# Insights into cellular and molecular mechanisms of normal and malignant hematopoiesis from mouse models

**Inauguraldissertation**

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Marwa Almosailleakh

State of Kuwait

**Basel, 2020**

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Professor Dr. med. Jürg Schwaller

Professor Dr. phil. Nat. Timm Schroeder

Basel, den 19.11.2019

Prof. Dr. Martin Spiess, Dean

دواؤك منك ولا تبصر... وداؤك منك ولا تشعر

أتزعم أنك جرم صغير ... وفيك انطوى العالم الأكبر

فأنت الكتاب المبين ...الذي بأحرفه يظهر المضمر

وما حاجة لك من خارج...وفكرك فيك وما تصدر

الإمام علي بن أبي طالب (ع)

Your malady is from you, yet you do not see it, your medicine is within you yet you do not perceive it. You claim to be a small entity, yet within you is enfolded the entire universe. You are the evident book, by whose alphabets the hidden becomes manifest. You have no need to look beyond yourself, what you seek is within you, if only you reflect.

**Imam Ali Bin Abi Taleb (AS)**

To My Family

**TABLE OF CONTENTS**

# 1. Summary

Acute myeloid leukemia (AML) refers to a group of aggressive rare heterogenous hematological disease that results from clonal transformation of immature hematopoietic progenitors. Leukemic transformation of cells occurs due to the acquisitions of genetic aberrations that confer a proliferative and survival advantage, and diminished hematopoietic differentiation. Identification of key alterations and mechanisms in the epigenetic network leading to cellular transformation and disease progression may provide novel intervention points for therapy.

To better understand the role of the NSD1 H3K36 histone methyltransferase that is recurrently mutated in AML, our lab established conditional KO mice. All *Nsd1*$^{-/-}$ mice developed a lethal disease mimicking human acute erythroleukemia. By performing several comparative analyses, I was able to show that the activity of the SET-domain of *Nsd1* is indispensable for terminal erythroid cell maturation. Our accumulative observations suggest a model where the NSD1 catalytic activity is essential for creating an erythroid promoting genome-wide epigenetic environment that modulates the activity of the GATA1 erythroid transcriptional master regulator (**Chapter I**).

To study the cellular origin of acute leukemia driven by MLL rearrangements, our lab established an inducible transgenic mouse model. Activation of the MLL-AF9 or MLL-ENL fusions led to fully reversible leukemia phenotype with some phenotypic differences. My work suggested that in contrast to MLL-AF9, the MLL-ENL fusion mostly targets multipotent progenitors leading to leukemia with myeloid and lymphoid characteristics. In addition, MLL-ENL expression level equal or exceeding the wild-type MLL seemed to be critical for leukemic transformation (**Chapter II & appendix 1**). MLL-fusions driven leukemia originating from hematopoietic stem cells (HSCs) are characterized by elevated expression of the transcription factors *ERG* and *EVI1*. To better understand their role in the disease, I attempted to establish a transgenic fluorescent ERG reporter. Genome editing in zygotes allowed us to establish ERG-2A-YFP mice, however, I never detected the fluorescent signals, most likely due to low expression (**Chapter III**).

Transgenic mouse models have been extensively used for delineating human underlaying leukemogenic mechanisms and the development of new therapeutic agents. I wrote a review about the various types of AML mouse models generated, and how they helped in our understanding of disease etiology and progression (**Appendix 2)**.

# 2. Introduction

## 2.1. Hematopoiesis

Hematopoiesis is the process by which the various types of blood cells are generated. Hematopoiesis first arises in the developing embryo when, due to its size increase, the demand for oxygen and nutrients exceeds the capability of passive diffusion. As early hematopoietic development is largely con cerned with providing sufficient oxygen and carbon dioxide supply to the developing tissues, red blood cells (RBCs) production is tightly linked to the development of the vascular system. Hence, due their harmonized functions, both the hematopoietic and cardiovascular systems are essential for fetal survival.

The ontogeny of the vertebrate hematopoietic system is a complex developmental process characterized by multiple waves (primitive and definitive), multiple origin sites [the yolk sac (YS), the aorta-gonad-mesonephros region (AGM), and the placenta] and shifting locations [the fetal liver (FL) and bone marrow (BM)] (**Fig. 1**)[1]. ''Primitive'' hematopoiesis, refers to the production of a transitory hematopoietic cell populations, such as primitive erythrocytes and some myeloid cells in the developing embryo after the onset of gastrulation. "Definitive" hematopoiesis, starts later in development and arises from definitive HSCs and gives rise to all mature blood (erythroid, myeloid and lymphoid) lineages in the adult organism. During postnatal life, the binary fate choice of HSCs self-renewal versus differentiation is very critical for the survival of the organism and thus, is tightly regulated to sustain the HSC pool size throughout life and prevent exhaustion.

### 2.1.1. Embryonic erythropoiesis: Sites and origin

Analysis of the erythroid cell population in the developing mammalian embryos revealed the presence of two distinct, temporally overlapping cell populations. The earliest population consisted of larger, nucleated cells and the later population consisted of smaller, enucleated RBCs. The transient presence of the larger, nucleated erythroid cells during early development was termed "primitive" erythropoiesis, while "definitive" erythropoiesis occurs during late fetal development and throughout postnatal life[2, 3].

During mammalian embryogenesis, the first wave of hematopoietic cells appears in the YS, starting from E7.25-E8.75 in the mouse[4], and within weeks 3-4 of human gestation[5]. Although predominantly erythroid, megakaryocytic and monocytic/macrophagic cells were

also detected[4, 6]. The emergence of primitive erythroid cells marks the beginning of hematopoiesis, and failure in primitive erythropoiesis invariably results in embryonic lethality[1].

The second “definitive” hematopoietic wave also initiates in the YS (E8.25 in mouse[4], and around 4 weeks in human[5]), comprises myelo-erythroid and immune-restricted progenitors, and emerges in conjunction with fully formed and functional vascular system[7, 8]. The third wave of hematopoiesis appears to be more complex, and arises from HSCs produced at multiple intra-embryonic sites, such as the AGM region (E10.5–11.5 in mouse) and its associated major vessels, the umbilical and vitelline arteries, and the placenta (**Fig.1A-B**)[9].

The mechanisms underlying YS-dependent definitive hematopoiesis are not as well-characterized as those involving the AGM region, where definitive HSCs are directly produced from endothelial precursors in a process termed “endothelial-to-hematopoietic transition” (EHT)[10, 11]. This trans-differentiation step involves the switching to a hematopoietic transcriptional program in selected endothelial cells, which then become hemogenic. This is followed by extensive morphological changes causing the breaking of tight junctions with neighboring endothelial cells and rounding up, and finally being released into the blood stream (**Fig. 1C**)[12]. Regardless of their site of origin, these early hematopoietic progenitor cells will reach and colonize the developing FL, thymus, spleen, and the BM (**Fig. 1B**).

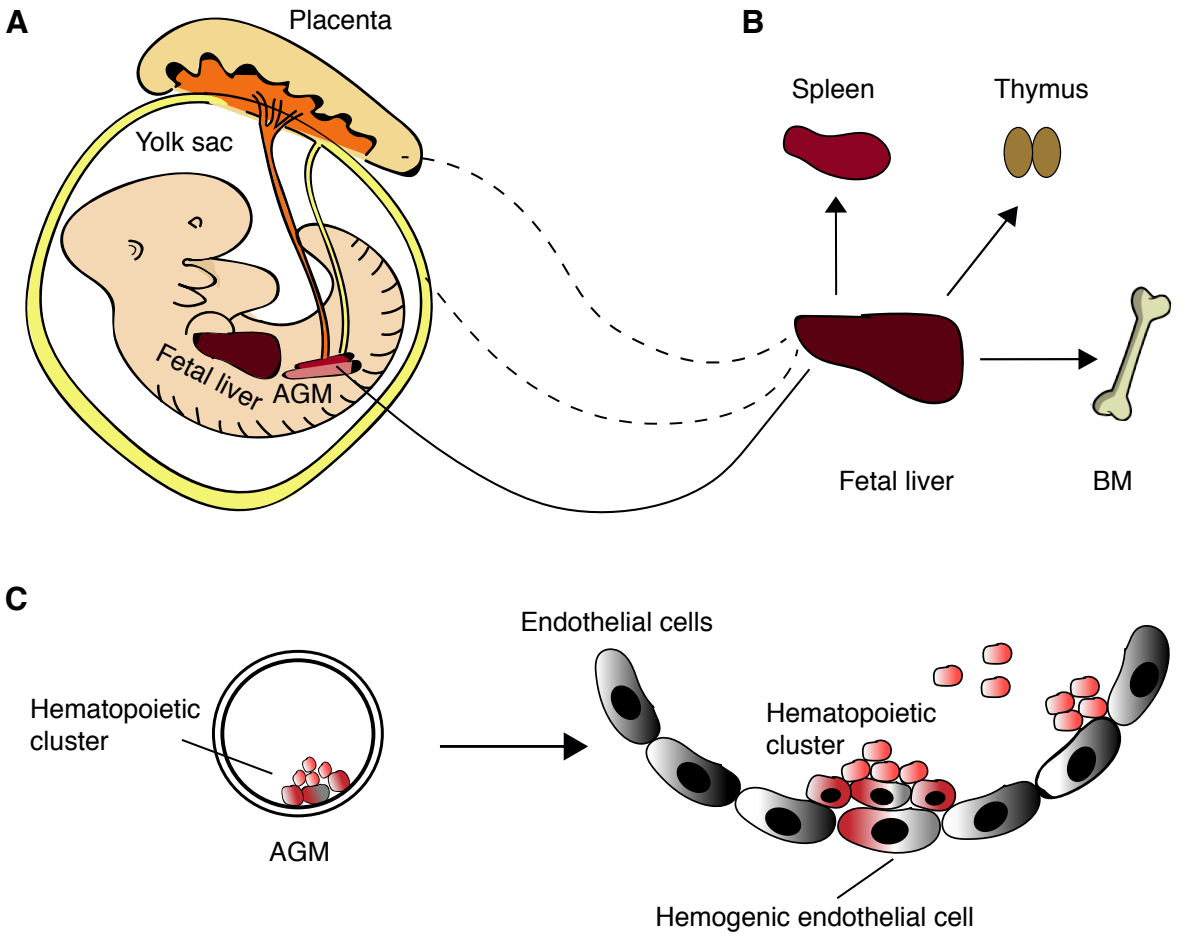



**Fig.1**: **Development of the murine hematopoietic system**. **(A)** Schematic representation of mouse embryo (E10.5) showing hematopoietic niches including the aorta-gonad-mesonephros region (AGM), yolk sac (YS), umbilical and vitelline arteries, placenta and liver. **(B)** Proposed contribution of different embryonic tissues (YS, placenta and AGM) to seeding the fetal liver and formation of definitive HSCs. At birth, fetal liver HSCs migrate and colonize the bone marrow (BM), spleen and thymus. The real contribution of the YS and placenta to definitive HSCs is still disputed (dotted lines), but not that of the AGM (solid line). **(C)** Hematopoietic cluster on the ventral wall of the AGM region (left), a magnified view (right) showing a hemogenic endothelial cell (flat morphology) transitioning to a hematopoietic cell (round morphology) (Adapted from [13])

