* **121**, 5015-5024 (2013).

287. M. Y. Zhang, S. C. Sun, L. Bell, B. A. Miller, NF-kappaB transcription factors are involved in normal erythropoiesis. *Blood* **91**, 4136-4144 (1998).
288. J. Zhang, L. Li, A. S. Baldwin, Jr., A. D. Friedman, I. Paz-Priel, Loss of IKKbeta but Not NF-kappaB p65 Skews Differentiation towards Myeloid over Erythroid Commitment and Increases Myeloid Progenitor Self-Renewal and Functional Long-Term Hematopoietic Stem Cells. *PLoS One* **10**, e0130441 (2015).
289. M. Grossmann *et al.*, The combined absence of the transcription factors Rel and RelA leads to multiple hemopoietic cell defects. *Proc Natl Acad Sci U S A* **96**, 11848-11853 (1999).
290. J. J. Liu, S. C. Hou, C. K. Shen, Erythroid gene suppression by NF-kappa B. *J Biol Chem* **278**, 19534-19540 (2003).
291. C. E. Bueso-Ramos *et al.*, Expression of constitutively active nuclear-kappa B RelA transcription factor in blasts of acute myeloid leukemia. *Hum Pathol* **35**, 246-253 (2004).
292. Z. Estrov *et al.*, Phenylarsine oxide blocks interleukin-1beta-induced activation of the nuclear transcription factor NF-kappaB, inhibits proliferation, and induces apoptosis of acute myelogenous leukemia cells. *Blood* **94**, 2844-2853 (1999).
293. J. Zhou, Y. Q. Ching, W. J. Chng, Aberrant nuclear factor-kappa B activity in acute myeloid leukemia: from molecular pathogenesis to therapeutic target. *Oncotarget* **6**, 5490-5500 (2015).
294. K. M. McNagny, T. Graf, E26 leukemia virus converts primitive erythroid cells into cycling multilineage progenitors. *Blood* **101**, 1103-1110 (2003).
295. F. Moreau-Gachelin, Lessons from models of murine erythroleukemia to acute myeloid leukemia (AML): proof-of-principle of co-operativity in AML. *Haematologica* **91**, 1644-1652 (2006).
296. P. Zhang *et al.*, PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood* **96**, 2641-2648 (2000).
297. S. Coulon *et al.*, Human erythroleukemia: is the two-hit model of mouse leukemogenesis valid in human disease? *Leukemia* **21**, 2212-2214 (2007).
298. N. Cervera *et al.*, Molecular characterization of acute erythroid leukemia (M6-AML) using targeted next-generation sequencing. *Leukemia* **30**, 966-970 (2016).
299. N. Cervera *et al.*, Acute erythroid leukemias have a distinct molecular hierarchy from non-erythroid acute myeloid leukemias. *Haematologica*, (2019).
300. A. S. Tsiftsoglou, I. S. Pappas, I. S. Vizirianakis, Mechanisms involved in the induced differentiation of leukemia cells. *Pharmacol Ther* **100**, 257-290 (2003).
301. S. J. England, K. E. McGrath, J. M. Frame, J. Palis, Immature erythroblasts with extensive ex vivo self-renewal capacity emerge from the early mammalian fetus. *Blood* **117**, 2708-2717 (2011).
302. A. Fagnan *et al.*, Human erythroleukemia genetics and transcriptomes identify master transcription factors as functional disease drivers. *Blood* **136**, 698-714 (2020).
303. E. M. Alexandrova *et al.*, Improving survival by exploiting tumour dependence on stabilized mutant p53 for treatment. *Nature* **523**, 352-356 (2015).
304. E. M. Alexandrova *et al.*, p53 loss-of-heterozygosity is a necessary prerequisite for mutant p53 stabilization and gain-of-function in vivo. *Cell Death Dis* **8**, e2661 (2017).
305. B. J. Aubrey, G. L. Kelly, A. Janic, M. J. Herold, A. Strasser, How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death Differ* **25**, 104-113 (2018).
306. W. W. Zheng *et al.*, EDAG positively regulates erythroid differentiation and modifies GATA1 acetylation through recruiting p300. *Stem Cells* **32**, 2278-2289 (2014).
307. Y. Li, H. Luo, T. Liu, E. Zacksenhaus, Y. Ben-David, The ets transcription factor Fli-1 in development, cancer and disease. *Oncogene* **34**, 2022-2031 (2015).
308. A. Hoelbl *et al.*, Stat5 is indispensable for the maintenance of bcr/abl-positive leukaemia. *EMBO Mol Med* **2**, 98-110 (2010).
309. D. Yan, R. E. Hutchison, G. Mohi, Critical requirement for Stat5 in a mouse model of polycythemia vera. *Blood* **119**, 3539-3549 (2012).
310. G. Yoshimoto *et al.*, FLT3-ITD up-regulates MCL-1 to promote survival of stem cells in acute myeloid leukemia via FLT3-ITD-specific STAT5 activation. *Blood* **114**, 5034-5043 (2009).
311. M. Maicas *et al.*, Functional characterization of the promoter region of the human EVI1 gene in acute myeloid leukemia: RUNX1 and ELK1 directly regulate its transcription. *Oncogene* **32**, 2069-2078 (2013).
312. R. Nilsson *et al.*, Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. *Cell Metab* **10**, 119-130 (2009).

313. A. C. Azim *et al.*, Human erythrocyte dematin and protein 4.2 (pallidin) are ATP binding proteins. *Biochemistry* **35**, 3001-3006 (1996).
314. S. P. Whitman *et al.*, FLT3 internal tandem duplication associates with adverse outcome and gene- and microRNA-expression signatures in patients 60 years of age or older with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood* **116**, 3622-3626 (2010).
315. M. J. Stankiewicz, J. D. Crispino, ETS2 and ERG promote megakaryopoiesis and synergize with alterations in GATA-1 to immortalize hematopoietic progenitor cells. *Blood* **113**, 3337-3347 (2009).
316. D. R. Pattabiraman, T. J. Gonda, Role and potential for therapeutic targeting of MYB in leukemia. *Leukemia* **27**, 269-277 (2013).
317. R. Endersby *et al.*, Hls5 regulated erythroid differentiation by modulating GATA-1 activity. *Blood* **111**, 1946-1950 (2008).
318. C. Tanikawa, C. Ri, V. Kumar, Y. Nakamura, K. Matsuda, Crosstalk of EDA-A2/XEDAR in the p53 signaling pathway. *Mol Cancer Res* **8**, 855-863 (2010).
319. T. Kawase *et al.*, PH domain-only protein PHLDA3 is a p53-regulated repressor of Akt. *Cell* **136**, 535-550 (2009).
320. F. Valenti *et al.*, Mutant p53 oncogenic functions are sustained by Plk2 kinase through an autoregulatory feedback loop. *Cell Cycle* **10**, 4330-4340 (2011).
321. Y. Zhang *et al.*, ZNF365 promotes stability of fragile sites and telomeres. *Cancer Discov* **3**, 798-811 (2013).
322. R. Quintens *et al.*, Identification of novel radiation-induced p53-dependent transcripts extensively regulated during mouse brain development. *Biol Open* **4**, 331-344 (2015).
323. X. Chen *et al.*, Construction and validation of the CRISPR/dCas9-EZH2 system for targeted H3K27Me3 modification. *Biochem Biophys Res Commun* **511**, 246-252 (2019).
324. Z. Wang *et al.*, BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. *Bioinformatics* **34**, 2867-2869 (2018).
325. X. J. Sun *et al.*, A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis. *Nature* **500**, 93-97 (2013).
326. T. Matsunaga *et al.*, Ex vivo large-scale generation of human platelets from cord blood CD34+ cells. *Stem Cells* **24**, 2877-2887 (2006).
327. E. Supper, S. Tahir, T. Imai, J. Inoue, N. Minato, Modification of Gene Expression, Proliferation, and Function of OP9 Stroma Cells by Bcr-Abl-Expressing Leukemia Cells. *PLoS One* **10**, e0134026 (2015).
328. C. Kwok, B. B. Zeisig, J. Qiu, S. Dong, C. W. So, Transforming activity of AML1-ETO is independent of CBFbeta and ETO interaction but requires formation of homo-oligomeric complexes. *Proc Natl Acad Sci U S A* **106**, 2853-2858 (2009).
329. H. Glimm, I. H. Oh, C. J. Eaves, Human hematopoietic stem cells stimulated to proliferate in vitro lose engraftment potential during their S/G(2)/M transit and do not reenter G(0). *Blood* **96**, 4185-4193 (2000).
330. M. Higuchi *et al.*, Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. *Cancer Cell* **1**, 63-74 (2002).
331. J. Deng *et al.*, Bruton's tyrosine kinase inhibition increases BCL-2 dependence and enhances sensitivity to venetoclax in chronic lymphocytic leukemia. *Leukemia* **31**, 2075-2084 (2017).
332. C. K. Lopez *et al.*, Ontogenic Changes in Hematopoietic Hierarchy Determine Pediatric Specificity and Disease Phenotype in Fusion Oncogene-Driven Myeloid Leukemia. *Cancer Discov* **9**, 1736-1753 (2019).
333. M. K. Foecking, H. Hofstetter, Powerful and versatile enhancer-promoter unit for mammalian expression vectors. *Gene* **45**, 101-105 (1986).
334. L. Damdindorj *et al.*, A comparative analysis of constitutive promoters located in adeno-associated viral vectors. *PLoS One* **9**, e106472 (2014).
335. S. Nagel *et al.*, Analysis of the nuclear distribution of the translocation t(8;21)-derived fusion protein AML1/ETO by confocal laser scanning microscopy. *J Hematother Stem Cell Res* **11**, 401-408 (2002).
336. A. Grande, C. Costas, J. Benavente, Subunit composition and conformational stability of the oligomeric form of the avian reovirus cell-attachment protein sigmaC. *J Gen Virol* **83**, 131-139 (2002).

337. P. J. Kolodziejski, M. B. Rashid, N. T. Eissa, Intracellular formation of "undisruptable" dimers of inducible nitric oxide synthase. *Proc Natl Acad Sci U S A* **100**, 14263-14268 (2003).
338. L. Wang *et al.*, The Leukemogenicity of AML1-ETO Is Dependent on Site-Specific Lysine Acetylation. *Science* **333**, 765 (2011).
339. D. Hildebrand, J. Tiefenbach, T. Heinzel, M. Grez, A. B. Maurer, Multiple regions of ETO cooperate in transcriptional repression. *J Biol Chem* **276**, 9889-9895 (2001).
340. S. Park *et al.*, Structure of the AML1-ETO eTAFH domain-HEB peptide complex and its contribution to AML1-ETO activity. *Blood* **113**, 3558-3567 (2009).
341. J. Li, C. Guo, N. Steinauer, J. Zhang, New insights into transcriptional and leukemogenic mechanisms of AML1-ETO and E2A fusion proteins. *Frontiers in biology* **11**, 285-304 (2016).
342. Y. Liu *et al.*, The tetramer structure of the Nervy homology two domain, NHR2, is critical for AML1/ETO's activity. *Cancer Cell* **9**, 249-260 (2006).
343. B. A. Hug, S. Y. Lee, E. L. Kinsler, J. Zhang, M. A. Lazar, Cooperative function of Aml1-ETO corepressor recruitment domains in the expansion of primary bone marrow cells. *Cancer Res* **62**, 2906-2912 (2002).
344. B. Lutterbach *et al.*, ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. *Mol Cell Biol* **18**, 7176-7184 (1998).
345. J. M. Amann *et al.*, ETO, a target of t(8;21) in acute leukemia, makes distinct contacts with multiple histone deacetylases and binds mSin3A through its oligomerization domain. *Mol Cell Biol* **21**, 6470-6483 (2001).
346. Y. Bartel, M. Grez, C. Wichmann, Interference with RUNX1/ETO leukemogenic function by cell-penetrating peptides targeting the NHR2 oligomerization domain. *Biomed Res Int* **2013**, 297692 (2013).
347. J. Schanda *et al.*, Suppression of RUNX1/ETO oncogenic activity by a small molecule inhibitor of tetramerization. *Haematologica* **102**, e170-e174 (2017).
348. S. Singh *et al.*, Mutant p53 establishes targetable tumor dependency by promoting unscheduled replication. *J Clin Invest* **127**, 1839-1855 (2017).
349. F. Mantovani, L. Collavin, G. Del Sal, Mutant p53 as a guardian of the cancer cell. *Cell Death & Differentiation* **26**, 199-212 (2019).
350. D. Walerych, M. Napoli, L. Collavin, G. Del Sal, The rebel angel: mutant p53 as the driving oncogene in breast cancer. *Carcinogenesis* **33**, 2007-2017 (2012).
351. S. Boettcher *et al.*, A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. *Science* **365**, 599-604 (2019).
352. S. S. Mello, L. D. Attardi, Not all p53 gain-of-function mutants are created equal. *Cell Death & Differentiation* **20**, 855-857 (2013).
353. Z. Zuo *et al.*, Acute myeloid leukemia (AML) with erythroid predominance exhibits clinical and molecular characteristics that differ from other types of AML. *PLoS One* **7**, e41485 (2012).
354. P. S. Ko *et al.*, The uniqueness of morphological features of pure erythroid leukemia in myeloid neoplasm with erythroid predominance: A reassessment using criteria revised in the 2016 World Health Organization classification. *PLoS One* **12**, e0172029 (2017).
355. D. Brown *et al.*, A PMLRARalpha transgene initiates murine acute promyelocytic leukemia. *Proc Natl Acad Sci U S A* **94**, 2551-2556 (1997).
356. A. J. Becker, E. A. McCulloch, L. Siminovitch, J. E. Till, The Effect of Differing Demands for Blood Cell Production on DNA Synthesis by Hemopoietic Colony-Forming Cells of Mice. *Blood* **26**, 296-308 (1965).
357. R. Schofield, A comparative study of the repopulating potential of grafts from various haemopoietic sources: CFU repopulation. *Cell Tissue Kinet* **3**, 119-130 (1970).
358. G. Genovese *et al.*, Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371**, 2477-2487 (2014).
359. S. Jaiswal *et al.*, Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371**, 2488-2498 (2014).
360. K. Barbosa, S. Li, P. D. Adams, A. J. Deshpande, The role of TP53 in acute myeloid leukemia: Challenges and opportunities. *Genes Chromosomes Cancer* **58**, 875-888 (2019).
361. N. Allende-Vega, M. Villalba, Metabolic stress controls mutant p53 R248Q stability in acute myeloid leukemia cells. *Sci Rep* **9**, 5637 (2019).
362. B. M. Dulmovits, J. Hom, A. Narla, N. Mohandas, L. Blanc, Characterization, regulation, and targeting of erythroid progenitors in normal and disordered human erythropoiesis. *Curr Opin Hematol* **24**, 159-166 (2017).

363. A. Hunt, M. Fischer, M. E. Engel, S. W. Hiebert, Mtg16/Eto2 contributes to murine T-cell development. *Mol Cell Biol* **31**, 2544-2551 (2011).
364. B. van Riel *et al.*, A novel complex, RUNX1-MYEF2, represses hematopoietic genes in erythroid cells. *Mol Cell Biol* **32**, 3814-3822 (2012).
365. M. X. Xiu, Y. M. Liu, The role of oncogenic Notch2 signaling in cancer: a novel therapeutic target. *Am J Cancer Res* **9**, 837-854 (2019).
366. S. Liu *et al.*, Targeting STAT5 in hematologic malignancies through inhibition of the bromodomain and extra-terminal (BET) bromodomain protein BRD2. *Mol Cancer Ther* **13**, 1194-1205 (2014).
367. H. Ikeda *et al.*, Expression and functional role of the proto-oncogene c-kit in acute myeloblastic leukemia cells. *Blood* **78**, 2962-2968 (1991).
368. K. W. Yip, J. C. Reed, Bcl-2 family proteins and cancer. *Oncogene* **27**, 6398-6406 (2008).
369. L. Zhao, P. Ye, T. J. Gonda, The MYB proto-oncogene suppresses monocytic differentiation of acute myeloid leukemia cells via transcriptional activation of its target gene GFI1. *Oncogene* **33**, 4442-4449 (2014).
370. P. J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, (2017).
371. P. D. Jeffrey, S. Gorina, N. P. Pavletich, Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science* **267**, 1498-1502 (1995).
372. M. E. Valk-Lingbeek, S. W. Bruggeman, M. van Lohuizen, Stem cells and cancer; the polycomb connection. *Cell* **118**, 409-418 (2004).
373. A. Sparmann, M. van Lohuizen, Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**, 846-856 (2006).
374. R. Margueron, D. Reinberg, The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349 (2011).
375. M. Wienken *et al.*, MDM2 Associates with Polycomb Repressor Complex 2 and Enhances Stemness-Promoting Chromatin Modifications Independent of p53. *Mol Cell* **61**, 68-83 (2016).
376. Y. Zhao *et al.*, EZH2 cooperates with gain-of-function p53 mutants to promote cancer growth and metastasis. *EMBO J* **38**, (2019).
377. G. Kuser-Abali *et al.*, An EZH2-mediated epigenetic mechanism behind p53-dependent tissue sensitivity to DNA damage. *Proc Natl Acad Sci U S A* **115**, 3452-3457 (2018).
378. M. T. McCabe, C. L. Creasy, EZH2 as a potential target in cancer therapy. *Epigenomics* **6**, 341-351 (2014).
379. M. Basu, M. W. Khan, P. Chakrabarti, C. Das, Chromatin reader ZMYND8 is a key target of all trans retinoic acid-mediated inhibition of cancer cell proliferation. *Biochim Biophys Acta Gene Regul Mech* **1860**, 450-459 (2017).
380. S. A. Wang, R. P. Hasserjian, Acute Erythroleukemias, Acute Megakaryoblastic Leukemias, and Reactive Mimics: A Guide to a Number of Perplexing Entities. *Am J Clin Pathol* **144**, 44-60 (2015).
381. C. Thiollier *et al.*, Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models. *J Exp Med* **209**, 2017-2031 (2012).
382. A. T. Merryweather-Clarke *et al.*, Global gene expression analysis of human erythroid progenitors. *Blood* **117**, e96-108 (2011).
383. J. Chen, C. S. Kremer, T. P. Bender, A Myb dependent pathway maintains Friend murine erythroleukemia cells in an immature and proliferating state. *Oncogene* **21**, 1859-1869 (2002).
384. J. C. Law, M. K. Ritke, J. C. Yalowich, G. H. Leder, R. E. Ferrell, Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. *Leuk Res* **17**, 1045-1050 (1993).
385. A. P. Bracken, M. Ciro, A. Cocito, K. Helin, E2F target genes: unraveling the biology. *Trends Biochem Sci* **29**, 409-417 (2004).
386. I. M. Min *et al.*, The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell* **2**, 380-391 (2008).
387. D. F. Legler, M. Thelen, New insights in chemokine signaling. *F1000Res* **7**, 95 (2018).
388. C. Quivoron *et al.*, TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* **20**, 25-38 (2011).
389. L. S. Ludwig *et al.*, Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. *Cell Rep* **27**, 3228-3240 e3227 (2019).

390. A. Isphording *et al.*, YWHAE-FAM22 endometrial stromal sarcoma: diagnosis by reverse transcription-polymerase chain reaction in formalin-fixed, paraffin-embedded tumor. *Hum Pathol* **44**, 837-843 (2013).
391. S. L. Gooskens *et al.*, The clinical phenotype of YWHAE-NUTM2B/E positive pediatric clear cell sarcoma of the kidney. *Genes Chromosomes Cancer* **55**, 143-147 (2016).
392. M. Mancini, A. Toker, NFAT proteins: emerging roles in cancer progression. *Nat Rev Cancer* **9**, 810-820 (2009).
393. O. F. Bueno, E. B. Brandt, M. E. Rothenberg, J. D. Molkentin, Defective T cell development and function in calcineurin A beta -deficient mice. *Proc Natl Acad Sci U S A* **99**, 9398-9403 (2002).
394. S. Jauliac *et al.*, The role of NFAT transcription factors in integrin-mediated carcinoma invasion. *Nat Cell Biol* **4**, 540-544 (2002).
395. H. Medyouf *et al.*, Targeting calcineurin activation as a therapeutic strategy for T-cell acute lymphoblastic leukemia. *Nat Med* **13**, 736-741 (2007).
396. T. Zenz *et al.*, TP53 mutation and survival in aggressive B cell lymphoma. *Int J Cancer* **141**, 1381-1388 (2017).
397. P. Bucher *et al.*, Targeting chronic NFAT activation with calcineurin inhibitors in diffuse large B-cell lymphoma. *Blood* **135**, 121-132 (2020).
398. B. Tessoulin *et al.*, p53 dysregulation in B-cell malignancies: More than a single gene in the pathway to hell. *Blood Rev* **31**, 251-259 (2017).
399. W. Ding, Y. Tong, X. Zhang, M. Pan, S. Chen, Study of Arsenic Sulfide in Solid Tumor Cells Reveals Regulation of Nuclear Factors of Activated T-cells by PML and p53. *Sci Rep* **6**, 19793 (2016).
400. N. Shinmen *et al.*, Activation of NFAT signal by p53-K120R mutant. *FEBS Lett* **583**, 1916-1922 (2009).
401. G. Thiel, G. Cibelli, Regulation of life and death by the zinc finger transcription factor Egr-1. *J Cell Physiol* **193**, 287-292 (2002).
402. M. Bettini, H. Xi, J. Milbrandt, G. J. Kersh, Thymocyte development in early growth response gene 1-deficient mice. *J Immunol* **169**, 1713-1720 (2002).
403. P. Laslo *et al.*, Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell* **126**, 755-766 (2006).
404. K. Krishnaraju, B. Hoffman, D. A. Liebermann, Early growth response gene 1 stimulates development of hematopoietic progenitor cells along the macrophage lineage at the expense of the granulocyte and erythroid lineages. *Blood* **97**, 1298-1305 (2001).
405. J. Vogel *et al.*, Transgenic mice overexpressing erythropoietin adapt to excessive erythrocytosis by regulating blood viscosity. *Blood* **102**, 2278-2284 (2003).
406. T. Quaschning *et al.*, Erythropoietin-induced excessive erythrocytosis activates the tissue endothelin system in mice. *FASEB J* **17**, 259-261 (2003).
407. J. Vogel, M. Gassmann, Erythropoietic and non-erythropoietic functions of erythropoietin in mouse models. *J Physiol* **589**, 1259-1264 (2011).
408. N. Novershtern *et al.*, Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296-309 (2011).
409. L. Liu *et al.*, Mutated genes and driver pathways involved in myelodysplastic syndromes-a transcriptome sequencing based approach. *Mol Biosyst* **11**, 2158-2166 (2015).
410. H. Dolatshad *et al.*, Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia* **29**, 1798 (2015).
411. D. Gaidatzis, A. Lerch, F. Hahne, M. B. Stadler, QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31**, 1130-1132 (2015).
412. G. Korotkevich, V. Sukhov, A. Sergushichev, An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*, 060012 (2019).
413. C. The Gene Ontology, Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res* **45**, D331-D338 (2017).
414. A. Liberzon *et al.*, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
415. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
416. P. D. Kingsley *et al.*, Ontogeny of erythroid gene expression. *Blood* **121**, e5-e13 (2013).

417. E. Wingender, P. Dietze, H. Karas, R. Knuppel, TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* **24**, 238-241 (1996).
418. L. Gautier, L. Cope, B. M. Bolstad, R. A. Irizarry, affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-315 (2004).
419. M. R. Corces *et al.*, Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* **48**, 1193-1203 (2016).
420. J. D. Buenrostro *et al.*, Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490 (2015).
421. C. Strupp *et al.*, New proposals of the WHO working group (2016) for the diagnosis of myelodysplastic syndromes (MDS): Characteristics of refined MDS types. *Leuk Res* **57**, 78-84 (2017).