### 2.1.2. The Fetal liver: A major hematopoietic reservoir

The liver is the main fetal hematopoietic organ where hematopoietic cells undergo expansion and differentiation. However, as outlined above, it does not produce HSCs *de novo*, but is seeded by circulating hematopoietic cells. The first phase of FL seeding in the mouse embryo initiates at E9.5-10.5, as the liver becomes populated by myelo-erythroid progenitors that generate definitive erythroid cells. The primary wave of FL seeding by YS definitive progenitors is most likely achieved by traveling through the vitelline vessels, which are the first vascular connections to the FL (**Fig. 1A-B**). However, the first HSCs do not appear in the FL until E11.5. The second wave of FL-HSCs colonization most probably originates from the

AGM and the placenta via the umbilical vessels, which is the second major vascular circuit connected to the FL (**Fig. 1A-B**). After E12.5, the FL becomes the primary fetal organ where HSCs start expansion and differentiation. The maximum number of HSCs reaches of about 1000 by E15.5-16.5[14], thereafter, FL-hematopoietic potential declines, parallel to the establishment of the BM as the main HSCs reservoir in the developing newborn and adult (**Fig. 1B**)[15]. Interestingly, unlike the predominantly quiescent BM-HSCs, FL-HSCs are actively cycling, and outcompetes adult BM HSCs when transplanted into irradiated mice[16]. At all times the FL is rich in single-lineage progenitor cells, specifically in CFU-Es (colony-forming unit erythroid) and proerythroblasts, indicating its important role in producing differentiated blood cells.

### 2.1.3. Adult erythropoiesis: from the BM to peripheral blood

Erythropoiesis is the process by which, every second, 2.5 million new RBCs (erythrocytes) are generated in the BM and released into circulation. Erythrocytes represent the most common cell types in the adult blood (adult humans contain between 10 and 20 trillion circulating erythrocytes), and it occurs at a steady, but low basal rate with approximately 1% of cells daily turnover. However, erythrocytes production can substantially increase during times of acute or chronic stress, such as trauma often with blood loss, or infections. Healthy erythrocytes have an average lifespan of 120 days, during which time they are continuously monitored by resident macrophages that reside in the liver and spleen. Whilst sensing damaged, as well as aged erythrocytes, the macrophages execute the removal of these cells from the circulation[17]. Iron released from senescent erythrocytes is recycled for heme synthesis and subsequent incorporation into hemoglobin molecules in new erythrocytes. Due to their limited lifespan, and to ensure physiological homeostasis, new erythrocytes are constantly produced in the BM to replenish the blood stream.

HSCs reside in BM niches consisting of endothelial cells, osteoblasts, stromal cells, hematopoietic cells, and the extracellular matrix, which functions to both support their maintenance and differentiation[18] . A complex, multistep process, governs the transition of the multipotent HSCs to the mature erythrocytes, beginning in the BM and concluding by the release of the mature progeny to the peripheral blood[19]. The journey from the multipotent HSC to the committed erythrocytes occurs over a progression of multiple defined intermediary cell types that exhibit gradual decline in differentiation plasticity, and gain in

erythroid characteristics[20]. The first steps of erythroid differentiation involve the differentiation of HSCs into more committed erythroid progenitors (**Fig. 2A**). The second phase of erythroid maturation involves the differentiation of the lineage committed proerythroblasts to the enucleated reticulocytes in the BM (**Fig. 2A**). The new cells enter the circulation as reticulocytes that are still engaged in protein translation. This phase is characterized by the gradual accumulation of hemoglobin, progressive decrease in cell size and nuclear condensation as a prelude to nuclear polarization, ultimately resulting in enucleation[21]. The final phase of erythroid development involves extensive organelle shedding and the maturation of the reticulocytes into erythrocytes. In humans, it takes around one week for reticulocytes to complete the maturation process. It is during this phase that the erythrocyte acquires its biconcave shape through extensive membrane remodeling and will circulate in the blood stream until it is removed by macrophages [22]. Mature erythrocytes have a diameter of only 6-8μM, their small size and biconcave shape create a large surface area for gas exchange and allows the cells to enter the microcapillaries in the tissues.

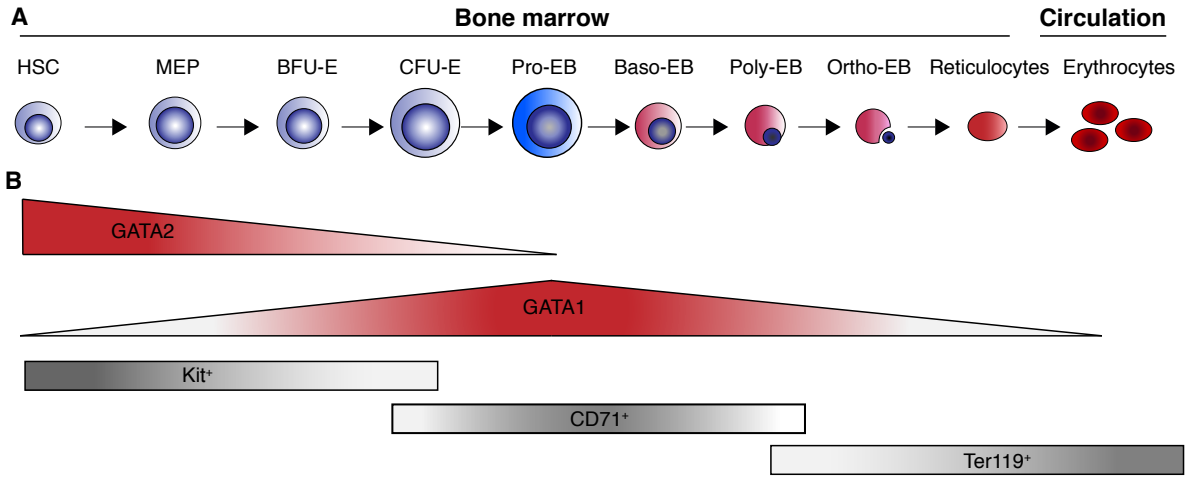



**Fig. 2**: **Erythroid differentiation in the mouse**. **(A)** Production of red blood cells (RBCs) involves the sequential differentiation of HSCs into mature erythrocytes via several intermediate cell stages. **(B)** The different stages of cellular identities are distinguished by dynamic changes in cell size, expression of key molecular regulators (e.g. GATA2, GATA1), and cell surface markers (Kit, CD71, Ter119). (Adapted from [13])

Terminal erythroid maturation occurs in specialized anatomic niches known as ''erythroblastic islands''. Erythroblastic islands are unique to mammalian erythropoiesis and consist of a central macrophage ''nurse'' cell, surrounded by erythroid cells at varying degrees of red cell maturation[23, 24](**Fig. 3**). The cells undergo 4 to 5 divisions to complete their maturation from proerythroblasts to erythrocytes. At the conclusion of terminal differentiation, the central macrophage removes the expelled nuclei by phagocytosis. The

central macrophage also functions to anchor erythroblasts within the island and provides the cellular interactions necessary to drive erythroid differentiation and proliferation, and directs the transfer of iron to erythroid progenitors for heme synthesis[25].

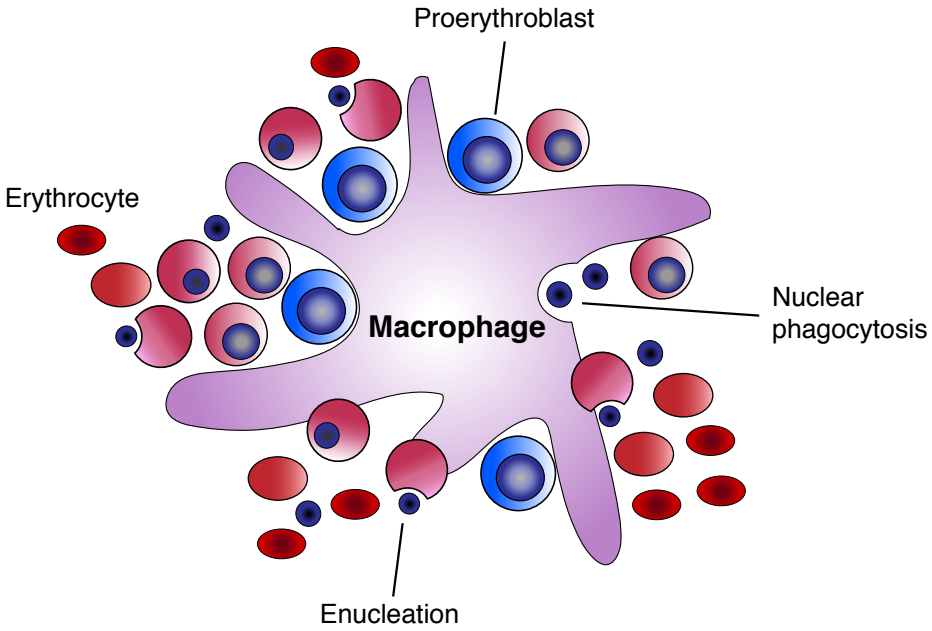



**Fig. 3**: **Differentiation processes occurring within the erythroblastic island**. The central macrophage (Pink) surrounded by erythroid cells at various stages of differentiations. Proerythroblasts are large cells with centrally located nuclei; further differentiated erythroblasts are smaller with more polarized nuclei. Following enucleation, expelled nuclei undergo phagocytosis by central macrophage (Adapted from [24]).

### 2.1.4. Transcriptional regulation of erythropoiesis

The traditional hierarchical model of hematopoiesis with a near-to-homogeneous HSC population residing at the apex has been progressively challenged by a model which proposes the existence of a heterogeneous cellular subpopulations, with predefined lineage fates [26, 27]. During development, and throughout life, hematopoietic progenitors respond to various external and intrinsic cues to differentiate, through plastic intermediate cell identities, to fully mature cells with specific physiological functions. Major molecular players influencing erythroid fate determination include classical hormones (e.g. thyroid hormones, androgens, corticosteroids), vitamins (e.g. vitamin B12 and folic acid), iron and regulators of iron metabolism (e.g. transferrin receptors-1 and -2), and hematopoietic associated growth factors such as stem cell factor (SCF) and erythropoietin (EPO).