List of abbreviations

ABL	Abelson Murine Leukemia Viral Oncogene Homolog 1
AEL	Acute Erythroid Leukemia
AEV	Avian Erythroblastosis Retrovirus
AF9	ALL1-Fused Gene From Chromosome 9 Protein
AGM	Aorta-Gonad-Mesonephros
ALL	Acute Lymphoid Leukemia
AMKL	Acute Megakaryoblastic Leukemia
AML	Acute Myeloid Leukemia
ASXL1	Additional Sex Combs Like 1, Transcripitonal Regulator
ATAC-seq	Assay for Transposase-Accessible Chromatin
ATRA	All-*Trans* Retinoic Acid
BAALC	Brain And Acute Leukemia, Cytoplasmic
Baso-E	Basophilic Erythroblast
BCL	B- cell lymphoma 2
BCOR	BCL6 Co-repressor
BFU-E	Burst Forming Units-Erythroid
BM	Bone Marrow
CBFA2T3	Core-Binding Factor, Runt Domain, Alpha Subunit 2; Translocated To 3
CEBPA	CCAAT Enhancer Binding Protein Alpha
CFU-GEMM	Colony Formation Unit – Granulocytes, Erythrocytes, Monocytes And Megakaryocytes
CFU-GM	Colony formation unit - Granulocytes and monocytes
ChiP-seq	Chromatin immunoprecipitation
CLP	Common Lymphoid Progenitor
CML	Chronic Myeloid Leukemia
CMP	Common Myeloid Progenitor
CTRL	Vector-control
DBD	DNA Binding Domain
DEG	Differentially Expressed Genes
DM	Differentiation Medium
DMSO	Dimethyl Sulfoxide
DNE	Dominant Negative
ENL	MLLT1 super Elongation Complex Subunit
EPO	Erythropoietin
EPOR	Erythropoietin Receptor
ERG	ETS Transcription Factor ERG
ES	Embryonic Stem
F-MuLV	Friend Murine Leukemia Virus
F36P	Acute myeloid leukemia M6 cell line
FAB	French-American-British
FAB-M6a	Erythroid/Myeloid Leukemia
FAB-M6b	Pure Erythroid Leukemia
FDR	False Discovery Rate
FL	Fetal Liver
FLI1	Friend Leukemia Integration-site 2
FLT3	FMS-Like Tyrosine Kinase 3
G-CSFR	Granulocyte Colony-Stimulating Factor Receptor
GDP	Guanosine Diphosphate
GFP	Green Fluorescent protein

GLIS2	GLIS Family Zinc Finger 2
GMP	Granulocyte-Monocyte-Progenitor
GOF	Gain-of-function
Gpa	Glycophorin A
GSEA	Gene Set Enrichment Analysis
H3K27	Histone 3, Lysine 27
H3K4	Histone 3, Lysine 4
HD3	Chicken erythroblastic cell line
HEL	Human Erythroleukemia cell line
HeLa	Cervical cancer cells derived from Henrietta Lacks
Hemgn	Erythroid differentiation-associated gene protein
HMBA	Hexamethylene Bisacetamide
HPC	Hematopoietic progenitor cells
HSC	Hematopoietic Stem Cell
HSP70	Heat Shock Protein 70
HSPC	Hematopoietic Stem and Progenitor cells
HUPKI	Human p53 Knock-In
i.f	Intrafemorally transplantation
i.v	Intravenously transplantation
Int-E	Intermediate Erythroblast
ITD	Internal-Tandem Duplication
JAK2	Janus Kinase 2
K562	Human erythroleukemia cell line
KMOE	Human erythroleukemia cell line
KMT2A	Lysine Methyltransferase 2A
KO	Knock-Out
Late-E	Late Erythroblast
LIC	Leukemic-Initiating Cell
LMPP	Lymphoid Myeloid Primed Progenitor
LOH	Loss Of Heterozygosity
LSC	Leukemic Stem Cell
M-CSFR	Macrophage Colony-Stimulating Factor Receptor
MC	Methylcellulose
MDM2	MDM2 Proto-Oncogene
MDS	Myelodysplastic Syndromes
Me3	Tri-methylation
MEL	Murine Erythroid Leukemia cell line
MEP	Myeloid-Erythroid Progenitor
MH1	MAD homology I Dwarfin-type
MM	Maintenance Medium
MOZ	Lysine Acetyltransferase 6A
MPN	Myeloproliferative Neoplasms
MPP	Multipotent Progenitor
MSCV	Murine Stem Cell Virus
MTG16	Myeloid Translocation Gene On Chromosome16
MYB	V-Myb Avian Myeloblastosis Viral Oncogene Homolog
N-E	NFIA-ETO2
NCoR	Nuclear receptor co-Repressor
NF-□B	Nuclear Factor-Kappa B
NFIA	Nuclear Factor 1 A
NGS	Next Generation Sequencing
NHR	Nervy Homology Regions

NK	Natural Killer Cell
NOS	Not Otherwise Specified
NPM1	Nucleophosmin1
NSD1	Nuclear Receptor Binding SET Domain Protein 1
NUP98	Nucleoporin 98
OCIM1	Human acute myeloid leukemia cell line
OCM2	Human acute myeloid leukemia cell line
Ortho-E	Orthochromatic Erythroblasts
P	P value
PCA	Principal Component Analysis
PEL	Pure Erythroid Leukemia
PHF6	PHD Finger Protein 6
PLT	Platelet
PML	Promyelocytic Leukemia
Poly-E	Polychromatophilic Erythroblasts
PRC2	Polycomb Repressive Complex 2
Pro-E	Proerythroblasts
PV	Polycythemia Vera
RARα	Retinoic Acid Receptor Alpha
RBC	Red Blood Cell
RELA	RELA Proto-Oncogene, NF-KB Subunit
RNA-seq	RNA high-throughput sequencing
RT-qPCR	Reverse Transcription qPCR
RUNX1	Runt-Related Transcription Factor 1
RUNX1T1	Core-Binding Factor, Runt Domain, Alpha Subunit 2; Translocated To, 1; Cyclin D-Related
SCF	Stem Cell Factor
SPM	Stem and Progenitor Medium
SRSF3	Serine And Arginine Rich Splicing Factor 3
STAG2	Stromal Antigen 2
TET2	Tet Methylcytosine Dioxygenase 2
TF1	Human erythroleukemia cell line
TGCA	The Genome Cancer Atlas
TIF2	Transcriptional Intermediary Factor 2
TP53	Tumor Protein P53
WB	Western Blot
WBC	White Blood Cell
WHO	World Health Organization
WT	Wild Type
WT1	WT1 Transcription Factor
Z-R	ZMYND8-RELA
ZMYND8	Zinc finger MYD-type containing 8

Appendices

Appendix I: Human erythroleukemia genetics and transcriptomes identify master transcription factors as functional disease drivers (*302*).

blood®
Regular Article

MYELOID NEOPLASIA

Human erythroleukemia genetics and transcriptomes identify master transcription factors as functional disease drivers

Alexandre Fagnan,[1,2] Frederik Otzen Bagger,[3-6,*] Maria-Riera Piqué-Borràs,[3,4,*] Cathy Ignacimouttou,[1,2,*] Alexis Caulier,[7,8] Cécile K. Lopez,[1,2] Elie Robert,[1,2] Benjamin Uzan,[9] Véronique Gelsi-Boyer,[10,11] Zakia Aid,[1,2] Cécile Thirant,[1,2] Ute Moll,[12,13] Samantha Tauchmann,[3,4] Amina Kurtovic-Kozaric,[14] Jaroslaw Maciejewski,[15] Christine Dierks,[16] Orietta Spinelli,[17] Silvia Salmoiraghi,[17,18] Thomas Pabst,[19] Kazuya Shimoda,[20] Virginie Deleuze,[21,22] Hélène Lapillonne,[23] Connor Sweeney,[24] Véronique De Mas,[25] Betty Leite,[26] Zahra Kadri,[27] Sébastien Malinge,[1,28] Stéphane de Botton,[1,2] Jean-Baptiste Micol,[1] Benjamin Kile,[29] Catherine L. Carmichael,[29] Ilaria Iacobucci,[30] Charles G. Mullighan,[30,31] Martin Carroll,[32] Peter Valent,[33,34] Olivier A. Bernard,[1,2] Eric Delabesse,[25] Paresh Vyas,[24] Daniel Birnbaum,[10,11] Eduardo Anguita,[35-38] Loïc Garçon,[7,8] Eric Soler,[21,22] Juerg Schwaller,[3,4,†] and Thomas Mercher[1,2,†]

[1]Unité 1170 (U1170), INSERM, Gustave Roussy, Université Paris Diderot, Villejuif, France; [2]Equipe Labellisée Ligue Nationale Contre le Cancer, Paris, France; [3]University Children's Hospital Beider Basel (UKBB), Basel, Switzerland; [4]Department of Biomedicine, University of Basel, Basel, Switzerland; [5]Center for Genomic Medicine, Copenhagen University Hospital, Copenhagen, Denmark; [6]Swiss Institute of Bioinformatics, Basel, Basel, Switzerland; [7]Equipe d'Accueil (EA) 4666, Hématopoïèse et Immunologie (HEMATIM), Université de Picardie Jules Verne (UPJV), Amiens, France; [8]Service Hématologie Biologique, Centre Hospitalier Universitaire (CHU) Amiens, Amiens, France; [9]Unité Mixte de Recherche 967 (UMR 967), INSERM–Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA)/Direction de la Recherche Fondamentale (DRF)/Institut de Biologie François Jacob (IBFJ)/Institut de Radiobiologie Cellulaire et Moléculaire (IRCM)/Laboratoire des cellules Souches Hématopoïétiques et des Leucémies (LSHL)–Université Paris-Diderot–Université Paris-Sud, Fontenay-aux-Roses, France; [10]U1068 and [11]UMR7258, Centre de Recherche en Cancérologie de Marseille, Centre National de la Recherche Scientifique (CNRS)/INSERM/Institut Paoli Calmettes/Aix-Marseille Université, Marseille, France; [12]Institute of Molecular Oncology, University Medical Center Göttingen, Göttingen, Germany; [13]Department of Pathology, Stony Brook University, Stony Brook, NY; [14]Clinical Center of the University of Sarajevo, University of Sarajevo, Sarajevo, Bosnia and Herzegovina; [15]Department of Translational Hematology and Oncologic Research, Cleveland Clinic Taussig Cancer Institute, Cleveland, OH; [16]Hämatologie, Onkologie und Stammzelltransplantation, Klinik für Innere Medizin I, Freiburg, Germany; [17]UOC Ematologia, Azienda Socio Sanitaria Territoriale Papa Giovanni XXIII Hospital, Bergamo, Italy; [18]FROM Research Foundation, Papa Giovanni XXIII Hospital, Bergamo, Italy;[19]Department of Oncology, Inselspital, University Hospital Bern/University of Bern, Bern, Switzerland; [20]Gastroenterology and Hematology, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan; [21]IGMM, University of Montpellier, CNRS, Montpellier, France; [22]Université de Paris, Laboratory of Excellence GR-Ex, Paris, France; [23]Centre de Recherche Saint Antoine (CRSA)–Unité INSERM, Sorbonne Université/Assistance Publique–Hôpitaux de Paris (AP-HP)/Hôpital Trousseau, Paris, France; [24]Medical Research Council Molecular Haematology Unit (MRC MHU), Biomedical Research Centre (BRC) Hematology Theme, Oxford Biomedical Research Centre, Oxford Centre for Haematology, Weatherall Institute of Molecular Medicine (WIMM), Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom; [25]Team 16, Hematology Laboratory, Center of Research of Cancerology of Toulouse, U1037, INSERM/Institut Universitaire du Cancer de Toulouse (IUCT) Oncopole, Toulouse, France; [26]Genomic Platform, Unité Mixte de Service - Analyse Moléculaire, Modélisation et Imagerie de la maladie Cancéreuse (UMS AMMICA), Gustave Roussy/Université Paris-Saclay, Villejuif, France; [27]Division of Innovative Therapies, UMR-1184, Immunologie des Maladies Virales, Auto-immunes, Hématologiques et Bactériennes (IMVA-HB) and Infectious Disease Models and Innovative Therapies (IDMIT) Center, CEA/INSERM/Paris-Saclay University, Fontenay-aux-Roses, France; [28]Telethon Kids Institute, Perth Children's Hospital, Nedlands, WA, Australia; [29]Australian Centre for Blood Diseases, Monash University, Melbourne, VIC, Australia; [30]Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN; [31]Hematological Malignancies Program, St. Jude Children's Research Hospital, Memphis, TN; [32]Division of Hematology and Oncology, University of Pennsylvania, PA; [33]Division of Hematology and Hemostaseology, Department of Internal Medicine I and [34]Ludwig Boltzmann Institute for Hematology and Oncology, Medical University of Vienna, Vienna, Austria; [35]Hematology Department, [36]Instituto de Medicina de Laboratorio (IML), and [37]Instituto de Investigación Sanitaria San Carlos, (IdISSC), Hospital Clínico San Carlos (HCSC), Madrid, Spain; and [38]Department of Medicine, Universidad Complutense de Madrid (UCM), Madrid, Spain

KEY POINTS

- **Transcriptomes cluster most AEL apart from other myeloid malignancies.**
- **Alterations of AEL erythroid master regulators impair GATA1 activity and induce the disease in mice.**

Acute erythroleukemia (AEL or acute myeloid leukemia [AML]-M6) is a rare but aggressive hematologic malignancy. Previous studies showed that AEL leukemic cells often carry complex karyotypes and mutations in known AML-associated oncogenes. To better define the underlying molecular mechanisms driving the erythroid phenotype, we studied a series of 33 AEL samples representing 3 genetic AEL subgroups including *TP53*-mutated, epigenetic regulator-mutated (eg, *DNMT3A*, *TET2*, or *IDH2*), and undefined cases with low mutational burden. We established an erythroid vs myeloid transcriptome-based space in which, independently of the molecular subgroup, the majority of the AEL samples exhibited a unique mapping different from both non-M6 AML and myelodysplastic syndrome samples. Notably, >25% of AEL patients, including in the genetically undefined subgroup, showed aberrant expression of key transcriptional regulators, including *SKI*, *ERG*, and *ETO2*. Ectopic expression of these factors in murine erythroid progenitors blocked in vitro erythroid differentiation and led to immortalization associated with decreased chromatin accessibility at GATA1-binding sites and functional interference with GATA1 activity. In vivo models showed development of lethal

erythroid, mixed erythroid/myeloid, or other malignancies depending on the cell population in which AEL-associated alterations were expressed. Collectively, our data indicate that AEL is a molecularly heterogeneous disease with an erythroid identity that results in part from the aberrant activity of key erythroid transcription factors in hematopoietic stem or progenitor cells. (*Blood*. 2020;136(6):698-714)

Introduction

Acute myeloid leukemia (AML) of the erythroid lineage (acute erythroleukemia [AEL] or AML-M6) accounts for 3% to 5% of AML patients and is inherently associated with poor outcome.[1-3] Although AEL can occur at any age, the majority of patients are >65 years, and the disease often occurs secondary to other neoplasms, including myeloproliferative neoplasms (MPNs) or myelodysplastic syndrome (MDS), or after cytotoxic cancer treatment. Two major morphological subtypes have been proposed: pure erythroleukemia (PEL; AML-M6b, also known as Di Guglielmo disease) with >80% of blasts committed to the erythroid lineage and AML-M6a characterized by the presence of both erythroid precursors and myeloid blasts.[1-3] The 2016 World Health Organization (WHO) classification integrated AML-M6a into MDSs or not otherwise specified AML (AML-NOS), but this classification remains a matter of debate.[4-6]

Functional studies have suggested that 2 to 5 genetic driver lesions on a background of preexisting alterations in hematopoietic stem or progenitor cells (HSPCs) might be sufficient to induce AML.[7,8] For AEL, earlier work showed that leukemic cells often have complex karyotypes, and targeted DNA sequencing revealed the presence of several known AML-associated mutations,[9-12] but AEL-driving molecular mechanisms remain incompletely understood and erythroleukemia-specific mutations have seldom been functionally validated. Strikingly, single or multiple *TP53* mutations have been shown to be a molecular hallmark of PEL.[13]

Normal erythroid differentiation is controlled by the activity of both extrinsic signaling factors, including erythropoietin (EPO) mediating its effects through the EPO receptor (EPOR) signaling pathways, and intrinsic multimeric transcription complexes.[14-16] The latter includes hematopoietic master regulators like GATA-binding protein 1 (GATA1), T-cell acute lymphocytic leukemia protein 1 (TAL1), LIM domain-only 2 (LMO2), CBFA2/RUNX1 partner transcriptional corepressor 3 (CBFA2T3, also known as ETO2), and LIM-domain-binding protein 1 (LDB1), thereafter broadly named GATA1 complexes, which can activate or repress transcription of target genes. These GATA1 complexes contribute to terminal erythroid differentiation through binding to gene loci and transcription of essential erythroid genes (eg, hemoglobin). This process is also regulated by Krüppel-like factor 1 (KLF1), which binds DNA next to the GATA1 complexes to coregulate erythroid genes.[17,18] To establish the erythroid differentiation program, functional synergism between these transcriptional complexes and the EPO/EPOR signaling is mediated by the presence of phosphorylated STAT5 binding in the neighborhood of GATA1 and KLF1.[19,20] Accordingly, mutations in these factors have been associated with altered erythropoiesis.[21,22] For example, *GATA1* mutations are associated with congenital erythroid hypoplasia (Diamond-Blackfan anemia [DBA]) or X-linked dyserythropoietic anemia.[23] Moreover, the identification of a NFIA-ETO2 fusion in pediatric PEL[24] suggests that an altered activity of these complexes may contribute to human erythroid leukemogenesis.

To better understand the molecular mechanisms that control the erythroid feature, we characterized the genetic and transcriptional landscape in leukemic cells from 33 AEL patients. We identified distinct molecular subgroups composed of patients carrying (1) *TP53* mutations, (2) various combinations of mutations previously found in AML and MDS such as *DNMT3A*, *TET2* or *IDH2*, and (3) those with none of these recurrent alterations. Comparative transcriptomics established an erythro/myeloid differentiation expression signature space that distinguished the majority of AEL cases from MDS or other AML forms. Notably, leukemic cells from >25% of AEL patients showed aberrant expression of key transcriptional regulators including *SKI*, *ERG*, and *ETO2*, which interfere with the activity of the erythroid master regulator GATA1. Combinatorial experimental expression in HSPC fractions induced lethal erythroid or mixed erythroid/myeloid diseases in mice phenocopying several aspects of the human disease, underlining their importance in the molecular pathogenesis of AEL.

Materials and methods

Patient samples

Fifty-eight human patient samples were obtained with the informed consent of the patient and approved by the local ethics committees in accordance with national ethics rules. AEL patient diagnostics were established according to the WHO 2008 classification and criteria described recently.[16] Cytogenetic risk groups were defined according to the revised International Prognostic Scoring System (IPSS-R).[25] Mononuclear cell fractions were obtained from patient blood or bone marrow (BM) samples by Ficoll gradient, and frozen in fetal bovine serum (FBS; Gibco) supplemented with 10% dimethyl sulfoxide (DMSO). DNA and RNA extraction were done on fresh or frozen samples. DNA was extracted using bulk or sorted cells from patient samples (n = 7, $CD36^+$ for blast cell population and $CD3^+$ or $CD19^+$ for non-neoplastic cell populations) or from xenograft-amplified samples (n = 4). RNA was extracted from patient samples (n = 22) and xenograft-amplified samples (n = 7) from bulk or sorted cells ($CD36^+$ or $CD45^+$ cells), respectively. We obtained appropriate sequencing material for 33 patients (11 paired patient samples for exome sequencing and 29 patient samples for RNA sequencing).

Murine models

C57BL/6JOlaHsd mice (named *C57BL/6J*) were purchased from Envigo and *NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ* (NSG) mice from The Jackson Laboratory (005557). *TP53$^{R248Q/+}$* knock-in mice were described previously.[26] To generate double transgenic TET2$^{-/-}$/GATA1s mice, we intercrossed *Tet2$^{-/-}$* and *Gata1$^{\Delta e2}$* (here named *Gata1s*) mice.[27,28] Mice were maintained at the Gustave Roussy preclinical facility and all experiments

were approved by the French National Animal Care and Use Committee (CEEA 26: projects 2017-082-12726 and 2017-084-12799).

Flow cytometry and cell sorting

Antibodies used for flow cytometry are listed in supplemental Table 1 (available on the *Blood* Web site). Cells were stained in 1× phosphate-buffered saline (PBS) supplemented with 2% FBS at 4°C for 30 minutes and washed prior to analysis. Whole BM or spleen cells were analyzed without red blood cell lysis. For cell sorting, total BM cells underwent red blood cell lysis. To obtain HSPCs, total BM was depleted of all major hematopoietic cell lineage (Lin^-) using the Mouse Hematopoietic Progenitor (Stem) Cell Enrichment Set (Becton Dickinson [BD]]. Progenitor populations were further purified by fluorescence-activated cell sorting (FACS) according to the following phenotypes: Hematopoietic stem cells (HSC) were defined as $Lin^-/Sca1^+/KIT^+/CD34^-/CD48^-$, megakaryocytic-erythroid progenitors (MEPs) were defined as $Lin^-Sca1^-/KIT^+/CD34^-/CD16/32^-$ and granulocyte-macrophage progenitors (GMP) were defined as $Lin^-/Sca1^-/KIT^+/CD34^+/CD16/32^+$. To obtain mouse erythroid progenitors, BM cells were first depleted using biotin-conjugated antibodies against CD3, B220, Gr-1, and CD11b (BD) followed by FACS according to the population described as $CD71^+/Ter119^+/KIT^+$. Flow cytometric analysis was performed using ARIAII, CANTO-II, or CANTO-X instruments (BD), and data were analyzed using the FlowJo software (Flowjo 9.3.2).

Cell culture

Mouse erythroid progenitor cells were expanded in StemSpan serum-free expansion medium (SFEM; Stem Cell Technologies) supplemented with penicillin (100 U/mL)-streptomycin (100 μg/mL), murine stem cell factor (mSCF) (10 ng/mL), murine interleukin 3 (mIL3) (10 ng/mL), mIL6 (10 ng/mL), human EPO (hEPO) (2 U/mL), 0.4% cholesterol, and dexamethasone (10^{-6} M). Mouse erythroleukemia (MEL) cells were maintained in RPMI 1640 (Gibco) supplemented with 10% FBS, penicillin (100 U/mL)-streptomycin (100 μg/mL) and 2 mM L-Glutamine (Gibco). Murine G1E cells, a generous gift from M. Weiss,[29] were maintained in Iscove modified Dulbecco medium (Gibco) supplemented with 15% FBS, penicillin (100 U/mL)-streptomycin (100 μg/mL), mSCF (10 ng/mL), hEPO (2 U/mL), monothioglycerol (4.5×10^{-5} M), and 2 mM L-Glutamine (Gibco). Murine growth factor-dependent Ba/F3 cells was maintained in RPMI 1640 (Gibco) supplemented with 10% FBS, penicillin (100 U/mL)-streptomycin (100 μg/mL), mIL3 (10 ng/mL), and 2 mM L-Glutamine (Gibco). Human embryonic kidney (HEK-293T) cells were grown in Dulbecco modified Eagle medium (Gibco) supplemented with 10% FBS, penicillin (100 U/mL)-streptomycin (100 μg/mL), and 2 mM L-Glutamine (Gibco).

Retroviral constructs, particle production, and cell transduction

The *SKI* complementary DNA (cDNA) was a kind gift from Suzana Atansoski (Basel, Switzerland). The other cDNAs were synthesized. All cDNAs were cloned into retroviral *pMSCV-IRES-EGFP* or *-mCherry* backbones. *GATA1* cDNA was cloned into lentiviral *pLT3-GEPIR-IRES-EGFP* expression vector. For retroviral or lentiviral particles production, HEK-293T cells were plated 1 day before cotransfection with the expression constructs coexpressing EGFP or mCherry and cDNA using the X-tremeGENE-9 DNA Transfection Reagent (Roche) or jetPRIME reagent (Polyplus transfection), respectively, according to the manufacturer's recommendations. Culture media were changed 24 hours posttransfection and supernatants containing viral particles were harvested 48 hours and 72 hours posttransfection. Murine cells were transduced by spinoculation (90 minutes at 2500 rpm, 33°C) with supernatants containing viral particles supplemented with 5 μg/mL polybrene in 7.5 mM HEPES buffer.

BM transplantation

Total BM (0.4×10^6 cells) and/or transduced progenitor cells were transplanted through IV injection in lethally (9.5 Gy) or sublethally (5 Gy) irradiated 8- to 10-week-old *C57BL/6J* recipient mice.

RNA extraction and RT-qPCR

RNA was extracted using a RNeasy Mini kit (Qiagen) or AllPrep DNA/RNA Mini kit (Qiagen), according to the manufacturer's recommendations and quantified using NanoDrop (ThermoScientific). Reverse transcription (RT) was performed using SuperScript II (Invitrogen). Quantitative polymerase chain reaction (qPCR) was performed using SYBR Select Master mix or TaqMan Gene Expression Master mix (Applied Biosystems) on a 7500HT Fast Real-Time PCR System (Applied Biosystems) following the manufacturer;'s recommendations. Primer sequences are listed in supplemental Table 2.

Whole-exome sequencing

Whole-exome sequencing was conducted as described previously[30] on paired-samples from 11 patients. DNA from sorted $CD3^+$ or $CD19^+$ nonneoplastic cells was used for exome capture using SureSelect All Exon V4 or V5 kits (Agilent Technologies). We performed paired-end sequencing (100 bp) using HiSeq2000 sequencing instruments at Gustave Roussy genomic platform. Reads were mapped to the reference genome hg19 using the Burrows-Wheeler Aligner (BWA) alignment tool version 0.7.10. PCR duplicates were removed using Picard tools–Mark Duplicates (version 1.119). Local realignment around indels and base quality score recalibration was performed using GATK 3.3 (Genome Analysis Tool Kit). Reads with a mapping quality score $< 30 < 20$ were removed. Somatic single-nucleotide variations (SNVs) and indels were called in the leukemic sample using Varscan (v2.3.7) by comparison with the paired nonneoplastic samples for exomes, and by comparison with the reference genome for RNA-seq. For candidate somatic mutations, the variants were adopted as candidate mutations when *P* value was <.001 and allele frequency was <.1 in the reference sample. Variants were annotated with Annovar (v141112). We excluded synonymous SNVs, variants located in intergenic, intronic, untranslated regions and noncoding RNA regions. The mean coverage in the targeted regions was, respectively, 85,4× and 91,2× for leukemic and nonneoplastic samples. The functional variants were predicted using the open platform Cancer Genome Interpreter[31] (CGI) and only known-variants or predicted driver variants were confirmed through visualization with IGV (2.3.88) and finally kept in this study.

RNA sequencing

RNA sequencing (RNA-seq) was performed as described.[27] Sequences were aligned to the reference genome with TopHat2 version 2.0.9 using the following parameters: –bowtie1–fusion-search–library-type fr-firststrand–read-realign-edit-dist 0 -p 8 -r 50 (or 2.0.14 for mice data sets) and Bowtie1 version 1.0.0. The

number of reads per genes (RefSeq database) was counted with HTSeq-count version 0.5.4p5 using the "union" mode. The counts were then normalized with the DESeq2 method, which takes into account the library size of each sample.

ATAC sequencing

ATAC-sequencing (ATAC-seq) analysis has been previously described[32] Briefly, after lysis of 50 000 cells, transposition, and purification steps, the transposed DNA fragments were amplified by PCR (12 cycles) using adapters from the Nextera Index Kit (Illumina). PCR purification was performed using Agencourt AMPure XP magnetic beads (Beckman Coulter) to remove large fragments and remaining primers. Library quality was assessed using an Agilent 2100 Bioanalyzer using a High Sensitivity DNA Chip (Agilent Technologies). Libraries were sequenced using NovaSeq 6000 sequencer (Illumina; 50 bp paired-end reads). Quality control of reads was performed using FastQC 0.11.7 and multiQC 1.5. The reads were aligned to the reference genome mm10 with bwa (aln 0.7.17). After alignment, we removed reads mapping to the mitochondrial genome, PCR duplicate reads, and reads with a mapping quality lower than 20 using samtools (v 1.9). Final read counts for all mouse data sets ranged from 42 to 128 million reads. Mapped reads were normalized to bins per million and were converted to bigwig format using deeptools (v3.3.0). Peak calling, differential analysis, annotation, and motif analysis were performed using macs2 (v 2.1.1) and homer (v4.10.4, annotatePeak.pl and findMotifsGenome.pl).

PCA of data from differentiation map

To define the hematopoietic space described (Figure 3A), we built a principal component analysis (PCA) of cell types from Differentiation Map [DMAP][33] (DMAP_PCA), excluding NK cells, B cells, T cells, and dendritic cells. As features we included ranks of differentially expressed genes (DMAP_DE) (FDR $<$ 0.05, logFC $>$ 2) determined with limma.[34] A Loess regression line was fitted in PCA space to erythroid cells (all erythrocytes and MEP), and Myelocytes (HSC, CMP, GMP, GRAN, MONO, BASO, and EOS). New data points (NP) was projected into the DMAP_PCA space as a dot product between scaled NP vector and DMAP_PCA rotation. By these calculation, we applied the same transformation to NP and added them to DMAP_PCA without recalculation of principal components.

Transcription factor activity inference

For gene regulatory network inference, the ARACNe-AP software was used to infer a Gene Regulatory Network using scRNA-seq data from healthy human progenitors to predict a list of target genes for each transcription factor (TF).[28,29] ARACNe was run over the log2 normalized counts in bootstrap mode (100 iterations), with a *P* value threshold of 1e-8 and a custom curated list of 2171 TFs. Therefore, the activity of each TF in a normal context was computed in a network. For each AEL sample, TF activities were inferred by interrogating this network with AEL transcriptome data and expressed as Normalized Enrichment Score (NES) using the R library viper, as described in the bioconductor package manual.[30] NES were used to test differential activity by Student *t* test and *P* value correction by Benjamini-Hochberg (FDR cutoff at 0.05). Differentially activated gene lists were established by PCA analysis using predicted activated gene matrix (previously computed using ARACNE and VIPER algorithm), then genes driving PCA dimensions were identified and ranked by contribution (using FactoMineR v1.41 and factoextra v1.0.5 R packages). Finally, heatmap of activated genes was obtained by plotting the top 50 most contributed genes from the first PCA dimension (using pheatmap v1.0.12 R package).

Data sets

Sequencing data were deposited into EBI - Array-Express under the accession E-MTAB-9012 (ATAC-seq) and European Genome-Phenome Archive (EGA) under the accession EGAS00001004203 (Exome/RNA-seq). Available GATA1 ChIP-seq on mouse erythroblasts were obtained from ENCODE (GSE36029; SRA accession: SRR492437) and available ATAC-seq data sets from mouse MEP, CFU-E, and proerythroblasts were previously published.[35]

Statistical analysis

Statistical significance was calculated using Prism (version 6.0a) and is indicated as *P* values (Student *t* test except when otherwise specified). $*P < .05$, $**P < .01$, $***P < .001$.

Results

Molecular alterations in AEL patients

We collected samples from 58 AEL patients, including 34 adults $>$60 years, 14 between 40 and 59 years, 8 young adults (21-39 years), and 2 pediatric patients. According to the 2008 WHO classification, 33 patients were diagnosed with de novo AEL, including 29 AML-M6a and 4 AML-M6b; 20 patients were diagnosed with AML-M6a secondary to MDS/CML/ALL; 1 with AML-M6b secondary to plexus choroid carcinoma; and a more precise diagnosis was lacking for 4 patients (supplemental Figure 1A; supplemental Table 3). Thereafter, the term "AEL" was used for all patients. Several AEL samples lacking sufficient number of viable cells, were expanded by xenografting them in *NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ* (NSG) mice. This approach provided additional leukemic material to isolate RNA (7 patients) and DNA (4 patients). Together, we obtained appropriate sequencing material for 33 patients and performed exome sequencing on 11 paired leukemic and non-neoplastic (either CD3$^+$ or CD19$^+$ cells from the same patient) samples and RNA sequencing of 29 leukemic samples. Combining exome and RNA-seq data, we identified sequence variants with predicted functional consequences in 62 genes (Figure 1).

These data, including high variant allele frequency (supplemental Figure 1B), support classification of patients into 3 molecular subgroups (Figure 1). Subgroup 1, presenting with *TP53* mutations (n = 12, 36.3% of patients), had in average 4.41 mutations per sample and was associated with both a higher cytogenetic risk and a poorer outcome (Figure 1; supplemental Figure 1C-F). Subgroup 2 (n = 11, 33.3%) mostly presented with *TET2* nonsense mutations (n = 8) and *DNMT3A* mutations (n = 5), including 2 patients with both *TET2* and *DNMT3A* mutations, and had in average 5.72 mutations per sample. Several patients with *TET2* and/or *DNMT3A* mutations also carried *SRSF2$^{P95H/R}$* or *IDH2^{R140Q}* mutations. Of note, in the only sample presenting both a *TET2* and an *IDH2* mutation, the variant allele frequencies were 60% and 13%, respectively (data not shown), possibly reflecting 2 independent clones. Interestingly, 1 case (#17) of

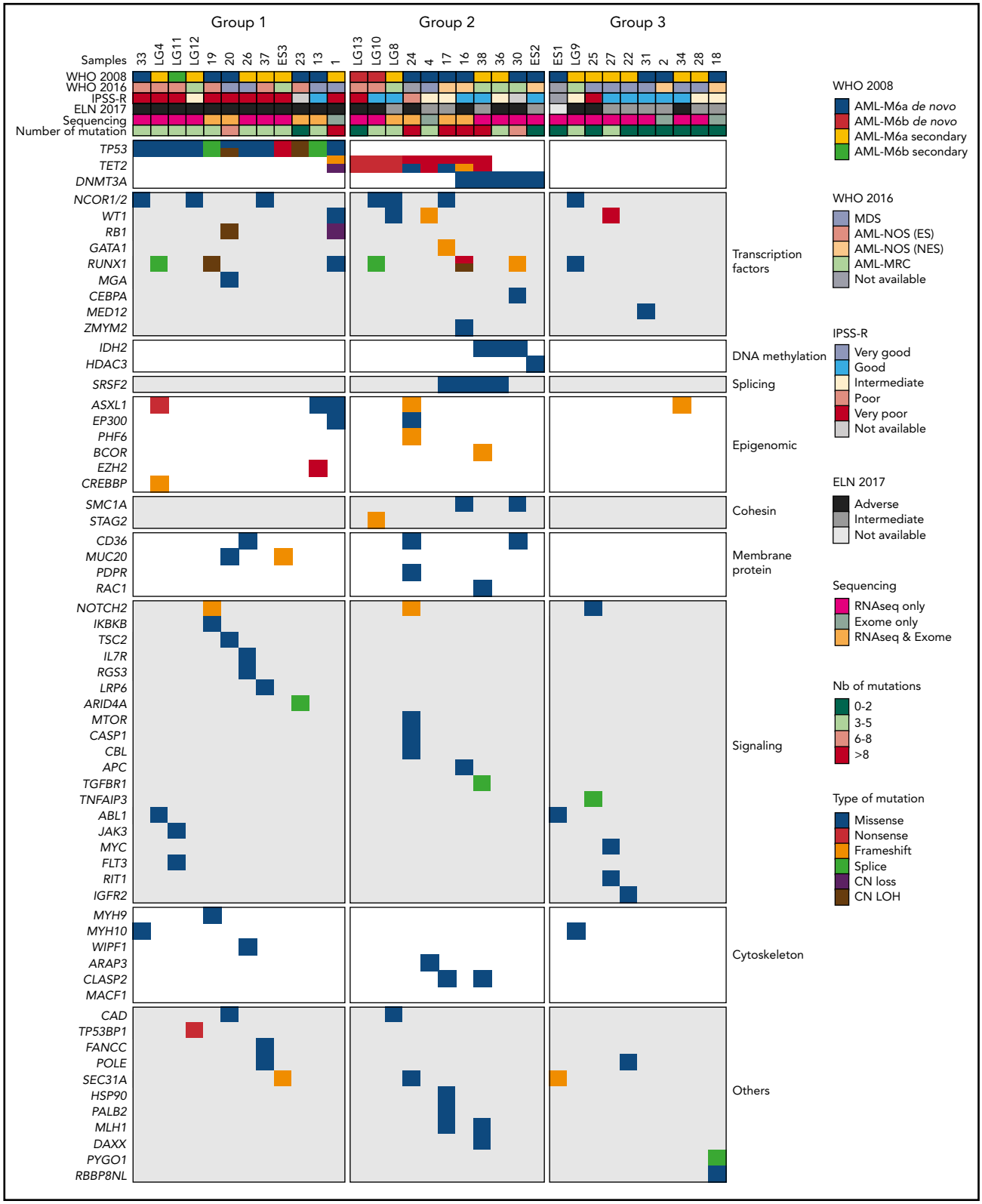


Figure 1. Genetic alterations of a human AEL cohort. Heatmap representing 3 molecular AEL subgroups: TP53-mutated (group 1), epigenetic modifier-mutated (group 2), and others (with no recurrent mutation; group 3), according to WHO 2008 and 2016 diagnosis classifications, cytogenetic risk group (IPSS-R), ELN 2017 risk stratification, methods of sequencing, number of predicted driver mutations, and type of predicted driver mutations.

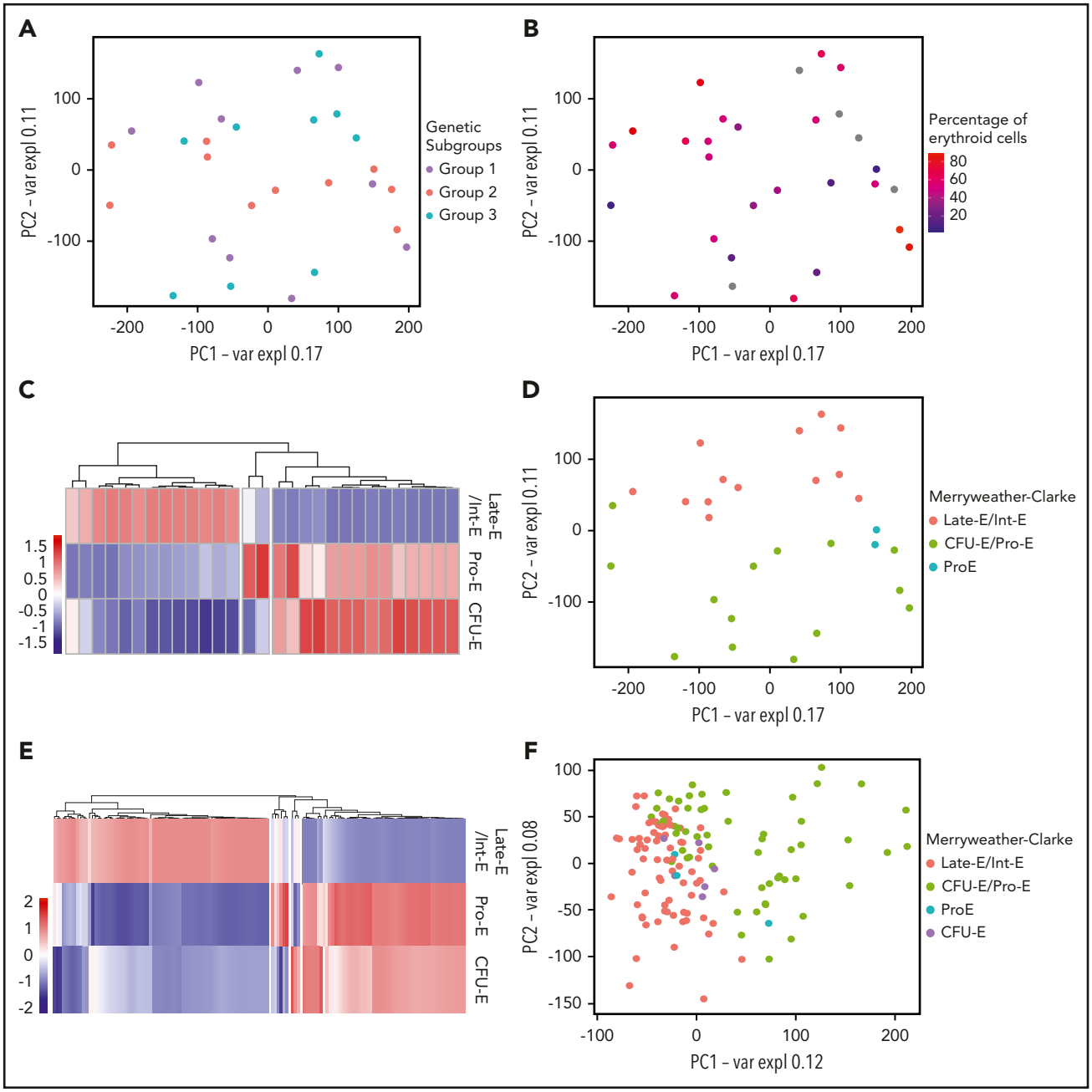


Figure 2. AEL transcriptomes reflect erythroid differentiation rather than genetic mutations. (A) PCA of AEL patient samples based on gene expression. Each point represents 1 sample, colored according to molecular subgroup. (B) PCA of AEL patient samples based on gene expression. Samples are colored according to the amount (percentage) of erythroid cells in the patient's BM at diagnosis. (C) Heatmap showing correlations between AEL patient samples and the Merryweather-Clarke et al erythroid expression signatures.[38] Merryweather-Clarke et al described the expression profiles of enriched human colony-formation unit-erythroid (CFU-E; CD71^{+}CD235^{-}), pro-erythroblasts (Pro-E; CD71^{+}CD235^{low}), intermediate (Int-E; CD71^{+}CD235^{high}), and late (Late-E; CD71^{low}CD235^{high}) erythroblast, based on surface cell marker expression. AEL patient-derived expression signatures clustered in an unsupervised manner into the 4 "Merryweather-Clark groups" according to the differentiation stage. (D) PCA of AEL patient samples, colored according to differentiation stage groups defined by Merryweather-Clarke et al. (E) Heatmap of correlations between AEL samples from Iacobucci et al and the Merryweather-Clarke et al data set, clustered in an unsupervised manner into 4 groups according to differentiation stage: CFU-E (CD71^{+}CD235^{-}), proerythroblasts (Pro-E; CD71^{+}CD235^{low}), intermediate (Int-E; CD71^{+}CD235^{high}), and late (Late-E; CD71^{low}CD235^{high}) erythroblast. (F) PCA of AEL samples from Iacobucci et al[36] colored according to differentiation stage groups found when compared with Merryweather-Clarke et al data set.

subgroup 2 harbored a *TET2* loss-of-function mutation and a *GATA1* mutation, predicted to encode the short isoform *GATA1s*. Additional mutations affected transcription factors (eg, *WT1*, *RUNX1*), epigenetic regulators (eg, *ASXL1*, *EP300*, *BCOR*), signaling mediators (eg, *NOTCH2*, *IL7R*), and other genes in this group of patients. Finally, subgroup 3 (n = 10, 30.4%) contained samples without *TP53* or epigenetic variants. On average, these AEL showed 1.60 mutations per sample, a significantly lower value than for subgroups 1 and 2 (Figure 1; supplemental Figure 1G-H). Overall, our data confirmed that AEL is a molecularly heterogeneous disease characterized by a high prevalence of genetic variants in *TP53* and epigenetic regulators comparable to other published cohorts.[11,36]

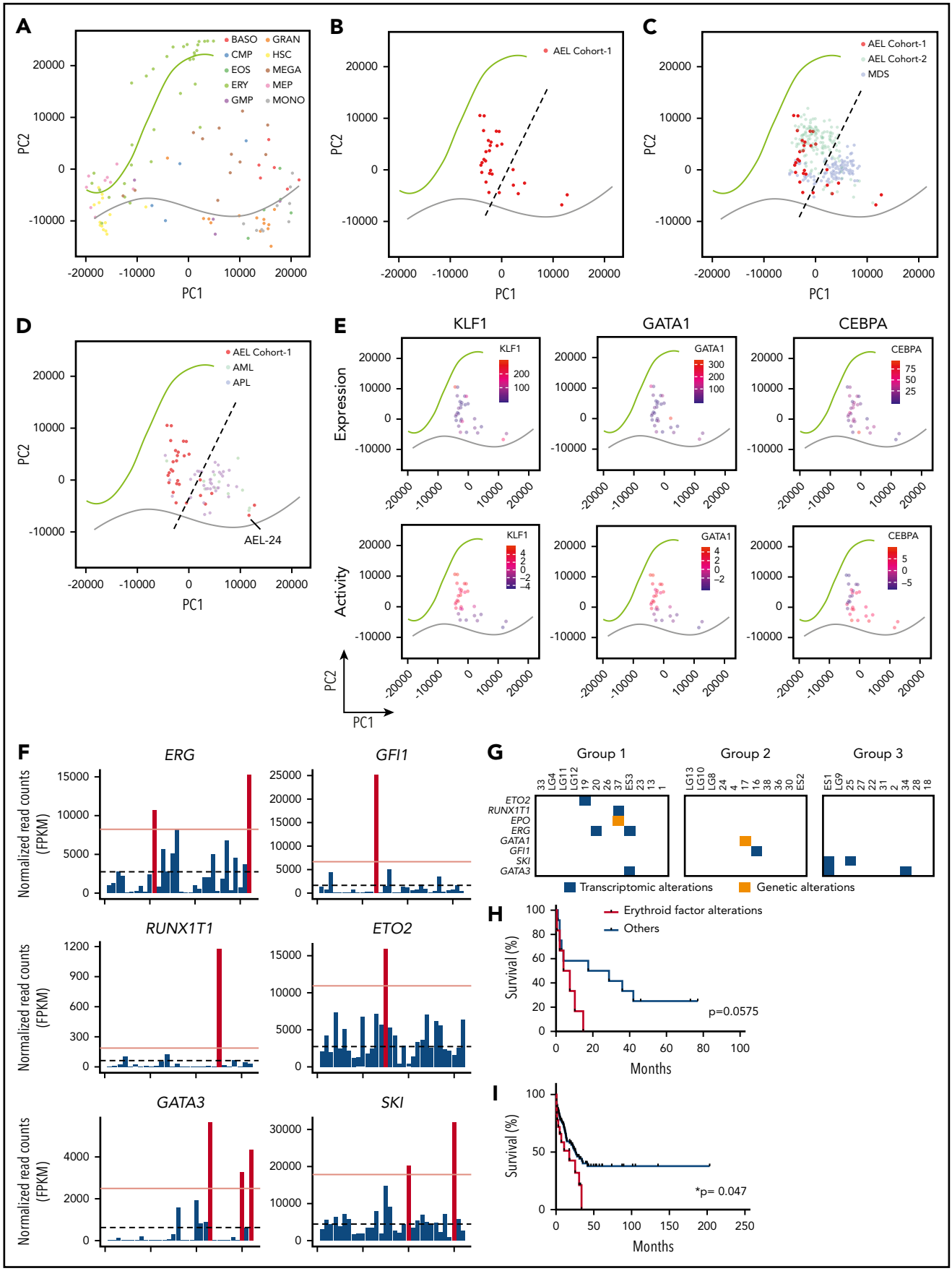


Figure 3.

AEL gene-expression signatures correlate with erythroid differentiation

As the heterogeneous genetic alterations did not provide any strong rationale for the erythroid phenotype of these leukemia, we investigated the erythroid feature by comparing gene-expression signatures (GES). PCA did not reveal any significant correlation between the GES and the 3 previously identified molecular subgroups (Figure 2A). Similarly, the percentage of erythroblasts in the patient BM at diagnosis was poorly reflected by GES (Figure 2B).