The process of lineage choice and differentiation is determined by the balance of transcription factors (TFs) that mediate differentiation along one hematopoietic lineage and

restrict differentiation along another hematopoietic lineage. TFs work in a combinatorial manner, together with cofactors and chromatin modifiers, to activate and/or repress their target genes, leading to the creation of a unique transcriptional program for each cell type. During erythropoiesis, the first step in lineage determination involves restriction to either the myeloid or the lymphoid branch. Several of the TFs that are preferentially expressed in cells destined to one of those fates have been identified, but few have been extensively studied as well as PU.1 and the GATA family of TFs. PU.1 mediates differentiation of HSCs along the lymphoid and myeloid lineages and inhibits differentiation of erythroid cells, on the other hand, GATA family members (GATA1 and GATA2) have the opposite activity and promote differentiation along the myelo-erythroid lineage[28, 29].

## 2.2. Acute myeloid leukemia (AML)

AML is a highly heterogenous hematological disease of aberrant clonal proliferation and differentiation of myeloid stem and progenitor cells. The disease can arise in healthy individuals *de novo*, or in patients with prior hematological disorders, such as myelodysplastic disorder syndrome (MDS) and myeloid proliferative neoplasm (MPN), or as a result of prior therapy (e.g., chemotherapy with topoisomerase II, alkylating agents, or radiation)[30, 31]. However, in most cases, AML is *de novo* diagnosed in previously otherwise healthy individuals. If untreated, accumulation of poorly differentiated myeloid cells in the BM leads to impaired normal hematopoiesis, leukocytosis, thrombocytopenia and anemia, hematopoietic failure and eventually death secondary to an infection or bleeding[32, 33].

The underlying genomic aberrations associated with AML has been thoroughly investigated since the 1970s, initiated by the examination of chromosomal karyotypes of patients' cells[33]. This has led to the identification, and later, the molecular characterization of several prevalent chromosomal rearrangement such as the t(8;21)(q21;q22) fusing the core-binding factor AML (RUNX1) to the core-repression ETO (RUNX1-RUNX1T1)[34], and the t(15;17)(q22;q12), fusing the promyelocytic leukemia (PML) to the retinoic acid receptor alpha (RAR$\alpha$), resulting in the production of a PML-RAR$\alpha$ chimeric protein[35]. One of the most interesting primary observations made from these studies, was that the majority of molecular aberrations detected involved at least one known regulator of normal hematopoiesis. Currently, more complex, recurring, and/or rare alterations including cytogenetically-silent translocations, point mutations in metabolic regulators and small copy number changes, are

accurately detected at high resolution utilizing next generation sequencing (NGS) technologies[33]. Furthermore, the rapid development of cloning and genome editing (GE) techniques has enabled scientists to generate various genetic cellular or animal models expressing the suspected proto-oncogenes, thus facilitating the *in vivo* study of their functional contribution to AML onset[36].

Modeling AML-associated lesions in animals revealed that leukemogenesis follows specific and multistep trajectories. This has led to the conception of the two-hit model , which divided the various mutations associated with AML into functional classes, including "class I mutations" (e.g. $FLT3^{ITD}$, $FLT3^{TKD}$, $KIT^{mut}$ & $NRAS^{mut}$), which activate signal transduction pathways and confer pro-proliferative and survival advantage, and "class II mutations" (e.g. $RUNX1^{mut}$, *PML-RAR*$\alpha$, *CBF*$\beta$*–MYH11* & *RUNX1-RUNX1T1* which impair normal hematopoietic differentiation[37] (**Fig. 4**). While neither of these mutations are sufficient to cause leukemia in isolation, together, they cooperate to trigger the disease. The initiating aberration in these AMLs are thought to be class II mutations, whereas class I mutations are typically a later event. The initial mutation ''1st hit'' can occur in progenitor cells in different developmental stages (**Fig 4A**), leading to the formation of cells blocked in differentiation, however, still retain certain aspects of their former identity. When a second mutation ''2nd hit'' occurs in the these cells (**Fig. 4B**), they become fully transformed into a leukemic blast with characteristic influenced by their ''origin'' (**Fig. 4C**)[38]. This model provided a general framework to view the etiology of AML as a disease in which differentiation is blocked and proliferation is increased. Although not all detected mutations can be categorized into one of the two classes, nevertheless they are thought to produce equivalent effects.

In a recent comprehensive whole-genome NGS study conducted on a cohort of 1540 AML patients, Papaemmanuil and colleagues identified 5234 leukemic driver mutations in 76 genes [33]. Genetic abnormalities detected in AML samples can be divided into distinct categories according to gene functions: fusions involving myeloid transcription factors (*PML-RAR*$\alpha$, *MYH11-CBF*$\beta$, & *RUNX1-ETO*), DNA methylation genes (*DNMT3A*, *TET2* & *IDH1/2*), chromatin modifiers (*MLL-fusions*, *ASXL1*, *EZH2* & *MLL-PTD*), spliceosome genes (*SF3B1*, *SRSF2* & *U2AF1*), cohesins (*SMC1A*, *SMC3* & *STAG2*), signal transduction genes (*FLT3*, *NRAS* & *KIT*), *NPM1* mutations, and chromosomal aneuploidy (e.g. complex karyotype, -5/5q, -7/7q, -12/12p &TP53)[33, 39, 40].

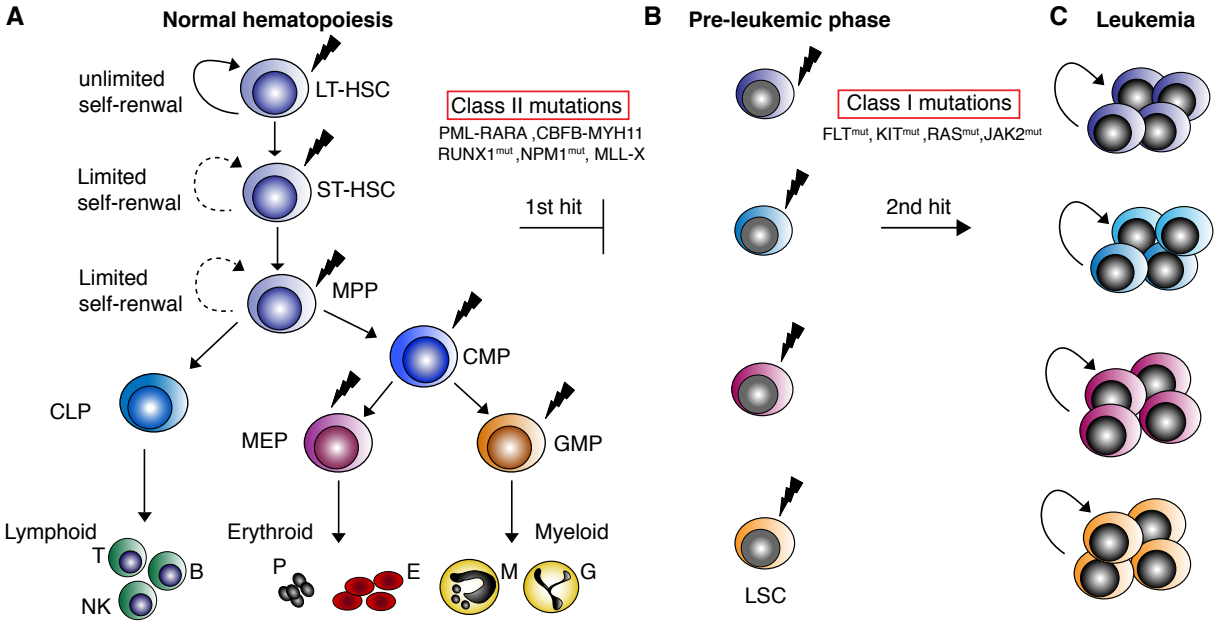



**Fig. 4: A molecular and cellular model of AML**. **(A)** In normal hematopoiesis, quiescent long-term repopulating HSCs (LT-HSC) with unlimited self-renewal capacity (solid circle arrow) give rise to short-term HSC (ST-HSC), and multipotent progenitors with (MPPs) with limited self-renewal capacity (doted circle arrow). MPPs further differentiate toward lineage-restricted progenitors, such as common lymphoid and myeloid progenitors (CLPs, CMPs) that have lost self-renewal capacity. These lineage restricted progenitors undergo extensive proliferation to produce all components of mature blood cell types. **(B)** The formation of a leukemic stem cell (LSC) in myeloid leukemia may result from mutations in cells in different stages of the hematopoietic hierarchy (cell of origin). In AML, type II mutations are associated with block hematopoietic differentiation, leading to the accumulation of progenitor cells, insufficient hematopoiesis, and formation of a ''preleukemia phase''. **(C)** A second mutational hit from type I mutations confer a proliferative and/survival advantage to progenitor cells (solid circle arrow) resulting in an excessive uncontrolled clonal expansion, and the development of leukemic clones; GMP, granulocyte-monocyte progenitors; E, erythrocyte; G, granulocyte; NK, natural killer cell; M, monocyte; P, platelet; T, T cell (Adapted from [41]).