Because the AEL WHO classification is based on the number of erythroid and myeloid blasts present in the BM, we used a digital cellular deconvolution method (xCell) to compute a GES-based enrichment in erythroid, myeloid, and other hematopoietic cell types (supplemental Figure 2A).[37] The majority of samples had a prominent "erythrocyte" signature (n = 20), whereas some AEL samples presented a higher signal for the immature (MPP, CMP, GMP) or mature myeloid (monocyte, neutrophil) signatures (n = 9). To further explore the link between AEL transcriptomes and different stages of human erythroid maturation, we compared GES from the patients with those obtained experimentally after in vitro differentiation of human peripheral blood mononuclear cells into colony-forming unit erythroid (CFU-E; $CD71^{+}CD235^{-}$), proerythroblasts (Pro-E; $CD71^{+}CD235^{low}$), intermediate (Int-E; $CD71^{+}CD235^{high}$), and late erythroblasts (Late-E; $CD71^{low}CD235^{high}$)[38] and observed clustering according to these maturation stages (Figure 2C-D). Importantly, transcriptomes from an independent larger cohort of AEL patients[36] clustered similarly (Figure 2E-F).

Together, AEL gene expression programs are influenced by the erythroid differentiation stages rather than by the presence of particular genetic lesions, suggesting that the erythroid identity in human AEL relates to the cellular origin and the activity of transcriptional regulators driving cellular differentiation.

A transcriptome-based space maps AEL, MDS, and other AML to erythroid- and myeloid-lineage trajectories

As the 2016 WHO classification assigns most cases previously diagnosed as AEL to MDS or other AML,[7] we aimed at designing a transcriptome-based space that is able to distinguish AEL from MDS and other non-AEL AML subtypes. To this end, we retrieved cellular signatures from the DMAP database[32] and computed erythroid and myeloid differentiation expression trajectories (Figure 3A). As expected, relative differential gene expression clustered our AEL samples between the erythroid and myeloid trajectories. The majority of cases (n = 25) mapped closer to the erythroid axis whereas the rest mapped closer to the myeloid trajectory (n = 7) and closer to MDS transcriptomes (Figure 3B-C). Likewise, the samples from a recently published large AEL cohort[36] mostly clustered apart from MDS samples[39,40] (Figure 3C) and apart from non-M6 AML samples[41] (Figure 3D). Notably, AEL samples mostly projected between HSC and mature erythroid cells supporting that only part of the maturation-associated erythroid program is expressed in these samples. Interestingly, among our AEL samples that mapped closer to the myeloid axis and other AML samples, sample 24 showed a high expression of *SPI1* (Figure 3D; supplemental Figure 2B) also seen in other AML subtypes and was actually independently reclassified as AML-M5 by clinicians during the course of this study. These data support the idea that the transcriptional programs of the majority of AEL cases differs from those of MDS and other AML subtypes. They also support the existence of an overlapping continuum between these entities and the WHO-2016 reclassification of some AEL cases as AML-NOS.

Expression and activities of erythroid regulators in AEL

Myelo/erythroid differentiation is controlled by expression and activity of a relatively small group of transcription factors. Using the ARACNe and VIPER packages[42-44] and a large data set from human healthy progenitor cell transcriptomes,[45] we computed the activity of transcription factors and inferred lists of putative target genes (supplemental Figure 3A-B). Interestingly, we observed a gradual decrease in expression of erythroid transcription factors (eg, *KLF1*, *GATA1*, *NFE2*, *TAL1*, *NFIA*) and their predicted activity when going from the erythroid to the myeloid trajectories and an inverse correlation with myeloid factors (eg, *CEPBA* and *SPI1*) (Figure 3E; supplemental Figure 3C). This finding indicates that AEL is characterized by the transcriptional proximity to the normal erythroid lineage trajectory and by the relative activity of master transcription factors that control erythroid differentiation.

These data led us to hypothesize that some AEL cases might be driven by aberrant expression and activity of erythroid transcription factors. We focused on factors known to be predominantly expressed during erythroid differentiation and/or to control the activity of the GATA1 erythroid master regulator.[14] Using a threshold of fourfold higher expression level than the average, we observed that some AEL patients indeed expressed abnormally high levels of *ERG* (n = 2), *GFI1* (n = 1), *RUNX1T1* (= 1), and *ETO2* (n = 1) (Figure 3F). *GATA3*, which enforced expression previously resulted in erythroid bias,[46] was also highly

Figure 3. Transcriptome-based mapping and aberrant erythroid regulators in AEL. (A) PCA of data from Differentiation Map (DMAP)[33] with regression to cell types in an erythroid and myeloid compartment. The regression line fits to erythroid (green) and myeloid (gray) cells in the PCA space of genes significantly (FDR > 0.05, LogFC > 2) segregating each hematopoietic population: basophiles (BASO), common myeloid progenitor cells (CMP), eosinophil (EOS), erythrocytes (ERY), granulocyte-monocyte progenitor cells (GMP), granulocytes (GRAN), hematopoietic stem cells (HSC), megakaryocytes (MEGA), megakaryocytes-erythroid progenitor cells (MEP), and monocytes (MONO). (B) PCA with regression lines from plot (A) with projection of AEL patient samples. (C) PCA with regression lines from plot (A) with projection of AEL patient samples from our cohort (AEL cohort 1), from Iacobucci et al[36] (AEL cohort 2) and MDS samples.[39,40] (D) PCA with regression lines from plot (A) with projection of our AEL patient samples, AML and APL samples from the Blueprint consortium database. (E) PCA with regression lines from plot (A) with projection of AEL patient samples colored with KLF1, GATA1, and CEBPA expression and predicted activity. (F) Histogram representation of *ERG*, *GFI1*, *RUNX1T1*, *ETO2*, *GATA3*, and *SKI* gene expression in AEL patients. Positive patient samples (red bars) were defined as presenting an expression above the threshold set as fourfold the average of AEL samples. Dotted bars represent the average expression of AEL samples. (G) Table indicate patient samples presenting with genetic alteration (orange) or transcriptional alteration (blue) of GATA1-associated genes[67-70] in 3 molecular subgroups of AEL: TP53-mutated, epigenetics, and others. (H) Kaplan-Meier survival plot of AEL patients grouped according to the presence (or absence) of genetic or transcriptional alterations defined in panel G. *P* value using log-rank Mantel-Cox test is indicated. (I) Kaplan-Meier survival plot of AEL patients from Iacobucci et al[36] and grouped according to the presence (or not) of genetic or transcriptional alterations defined in panel G. *P* value using log-rank Mantel-Cox test is indicated.

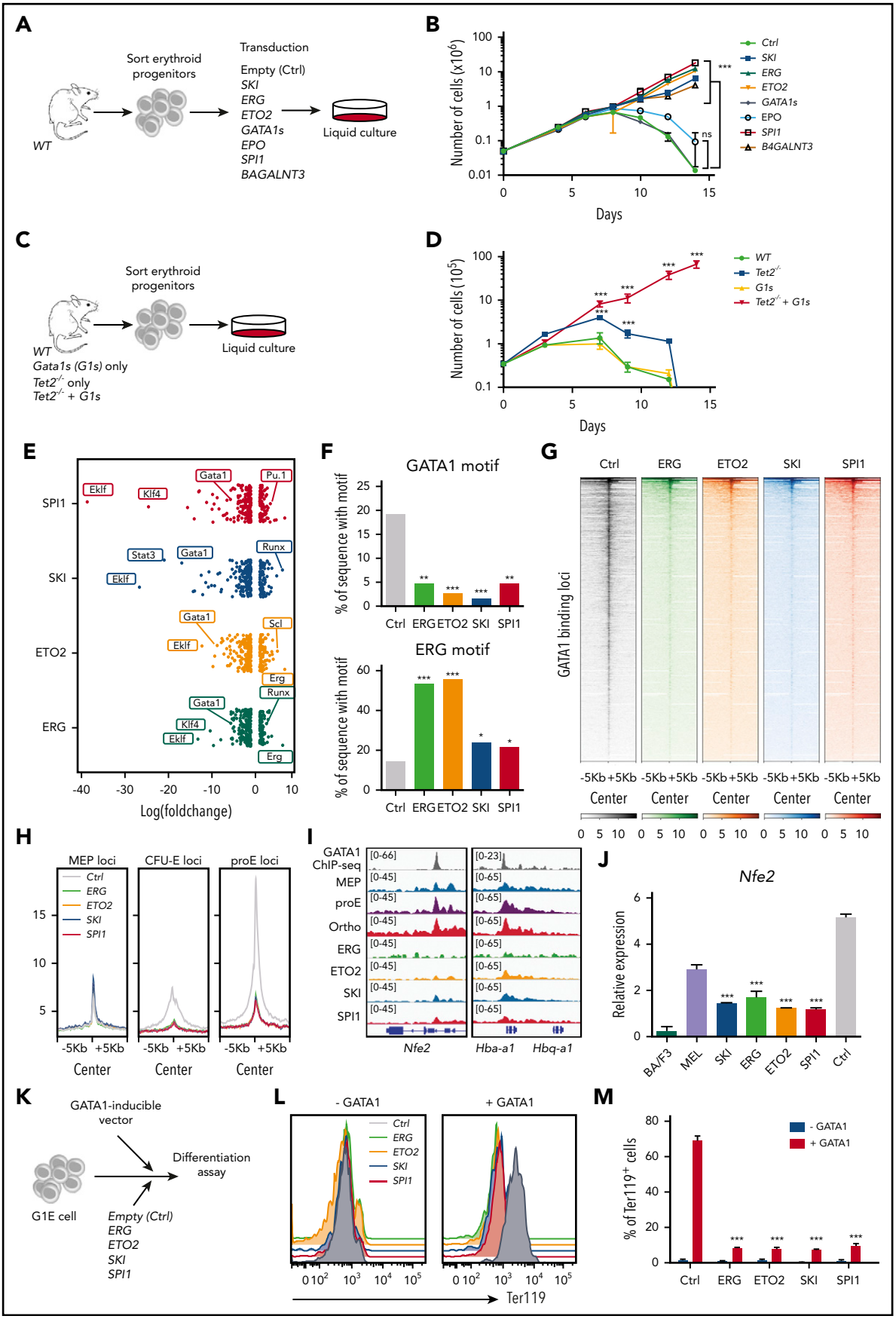


Figure 4.

expressed in 3 samples. Notably, we also found high expression of the transcriptional corepressor *SKI* (v-Ski avian sarcoma viral oncogene homolog) in 2 patients from molecular subgroup 3 (Figure 3F). Interestingly, *v-Ski* was previously reported to transform chicken erythroid cells, and to directly interact with *GATA1* to repress erythroid differentiation.[47-50] Our findings suggest that *SKI* not only influences experimental erythroid differentiation but could also contribute to human AEL pathogenesis.

The search for fusion transcripts using RNA-seq data revealed additional alterations, including 1 in-frame *BCR-ABL1* fusion gene (expected in this secondary-to-CML sample) and 3 novel out-of-frame fusion transcripts, notably 2 of them in a *TP53*-mutated context (supplemental Figure 4A-B). Sample 37 harbored an out-of-frame fusion of *YWHAE* (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein ε) with *EPO* and sample ES3 showed an out-of-frame fusion of *HSD17B11* (hydroxysteroid 17-β dehydrogenase 11) with *B4GALNT3* (β-1,4-*N*-acetyl-galactosaminyltransferase 3). These fusions were associated with ectopic expression of *EPO* and *B4GALNT3*, respectively. Interestingly, the *YWHAE-EPO*$^{+}$ sample also showed high *EPOR*, suggesting an autocrine mechanism of proliferation/survival in this case (supplemental Figure 4C). Although the role of *B4GALNT3* in human erythropoiesis remains unclear, an out-of-frame fusion leading to overexpression of *B4GALNT3* has been previously reported in thyroid carcinoma.[51] Sample ES1 from subgroup 3 presented with an out-of-frame fusion targeting the middle of the *DNMT3B* locus and associated with lower *DNMT3B* expression as compared with other samples, supporting *DNMT3B* inactivation. Notably, *Dnmt3a* and *Dnmt3b* expression were previously reported to be tightly controlled during erythroid maturation in mice.[52]

Overall, genetic and transcriptional alterations in erythroid regulators, including physical or functional interactors of the GATA1 transcriptional complexes were found in 9 of 33 patients (27%) (Figure 3G). Notably, these patients showed a trend toward poorer overall survival (Figure 3H), which became significant upon analysis of a larger AEL data set (Figure 3I).[36] Together, transcriptome analysis revealed that the majority of AEL samples are significantly different from MDS and other AML subtypes and that AEL frequently presents with epigenomic alterations that converge on factors interfering with GATA1 activity.

Overexpression of AEL-associated GATA1-interfering factors transforms mouse erythroid progenitors

To functionally test whether the aberrant expression of GATA1-interfering factors identified in AEL samples may contribute to the transformation of the erythroid lineage, we explored the consequences of ectopic expression of *SKI*, *ERG*, *ETO2*, *GATA1s*, *EPO*, *SPI1*, and *B4GALNT3* on murine erythroid progenitors. FACS-purified KIT^{+}CD71^{+}Ter119^{+} cells were transduced with retroviruses encoding these genes and grown in vitro (Figure 4A). In contrast to vector-transduced controls that proliferated for only ~7 days, ectopic expression of *ERG*, *SPI1*, *ETO2*, *SKI*, and *B4GALNT3* significantly maintained proliferation of erythroid cells presenting with an immature CD71^{+}KIT^{+}Ter119^{-} phenotype and a proerythroblast morphology for >30 days (Figure 4B; supplemental Figure 5A). Although the precise comparison between the overexpression level observed in human AEL samples and those achieved in murine models is technically challenging in this setting, a similar range of overexpression was observed for *ERG*, *ETO2*, *SKI*, and *B4GALNT3* (supplemental Figure 5B). Notably, ectopic expression of *EPO* or *GATA1s* alone was not sufficient to expand erythroblasts longer than 10 days (Figure 4B).

To address whether a cooperation between *Tet2*-inactivating and *Gata1s* mutations could transform erythroblasts in vitro, we purified erythroid progenitors from wild-type, *Tet2*-deficient[27] (thereafter named *Tet2*$^{-/-}$), *Gata1*$^{\Delta e2}$ knock-in[28] (thereafter named *Gata1s*), and double *Tet2*$^{-/-}$+Gata1s transgenic mice and compared their proliferation (Figure 4C). Although *Gata1s*- or *Tet2*$^{-/-}$-only erythroblasts did not expand for >10 to 15 days, *Tet2*$^{-/-}$+*Gata1s* erythroblasts proliferated >2 months and exhibited an erythroid morphology (Figure 4D; supplemental Figure 5C).

Figure 4. AEL-associated transcription factors transform erythroid progenitors and impair GATA1 activity. (A) Experimental design: mouse erythroid progenitors (CD71^{+}Ter119^{+}KIT$^{+/low}$) were sorted from lineage marker–depleted BM, transduced with retrovirus encoding *SKI*, *ERG*, *ETO2*, *GATA1s*, *EPO*, *SPI1*, or *B4GALNT3* combined with *IRES-GFP* (GFP expression is a surrogate marker for transgene expression), or an empty vector (Ctrl), and maintained in StemSpan SFEM with cytokines (mSCF, mIL3, mIL6, hEPO, cholesterol, and dexamethasone). (B) A total of 5 × 10^{4} transduced mouse erythroid progenitors were cultured for 15 days and viable cells were enumerated by trypan-blue exclusion. Mean plus or minus SD number of cells is represented. *SKI* (n = 5), *ERG* (n = 4), *ETO2* (n = 3), *GATA1s* (n = 3), *EPO* (n = 3), *SPI1* (n = 6), *B4GALNT3* (n = 3), or empty vector (Ctrl) (n = 5). (C) Experimental design: erythroid progenitors (CD71^{+} Ter119^{+} KIT$^{+/low}$) from WT, *Tet2*$^{-/-}$, *Gata1s* (G1s), or *Tet2*$^{-/-}$+ *Gata1s* (*Tet2*$^{-/-}$+*G1s*) mice were sorted from lineage marker–depleted BM and maintained for 15 days in StemSpan SFEM with cytokines (mSCF, mIL3, mIL6, hEPO, cholesterol, and dexamethasone). (D) A total of 3.5 × 10^{4} sorted erythroblasts from WT, *Tet2*$^{-/-}$, *Gata1s* (G1s), or *Tet2*$^{-/-}$+ *Gata1s* (*Tet2*$^{-/-}$+*G1s*) mice were grown in liquid cultures over 15 days and viable cells were counted by trypan-blue exclusion. Mean plus or minus SD (n = 3) is shown. (E) Dot-plot showing the log(fold changes) of the percentage of sequence for a given motif found under ATAC-seq peaks, between normal and transformed erythroblasts (expressing ERG, ETO2, SKI, and SPI1). (F) Histogram representation of the percentage of sequence with GATA1 (top) or ERG (bottom) motif found under ATAC-seq peaks of normal (Ctrl) and transformed erythroblasts. Statistical differences were calculated using the χ^2 test. For GATA1 motif, Ctrl vs ERG: $P = .0013$; Ctrl vs ETO2: $P = .0001$; Ctrl vs SKI: $P < .00001$; Ctrl vs SPI1: $P = .0032$. For ERG motif, Ctrl vs ERG: $P = .00001$; Ctrl vs ETO2: $P < .00001$; Ctrl vs SKI: $P = .0498$; Ctrl vs SPI1: $P = .0415$. (G) Heatmap representing the hierarchical clustering of ATAC-seq signals, performed using normal (Ctrl) and transformed erythroblast by either ERG, ETO2, SKI, or SPI1, focused on GATA1-binding sites in normal Ter119^{+} erythroblast (ENCODE). Heatmaps were focused on peak centers with ±5 kb. (H) Profile plot representing ATAC-seq signals performed using normal (Ctrl) and transformed erythroblast expressing either ERG, ETO2, SKI, or SPI1 on ATAC-seq specific peaks previously identified in mouse MEP, CFU-E, or proerythroblasts (proE).[35] Profile-plot were focused on peaks centers with ±5 kb. (I) Visualization of GATA1 ChIP-seq peaks performed in normal erythroblast (ENCODE, first lane, gray) and ATAC-seq peaks of normal MEP, proerythroblast, orthochromatic erythroblast, and transformed erythroblasts expressing ERG, ETO2, SKI, or SPI1, focused on *Nfe2*, *Hba-a1*, and *Hbq-a1* genes, using IGV software (v 2.3.88). (J) Quantitative *Nfe2* mRNA expression measured by RT-qPCR in WT erythroblast (Ctrl) or erythroblast transformed with either *SKI*, *ERG*, *ETO2*, or *SPI1* overexpression. Expression levels were also compared with the erythroleukemia MEL or the Ba/F3 myelolymphoid cell line. (K) Experimental design: G1E cell lines were cotransduced with a GATA1 doxycycline-inducible vector and with *MSCV*-vector expressing either ERG, ETO2, SKI, SPI1, or empty. (L) Flow cytometry histogram analysis of Ter119 expression in G1E cells expressing ERG (green), ETO2 (orange), SKI (blue), SPI1 (red), or empty control (gray), without (left) or with (right) induction of GATA1 expression. (M) Histogram representation of Ter119 expression detected by flow cytometry analysis of G1E cells expressing ERG, ETO2, SKI, SPI1, or empty control, without (gray) or with (red) induction of GATA1 expression. Statistical significance (in panels B, D, J, and M) is indicated as *P* values (Student *t* test except when otherwise specified). *$P < .05$; **$P < .01$; ***$P < .001$.

Collectively, these data demonstrate that ectopic expression of *ERG*, *ETO2*, *SKI*, or the combination of *Tet2* loss-of-function and *Gata1s* mutations can efficiently immortalize murine erythroblasts in vitro.

Aberrantly expressed AEL-associated transcriptional regulators interfere with GATA1 chromatin accessibility and function

To better understand how aberrantly expressed transcription factors (ERG, ETO2, SKI, SPI1) immortalize erythroblasts, we studied chromatin accessibility by ATAC-seq. Motif analysis revealed a lower representation of GATA1 and KLF1 (also known as EKLF) motifs and a global increase of ETS-associated motifs (including ERG and SPI1 motifs) in all transformed cells compared with Ctrl (Figure 4E-F; supplemental Figure 5D).

To investigate chromatin accessibility at erythroid GATA1-binding sites and at sites that are regulated at specific stages of healthy erythroid differentiation, we interrogated previously published data sets.[35] Interestingly, ERG-, ETO2-, SKI-, and SPI1-overexpressing erythroblasts showed a decreased chromatin accessibility at erythroid GATA1-binding sites compared with vector-transduced control cells (Figure 4G). ERG-, ETO2-, SKI-, and SPI1-expressing erythroblasts also showed a decreased chromatin accessibility at sites open in healthy CFU-E and proerythroblasts, whereas there was no difference in chromatin accessibility at sites open in less differentiated MEP cells (Figure 4H; supplemental Figure 5E). Notably, these observations correlated with a decreased chromatin accessibility and mRNA expression at GATA1-controlled erythroid genes such as nuclear factor erythroid 2 (*Nfe2*) or hemoglobin A1 (*Hba-a1*) (Figure 4I; supplemental Figure 5F).

Finally, we investigated the consequence of aberrant expression of these transcription factors on GATA1 activity in the GATA1-deficient G1E erythroid cell line in which terminal erythroid maturation can be induced by expression of exogenous *Gata1*[29] (Figure 4K). As expected doxycycline-induced *Gata1* expression restored G1E erythroid differentiation with upregulation of Ter119 expression (supplemental Figure 5G). In contrast, ectopic expression of ETO2, ERG, SKI, or SPI1 significantly inhibited GATA1-induced differentiation (Figure 4L-M).

Collectively, these data indicate that aberrant expression of ETO2, ERG, SKI, and SPI1 functionally interferes with GATA1 activity and restrains GATA1-dependent erythroid differentiation consistent with impaired differentiation observed in primary human AEL cells.

In vivo modeling of AEL from immortalized erythroblasts

We used complementary strategies to model in vivo the leukemogenic potential of AEL-associated alterations of the different molecular subgroups. First, to ascertain that in vitro–transformed erythroblasts can induce disease in vivo, we injected them into irradiated syngeneic recipients. ERG-, ETO2-, SKI-, or *Tet2*$^{-/-}$+*Gata1s*-transformed, but not SPI1- or B4GALNT3- transformed, cells rapidly induced a fully penetrant fatal disease characterized mostly by the accumulation of CD71^{+}Ter119^{-} and few CD71^{+}Ter119^{+} blasts lacking expression of myeloid markers (supplemental Figure 6A-C). Histopathological analysis of symptomatic mice showed infiltration of BM, spleen, and livers by erythroblasts expressing nuclear GATA1 (supplemental Figure 6D-E). These results show that some epigenomic alterations found in human AEL have the potential to immortalize murine erythroblasts, which can then induce an AEL-like disease.

In vivo modeling of functional cooperation between AEL-associated alterations

Next, to define in vivo transforming capacities starting from healthy hematopoietic progenitors, we obtained oncogene-expressing Lin^{-} HSPCs (either by retroviral transduction or by breeding transgenic models) and assessed disease development upon engraftment into lethally irradiated recipients. Based on our observations that AEL subgroups 1 and 2 showed frequent cooccurrence of mutations (Figure 1B) and that GATA1 activity is targeted either directly (*TET2*+*GATA1s* and *IDH2*+*GATA1* mutations in another cohort[36]) or through associated factors (*TP53*+*ERG*^{high}), we investigated these 2 representative potential functional cooperation schemes.