### 2.2.2. Classification of AML

Due to the heterogenous nature of AML, a group of morphologists from France, the United States, and Great Britain developed a classification system designed to quantify and standardize definitions of clinically and biologically distinct subtypes of AML and acute lymphoblastic leukemia (ALL), known today as the French–American–British (FAB) classification system. Established in 1976, the system outlines eight broad AML subtypes (M0 through M7)[32] based on the morphological and cytochemical features of leukemic cells obtained from patients' peripheral blood or BM. The FAB nomenclature assigned to each AML subtype reflects the normal marrow element that the blast most closely resembles (**Table 1**).

| FAB subtype | Name | Adult AML patients (%) |
|---|---|---|
| **M0** | Undifferentiated acute myeloblastic leukemia | 5% |
| **M1** | Acute myeloblastic leukemia with minimal maturation | 15% |
| **M2** | Acute myeloblastic leukemia with maturation | 25% |
| **M3** | Acute promyelocytic leukemia | 10% |
| **M4** | Acute myelomonocytic leukemia | 20% |
| **M4eos** | Acute myelomonocytic leukemia with eosinophilia | 5% |
| **M5** | Acute monocytic leukemia | 10% |
| **M6** | Acute erythroid leukemia | 5% |
| **M7** | Acute megakaryocytic leukemia | 5% |

**Table 1**: The French-American-British (FAB) classification of myeloid malignancies (M0-M7), and their corresponding frequencies in adult patients. (adapted from https://www.cancer.org/cancer/acute-myeloid-leukemia)

In 2001, the World Health Organization (WHO) introduced a new classification system, which was subsequently updated in 2008[42] and in 2016[43], in an effort to integrate the latest advances made in the diagnosis and management of AML. In addition to clinical, morphological, and histochemical information, the WHO classification of AML also incorporates prognostic, immunophenotypic, and genetic data to define six major disease entities (AML with recurrent genetic abnormalities; AML with myelodysplasia-related features; therapy-related AML; AML not otherwise specified; myeloid sarcoma; and myeloid proliferation related to Down syndrome)[43, 44], which are further divided into 11 subtypes among cases of AML with recurrent genetic abnormalities (**Table 2**). The mutated *NPM1* and *CEBP*$\alpha$ subtypes were first introduced as a provisional entries in the 2008 revision[42], while the *BCR-ABL1*$^{+}$, and AML with mutated *RUNX1* were included in 2016 revision[43]. The classification also distinguishes between genetic abnormalities associated with *de novo* AML from AML with features and history of myelodysplasia-related cytogenetic abnormalities such as monosomy 5 or 7, and deletion 5q or 7q[32].

| Categories | Observed alterations |
|---|---|
| **AML with recurrent genetic abnormalities** | t(8;21) (q22;q22.1); *RUNX1-ETO*<br>inv(16) (p13.1q22) or t(16;16)(p13.1;q22); *CBFβ-MYH11*<br>*PML-RARα* (APL)<br>t(9;11) (p21.3;q23.3); *MLLT3-KMT2A*<br>t(6;9) (p23;q34.1); *DEK-NUP214*<br>inv(3) (q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM (EVI1)*<br>Mutated *NPM1*<br>Biallelic mutations of *CEBPα*<br>t(1;2) (p13.3;q13.3); *RBM15 -AMKL1*<br>AML with *BCR-ABL1* (provisional entry)<br>Mutated *RUNX1* (provisional entry) |
| **AML with myelodysplastic changes & Therapy-related neoplasms** | AML with minimal differentiation<br>AML without maturation<br>AML with maturation<br>Acute myelomonocytic leukemia<br>Acute monoblastic/monocytic leukemia<br>Pure erythroid leukemia<br>Acute megakaryoblastic leukemia<br>Acute basophilic leukemia<br>Acute panmyelosis with myelofibrosis |
| **Myeloid sarcoma**<br>**Myeloid proliferation related to Down syndrome** | Transient abnormal myelopoiesis (TAM)<br>Myeloid leukemia associated with Down Syndrome |

**Table 2: The 2016 revised WHO classification of myeloid neoplasms with diagnostic genetic aberration**. (adapted from[43]).

### 2.2.3. Leukemia from the erythroid lineage

Myeloid neoplasms with erythroid predominance (MN-EP) is a broad term used to described a heterogenous group of malignancies affecting the myeloid-erythroid linage of the hematopoietic system, including *polycythemia vera* (PV), MDS with prominent erythroid predominance (previously acute erythroid leukemia or AEL) and ‘’pure erythroid leukemia” (PEL). PEL is characterized by an uncontrolled proliferation of maturation-arrested early erythroid precursors, mainly proerythroblasts[45]. It is a very rare disease, representing < 1% of all AML cases.

Due to poor reproducibility of non-erythroid blast counts, and as an attempt to achieve consistency in expressing blast percentages across all myeloid neoplasms, the WHO has constantly revised its diagnostic criteria for PEL [43, 46-48]. In the latest 2016 update of the WHO classification, myoblasts are always counted as a percentage of total BM cells. Thus, AEL diagnosis has been replaced with AML, not otherwise specified (AML-NOS) when absolute myeloblasts ≥ 20% of all nucleated marrow cells, unless molecular or cytogenic markers identify another AML subtype, or an MDS when absolute myeloblasts < 20% of all nucleated marrow cells but ≥ 20% of non-erythroid cells (MDS with excess blasts). A diagnosis of PEL remains reserved for cases with >80% immature erythroid precursors, ≥30% proerythroblasts and total blast cell percentage <20%[43] (**Table 3**). This change was based on the close resemblance of acute erythroid leukemia to MDS in terms of its clinical, morphological, and genetic abnormalities.

| Neoplasm | Key features/Criteria |
|---|---|
| **Polycythemia Vera** | *JAK2* $^{V617F}$ or *JAK2* exon 12 mutations and<br>Myeloblasts < 20% and<br>Increased red cell mass (RCM): > 25% above predicted normal value<br>WHO criteria for PV fulfilled |
| **Myelodysplastic syndrome with erythroid predominance** | Bone marrow cell dysplasia<br>≥50% erythroid precursors in BM and<br>Myeloblasts >20% but ≥20% of non-erythroid |
| **AML with myelodysplasia-related changes & AML, NOS (non-erythroid subtype)** | ≥50% erythroid precursor in BM<br>Myoblast ≥20% |
| **Pure erythroleukemia** | >80% of all BM cells are immature erythroid precursors with ≥30% proerythroblasts<br>Myeloblasts <20% |

**Table 3: The latest WHO classification for erythroid malignancies**. PV: polycythemia vera, BM: bone marrow (Adapted from[49]).

### 2.2.4. Molecular genetics of erythroid leukemia

Erythroleukemia is mostly diagnosed in people older than 50 years, however, rare cases were reported with erythroleukemia occurring in very young children[50-52]. Unfortunately, in contrast to erythroid MPN, and despite a growing list of reported mutations, very little is known about the molecular mechanisms mediating the differentiation blockage, enhanced survival and proliferation of proerythroblasts leading to initiation and progression of PEL in patients. Earlier studies, conducted in the 50s and 70s of the last century, showed that particular viruses, such as the avian erythroblastosis virus and Friend's spleen focus-forming virus, can induce malignancy resembling erythroid leukemia in susceptible animal models[53-55], however, no evidence of viral origin for the human disease has been identified so far.

Cytogenetic analysis of PEL cases usually shows a complex karyotype, with frequent abnormalities including complete or partial monosomy of chromosomes 5 and 7[56]. Two particular chromosomal translocations t(1;16)(p31;q24) and t(11;20)(p11;q11) have been reported in pediatric PEL resulting in expression of *NFIA-CBFA2T3*[50, 52] or *ZMYND8-*

*RELA*[51] fusion genes respectively. Although no unique genetic mutations have been described in PEL, *TP53* mutations seem to be the most prevalent [33, 45, 57]. Mutations commonly found in other types of AML, such as *FLT3*, *NPM1*, and *CEBP$\alpha$*, do not show a particular association with the PEL phenotype, which indicates that *TP53* mutations may play a role in the pathogenesis of PEL.

### 2.2.5. Leukemic cell of origin

Compared to solid cancers, leukemia development in is associated with very few molecular aberrations. However, the disease is highly heterogeneous with varying clinical manifestations and prognosis. The WHO prognostic categorization of AML into favorable, intermediate and poor outcome, is based on cytogenetic and molecular alterations detected at diagnosis in patients' samples. Combining data from interdisciplinary studies have linked specific molecular lesions, to particular cellular morphology, immune-phenotype, gene expression signature, and ultimately patients' response to treatment and overall survival.

The mammalian hematopoietic system is organized into a dynamic hierarchy of hemopoietic cells that gradually lose the ability to self-renewal and proliferate as they commit toward a lineage specific differentiation (**Fig. 4**). Several lines of evidence are suggesting that, within leukemia subtypes carrying comparable genetic lesions, the difference in leukemic characteristics is derived from different 'cell of origin' (**Fig. 4A**)[58, 59]. The term 'cell of origin' was coined for the cells which acquire the first oncogenic hit that resulted in the initiation of cancer[60]. One of the first results that support this hypothesis came from the observation that unlike the AML-associated *MOZ-TIF2* fusion gene, expression of the *BCL-ABL* fusion associated with chronic myeloid leukemia (CML) was not able to transform committed progenitor cells (CMPs and GMPs), and transformation and disease initiation could only be achieved from uncommitted HSCs[61]. In contrast, MLL-rearrangements (MLL-r) were able to transform and initiate leukemia is both uncommitted HSC, and progenitor[58, 59, 62, 63], however, differences in latency, aggressiveness and treatment resistance were observed when the disease was initiated from HSCs or GMPs[58, 59]. Specifically, even within the long-term HSCs (LT-HSCs) population, different cells seem to respond to MLL-AF9 transformation differently, as between 10-20% of MLL-AF9$^{+}$ LT-HSC transplanted mice developed a particularly aggressive disease with very short latency[59]. Comparative transcriptome analysis showed distinct gene expression signatures, with the aggressive LT-HSC-derived AML

blasts expressed higher levels of the TF ecotropic viral integration site 1 (*Evi1*) and the TF ETS-related (*Erg*) genes. *Evi1* is a member of the zinc finger family of TFs known to be critical regulators during the development of embryos' hematopoietic system and the functional integrity of adult HSCs[64, 65]. Based on several published bone marrow reconstitution studies in mice, *Evi1* expression appears to identify a distinct population of LT-HSC with long repopulating capacity [66, 67]. Notably, higher expression of *Evi1* in the HSC compartment seems to predict poor clinical outcome in myeloid malignancies[68, 69]. *Erg* belongs to the ETS family of TFs, and is also an essential regulator of hematopoietic commitment and maintenance[70]. HSCs from mice heterozygous for a loss-of-function *Erg* mutation were incapable of competing with normal cells for long-term reconstitution[71]. Loss of ERG protein also caused failure in HSC maintenance and instant depletion of definitive hematopoiesis[72]. On the other hand, similarly to *Evi1*, high expression of *Erg* is also associated with poor prognosis in particular subgroup of acute myeloid (AML) and acute T-lymphoblastic leukemia (T-ALL)[73].

## 2.3. Epigenetic regulation of leukemia

The human genome contains more than 20,000 protein coding genes, however, only a fraction of these genes is expressed in each of the approximately 200 cells types found in the adult human at any time point. Epigenetic refers to the heritable changes in gene expression, without actual changes in the DNA sequence. Epigenetic modifications, mainly DNA methylation and histone posttranslational modifications, have been implicated in fundamental biological process, such as DNA replication and repair, recombination, cell cycle progression and transcription[74]. DNA methylation occurs at CpG dinucleotides (CpG islands), by the addition of a methyl group to the C5 position of cytosine, and the reaction is established by DNA methyltransferases (DNMTs).