Previous work has shown that both *Tet2* loss-of-function and *Gata1s* alter erythroid differentiation but do not induce bona fide leukemia in vivo alone[27,28,53-57] (supplemental Figure 7A). To address functional cooperation, we transplanted *Tet2*$^{-/-}$+*Gata1s* Lin^{-} HSPCs into lethally irradiated recipients (Figure 5A). As opposed to recipients of *Tet2*$^{-/-}$-only cells, recipients of *Tet2*$^{-/-}$+*Gata1s* cells developed a rapid and fully penetrant lethal disease associated with high WBC, anemia, thrombocytopenia, and splenomegaly (Figure 5B-C; supplemental Figure 7B). Flow cytometry analysis indicated that leukemic blasts were primarily CD11b^{+}Gr1^{+} myeloid cells (Figure 5D) and histopathological analysis confirmed that BM and spleen were highly infiltrated by blasts with myeloid features (supplemental Figure 7C). Notably, we also observed emperipolesis that was previously described in murine GATA1s models. Together, these data demonstrate that *Tet2* loss of function cooperates with *Gata1s* mutation to promote an AML-like phenotype in vivo.

TP53-mutated AEL samples are associated with other alterations, including aberrant expression of the transcription factor ERG (Figure 3F). Most AEL-associated TP53 alterations are DNA-binding missense mutations[13] including *TP53*^{R248Q}.[26] To address functional cooperation, we transplanted *TP53*^{R248Q} Lin^{-} HSPCs transduced with an *ERG*-expressing retrovirus. Because ectopic *ERG* expression in adult murine hematopoiesis was shown to primarily induce T-cell leukemia[58,59] (supplemental Figure 7D-E), we assessed the long-term consequences of high *ERG* expression specifically in erythroid progenitors by transplanting purified ERG-transduced (GFP^{+}) wild-type or *TP53*^{R248Q} erythroblasts obtained from primary recipients, into secondary recipients (Figure 5E). All recipients of *TP53*^{R248Q} erythroblasts overexpressing *ERG* developed a fatal leukemia with a median survival of 60 days, whereas recipients of *ERG*-expressing wild-type erythroblasts developed disease after 4 months (Figure 5F). The *TP53*^{R248Q}+*ERG*-induced disease was characterized by anemia, thrombocytopenia (Figure 5G), and the accumulation of CD71^{+}Ter119^{+} erythroid and to a lesser extent CD11b^{+}Gr1^{+} myeloid progenitors in the BM (Figure 5H), with infiltration in spleen and liver (supplemental Figure 7F). These data indicate that an AEL-associated *TP53* DNA-binding mutation cooperates with aberrantly high *ERG*

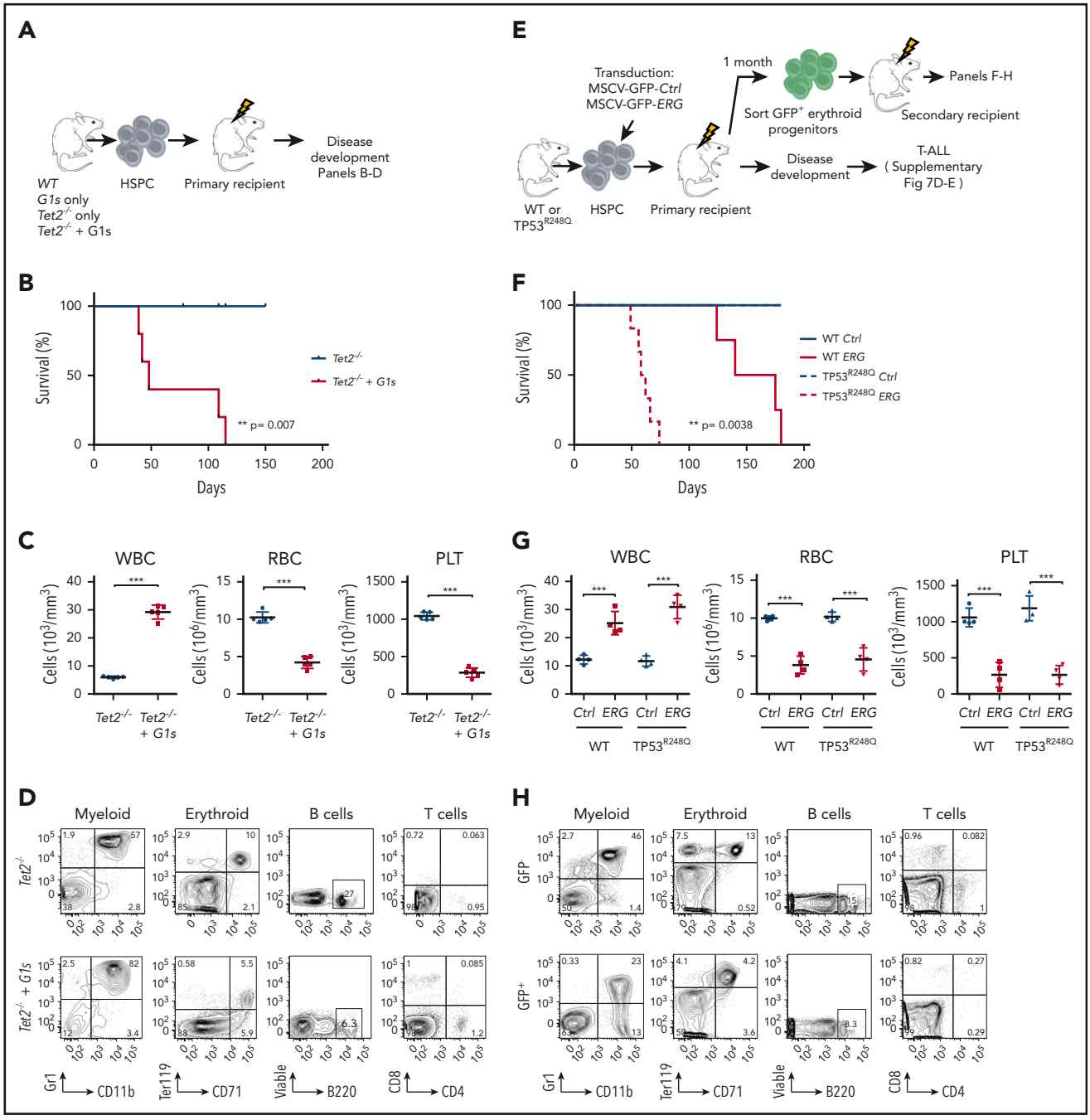


Figure 5. In vivo modeling of functional cooperation between AEL-associated alterations. (A) Experimental design: mouse CD45.2$^+$ HSPCs from *WT*, *G1s*, *Tet2*$^{-/-}$, or *Tet2*$^{-/-}$+ *G1s* mice were sorted and injected into sublethally irradiated CD45.1$^+$ recipients followed for disease development. (B) Kaplan-Meier plot of diseased recipients of HSPC cells from *Tet2*$^{-/-}$ (n = 5) or *Tet2*$^{-/-}$+*G1s* (n = 5) mice. (C) Peripheral blood counts (WBC, RBC, and PLT) of immunodeficient mice engrafted with HSPC cells from *Tet2*$^{-/-}$ (n = 5) or *Tet2*$^{-/-}$+*G1s* (n = 5) mice. (D) Flow cytometry analysis of myeloid cells (Gr1$^+$CD11b$^+$), erythroid progenitors (CD71$^+$ Ter119$^+$), B cells (B220$^+$), and T cells (CD4$^+$CD8$^+$) gated for CD45.2$^+$ donor cells from *Tet2*$^{-/-}$ or *Tet*$^{-/-}$+*G1s* mice. (E) Experimental design: HSPCs from WT or TP53^{R248Q} mice were retrovirally transduced with *ERG* or an empty vector (*Ctrl*), containing both an *IRES-GFP* expression cassette, and transplanted into WT primary recipients. Top, Primary recipients of *ERG*-transduced cells developed T-ALL (supplemental Figure 6). Bottom, to investigate cooperation specifically in erythroid progenitors, *ERG*-expressing erythroid progenitors (GFP$^+$CD71$^+$Ter119$^+$KIT$^{low/+}$) were sorted from primary recipient mice and transplanted into secondary recipients to follow disease development. (F) Kaplan-Meier plot of diseased secondary recipient mice (described in panel A). WT (*Ctrl* or *ERG*): n = 4, TP53^{R248Q} (*Ctrl* or *ERG*): n = 6. (G) Peripheral white blood cell (WBC), red blood cell (RBC), and platelet (PLT) counts of diseased mice transplanted with WT or TP53^{R248Q} HSPC retrovirally transduced with *ERG* or empty vector (*Ctrl*). WT (*Ctrl* or *ERG*): n = 4, TP53^{R248Q} (*Ctrl* or *ERG*): n = 4. (H) Flow cytometry analysis of myeloid cells (CD11b$^+$Gr-1$^+$), erythroid progenitors (CD71$^+$Ter119$^+$), B cells (B220$^+$), and T cells (CD4$^+$CD8$^+$) from diseased secondary recipients of TP53^{R248Q} HSPCs retrovirally transduced with *ERG* gated for GFP$^+$ or GFP$^-$ cells. Statistical significance (in panels B, C, F, and G) is indicated as *P* values (Student *t* test except when otherwise specified). *$P < .05$; **$P < .01$; ***$P < .001$.

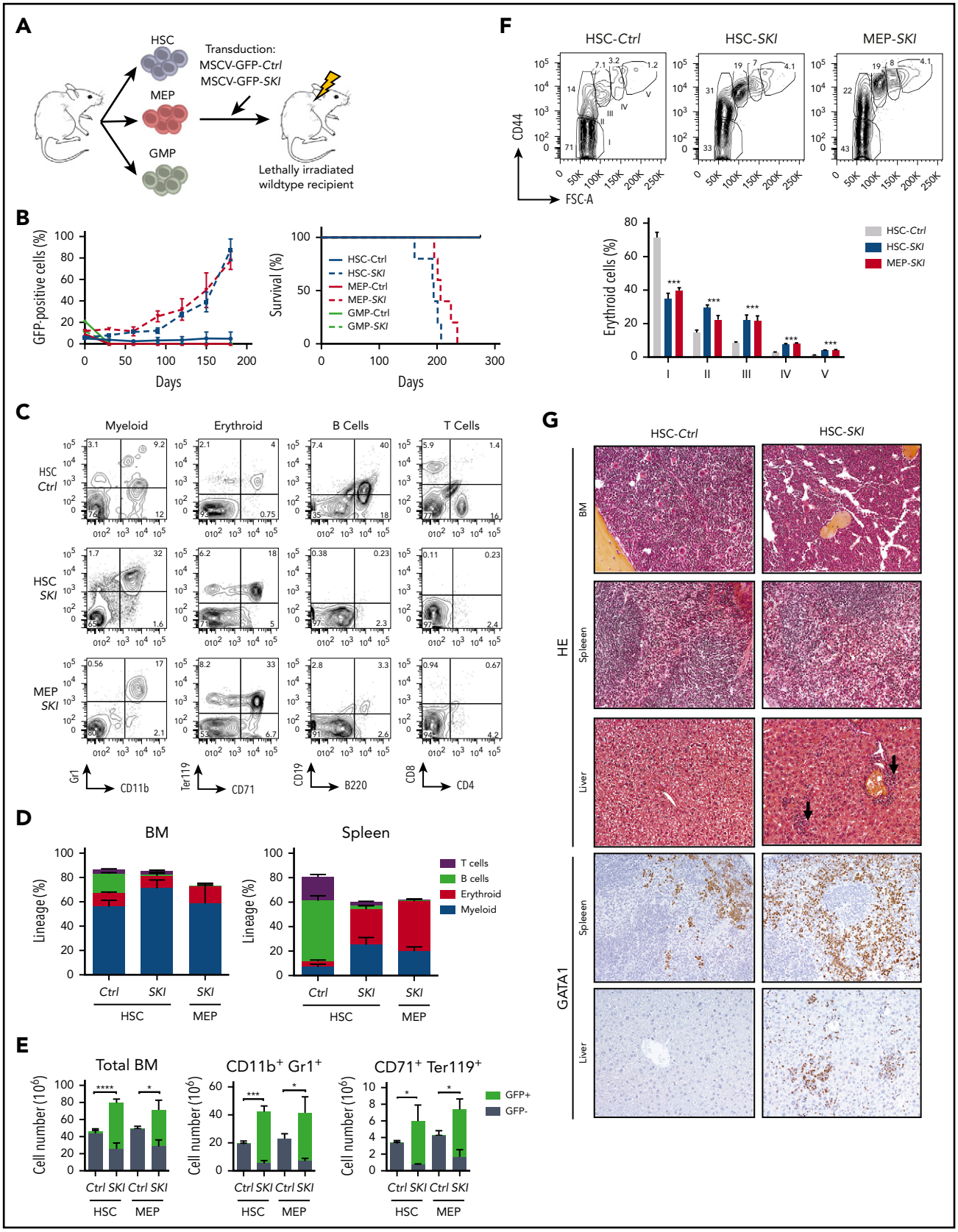


Figure 6. In vivo modeling of erythroid transformation by aberrant SKI expression. (A) Experimental design: hematopoietic stem cells (HSC), granulocyte-monocyte progenitors (GMP), and megakaryocyte-erythroid progenitors (MEP) were sorted from WT mice, retrovirally transduced with *SKI* or an empty vector (*Ctrl*) carrying an *IRES-GFP* expression cassette, and injected into lethally irradiated recipients. (B) Left, GFP$^+$ cells (percentage) in the BM of transplanted mice. Mean plus or minus SD is shown (n = 5 per group). Right, Kaplan-Meier plot of disease symptoms in transplanted mice. Only mice transplanted with *SKI*-transduced HSCs and MEP developed disease. (C) Flow cytometry

expression to enhance the proliferative capacity of erythroid progenitors leading to leukemia with several features of the human disease.

Taken together, these results demonstrate that mutation associations in human AEL functionally cooperate to induce murine AML-like leukemia in vivo. Notably, although both combinations could readily induce an AEL-like phenotype when expressed in erythroid-restricted progenitors, their expression in HSPCs led to other mostly mixed leukemia phenotypes. Therefore, these data also suggest that the target cell in which these mutations are active impacts the disease phenotype.

In vivo modeling of erythroid transformation by aberrant SKI expression

To further explore the relevance of the cell context for consequences of the transcriptional alterations, we investigated in vivo disease development upon high expression of *SKI*, which was observed in 2 AEL samples of subgroup 3 (Figure 3G). First, we confirmed previous work[49] showing that transplantation of Lin^- HSPCs retrovirally overexpressing *SKI* induced a lethal disease (supplemental Figure 8A) characterized by anemia, thrombocytopenia, and increased myeloid cells in the periphery (supplemental Figure 8B) associated with hypercellular BM and spleens showing high percentage of mostly myeloid or erythroid transgene-expressing cells (supplemental Figure 8C-E). Notably, GFP detection in all 3 myeloid, erythroid, and platelet lineages (supplemental Figure 8F) suggested that *SKI* overexpression may affect early multipotent stem or progenitor cells.

To investigate whether the transforming activity of SKI depends on the hematopoietic target cell, we purified, transduced and transplanted long-term multipotent HSC, erythroid-enriched (MEP), or myeloid-committed granulocyte-macrophage (GMP) progenitors (Figure 6A). Three weeks posttransplant, transduced cells were detectable in the blood for all groups but engraftment in the BM was only observed in recipients from *SKI*-transduced HSCs and MEPs that later developed symptomatic diseases characterized by anemia, thrombocytopenia, (supplemental Figure 8H) and presenting with both $CD11b^+Gr1^+$ myeloid and $CD71^+Ter119^+$ erythroid features (Figure 6B-E). *SKI*-transduced GMP recipients did not develop disease. Symptomatic recipients of HSCs or MEPs showed an increase in basophilic, polychromatophilic, and orthochromatic erythroblasts and in reticulocytes associated with a relative decrease in mature red cells (Figure 6F), suggesting that *SKI* delays but does not fully block erythroid differentiation in vivo. Histopathological analyses confirmed BM hypercellularity and revealed infiltration of erythroid cells in the spleen and liver (Figure 6G; supplemental 8I).

Taken together, these data indicate that high *SKI* expression transforms HSCs and MEPs, but not myeloid-restricted progenitors like GMPs. Although aberrant *SKI* expression in erythroid-restricted progenitors leads to pure erythroid proliferation, expression in more immature HSPCs resulted in increased self-renewal capacity with aberrant differentiation toward both myeloid and erythroid lineages indicative of an AEL/MDS-like disease.

Discussion

AEL is an aggressive human cancer, often difficult to diagnose due to its close resemblance to other forms of hematopoietic malignancies presenting with variable compositions of cells with erythroid features, like MDS or certain AML subtypes. Here, we describe novel features of AEL that shed light on the pathophysiology of this disease. First, our data indicate that the majority of human AEL exhibit a unique erythroid transcriptional signature that differs from those found in patients with non-M6 AML or MDS without prominent erythropoiesis. Second, aberrant expression of various transcriptional regulators known to modulate GATA1 activity was frequently found in AEL and may represent a common molecular module that controls erythroid differentiation. Third, in vivo models demonstrate that the relative composition of the erythroid and myeloid features is strongly dependent on the hematopoietic target cell in which a driving oncogene is expressed, providing a basis for a better understanding of the highly heterogeneous clinical appearance of AEL.

The genomic lesions described here are in line with previous reports, including the largest genetic landscape study of human erythroleukemia to date,[9-12,36] and support classification of AEL patients into molecular subgroups. In our study, 3 subgroups were identified, including patients with *TP53* mutations (36.3% of cases), patients with mutations in epigenetic regulators previously associated with clonal hematopoiesis of indeterminate potential (CHIP) and MDS (eg, *DNMT3A*, *TET2*, and *IDH1/2* mutations) (33.3% of cases) and another group of patients presenting with none of these recurrent alterations (30.4% of cases). Although TP53 mutations and epigenetic mutations are not mutually exclusive, their frequencies within AEL samples are similar.[36] In our sequenced AEL cases, we did not detect the other recently described subgroups, including those with *NUP98*, *KMT2A*, and other in-frame fusions,[36] which could reflect the limited number of pediatric patients in our cohort. Also, consistent with the frequent association between *FLT3* mutations and *NPM1* or *KMT2A* alterations,[33] our patient cohort lacked samples with *FLT3* or *NPM1* mutations. For samples sequenced with RNA-seq only, we cannot exclude the possibility that some structural variants or low expressed mutated transcripts remained undetected.

In contrast to previous studies, we also found out-of-frame fusion transcripts associated with altered expression of 1 of the partner

Figure 6 (continued) analysis of myeloid cells ($CD11b^+Gr\text{-}1^+$), erythroid progenitors ($CD71^+Ter119^+$), B cells ($B220^+$), and T cells ($CD4^+CD8^+$) gated for viable GFP^+ cells, in spleens of mice transplanted with HSC (*Ctrl* vs *SKI*) or MEP (SKI) cells. (D) Percentage of myeloid cells ($CD11b^+Gr1^+$), erythroid progenitors ($CD71^+Ter119^+$), B cells ($B220^+$), and T cells ($CD4^+CD8^+$) within viable GFP^+ cells, in HSC-Control (*Ctrl*), HSC-SKI, and MEP-SKI in primary mice BM and spleen. Mean plus or minus SD is shown (n = 5 per group). (E) Number of GFP^+ and GFP^- total BM cells, myeloid cells ($CD11b^+Gr1^+$), and erythroid progenitors ($CD71^+Ter119^+$) in the BM of diseased primary recipients. Mean plus or minus SD is shown (n = 5 per group). (F) Flow cytometry analysis of terminal erythroid differentiation in the BM of diseased HSC (*Ctrl* and SKI) and MEP-SKI recipients, determined by forward scatter (FSC-A) and CD44 expression gated for viable $Ter119^+$ cells. Bottom panel, Mean plus or minus SD (n = 5 per group) of the percentage of each population. (G) Histopathology analysis of BM spleen and liver of mice transplanted with HSC-*Ctrl* (left row) or HSC-*SKI* (right row) stained with hematoxylin-eosin (top 6 photomicrographs; original magnification ×20) or with a GATA1 antibody (bottom 4 photomicrographs; original magnification ×20).

genes. For example, the fusion between *YWHAE* and *EPO* in a *TP53*-mutated patient was associated with ectopic expression of *EPO*, and the concomitant high expression of *EPOR* suggested an autocrine EPO/EPOR-signaling mechanism.[15] Interestingly, alterations of multiple signaling intermediates, including downstream of EPO/EPOR, were recently found in up to 48% of human AEL samples[36] and acquired activating KIT mutations were also essential to induce a bona fide erythroleukemia in a transgenic murine model,[60] indicating the importance of signaling alterations for efficient oncogenic transformation of the erythroid lineage.

As the vast majority of the AEL-associated mutations are also found in a wide spectrum of human myeloid malignancies, it is essential to gain insights into their functional role in the erythroid phenotype that leads to a diagnosis of AEL. Together with the molecular alterations targeting *EPO* (*YWHAE-EPO* fusion) and the erythroid transcription factor *GATA1* found here, the recently reported *APLP2-EPOR* and *MYB1-GATA1* fusion genes[36] further support the relevance of alterations in erythroid master regulators as underlying the erythroid phenotype in some AEL cases. However, most AEL do not present with erythroid-specific genetic alterations. Our patient-based transcriptional data, together with chromatin accessibility and functional analyses in cellular and in vivo models revealed that at least 25% of AEL cases present with transcriptional alterations ultimately interfering with GATA1 activity through direct or functional interaction within the GATA1 transcriptional complexes (eg, aberrantly expressed *ETO2*, *ERG*, *SKI*, *SPI1*). Although some of these transcriptional alterations (eg, *ERG*^{high} expression) were recently reported to have a genetic bases,[61,62] the origin of some others remains to be determined (eg, *SKI*^{high}). As reported previously,[60,63] we noted that ectopic *SPI1* expression was sufficient to immortalize erythroblasts in vitro but not to induce the disease in vivo, supporting the idea that cooperating alterations that have yet to be identified are required in *SPI1*^{high} human leukemia.

Although the basis for the erythroid phenotype remains to be demonstrated in many cases, several epigenomic AEL alterations may also functionally converge on aberrant activity of erythroid master regulators. Indeed, a novel signaling pathway based on JAK2-mediated phosphorylation of TET2 leading to interaction with KLF1 was recently reported.[64] Combined TET2 and DNMT3A inactivation was also reported to upregulate expression of *KLF1* and *EPOR* in HSCs.[55] Therefore, the concomitant *TET2* and *DNMT3A* mutations observed in 2 AEL patients and the presence of *TET2* and *GATA1s* mutations in another AEL sample support a functional synergism between alterations of KLF1 and GATA1 transcriptional programs leading to differentiation blockage. Based on these observations, we hypothesize that the erythroid phenotype in AEL results from a cooperation between genetic and transcriptional alterations. As proposed for other subtypes of leukemia, interference with the activity of altered erythroid master regulators, for example, through targeting of critical protein-protein interactions may therefore represent promising therapeutic strategies for AEL.[65,66]

Our observations also have implications for the classification of AEL patients into molecular and/or prognosis subgroups. Comparative analysis of AEL expression signatures with normal erythroid and myeloid differentiation indicated that AEL is heterogeneously spread along a differentiation-associated trajectory with some patient samples clustering next to progenitors retaining myeloid features and other patient samples clustering closer to the erythroid trajectory. Also, although several oncogenes (eg, *SKI*) can transform restricted erythroid lineages, they led to mixed erythroid/myeloid hematopoietic malignancies upon expression in multipotent murine progenitors. These data indicate that the relative composition of myeloid vs erythroid elements at time of diagnosis is not solely based on the type of mutations but likely also reflects the type of progenitor targeted by these mutations. Notably, the relationship between gene-expression signatures and normal differentiation trajectories was not clearly visible when comparing the reported immunophenotypes of the blasts, and no correlation was found with the different molecular subgroups. These data strongly suggest that, in some AEL patients, the erythroid phenotype maybe initiated either by strong mutations that interfere with erythroid differentiation, or by mutations that provide advantages to erythroid-restricted progenitors. Alternatively, in others, the erythroid phenotype may originate from mutations in multipotent progenitors with a subsequent epigenetic drift toward the erythroid lineage.

Taken together, our work provided insights into the molecular mechanisms of the erythroid identity in AEL. Future studies need to resolve, likely at the single-cell level, the clonal genetic and epigenomic heterogeneous architecture in prospectively collected fresh samples as a further step toward the development of specific therapies.

Acknowledgments

The authors are grateful to Olivier Hermine, Françoise Moreau-Gachelin, Michaela Fontenay, and Françoise Pflumio for expert advice and useful discussions. The authors also thank Suzana Atanasoski (Basel, Switzerland) for providing the human *SKI* cDNA, and Stéphanie Ranga and Michael Finnegan for handling human AEL samples and helping with cDNA cloning. One patient's samples were handled, conditioned, and stored by the FILOtheque (no. BB-0033-00073), Tumor Bank of the French Innovative Leukemia Organization (FILO) Group (Cochin Hospital, Paris, France).

This work was supported by Institut National du Cancer (PLBIO-2014-176 and PLBIO-2018-169 [T.M.]), Ligue Contre le Cancer (PhD grant [A.F.], Équipe Labelisée [T.M.]), Institut National du Cancer (INCa)-PlanCancer "Soutien pour la Formation" 2009-2013 (C.I.), Cancéropôle Ile-de-France (2014-2017 [C.K.L.]), Sites de Recherche Intégrée sur le Cancer (SIRIC)-SOCRATE (INCa-DGOS-INSERM 6043 [T.M.] and INCa-DGOS-INSERM_12551 [O.A.B.]), Fondation pour la Recherche Médicale (C.I., Z.A. [FRM-ING20150532273], and C.K.L.), Fondation de France (FdF-00057925 [C.T. and T.M.]), and the Gustave Roussy Genomic Core Facility (Taxe d'Apprentissage TA2018-ALFA [A.F.]). J.S. was supported by grants from the Swiss Cancer League (KFS-3487-08-2014), the Gertrude von Meissner Foundation (Basel, Switzerland), the San Salvatore Foundation (201525; Lugano, Switzerland), the Wilhelm Sander Foundation (2017.035.1; Munich, Germany) and the Swiss National Science Foundation (SNF, 31003A_173224/1). U.M. was supported by the National Institutes of Health, National Cancer Institute (2RO1CA176647) and the Stony Brook Foundation (Carol Baldwin Foundation). P. Valent was supported by the Austrian Science Fund (FWF) grant F4704-B20 and a stem cell grant of the Medical University of Vienna. P. Vyas was supported by the Bloodwise and Children with Cancer Specialist Programme (grant 13001), the National Institute for Health Research (NIHR) Oxford Biomedical Research Fund, and the Medical Research Council Molecular Haematology Unit (MRC MHU; MC_UU_12009/11). E.A. was supported by the Fundación Hay Esperanza. C.M. was supported by the Leukemia & Lymphoma Society Translational Research Program, National Institutes of Health and National Cancer Institute Outstanding Investigator Award R35 CA197695, and the American Lebanese Syrian Associated Charities

of St. Jude Children's Research Hospital. E.S. was supported by Fondation pour la Recherche Médicale (Equipe FRM DEQ20180339221), the ATIP-Avenir Program (Plan Cancer), and Labex EpiGenMed (Investissements d'Avenir Program, reference ANR-10-LABX-12-01).

Authorship

Contribution: A.F., M.-R.P.-B., C.I., A.C., C.K.L., B.U., Z.A., C.T., S.T., B.L., V.D., S.M., L.G., E.S., J.S., and T.M. performed and analyzed experiments; A.F., M.-R.P.-B., F.O.B., and E.R. performed bioinformatics analyses; V.G.-B., A.K.-K., J.M., C.D., O.S., S.S., C.S., V.D.M., T.P., K.S., H.L., S.d.B., J.-B.M., I.I., C.G.M., B.K., C.L.C., M.C., P. Valent, E.D., P. Vyas, D.B., and E.A. provided patient samples and clinical information; A.F., U.M., Z.K., S.M., O.A.B., D.B., E.A., L.G., E.S., J.S., and T.M. provided major intellectual inputs and/or reagents; T.M. and J.S. conceived and supervised the project and drafted the manuscript; and all authors revised and approved the final version of the manuscript.

Conflict-of-interest disclosure: C.G.M. received research funding from Abbvie, Loxo Oncology, and Pfizer; and speaking and travel fees from Illumina and Amgen. The remaining authors declare no competing financial interests.

ORCID profiles: A.F., 0000-0003-3319-9864; F.O.B., 0000-0003-0636-8845; C.K.L., 0000-0003-4473-6311; C.T., 0000-0003-3873-5240; S.T., 0000-0003-0336-9770; A.K.-K., 0000-0002-6063-7260; O.S., 0000-0003-0374-1536; V.D., 0000-0001-9872-1946; C.S., 0000-0002-3551-1009; S.M., 0000-0002-9533-7778; S.d.B., 0000-0002-8126-4942; C.G.M., 0000-0002-1871-1850; P. Valent, 0000-0003-0456-5095; E.D., 0000-0002-0928-0753; P. Vyas, 0000-0003-3931-0914; E.A., 0000-0003-1386-4943; E.S., 0000-0003-0521-7463; J.S., 0000-0001-8616-0096; T.M., 0000-0003-1552-087X.

Correspondence: Thomas Mercher, Institut Gustave Roussy, INSERM U1170, 39 rue Camille Desmoulins, 94800 Villejuif, France; e-mail: thomas.mercher@inserm.fr; or Juerg Schwaller, University Children's Hospital Basel (UKBB), Department of Biomedicine (DBM), University of Basel, ZLF-Laboratory 202, Hebelstrasse 20, CH-4031 Basel, Switzerland; e-mail: j.schwaller@unibas.ch.

Footnotes

Submitted 27 August 2019; accepted 25 March 2020; prepublished online on *Blood* First Edition 29 April 2020. DOI 10.1182/blood.2019003062.

*F.O.B., M.-R.P.-B., and C.I. contributed equally as second author.

†J.S. and T.M contributed equally as senior author.

Contact the corresponding authors for original data.

The online version of this article contains a data supplement.

There is a *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

REFERENCES

1. Hasserjian RP, Zuo Z, Garcia C, et al. Acute erythroid leukemia: a reassessment using criteria refined in the 2008 WHO classification. *Blood*. 2010;115(10):1985-1992.
2. Wang SA, Hasserjian RP. Acute erythroleukemias, acute megakaryoblastic leukemias, and reactive mimics: a guide to a number of perplexing entities. *Am J Clin Pathol*. 2015;144(1):44-60.
3. Boddu P, Benton CB, Wang W, Borthakur G, Khoury JD, Pemmaraju N. Erythroleukemia-historical perspectives and recent advances in diagnosis and management. *Blood Rev*. 2018;32(2):96-105.
4. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia [published correction appears in *Blood*. 2016;128(3):462-463]. *Blood*. 2016;127(20):2391-2405.
5. Arber DA. Revisiting erythroleukemia. *Curr Opin Hematol*. 2017;24(2):146-151.
6. Qiu S, Jiang E, Wei H, et al. An analysis of 97 previously diagnosed de novo adult acute erythroid leukemia patients following the 2016 revision to World Health Organization classification. *BMC Cancer*. 2017;17(1):534.
7. Ley TJ, Miller C, Ding L, et al; Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368(22):2059-2074.
8. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med*. 2016;374(23):2209-2221.
9. Grossmann V, Bacher U, Haferlach C, et al. Acute erythroid leukemia (AEL) can be separated into distinct prognostic subsets based on cytogenetic and molecular genetic characteristics. *Leukemia*. 2013;27(9):1940-1943.
10. Cervera N, Carbuccia N, Garnier S, et al. Molecular characterization of acute erythroid leukemia (M6-AML) using targeted next-generation sequencing. *Leukemia*. 2016;30(4):966-970.
11. Cervera N, Carbuccia N, Mozziconacci M-J, et al. Revisiting gene mutations and prognosis of ex-M6a-acute erythroid leukemia with regard to the new WHO classification. *Blood Cancer J*. 2017;7(8):e594.
12. Ping N, Sun A, Song Y, et al. Exome sequencing identifies highly recurrent somatic GATA2 and CEBPA mutations in acute erythroid leukemia. *Leukemia*. 2017;31(1):195-202.
13. Montalban-Bravo G, Benton CB, Wang SA, et al. More than 1 TP53 abnormality is a dominant characteristic of pure erythroid leukemia. *Blood*. 2017;129(18):2584-2587.
14. Kerenyi MA, Orkin SH. Networking erythropoiesis. *J Exp Med*. 2010;207(12):2537-2541.
15. Kuhrt D, Wojchowski DM. Emerging EPO and EPO receptor regulators and signal transducers. *Blood*. 2015;125(23):3536-3541.
16. Valent P, Büsche G, Theurl I, et al. Normal and pathological erythropoiesis in adults: from gene regulation to targeted treatment concepts. *Haematologica*. 2018;103(10):1593-1603.
17. Soler E, Andrieu-Soler C, de Boer E, et al. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev*. 2010;24(3):277-289.
18. Li L, Freudenberg J, Cui K, et al. Ldb1-nucleated transcription complexes function as primary mediators of global erythroid gene activation. *Blood*. 2013;121(22):4575-4585.
19. Gillinder KR, Tuckey H, Bell CC, et al. Direct targets of pSTAT5 signalling in erythropoiesis. *PLoS One*. 2017;12(7):e0180922.
20. Perreault AA, Benton ML, Koury MJ, Brandt SJ, Venters BJ. Epo reprograms the epigenome of erythroid cells. *Exp Hematol*. 2017;51:47-62.
21. Singleton BK, Frayne J, Anstee DJ. Blood group phenotypes resulting from mutations in erythroid transcription factors. *Curr Opin Hematol*. 2012;19(6):486-493.
22. Doshi BS, Abramowsky C, Briones M, Bunting ST. Concomitant a novel ALAS2 mutation and GATA1 mutation in a newborn: a case report and review of the literature. *Am J Blood Res*. 2014;4(1):41-45.
23. Crispino JD, Horwitz MS. GATA factor mutations in hematologic disease. *Blood*. 2017;129(15):2103-2110.
24. Micci F, Thorsen J, Panagopoulos I, et al. High-throughput sequencing identifies an NFIA/CBFA2T3 fusion gene in acute erythroid leukemia with t(1;16)(p31;q24). *Leukemia*. 2013;27(4):980-982.
25. Greenberg PL, Tuechler H, Schanz J, et al. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood*. 2012;120(12):2454-2465.
26. Alexandrova EM, Yallowitz AR, Li D, et al. Improving survival by exploiting tumour dependence on stabilized mutant p53 for treatment [published correction appears in *Nature*. 2015;527(7578):398]. *Nature*. 2015;523(7560):352-356.

27. Quivoron C, Couronné L, Della Valle V, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell*. 2011;20(1):25-38.

28. Li Z, Godinho FJ, Klusmann J-H, Garriga-Canut M, Yu C, Orkin SH. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat Genet*. 2005;37(6):613-619.

29. Weiss MJ, Yu C, Orkin SH. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol*. 1997;17(3):1642-1651.

30. Damm F, Mylonas E, Cosson A, et al. Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discov*. 2014;4(9):1088-1101.

31. Tamborero D, Rubio-Perez C, Deu-Pons J, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med*. 2018;10(1):25.

32. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*. 2015;109:21.29.1-21.29.9.

33. Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296-309.

34. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.

35. Heuston EF, Keller CA, Lichtenberg J, et al; NIH Intramural Sequencing Center. Establishment of regulatory elements during erythro-megakaryopoiesis identifies hematopoietic lineage-commitment points. *Epigenetics Chromatin*. 2018;11(1):22.

36. Iacobucci I, Wen J, Meggendorfer M, et al. Genomic subtyping and therapeutic targeting of acute erythroleukemia. *Nat Genet*. 2019;51(4):694-704.

37. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017;18(1):220.

38. Merryweather-Clarke AT, Atzberger A, Soneji S, et al. Global gene expression analysis of human erythroid progenitors. *Blood*. 2011;117(13):e96-e108.

39. Liu L, Wang H, Wen J, et al. Mutated genes and driver pathways involved in myelodysplastic syndromes—a transcriptome sequencing based approach. *Mol Biosyst*. 2015;11(8):2158-2166.

40. Dolatshad H, Pellagatti A, Fernandez-Mercado M, et al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells [published correction appears in *Leukemia*. 2015;29(8):1798]. *Leukemia*. 2015;29(5):1092-1103.

41. Chen L, Ge B, Casale FP, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell*. 2016;167(5):1398-1414.e24.

42. Lachmann A, Giorgi FM, Lopez G, Califano A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics*. 2016;32(14):2233-2235.

43. Alvarez MJ, Shen Y, Giorgi FM, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet*. 2016;48(8):838-847.

44. Madan V, Kanojia D, Li J, et al. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat Commun*. 2015;6:6042.

45. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. 2017;19(4):271-281.

46. Chen D, Zhang G. Enforced expression of the GATA-3 transcription factor affects cell fate decisions in hematopoiesis. *Exp Hematol*. 2001;29(8):971-980.

47. Larsen J, Meyer S, Steinlein P, Beug H, Hayman MJ. Transformation of chicken bone marrow cells by the v-ski oncogene. *Oncogene*. 1993;8(12):3221-3228.

48. Ueki N, Zhang L, Hayman MJ. Ski negatively regulates erythroid differentiation through its interaction with GATA1. *Mol Cell Biol*. 2004;24(23):10118-10125.

49. Singbrant S, Wall M, Moody J, et al. The SKI proto-oncogene enhances the in vivo repopulation of hematopoietic stem cells and causes myeloproliferative disease. *Haematologica*. 2014;99(4):647-655.

50. Muench DE, Ferchen K, Velu CS, et al. SKI controls MDS-associated chronic TGF-β signaling, aberrant splicing, and stem cell fitness. *Blood*. 2018;132(21):e24-e34.

51. Costa V, Esposito R, Ziviello C, et al. New somatic mutations and WNK1-B4GALNT3 gene fusion in papillary thyroid carcinoma. *Oncotarget*. 2015;6(13):11242-11251.

52. Shearstone JR, Pop R, Bock C, Boyle P, Meissner A, Socolovsky M. Global DNA demethylation during mouse erythropoiesis in vivo. *Science*. 2011;334(6057):799-802.

53. Hollanda LM, Lima CSP, Cunha AF, et al. An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. *Nat Genet*. 2006;38(7):807-812.

54. Ge L, Zhang RP, Wan F, et al. TET2 plays an essential role in erythropoiesis by regulating lineage-specific genes via DNA oxidative demethylation in a zebrafish model. *Mol Cell Biol*. 2014;34(6):989-1002.

55. Zhang X, Su J, Jeong M, et al. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet*. 2016;48(9):1014-1023.

56. Yan H, Wang Y, Qu X, et al. Distinct roles for TET family proteins in regulating human erythropoiesis. *Blood*. 2017;129(14):2002-2012.

57. Moran-Crusio K, Reavie L, Shih A, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*. 2011;20(1):11-24.

58. Thoms JAI, Birger Y, Foster S, et al. ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer. *Blood*. 2011;117(26):7079-7089.

59. Carmichael CL, Metcalf D, Henley KJ, et al. Hematopoietic overexpression of the transcription factor Erg induces lymphoid and erythro-megakaryocytic leukemia. *Proc Natl Acad Sci USA*. 2012;109(38):15437-15442.

60. Kosmider O, Denis N, Lacout C, Vainchenker W, Dubreuil P, Moreau-Gachelin F. Kit-activating mutations cooperate with Spi-1/PU.1 overexpression to promote tumorigenic progression during erythroleukemia in mice. *Cancer Cell*. 2005;8(6):467-478.

61. Takeda J, Yoshida K, Nannya Y, et al. Novel molecular pathogenesis and therapeutic target in acute erythroid leukemia [abstract]. *Blood*. 2019;134(suppl 1):914.

62. Adélaïde J, Cervera N, Guille A, et al. Gains of EPOR and ERG genes in adult erythroleukaemia. *Br J Haematol*. In press;

63. Rimmelé P, Kosmider O, Mayeux P, Moreau-Gachelin F, Guillouf C. Spi-1/PU.1 participates in erythroleukemogenesis by inhibiting apoptosis in cooperation with Epo signaling and by blocking erythroid differentiation. *Blood*. 2007;109(7):3007-3014.

64. Jeong JJ, Gu X, Nie J, et al. Cytokine-regulated phosphorylation and activation of TET2 by JAK2 in hematopoiesis. *Cancer Discov*. 2019;9(6):778-795.

65. Wichmann C, Becker Y, Chen-Wichmann L, et al. Dimer-tetramer transition controls RUNX1/ETO leukemogenic activity. *Blood*. 2010;116(4):603-613.

66. Thirant C, Ignacimouttou C, Lopez CK, et al. ETO2-GLIS2 hijacks transcriptional complexes to drive cellular identity and self-renewal in pediatric acute megakaryoblastic leukemia. *Cancer Cell*. 2017;31(3):452-465.

67. Han GC, Vinayachandran V, Bataille AR, et al. Genome-wide organization of GATA1 and TAL1 determined at high resolution. *Mol Cell Biol*. 2015;36(1):157-172.

68. Wadman IA, Osada H, Grütz GG, et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J*. 1997;16(11):3145-3157.

69. Tsang AP, Visvader JE, Turner CA, et al. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*. 1997;90(1):109-119.

70. Schuh AH, Tipping AJ, Clark AJ, et al. ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol Cell Biol*. 2005;25(23):10235-10250.

Supplemental Figures for:

Human erythroleukemia genetics and transcriptomes identify master transcription factors as functional disease drivers

Alexandre Fagnan, Frederik Otzen Bagger, Maria-Riera Piqué-Borràs, Cathy Ignacimouttou, Alexis Caulier, Cécile K. Lopez, Elie Robert, Benjamin Uzan, Véronique Gelsi-Boyer, Zakia Aid, Cécile Thirant, Ute Moll, Samantha Tauchmann, Amina Kurtovic-Kozaric, Jaroslaw Maciejewski, Christine Dierks, Orietta Spinelli, Silvia Salmoiraghi, Thomas Pabst, Kazuya Shimoda, Virginie Deleuze, Hélène Lapillonne, Connor Sweeney, Véronique De Mas, Betty Leite, Zahra Khadri, Sébastien Malinge, Stéphane de Botton, Jean-Baptiste Micol, Benjamin Kile, Catherine L. Carmichael, Ilaria Iacobucci, Charles Mullighan, Martin Caroll, Peter Valent, Olivier A. Bernard, Eric Delabesse, Paresh Vyas, Daniel Birnbaum, Eduardo Anguita, Loïc Garçon, Eric Soler, Juerg Schwaller, Thomas Mercher

SUPPLEMENTAL FIGURES

Supplemental Figure 1.

Association with clinical AEL patient outcomes

(A) Patient distribution according to age (left): older adults (>60 years, n=34), adults (40-59 years, n=14), young adults (21-39 years, n=8) and pediatric patients (0-20 years, n=2) or according to WHO 2008 diagnosis (right): AML-M6a *de novo* (n=29), AML-M6a secondary (n=20), AML-M6b *de novo* (n=4) or AML-M6b (n=1).

(B) Variant allele frequency of recurrently mutated genes in AEL patients.

(C-F) Kaplan-Meier survival plots of AEL patients according to (C) genomic subgroup: Group1 (n=10), Group 2 (n=7), Group3 (n=6); (D) IPSS-R cytogenetic risk group: Very good/Good (n=24), Intermediate (n=4), Very poor/Poor (n=20); (E) WHO 2008 diagnosis: M6a *de novo* (n=29), M6b (n=5), M6a secondary (n=20); (F) age at diagnosis: Young (n=8), Adult (n=14), Older adult (n=34). p-value using Log rank Mantel-Cox test are indicated.

(G-H) Histogram representation of the number of mutations per patient according to genomic subgroups (G) in our AEL cohort [Group 1: n=12, Group 2: n=11, Group 3: n=10] and (H) in the Iacobucci *et al.* AEL cohort[33] [TP53: n=48, Others with epigenetic: n=25, Others without epigenetic: n=29]. The "TP53" and "Others" groups were extracted directly from Iacobucci *et al.*[33], while the sub-classification into "with epigenetic" and "without epigenetic" was a re-analysis of the data.

Supplemental Figure 2.

Hematopoietic lineage enrichment profile of AEL

(A) Heatmap representation of hematopoietic lineage enrichment score profile obtained using xCell software[34] on AEL RNAseq data.

(B) *SPI1* expression in AEL patients (normalized number of counts are represented). The red bar highlights sample that expressed more than 3-fold the highest expression observed in either normal human BFU-E or CFU-E.

Supplemental Figure 3.

Transcription factor activity inference

(A) Schematic representation of the generation of a transcription factor (TF) network using the ARACNe algorithm to compute a human HSPC network and the VIPER algorithm to compute the activity of TF based on our AEL patient transcriptome signatures.

(B) Heatmap representation of the top 50 most differentially transcription factors activated in AEL patient samples. Differential activated gene lists were established by PCA analysis using predicted activated gene matrix (previously computed using ARACNE and VIPER algorithm), then genes driving PCA dimensions were identified and ranked by contribution (using FactoMineR v1.41 and factoextra v1.0.5 R packages). Finally, heatmap of activated genes was obtained by plotting the top 50 most contributed genes from the first PCA dimension (using pheatmap v1.0.12 R package).

(C) PCA with regression lines from plot in Figure 3A with projection of AEL patient samples colored with SPI1, NFIA and TAL1 expression and predicted activity.

Supplemental Figure 4.

Fusion transcripts in human AEL

(A) Heatmap representation of patients carrying out-of-frame (green) or in-frame (red) fusion transcripts according to their molecular subgroup.

(B) Sequences of the breakpoint *YWHAE-EPO* and *HSD17B11-B4GALNT3* fusion transcripts. The number of reads supporting the fusion are indicated.

(C) Paired histogram representation of *YWHAE* and *EPO, HSD17B11* and *B4GALNT3, LINC01* and *DNMT3B* gene expression in AEL patients, normal human BFU-E and CFU-E. Patients with red bars indicate those presenting with the fusion transcripts.

Supplemental Figure 5.

ERG, ETO2, SKI and SPI1 impairs chromatin accessibility at GATA1 binding site and functionally impairs GATA1 activity

(A) Flow cytometry analysis of CD71 and Ter119 expression (top) and May Grunwald Giemsa stained cytospots (bottom) of mouse erythroblast transformed with either expression of *SKI, ERG, ETO2, SPI1* and *B4GALNT3* overexpression or with cooperation between *Tet2*$^{-/-}$ + *G1s.*

(B) Histogram representation of the foldchanges expression of human AEL samples (fold-changes computed by comparing impacted patient sample to the mean of non-impacted patient samples) and transduced mouse erythroblast (fold-changes computed by comparing cells transduced with the indicated factor to Ctrl) for *SKI*, *ERG*, *ETO2*, *SPI1*, *B4GALNT3* and *EPO* mRNA expression.

(C) Histogram representation of *SKI, ERG, ETO2, SPI1* and *B4GALNT3* mRNA expression measured by RT-qPCR in normal mouse erythroblast or erythroblast transformed with either expression of *SKI, ERG, ETO2, SPI1* and *B4GALNT3* overexpression (n=3 for all conditions). Expression levels were also compared to the erythroleukemia MEL or the Ba/F3 myelo-lymphoid cell line.