Histones are small basic proteins involved in the spatial organization of DNA within the nucleus. Histone tails are covalently modified, for example by methylation, phosphorylation, acetylation, SUMOylation, and ubiquitylation. These epigenetic, posttranslational modifications (PTM) of histones are highly controlled, governed by three categories of enzymatic proteins: ''writers'' are involved in the addition of chemical groups to histone tails, ''erasers'' involved in the removal of these chemical groups, and ''readers'' are proteins that specifically recognize these histone modifications (**Fig. 5**). The balance between competing

enzymes such as the histone lysine methyltransferase (HKMTs) and histone lysine demethylase (HKDMs), and the histone acetyltransferase (HATs) and histone deacetylases (HDACs) create a tight circle of gene expression regulation ensuring appropriate activation and repression of genes[75, 76]. The genome-wide localization of epigenetic regulator elements, such of histone PTMs, DNA methylation and chromatin regulatory proteins, in a variety of biological processes has allowed scientist to begin defining ‘’epigenomic landscapes’’ and relate them to different cellular phenotypes[77].

Lysine methylation is a prominent feature of the post translational histone modifications in the regulation of chromatin structure and function. HKMTs target specific residues on histones H3 and H4 tails for modifications, transferring one, two or three methyl groups to lysines. Methylation of histone H3K4, H3K36, and H3K79 is associated with transcriptional activations, while dimethylation and trimethylation of H3K9 and H3K27 are associated with transcriptional repression. Methylation at specific lysine residues is catalyzed by specific methyltransferases: H3K4 methylation is catalyzed by the methyltransferase MLL1 (KMT2A) and MLL2 (KMT2D), H3K79 is methylated by DOT1L, H3K27 methylation is catalyzed by EZH1, and H3K36 methylation is performed by SETD2 and NSD HKMT family.

Recurrent genetic alterations occurring in AML are found at the level of various genes involved in the epigenetic control of the DNA and histone methylome. In fact, genes involved in DNA methylation (*DNMT3*, *TET2, IDH1* & *IDH2*) or histone methylation and demethylation (*NSD*, *EZH2*, *MLL* & *DOT1L*) are frequently mutated in AML. These observations have strongly supported a major role of dysregulated epigenetic regulatory processes in leukemia onset and development[33, 39, 40].

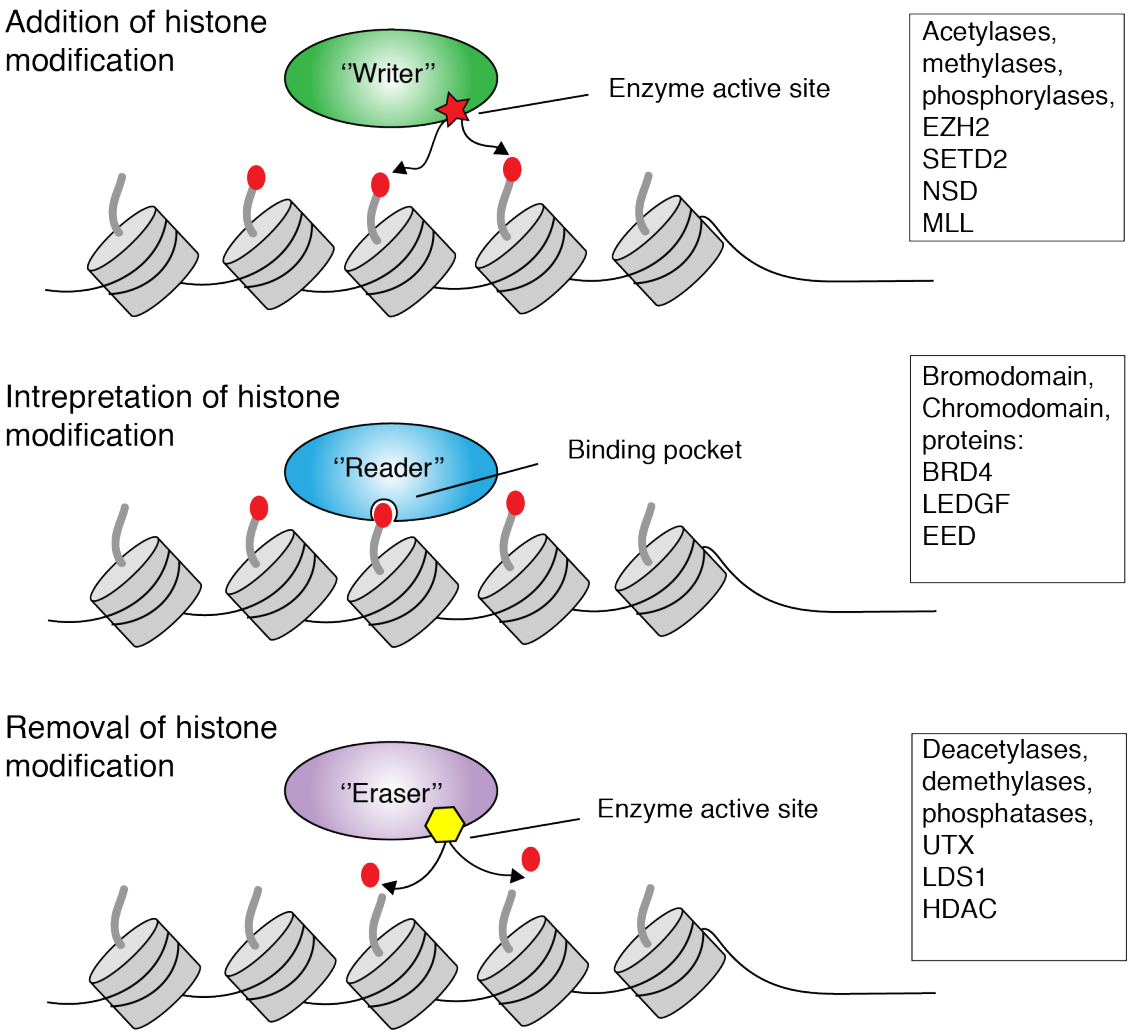



**Fig. 5: Classes of histone modification enzymes**. The addition of histone post-translational modifications is catalyzed by a class of enzymes known as chromatin "writers". Chromatin writers can various molecules moieties to histone tails including methyl, acetyl and phosphate groups. Chromatin ''readers'' contain unique binding domains, such as chromo-, bromo-, and PHD-finger domains to bind histones marked by specific modifications, or a combination of modifications. Finally, chromatin "erasers" catalyze the removal of histone modifications, thereby reversing their biochemical effects (modified from [78]).

### 2.3.1. The mixed lineage leukemia (MLL1) lesions in leukemia

The *MLL* (11q23) histone methyltransferase mutations are recurrently detected in acute leukemia, including AML, acute lymphoblastic leukemia (ALL), and mixed lineage leukemias (MLL). *MLL* translocations are predominantly observed in pediatric leukemia (80%)[79], however, they are also detected in around 5-10% of adult primary or secondary AML cases[79]. *MLL-r* are frequently associated with an aggressive form of the disease, and poor prognosis. MLL is a large multidomain protein with several putative DNA-binding domains at the N-terminus, and a C-terminal SET [Su(var) 3-9, Enhancer-of-zeste, Trithorax] domain specific for H4K3me1/2/3. The most frequent MLL fusion partners, present in around

80% of *MLL-r* leukemias, are members of genes encoding super elongation complex nuclear proteins AF4, AF9, AF10, AF6 and ENL[79]. Generally, MLL1 fusion proteins lose the catalytic SET methyltransferase domain, but do retain their DNA-binding motifs that target genes. In addition to being involved in translocations, the *MLL* gene is also involved in other aberrations such as partial tandem duplications (*MLL-PTD*), consisting of an in-frame repetition of *MLL* exons[80].

MLL plays a critical role in HSCs self-renewal by regulating expression of the developmentally important homeobox *Hox* cluster genes[81, 82]. Loss of *Mll* results in embryonic lethality in mice, and altered *Hox* gene expression[83]. These *Mll*$^{-/-}$ embryos exhibit defects in YS hematopoiesis, with reduced proliferation and/or survival of hematopoietic progenitors[84, 85] and defective HSCs activity in the AGM region[86]. The *HOX* gene family is a highly conserved group of homeodomain-containing TF that play a role in promoting HSC self-renewal, and which downregulation correlates with terminal differentiation[87]. A hallmark of *MLL-r* leukemia is high expression of the *HOXA* cluster genes and *MEIS1*. DOT1L was found to interact with several recurrent *MLL* translocation partners such as AF9, ENL, and AF10, suggesting that MLL1 fusion proteins may directly recruit DOT1L to MLL1 fusion target loci through the C-terminal portion of the chimeric proteins[88]. Later studies using loss-of-function mouse models reported that leukemias driven by MLL1 fusion proteins are selectively dependent on DOT1L for leukemia initiation and maintenance[89]. DOT1L catalyzes the methylation of H3K79, a histone mark associated with transcriptional activation. Thus, the consensus mechanism of *Mll-r* cellular transformation is thought to occur by the inappropriate recruitment of DOT1L by MLL-r fusions to the *HoxA* gene locus, causing a misplaced enrichment in H3K27 methylation at the gene locus, coupled with uninterrupted transcription. This subsequently leads to increased *HoxA* target gene expression and block in myeloid lineage differentiation.

### 2.3.2. The NUP98-NSD1 fusion

A cytogenetically silent chromosomal translocation t(5;11)(q35;p15.5) leading to the expression of a NUP98-NSD1 fusion has been detected in pediatric MDS and AML with poor prognosis[90-92]. The fusion contains the FG-repeat domain of NUP98, a nucleoporin protein family member that can interact with the histone acetyltransferase CBP/EP300, and the carboxyl terminal of the nuclear receptor binding SET domain protein (NSD1), that retain the

five PHD fingers, the Cys-His-rich domai n (C5HCH), one PWWP domain, and the catalytic SET domain (**Fig. 6**). Retroviral infection of NUP98-NSD1 in hematopoietic cells enhanced expression of *HoxA7*, *HoxA9*, and *Meis1* proto-oncogenes[93]. Mechanistically, several lines of evidence suggest that the NUP98-NSD1 fusion protein is directly recruited to the distal *HoxA* cluster, leading to H3K36 methylation and histone acetylation. This chromatin setting could prevent the recruitment of EZH2 (H3K27me3) and PcG (the Polycomb group- conserved transcriptional repressors) proteins leading to a block of differentiation, progenitor immortalization, and ultimately to tumorigenesis[93]. Structure/function analysis of the NUP98-NSD1 fusion showed a requirement for the FG repeats of NUP98 for recruitment of the CBP/EP300 histone acetylation complex, and the NSD1-SET and PHD domains for H3K36 methyltransferase activity and DNA binding at the *HoxA* locus, for leukemic transformation[93]. Co-transduction of BM-derived progenitor cells with retroviruses expressing *NUP98-NSD1* fusion and the *FLT3-ITD* mutation resulted in development of AML in mice after a very short latency[94]. Notably, AML patients (82%) with concomitant *NUP98-NSD1* and *FLT3-ITD* had a worse outcome than those harboring *NUP98-NSD1* only[95], suggesting a potent cooperation between NUP98-NSD1 and FLT-ITD during leukemic transformation that could be targeted therapeutically with FLT3 inhibitors[96, 97].

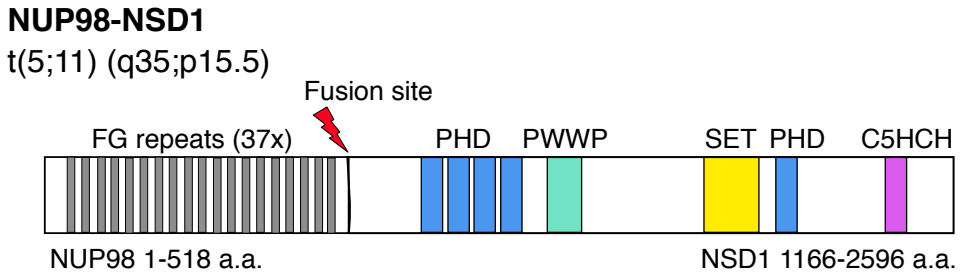



**Fig. 6**: **The NUP98-NSD1 fusion protein**. In-frame genomic rearrangement leads to the fusion of the C-terminus of NSD1 (amino acids 1116-2596) to NUP98 (amino acids 1-518). Three domains of the fusion play a critical role in leukemic transformation, the FG (Phenylalanine, Leucine) repeats of NUP98, the fifth PHD finger, and SET domain of NSD1.

### 2.3.3. The nuclear receptor binding SET domain protein (NSD) family

The NSD family of HMTase is a phylogenetically distinct subfamily of HKMTases comprised of three nuclear receptor SET domain containing 1, 2 and 3 proteins (NSD1, NSD2 & NSD3). The *NSD1* (*KMT3B*) gene is comprised of 24 exons encoded on chromosome 5q35, *NSD2* maps to chromosome 4p16 (MMSET/WHSC1), and *NSD3* on chromosome 8p12 (NSD3/WHSC1L1)[98]. The full-length members of the NSD family contain multiple protein

domains, such as the evolutionarily conserved catalytic SET domain responsible for the HMTase activity[99-101], two PWWP (Proline-tryptophan-tryptophan-proline) domains that are critical for binding methylated histone H3 as well as DNA, and plant homeodomain (PHD) zinc fingers important for interactions with other methylated histones (**Fig. 7**) [102].

NSD1, NSD2, and NSD3 function by adding methyl groups to H3K36[98, 101, 103, 104]. H3K36 is found in non-, mono-, di-, and tri-methylated forms (me1, me2 and me3 respectively), and this methylation is associated with transcription of active euchromatin[105]. The various methylated form of H3K36 may have different biological functions depending on the organism or cellular context. In humans, there is a preference for H3K36me1 at active promoters, and this mark is detected in active regions of the β-globin locus, while the H3K36me3 along the same locus body is broadly associated with active transcription[106]. Experimental evidence suggests that NSD1, NSD2 and NSD3 play non-redundant roles during development as genetic deletion of either *Nsd1* or *Nsd2* is lethal in mice[107]. Thus, alterations or amplifications of *NSD1*, *NSD2* and/or *NSD3* that dysregulate H3K36 methylation marks have profound effects on cell growth and differentiation and are linked to several developmental defects and diseases. In addition, overexpression, gain of function somatic mutations, and translocation of NSDs have been reported to frequently occur in variety of cancers (**Table 4**).

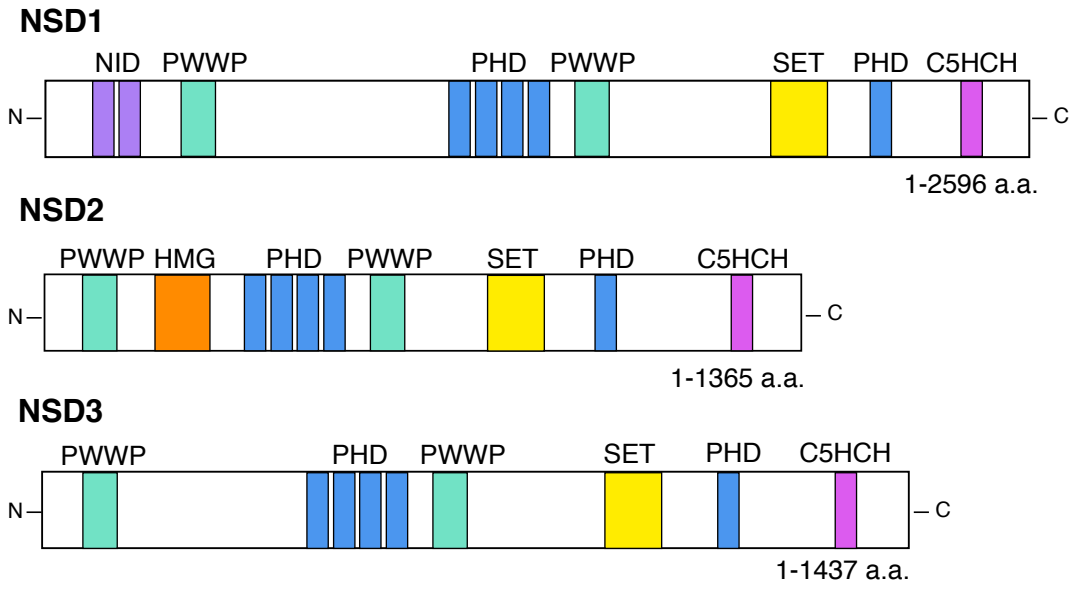



**Fig. 7**: **NSD family protein domain structure**. NID (nuclear repressor interacting domain), PWWP (Proline-tryptophan-tryptophan-proline), PHD (plant homeodomain zinc fingers), SET (Su(var) 3-9, Enhancer-of-zeste, Trithorax), C5HCH (Cys-His), and HMG (high mobility group). NSD1 is the longest member with 2596 amino acids land the only one that contains an NID domain, which facilitate its binding to corepressors and coactivators, such as thyroid hormone (TR), and estrogen receptors (ER).

| NSD | Disease | Alteration | Ref |
|---|---|---|---|
| **NSD1** | Acute Myeloid leukemia (AML) | *NUP98-NSD1*, t(5;11) (q35;p15.5) fusion & 5q35 monosomy | **[90-92, 108]** |
| | Myeloid dysplastic syndrome (MDS) | *NUP98-NSD1*, t(5;11) (q35;p15.5) fusion & 5q35 monosomy | **[109, 110]** |
| | Multiple myeloma (MM) | Overexpression | **[111]** |
| | Lung cancer | Inactivating mutations and deletions | **[112]** |
| | Neuroblastomas | Sotos associated, inactivating mutations | **[113]** |
| | Glioblastomas | Sotos associated, inactivating mutations | **[113]** |
| | Head and neck squamous carcinoma (HNSC) | Inactivating mutations and deletion | **[112, 114, 115]** |
| | Skin cancer | Reduced expression | **[116]** |
| | Sotos /Weaver syndrome | Haploinsufficiently, Intragenic mutations & 5q35 microdeletions. Inactivating mutations | **[117, 118]** |
| **NSD2** | Acute lymphoblastic leukemia (ALL) | Substitutions, activating mutations | **[119]** |
| | Multiple myeloma (MM) | *FGFR3-NSD2*, t(4;14) (p16;q32) fusion | **[120]** |
| | Prostate cancer | Overexpression | **[121]** |
| | Breast cancer | Overexpression | **[122]** |
| | Glioblastoma multiform | Overexpression | **[123]** |
| **NSD3** | Acute myeloid leukemia (AML) | *NUP98-NSD3*, t(8;11) (p11.2;p15) fusion | **[124]** |
| | Lung cancer | *NSD3-NUT*, t*(*8*;*15*)(p12;*q15*)* fusion | **[125, 126]** |
| | Breast cancer | Amplification, overexpression | **[127]** |

**Table 4: Malignancies associated with genetic aberration in NSD family of methyltransferases**.

### 2.3.3.1. NSD1

NSD1 was originally isolated in a yeast two hybrid screen for proteins associated with the ligand-binding domain of the retinoic acid receptor alpha (RAR$\alpha$)[99, 128]. It contains two nuclear steroid receptor interaction domains (NID-L and NID+L) that regulate the function of retinoic acid, thyroid, retinoid X, and estrogen nuclear receptors [99]. Subsequent studies

revealed that the SET domain of NSD1 methylated H3K36 and H4K20 *in vitro* [107]. However, more recent experiments suggest that the enzyme is specific for H3K36[101]. More importantly, studies using defined nucleosome substrates with various forms of methylated histone H3K36 suggested that NSD1 is a di-methylase specific for H3K36[129]. Deletion of NSD1 reduce the levels of H3K36me1, 2 and 3, suggesting that NSD1 is mono/di-methylase and that this modification serves as a substrate for trimethylation by SETD2[103]. In addition, depletion of NSD1 reduced RNA polymerase II (RNAPII) promoter occupancy and inhibited the transition of RNAPII from an initiation to an elongation competent state [103].

NSD1 is critical for normal murine embryonic development. *Nsd1*$^{-/-}$ embryos are able to gastrulate and initiate mesoderm formation, but fail at embryonic day E6.5 due to high levels of apoptosis[107]. Germline defects (mostly loss-of-function mutations) of *NSD1* are linked to Sotos syndrome, as well as a unique variant of Weaver syndrome[118]. Characterized as overgrowth disorders, affected patients experience pre-/postnatal overgrowth, enhanced bone age, neurodevelopmental delay, and an enhanced risk for cancer[130]. Interestingly, recent reports have shown that DNA isolated from Sotos patients is hypomethylated with distinct methylation signature comparable to those observed during ageing[131, 132]. In addition to histone H3, non- histone targets of NSD1 have been reported. The carboxyl-terminus of NSD1 contains a unique PHD finger region termed the PHDvC5HCH domain that interacts with the transcription repressor Nizp1. Mutations in the NSD1 PHDvC5HCH domain may interfere with NIsp1 transcription repression[133]. NSD1 has also been reported to coimmunoprecipitate with the NF-κB subunit p65, and to be required for p65 methylation at lysine residues K218 and K221[134]. However, whether p65 is a direct substrate of NSD1 has yet to be confirmed *in vivo*.

Aberrant NSD1 expression has been associated with many cancer pathologies, and tumors occur in 3% of patients diagnosed with Sotos syndrome[135]. In addition, NSD1 was inactivated via CpG islands-promoter hypermethylation in neuroblastomas and gliomas [113]. This transcriptional silencing was associated with diminished methylation of H3K36 and H4K20. NSD1 expression was also downregulated during the progression from non-metastatic to metastatic melanoma[111], and a transposon screen for frequently occurring mutations in skin tumors of mice revealed that NSD1 is among those gene with significantly decreased expression during tumor development (**Table 4**)[116] .