(D) Dot-plot representation of motifs analysis under ATAC-seq peaks of normal (Ctrl) or transformed erythroblast (ERG, ETO2, SKI and SPI1). Each dot represents the log(p.value) for a given motif and increased ranked in abscise by log(p.value)

(E) Heatmap representing the hierarchical clustering of ATAC-seq signal, performed using normal (Ctrl) and transformed erythroblast by either ERG, ETO2, SKI or SPI1, focused on specific open chromatin region of MEP (top), CFU-E (middle) and pro-erythroblast (proE; bottom). Heatmaps were focused on peaks centers with +/-5Kb.

(G) Quantitative *Gata1* mRNA expression measured by RT-qPCR in WT erythroblast (Ctrl) or erythroblast transformed with either *SKI, ERG, ETO2, SPI1* or *B4GALNT3* overexpression. Expression levels were also compared to the erythroleukemia MEL or the Ba/F3 myelo-lymphoid cell line (n=3 for all conditions).

(G) Flow-cytometry analysis histogram of Ter119 expression in G1E cells expressing empty control without (grey) or with (red) induction of GATA1 expression by doxycycline induction.

Supplemental Figure 6.

Transplantation of *in vitro*-transformed erythroblasts

(A) Kaplan-Meier plot of mice transplanted with mouse erythroblasts retrovirally expressing *SKI*, *ERG, ETO2, SPI1or B4GALNT3*, expanded over 2 months *in vitro*. n=4 for all groups. Only erythroblast expressing ERG, ETO2 or SKI were able to engraft in mice and promote disease.

(B) Flow cytometry analysis of CD71 and Ter119 (top) or CD11b and Gr1 (bottom) surface expression, gated on transgene-expressing cells (GFP^+), in mice injected with 10^6 mouse erythroblasts expressing *SKI, ERG* or *ETO2*.

(C) Flow cytometry analysis of myeloid cells ($CD11b^+Gr1^+$), erythroid progenitors ($CD71^+$ $Ter119^+$), B cells ($B220^+$ $CD19^+$) and T cells ($CD4^+CD8^+$) gated for $CD45.2^+$ cells in the BM of immunodeficient recipients ($CD45.1^+$) engrafted with $CD45.2^+$ erythroblasts from *Tet2*$^{-/-}$+*G1s* mice.

(D) Histopathological analysis of BM and spleen of mice engrafted with normal erythroblasts or transformed erythroblast either expressing SKI, *ERG* or *ETO2* or from *Tet2*$^{-/-}$ + *G1s* mice, previously expanded *in vitro* over 2 months. Staining were performed using H&E.

(E) Histopathology analysis of Bone marrow (BM, left) and spleen (rigth) of $CD45.1^+$ mice engrafted with transformed erythroblast expressing an empty vector (top) or SKI (bottom).

Supplemental Figure 7.

Modeling AEL-associated oncogenic cooperation in mice

(A) Flow cytometry analysis of myeloid cells ($CD11b^+Gr1^1$), erythroid progenitors ($CD71^+$ $Ter119^+$) and erythroid differentiation (forward scatter area: FSC-A, CD44) in BM of WT, *Tet2*$^{-/-}$+*Gata1*$^{+/Y}$ (*Tet2*$^{-/-}$), *Tet2*$^{+/+}$+*Gata1*$^{Gata1s/Y}$ *(Gata1s)* or *Tet2*$^{-/-}$+*Gata1*$^{Gata1s/Y}$ (*Tet2*$^{-/-}$+*Gata1s)* male transgenic mice at 8 months.

(B) Histogram representation of spleen weight (left) and spleen photograph (right) of recipient mice engrafted with HSPC from *Tet2*$^{-/-}$ (n=5) and *Tet2*$^{-/-}$+*Gata1s* (n=4) mice.

(C) Histopathology of BM and spleen sections from recipient mice engrafted with HSPC from *Tet2*$^{-/-}$+*Gata1s* mice.

(D) Kaplan-Meier plot of lethally irradiated mice engrafted with HSPC cells from either WT or TP53^{R248Q} mice transduced with ERG or empty vector (Ctrl). WT: n=5, TP53^{R248Q}: n=6.

(E) GFP+ myeloid cells (Gr1^{+}CD11b^{+}), erythroid progenitors (CD71^{+} Ter119^{+}), B cells (B220^{+}) and T cells (CD4^{+}CD8^{+}) in recipients of WT or TP53^{R248Q} HSPC transduced with *ERG* or empty vector (*Ctrl*) at pre-leukemic or leukemic state.

(F) Histopathology of spleen and liver sections from secondary recipient mice engrafted with GFP^{+} erythroblast from primary recipient mice previously engrafted with HSPC from *TP53*^{R248Q} expressing ERG.

Supplemental Figure 8.

SKI overexpression in HSPC leads to a myeloid/erythroid disease in mice

(A) GFP^{+} cells (%) in BM, and Kaplan-Meier plot of lethally irradiated mice transplanted with 0.5x10^{6} lineage marker-depleted HSPC transduced with either murine *SKI* cDNA (n=5) or empty vector (*Ctrl*; n=5).

(B) Peripheral white blood cells (WBC), red blood cells (RBC), platelets (PLT), lymphocytes (Lym), monocytes (Mono), neutrophils and granulocytes (N/Gr) counts. Mean+/-SD is indicated (n=4).

(C) GFP+ myeloid cells (CD11b^{+}Gr1^{+}), erythroid progenitors (CD71^{+} Ter119^{+}), B cells (B220^{+}CD19^{+}) and T cells (CD4^{+}CD8^{+}) gated in the BM of mice transplanted with SKI or empty vector (Ctrl) transduced cells.

(D) GFP+ (%) myeloid cells (CD11b^{+}Gr1^{+}), erythroid progenitors (CD71^{+}Ter119^{+}), B cells (B220^{+}CD19^{+}) and T cells (CD4^{+}CD8^{+}) in the BM or spleen of mice transplanted with *SKI* or control (Ctrl) transduced cells. Mean+/-SD is indicated (Ctrl: n=5; SKI: n=5).

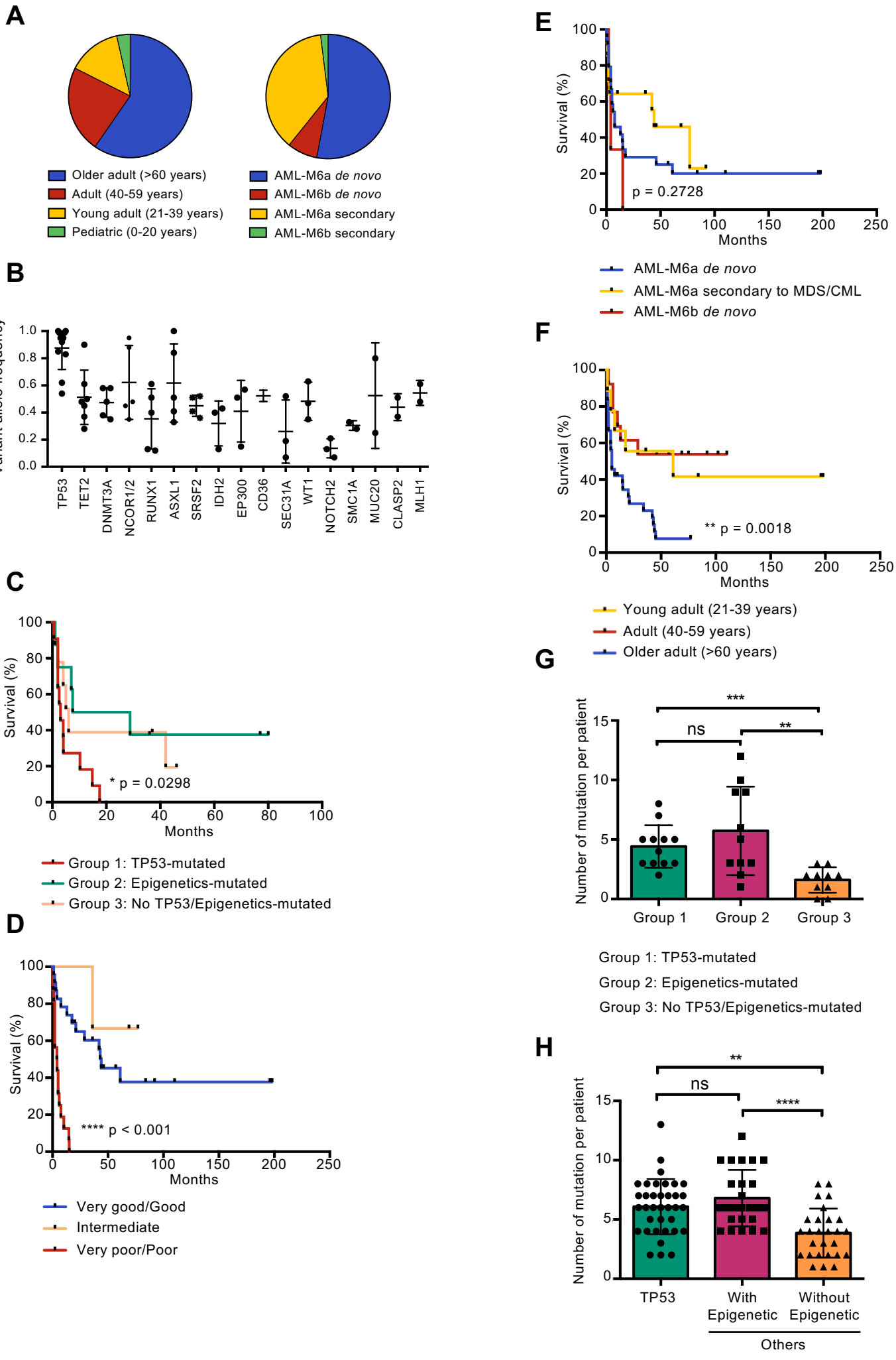
(E) Number of GFP^+ and negative myeloid cells ($CD11b^+Gr1^+$), erythroid progenitors ($CD71^+Ter119^+$), B cells ($B220^+CD19^+$) and T cells ($CD4^+CD8^+$) in BM or spleen of primary recipients. Mean+/-SD is indicated (Ctrl: n=5; SKI: n=5).

(F) Flow cytometry analysis of GFP^+ cells in peripheral blood: platelets (PLT), red blood cells (RBC) and white blood cells (WBC) of primary engrafted mice

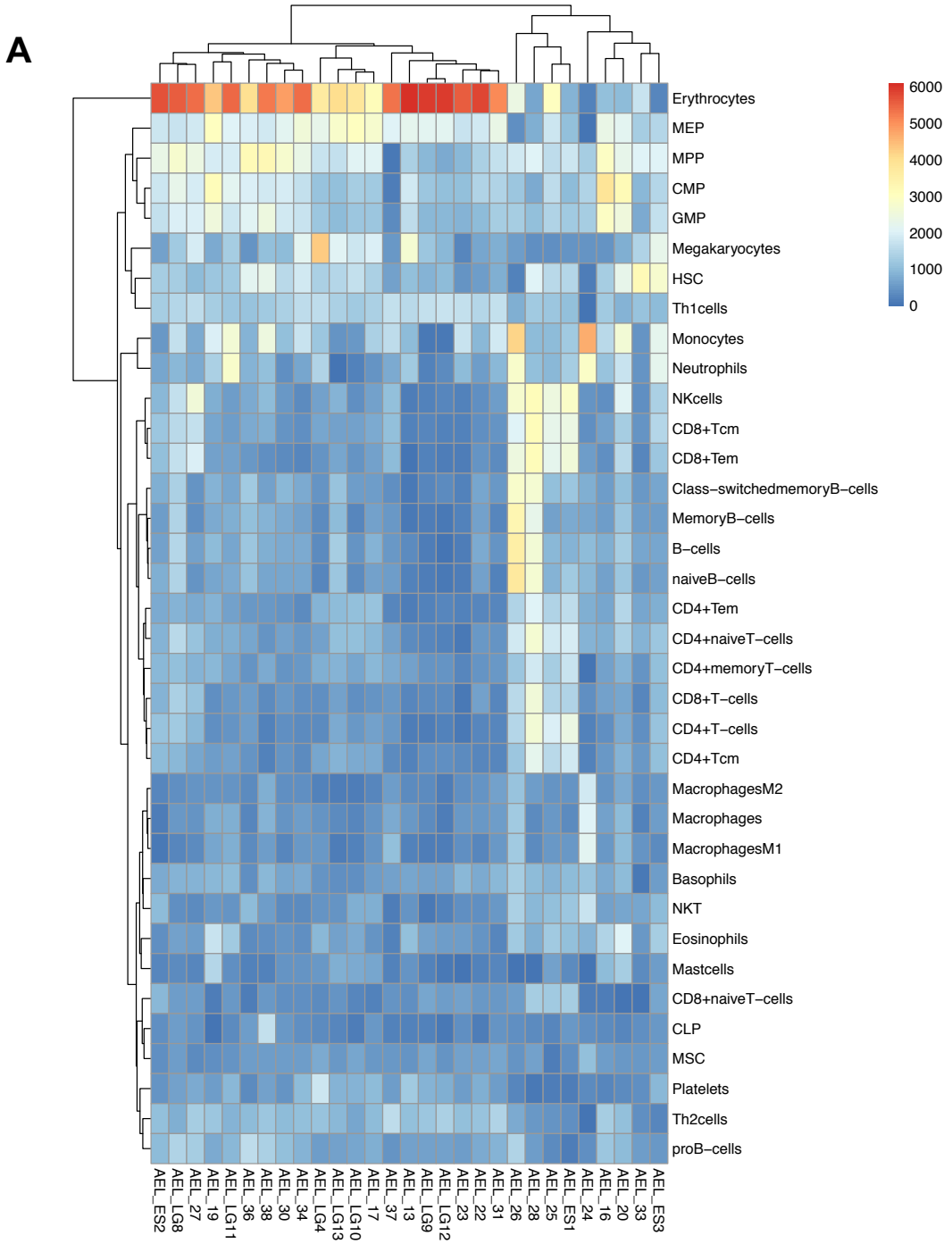
(G) GFP^+ (%) cells in peripheral blood (WBC, RBC, PLT) and BM in recipient mice engrafted with HSC, MEP and GMP expressing *SKI* or empty vector (*Ctrl*), 3 weeks post-engraftment. Mean+/-SD is indicated (HSC: n=5, GMP: n=5, MEP: n=4).

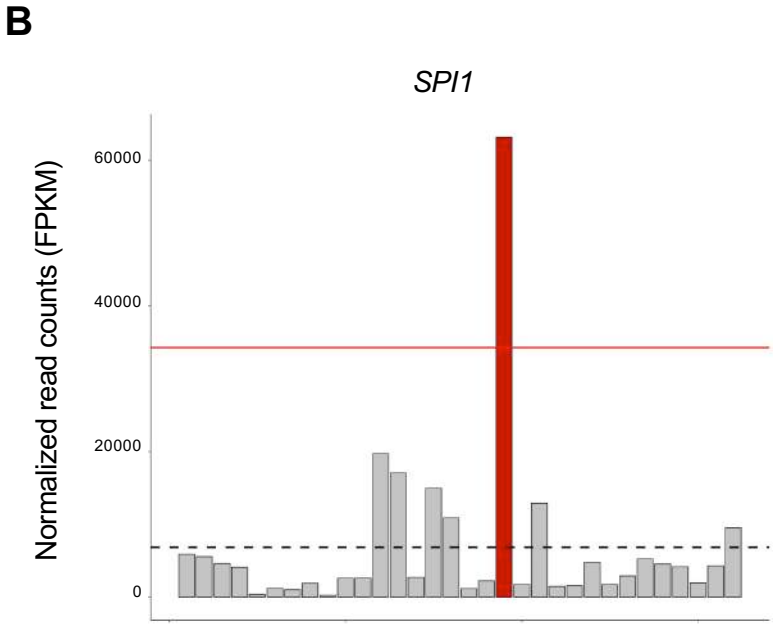
(H) Peripheral white blood cells (WBC), red blood cells (RBC) and platelet (PLT) counts in mice transplanted with HSC or MEP expressing *SKI* or empty vector (*Ctrl*), at time of sacrifice. Mean+/-SD is indicated (HSC: n=5; MEP: n=5).

(I) Histopathology of BM and spleen of MEP expressing *SKI* engrafted mice stained by HE or an GATA1 antibody.