In summary, NSD1 is an H3K36-specific mono- and dimethyltransferase that promotes or represses transcription, and is critical for normal growth and development. Aberrant

expression of NSD1 drives the pathobiology of Sotos syndrome and tumorigenesis. Although much work has been performed regarding the structure and function of NSD1, we are still clueless on the exact cellular context or mechanisms regulate the oncogenic properties of NSD1.

## 2.4. The GATA TF family

BM erythropoiesis is regulated by EPO (the kidney-derived cytokine erythropoietin), which is induced under oxygen deprivation (hypoxia) conditions and stimulates the terminal proliferation and differentiation of CFU-E progenitors. Binding of EPO to EPO receptors (EpoR) on the surface of erythroid progenitor cells triggers the activation of multiple intracellular signal transduction pathways, including JAK2/STAT5, AKT, and MAPK[136, 137]. In erythroid cells, the downstream activation of STAT5 and other EPO-regulated transcriptional regulators leads to the activation of a relatively small number of erythroid lineage-restricted transcription regulators, including, SCL/TAL1, KLF1, and the GATA family of TFs.

The expression profiles of the six structurally related GATA family members can be divided into hematopoietic (GATA1 to GATA3) and non-hematopoietic (GATA4 to GATA6). All members of the GATA family have highly conserved DNA-binding domains that bind to the 6-nucleatide consensus DNA sequence (A/T)GATA(A/G)(WGATAR) through two zinc fingers (**Fig. 8**)[138],[139]. The two zinc fingers domains interact with the major groove of the DNA at separate target sites, and each has a distinct function. The C-terminal zinc finger (C-ZF) binds to the GATA consensus sites, whereas the N-terminal zinc finger (N-ZF) promotes the binding of GATA to specific DNA sequences through the interaction with other zinc finger coregulators, such as FOG1 (Friend of GATA1)[140].

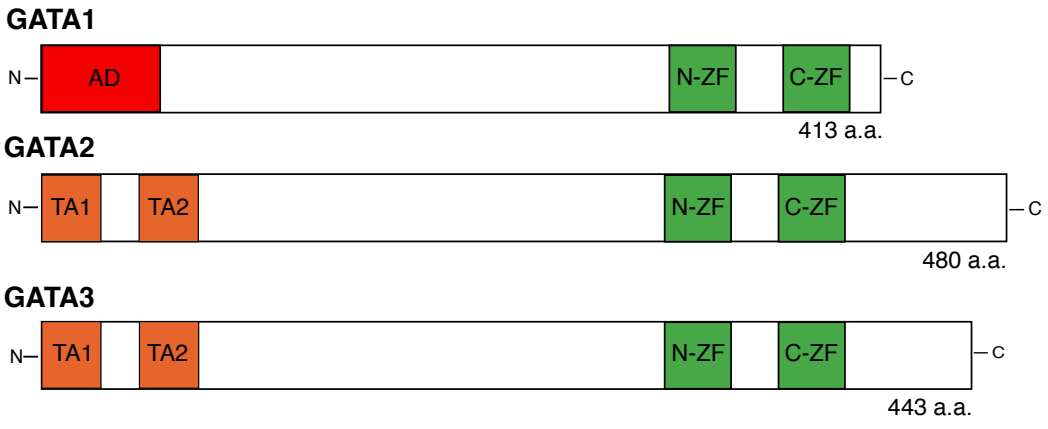


**Fig. 8**: **GATA family proteins**. All members contain two highly conserved zinc finger domains (N-ZF & C-ZF, green), however, only GATA1 contains an N terminal activation domain (AD, red). The N-ZF domain of GATA1 binds to the cofactor FAG1 and stabilizes the binding to specific DNA sequences. GATA2 and GATA3 contain two transactivation domains (TA1 & TA2, orange) which also facilitate the binding with transcription coregulators.

Two members of the GATA family, GATA1 and GATA2, control the expression programs of cells at different stages of erythropoiesis. GATA1 expression is indispensable for establishing the erythroid program, and is the master regulator of lineage commitment, differentiation, and survival of erythroid progenitors[141]. During erythroid differentiation, the highest GATA1 expression is detected at the proerythroblast stage. Thereafter, the expression of GATA1 decreases, via caspase-mediated cleavage, en route to maturation into erythrocytes[142] (**Fig. 2B**). In addition to erythroid precursors, GATA1 expression was also detected in megakaryocytes, eosinophils and mast cells[143]. Observations from multiple studies using different loss-of-function models have shown that dysregulation of GATA1 results in failed erythropoiesis[144-146]. *Gata1* deficient mice die *in utero* (E11) due to the failure of erythroid progenitors to mature beyond the proerythroblast stage[146]. Extensive cellular apoptosis is detected during definitive hematopoiesis, partially due to reduced expression of anti-apoptotic gene *Bcl-xL*[147], which was rescued by *Gata1* re-introduction[147]. Conditional knockout of *Gata1* during adult and stress erythropoiesis in mice, resulted in arrested maturation of cells at the proerythroblast stage, thrombocytopenia, and excessive proliferation of megakaryocytes in the spleen[148]. Despite high EPO levels, loss of *Gata1* in mice led also to reduction in the BM erythroid compartment, resembling the clinical condition aplastic crisis in humans (an acute form of pure red aplasia)[148]. Further evidence for the critical role for GATA1 in regulating erythroid gene expression came through the detection of *GATA1* mutations (germline and somatic) in several human blood disorders such as Diamond-Blackfan anemia (DBA), transient myeloproliferative disorder (TMD), and Down Syndrome associated acute megakaryoblastic leukemia (AMKL)[149].

In contrast to GATA1, GATA2 is expressed at high levels in HSCs, CMPs, and MEPs, and its expression gradually decreases on the onset of erythroid commitment (proerythroblast stage) (**Fig. 2B**)[138]. GATA2 expression is required for proliferation and survival of early hematopoietic cells, as primitive and definitive hematopoiesis is abolished when the *Gata2* gene is deleted. Homozygous deletion of *Gata2* results in E10-11 embryonic lethality from FL anemia, and loss of definitive HSCs[150]. Selective conditional knockout of *Gata2* in endothelial and hematopoietic resulted in deficient HSCs generation in the AGM, in addition

to a decline in their fitness and survival[151]. In addition, GATA2 appropriate function is also depended on the maintenance of its concentration within a critical physiological window. Overexpression of GATA2 blocked erythroid differentiation[152], and heterozygous *Gata2* knock out mice (*Gata2*$^{+/}$) had defective expansion and function of HSCs[153]. Consistent with the tight control of *Gata2* transcription, GATA2 protein has a short half-life of < 1h, which provides opportunities to rapidly change its cellular concentration[154]. Germline and sporadic heterozygous mutations in *GATA2* cause a complex immunodeficiency syndrome (Mono MAC) that often progresses to MDS and AML[155].

ES cells lacking the *Gata3* gene were able to contribute to the erythroid lineages, but not to T-cell lineages[156]. Thus, In the hematopoietic system, *Gata3* expression is essential for innate and adaptive lymphoid development, where it regulates the differentiation, maintenance, and survival of early and mature T-and natural killer cells[157]. More specifically, within the T-cell lineage, *Gata3* regulated the differentiation of T helper type 2 cells (Th2), by controlling genes that encode Th2 cytokines *Il4*, *Il5* and *Il13*[158, 159].

### 2.4.1. GATA1 cofactors interactions

GATA1 controls hematopoietic development by activating and repressing gene expression. During erythroid differentiation, GATA1 interacts with ubiquitous and lineage-restricted TFs, chromatin modifying/remodeling enzymes, and other coregulators to activates erythroid-specific genes and repress genes promoting proliferation. Transcriptional activation (e.g. *hemaglobin genes*) and repression (e.g. *Gata2, Kit & c-Myc*) require GATA1-selective interactions with diverse coregulators including FOG1, SCL/TAL1 complex, CREB-Binding Protein (CBP)/EP300, the nucleosome remodeling and deacetylase complex (NuRD), and the histone methyltransferase SetD8.

### 2.4.2. FOG1

GATA1 transcriptional activity acts largely via its interaction with the cofactor FOG1[160]. FOG1 protein contains nine zinc fingers, four of which are capable of mediating binding to the GATA1 N-ZF domain. During embryogenesis, FOG1 and GATA1 are co-expressed in YS and FL erythroid precursors, and targeted homozygous loss of FOG1 expression (*Zfpm1*$^{-/-}$) in mice resulted in maturation blockades in primitive and definitive proerythroblasts development[161]. FOG1 does not bind DNA directly, however it is recruited to the chromatin

by GATA1, and in return, it facilitates GATA1 chromatin occupancy at certain sites (**Fig. 9A**). During erythropoiesis, GATA1 modulates *Gata2* expression by binding upstream to *Gata2* promoter region and suppress its expression in a phenomenon termed ''the GATA switch'' (**Fig. 9B-C**). In cells lacking FOG1, GATA1 chromatin occupancy is reduced at *Gata2* locus sites, and is unable to repress its expression[162]. By contrast, repression of *Lyl1* gene was unaffected by loss of FOG1 in *Zfpm1*$^{-/-}$ cells[163] . Thus, although mostly FOG1-dependent, at specific subset of genomic loci, GATA1-mediated activation and repression can also occur in a FOG1 independent manner.

### 2.4.3. The SCL/TAL1 complex

The interaction of the basic helix -loop-helix TF SCL/TAL1 with GATA1 at specific GATA composite E-box elements, recruits bHLH protein E2A, and the non-DNA binding proteins, LIM-only protein 1 (LMO2), and LIM-domain-binding protein 1 (LDB1) as part of a multimeric protein complex. LMO2 mediates the complex assembly via binding to GATA1 at the N-ZF, and acting as a docking site for the other SCL/TAL1 proteins[164](**Fig. 9D**). LDB1 binds the LMO2 LIM domain and also contributes to the assembly of the SCL/TAL1–GATA complex. TAL1, LMO2, and LDB1 are essential for erythrocyte and megakaryocyte differentiation. Homozygous loss of *Ldb1* in mice results in severe defect in the expansion of the YS and absence of blood islands, leading to anemia and embryonic lethality at E9.5[165]. similar to *Ldb1* loss, deficient primitive and definitive hematopoiesis was also observed for *Lmo2* null mice[166]. The complex acts to modulates GATA1 activity via promoting the formation of higher-order chromatin loops at genomic locations (e.g. *Hb-*$\beta$), and these long-range interactions are believed to promote transcriptional activation[167].