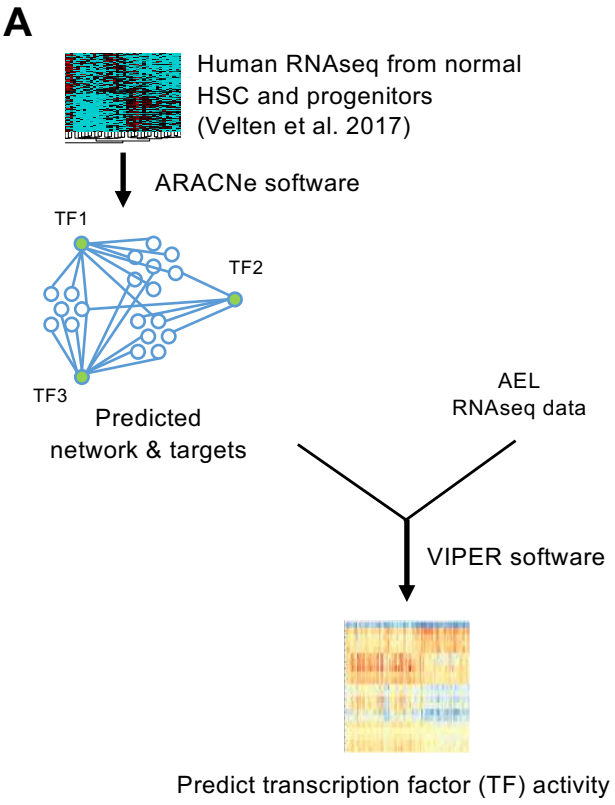


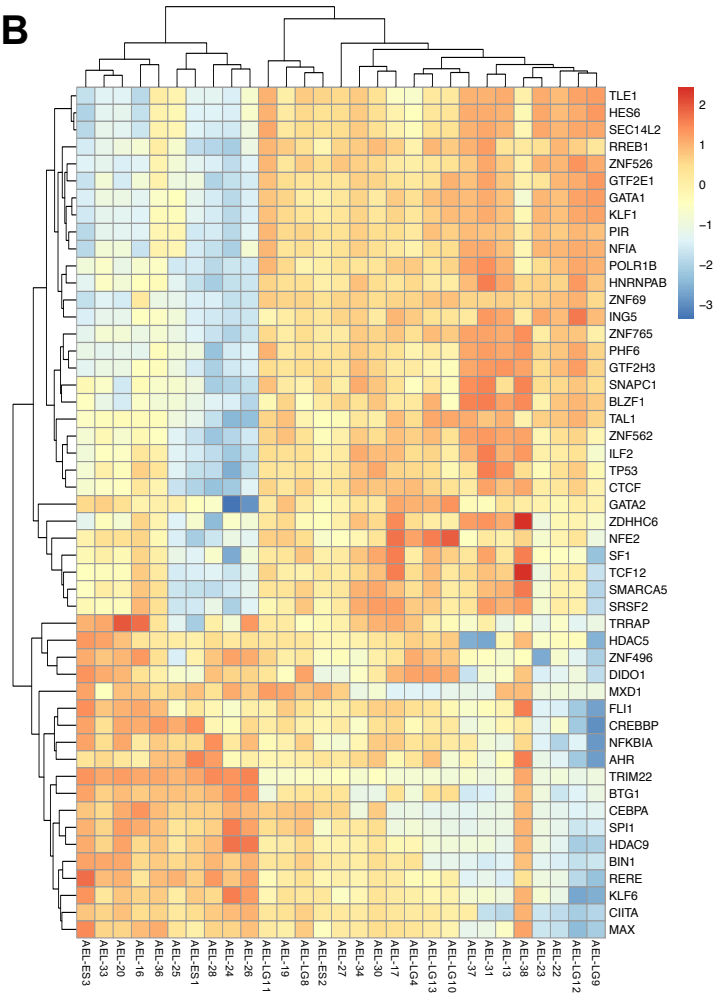
Supplemental Figure 1



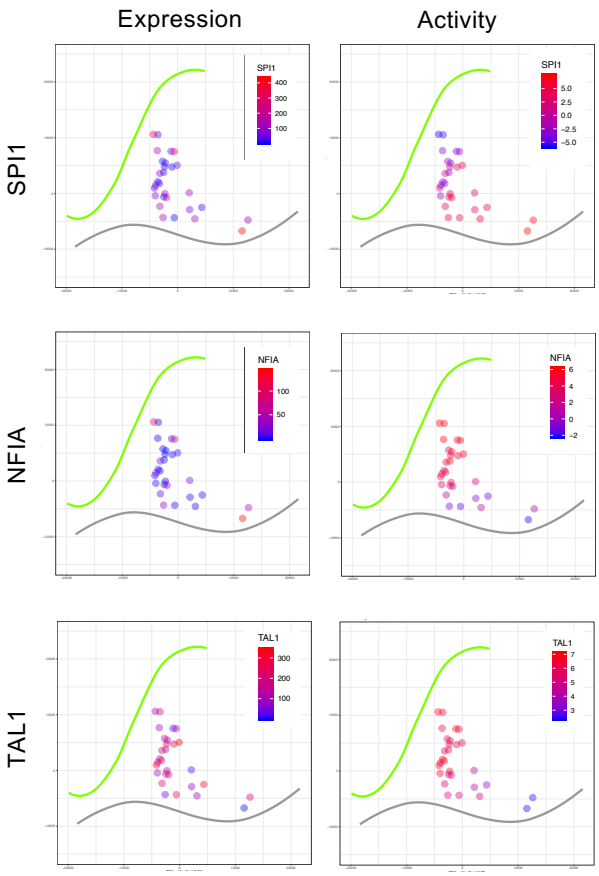


Supplemental Figure 2

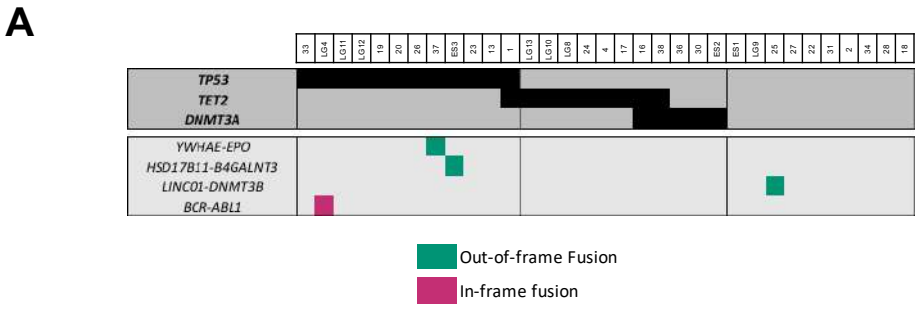


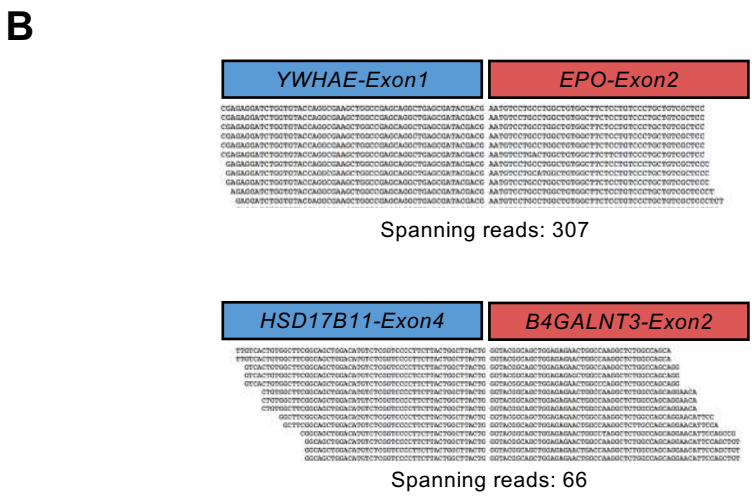


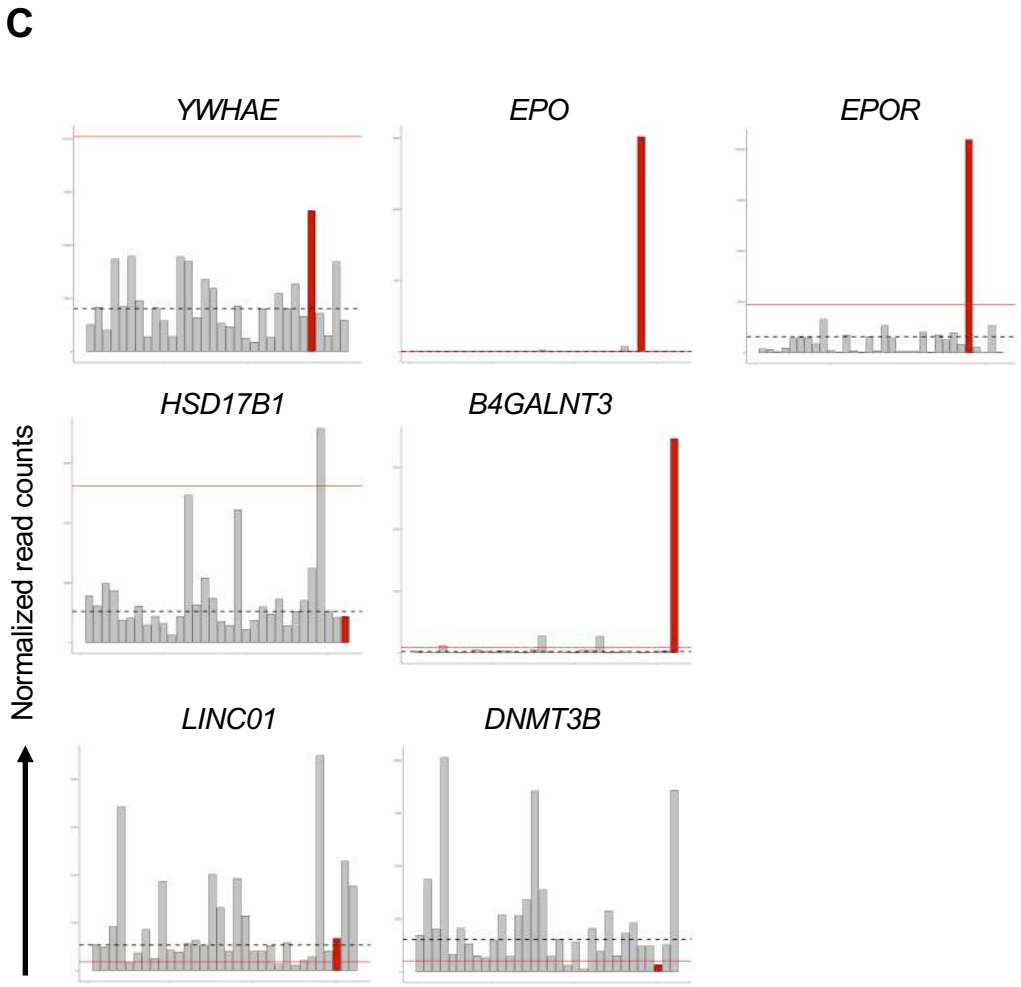
50 most differentially activated TF



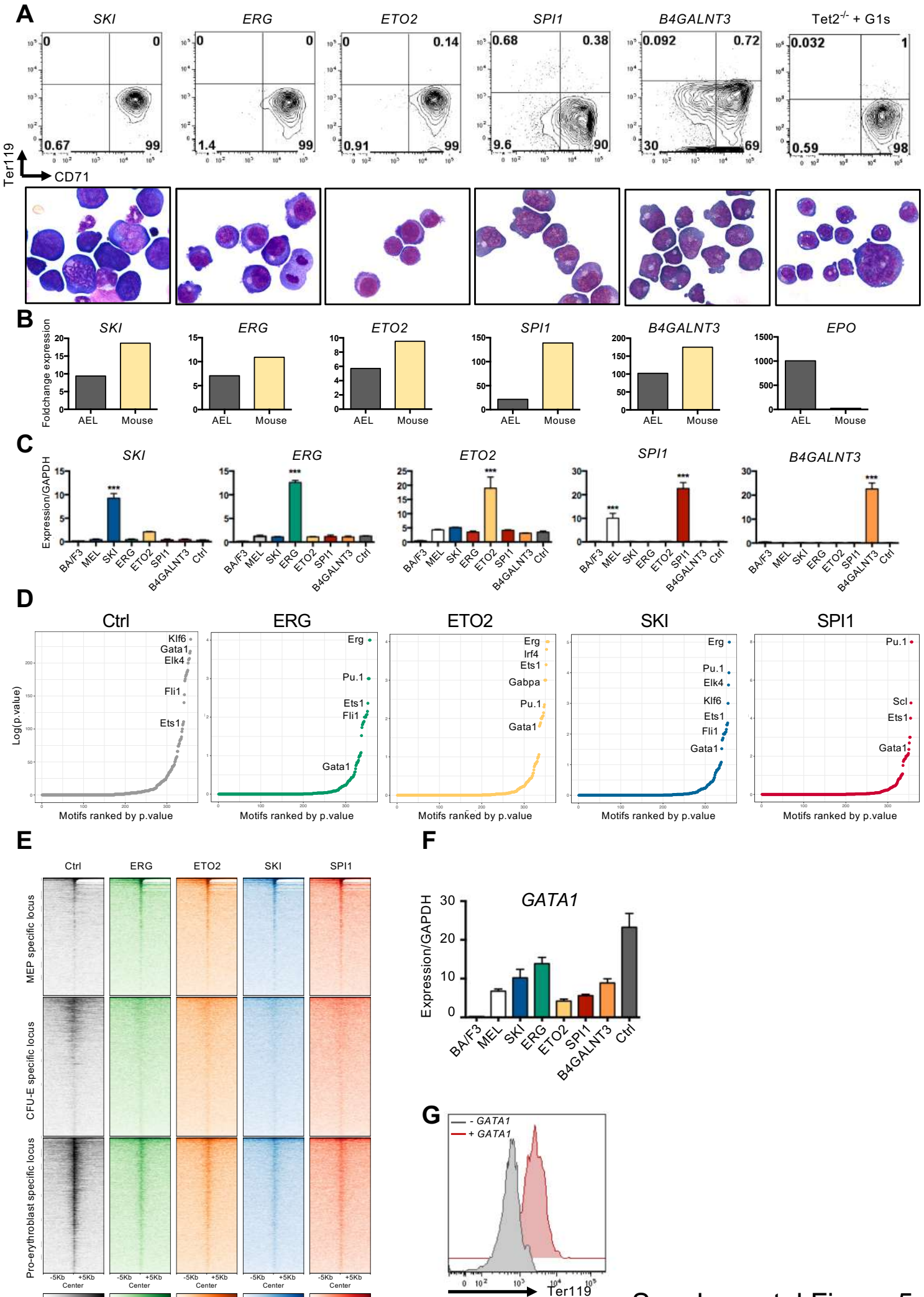
Supplemental Figure 3



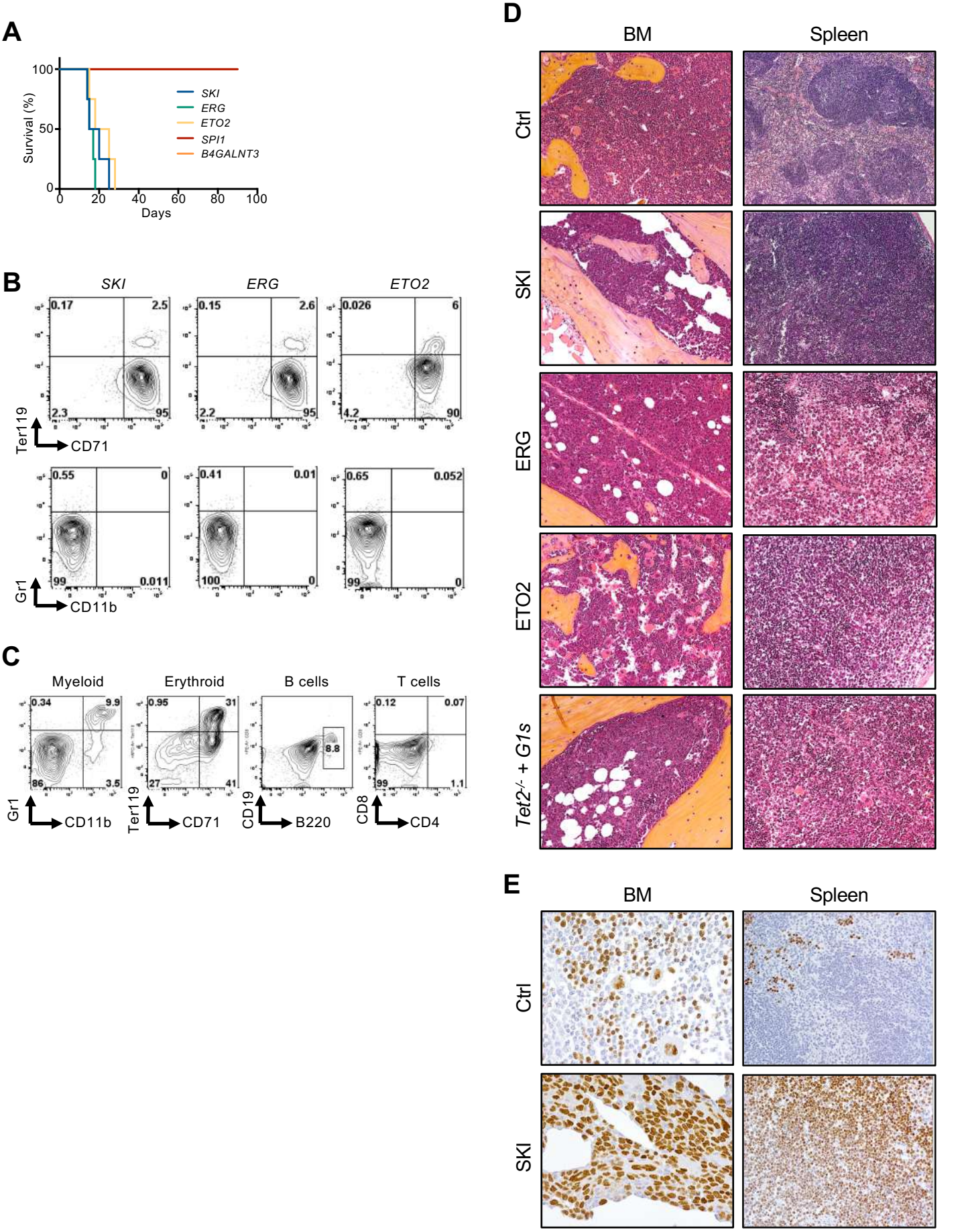




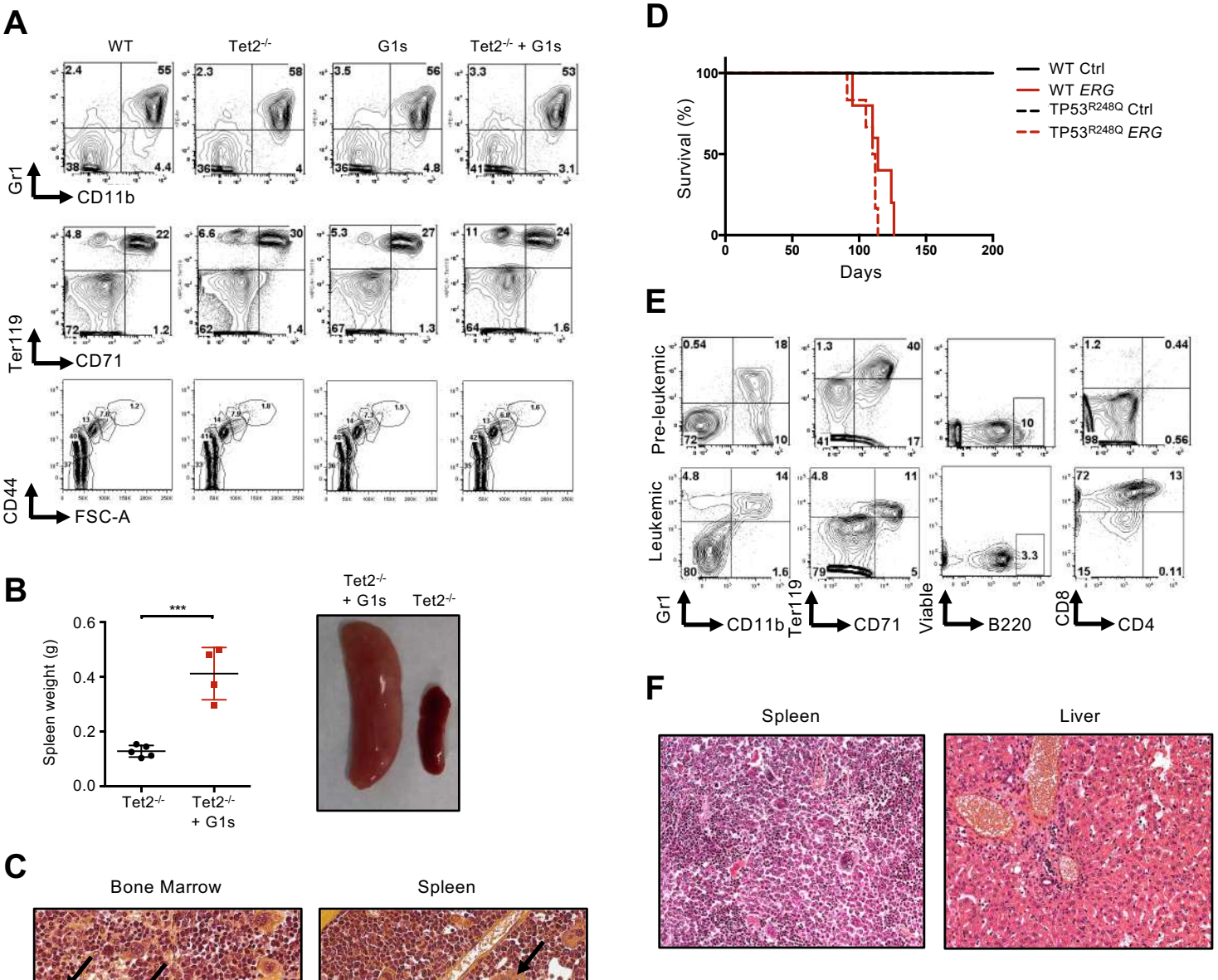
Supplemental Figure 4



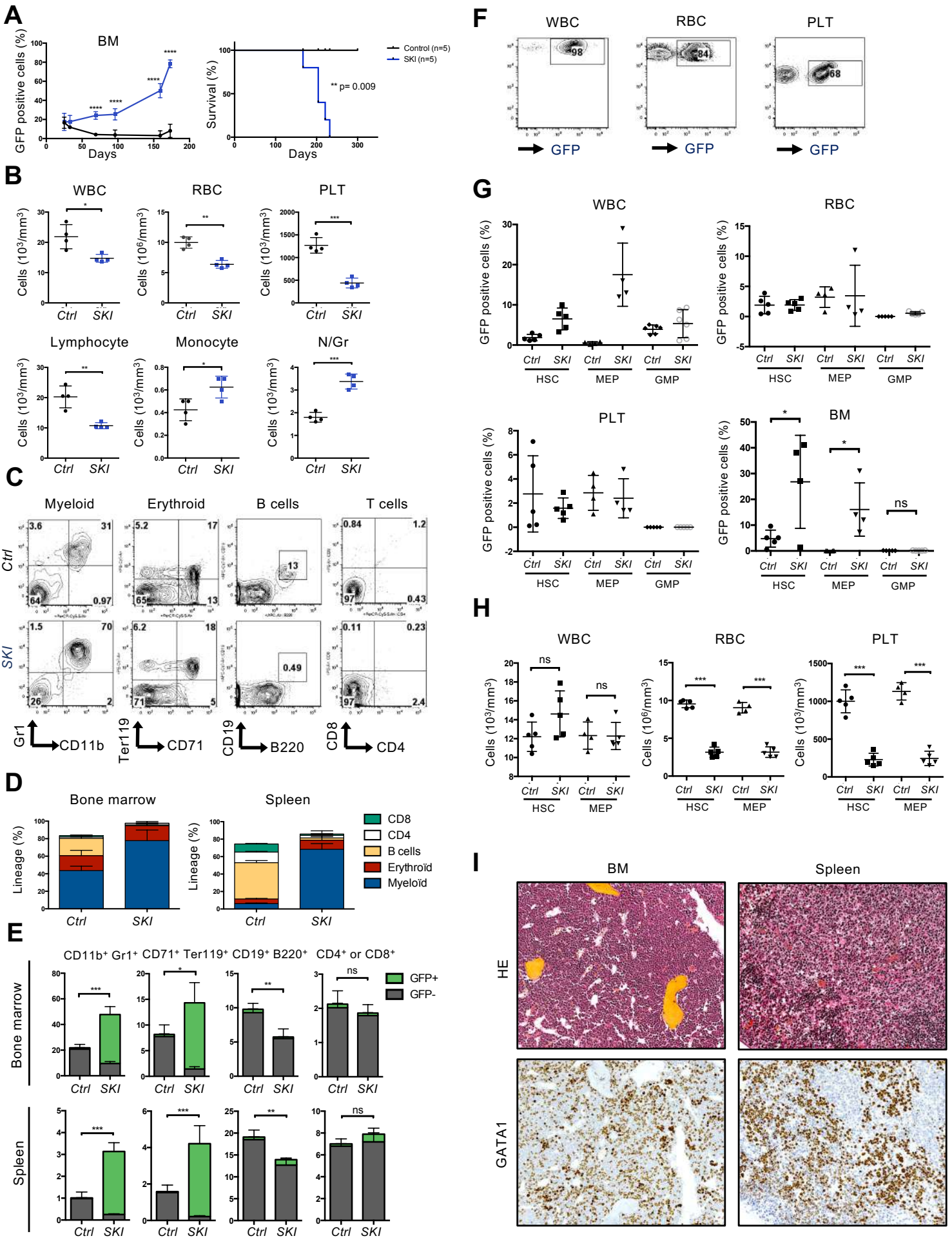
Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7



Supplemental Figure 8

Email
mriera.pique@gmail.com

Phone
+41786421791

Nationality
Spanish

Swiss permit
B permit

Link
www.linkedin.com/in/mariarierapiqueborras/

Languages

Spanish
Native proficiency

Catalan
Native proficiency

English
Professional working proficiency

German
Elementary, A.2

IT Skills

R language & RStudio
Database management (TCGA dataset, GEO datasets, BioMart, Ensembl, Genome Browser, etc)
Microsoft office
Affinity designer
GraphPad Prism
FlowJo

Maria-Riera Piqué Borràs

Independent and resourceful scientist with a strong background in hematology, rare diseases and bioinformatics acquired in three different labs. Hands-on experience in project management and mentoring of students as well as participation in collaboration projects.

Research Experience

- ### PhD Student

  University Children's Hospital Basel, Basel (Switzerland)
  06/2015 – 03/2020
  - Obtained a wealth of experience in the leukemia field, from in vivo mouse models, western blot and FACS, to RNA-, ChIP- and ATAC-sequencing preparation and bioinformatics analysis.
  - Studied novel fusion genes found in children with Acute Erythroid Leukemia (AEL).
  - Presented at national and international conferences.
  - Analyzed RNA-seq data from AEL patients.
  - Mentored and supervised a master student.

- ### Research Internship and Bioinformatic Analyst

  Centre for Genomic Regulation, Barcelona (Spain)
  10/2014 – 02/2015
  - Extracted RNA from human biofluids and prepared small RNA libraries to identify circulating small RNAs to be used as non-invasive biomarkers for communicable and non-communicable diseases.

- ### Bioinformatic Analyst

  Catalan Institute of Oncology, Barcelona (Spain)
  02/2014 – 09/2014
  - Analyzed RNA-seq data with R from breast cancer patients to find relevant groups according to their gene expression as well as to detect biomarkers that defined those groups.

- ### Research Internship

  CEXS-University Pompeu Fabra, Barcelona (Spain)
  02/2013- 06/2013
  - Studied Zebrafish inner ear and lateral line hair cell regeneration using immunohistochemistry, *in situ* hybridization assays and fluorescent microscopy.

- ### Research Internship

  Vall d'Hebron Barcelona Hospital Campus, Barcelona (Spain)
  07/2012 - 09/2012
  - Learned basic laboratory techniques to identify new markers of response to treatment for colorectal cancer patients.

Soft Skills

Teamwork
Problem-Solving
Conflict-handling
Mentoring
Pressure management
Collaboration

Education

- PhD in Medical-Biological Research
  University of Basel, Basel (Switzerland), 06/2015 – 03/2020
  Thesis title: "Molecular mechanisms of acute erythroid leukemia: learning from rare chromosomal translocations in pediatric patients"
- MSc in Omics Data Analysis
  University of Vic, Barcelona (Spain), 10/2013 – 09/2014
  Thesis title: "Identification of molecular subtypes and gene expression patterns of breast cancer analysing RNA-seq data"
- BSc in Biotechnology
  University of Vic, Vic (Spain), 09/2009 – 06/2013
  Thesis title: "Study of inner ear and lateral line hair cell regeneration"

Courses & Training list highlights

- LTK M.2: Course for Study Directors of Animal Experimentation
  University of Zurich, Basel (Switzerland), 2020
- Project Management for Researchers
  University of Basel, Basel (Switzerland), 2020
- Essentials in Drug Development & Clinical Trials
  Swiss Tropical & Public Health Institute, Basel (Switzerland), 2019
- Good Clinical Practice for Investigators and Study Teams
  Swiss Tropical & Public Health Institute, Basel (Switzerland), 2019
- LTK M.1: Introductory Course in Laboratory Animal Science
  University of Zurich, Basel (Switzerland), 2015

Conference & Award list highlights

- **European Hematology Association conference, 2018.**
- **American Society of Hematology (ASH) conference, 2019:** ASH Abstract Achievement Award and oral presentation (see publications).
- **European Hematology Association conference, 2020:** Poster abstract presentation.

Publications

- Rubio M, Bustamante M, Hernandez-Ferrer C, Fernandez-Orth D, Pantano L, Sarria Y, **Piqué-Borràs M**, *et al.* Circulating miRNAs, isomiRs and small RNA clusters in human plasma and breast milk. PLoS One 2018. 13(3):e0193527.
- Bezerra M. F., Lima A. S., **Piqué-Borràs M-R**, *et al.* Co-occurrence of DNMT3A, NPM1, FLT3 mutations identifies a subset of AML with adverse prognosis. Blood 2020; 135 (11):870875.
- Fagnan A, Otzen Bagger F, **Piqué-Borràs M-R**, *et al.* Human erythroleukemia genetics and transcription identify master transcription factors as functional disease drivers. Blood 2020; 136 (6):698714.
- **Piqué-Borràs M-R**, *et al.* Transformation Mechanisms of Nfia-ETO2 Fusion Gene Associated with Pediatric Pure Acute Erythroleukemia. Blood 2019; 134 (Supplement_1):532 (ASH Abstract). Manuscript ready for submission to Blood.

References

Prof. Dr. Juerg Schwaller, group leader of the Childhood Leukemia research group, University Children's Hospital Basel (Switzerland).
Email: j.schwaller@unibas.ch. Phone: +41612653517.

Dr. Frederik Otzen Bagger, head of Bioinformatics hos Rigshospitalet, Genomic Medicine (Copenhagen, Denmark). Principle Investigator and group leader.
Email: frederik.otzen.bagger@regionh.dk