### 2.4.4. The nucleosome remodeling and deacetylase (NuRD) complex

Specific histone modifications are often differentially enriched at genomic sites with distinct functions (enhancers vs promoters), and during different stages of embryonic development and lineage specification. The balance between the different histone PTM create a tight circle of gene expression regulation ensuring appropriate activation and repression of genes. The ubiquitously expressed nucleosome remodeling and deacetylase (NuRD) complex is unique among chromatin remodeling complexes in having two enzymatic functions. The complex contains an ATPase dependent nucleosome remodeling protein, the

chromodomain-helicase-DNA-binding proteins CHD3 or CHD2 (a.k.a. Mi2β or Mi2α respectively), and the histone deacetylase (HDAC1 and HDAC2) subunits[168]. FOG1 binds to the NuRD complex through a conserved N-terminal motif (**Fig. 9C**). Although previously thought to be exclusively required for GATA1-FOG1 mediated repression[169], FOG1-NuRD interaction has since been shown to also be essential for GATA1-FOG1 transcriptional activation during erythroid development[170]. Homozygous mice with triple point mutations in FOG1 that abrogates NuRD binding were defective in erythroid and megakaryocytic lineages. The mutated mice phenocopied germline mutations in the *Gata1* and *Fog1* genes, including anemia and macrothrombocytopenia. Gene expression analysis in erythroid and megakaryocytic cells indicted that both the activation and repression of GATA1-FOG1 dependent genes are negatively affected[170, 171]. -

### 2.4.5. SetD8

Setd8 (also known as KMT5A & PRSet7) is a histone methyltransferase identified in mammals that catalyzes histone H4 mono-methylation at lysine 20 (H4K20me1), and which expression is significantly higher in erythroblasts compared to most other cell types. Homozygous knockout of *Setd8* in mice results in pre-implantation lethality[172]. Targeted loss of *Setd8* expression in the hematopoietic system, resulted in defective primitive erythropoiesis, anemia starting at E9.5, and death from anemia by E12.5[173]. In addition, Stable *Setd8* knockdown in extensively self-renewing erythroblasts (ESREs), resulted in impaired erythroid maturation characterized by incomplete nuclear condensation, and lower rates of enucleation[174]. Although physical interaction between GATA1-SetD8 is yet to be proven, GATA1 and SetD8 can occupy overlapping chromatin regions, such as the critical regulatory elements in the *Gata2* locus[174]. GATA1 expression induces H4K20me1 extension upstream of the promoter and along gene body. SetD8 represses *Gata2* expression by catalyzing H4K20me1 at an intronic *cis*-element upstream of the *Gata2* 1S promoter (+ 9.5 site), and restricts Scl/TAL1 occupancy at this site[175]. SetD8 downregulation resulted in loss of H4K20me1 and gain of H4 acetylation at the *Gata2* promoter leading to an increase in GATA2 level in erythroid precursors and block in erythroid maturation. Lowering GATA2 expression in SetD8 knockdown cells did not rescue erythroid maturation, indicating the involvement of different genes/process in SetD8 regulation of erythroid maturation, such as regulation of cell cycle progression, and chromatin condensation[173, 175].

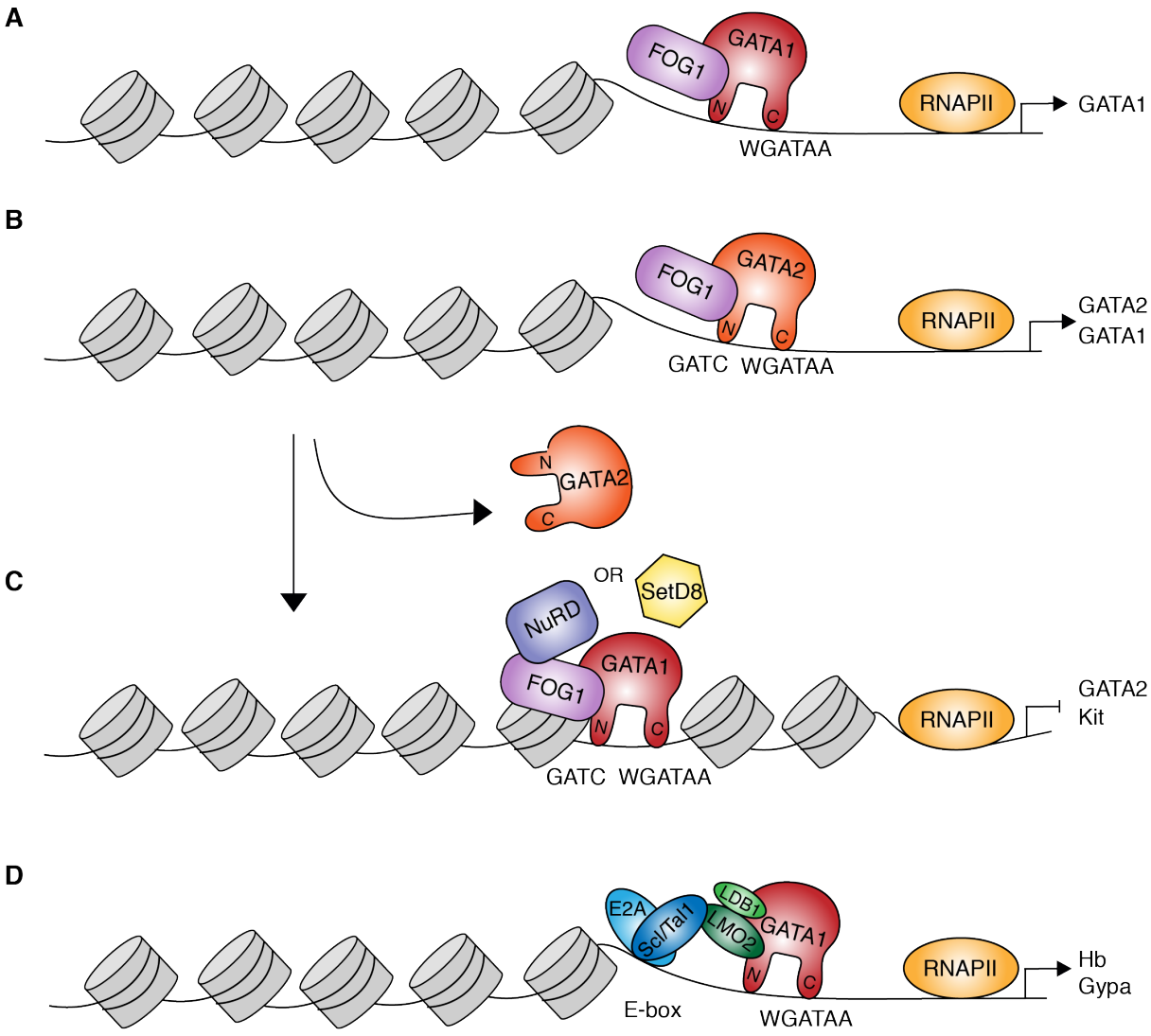



**Fig. 9**: **Model of GATA multiprotein complexes orchestrating expression of erythroid regulators**. **(A-B)** GATA1 and GATA2 factors bind to their corresponding locus, and positively regulate their own expression. However, only GATA2 also induces GATA1 expression. The factors bind DNA (WGATTA motif) via the C terminal zinc finger (C), while binding of FOG1 to the N terminal zinc finger (N), stabilizes GATA binding to DNA. **(C)** The "GATA-switch". On the onset of erythroid differentiation, and upon elevation of GATA1 levels, GATA1 displaces GATA2 at the *Gata2* locus promoter. The switch is facilitated by FOG1, and is ensured by a lower GATA2 protein stability compared to GATA1[154]. Other coregulators such as the NuRD complex (Histone deacetylation) and SetD8 (H4K20me1) are also suggested to play a role in the GATA2 displacement/repression[170, 174]. **(D)** Occupancy data of the of GATA-1, SCL/TAL1, LMO2 and LDB1 complex maps with gene expression profiling data[176]. This suggests that the complex is largely involved in activating the expression of erythroid associated genes (hemoglobin, *Hb*, and Glycophorin, *Gypa*). RNAPII (RNA polymerase II) (Adapted from [177].

Despite extensive research into the GATA family function and regulation during development and hematopoiesis, a comprehensive overview of how GATA factors interact with specific coregulators in a cell-type and developmental-stage-specific manner remains

elusive. In addition to instructive extracellular signaling, commencement of erythroid differentiation requires GATA factor and coregulator appropriate expression and loci occupancy, achieved through the modulations of expression/inhibition of overlapping TFs, and suitable chromatin marks.

## 2.5. GATA1 dysregulation in red blood cell disorders

The X-linked *GATA1* gene encodes a zinc finger TF expressed in multiple cell types of the hematopoietic system, such as in erythroid, megakaryocytic, eosinophilic, dendritic, and mast cells[138, 148, 178, 179]. During embryogenesis, *Gata1*$^{-/-}$ proerythroblasts fail to differentiate further into mature erythrocytes, leading to embryonic mortality at E11 due to severe anemia [146]. Female mice with reduced *Gata1* expression (*GATA1*$^{1.05/X}$) frequently develop a hematopoietic disorder resembling myelodysplastic syndrome with a splenic accumulation of proerythroblasts expressing low levels of GATA1[180]. Ultimately, the mice develop two distinct type of acute leukemias, a late-onset B-cell lymphoproliferative disease or an earlier erythroleukemia-like disease. Interestingly, mice with complete loss of GATA1 expression (*GATA1*$^{null/X}$) never developed leukemia, which suggests that low GATA1 expression, sufficient to support survival and proliferation but not differentiation, leads to the accumulation of progenitors that could permissible for an oncogenic stimulus. In humans, point mutations at a splice site of the *GATA1* gene, abrogating the production of the full-length form of the protein (GATA1-L), were recently implicated in pathogenesis of DBA [181]. In addition, amino acid substitution mutations in N-ZF of GATA1 (V205M[182], G208S[183] and D218Y[184]), which disrupt the GATA1-FOG1 binding affinity, and others (R216Q[185] and R216W[186]) impairing GATA1 binding to DNA and SCL/TAL1 complex recruitment[187], were detected in a spectrum of blood cell disorder, such as severe or moderate thrombocytopenia, with or without pronounced anemia, and erythropoietic porphyria [149].

# 3. Chapter I: NSD1 role in erythroid differentiation

## 3.1. Background: Role of *Nsd1* in normal hematopoiesis

Being part of an AML-associated fusion gene raises the question about the role NSD1 in normal hematopoiesis. To address that, our lab performed shRNA-based knockdown experiments and found that reduction of NSD1 expression in human $CD34^+$ hematopoietic cells altered their clonogenic growth and resulted in accumulation of erythroid progenitor cells. Based on a previously reported conditional allele, our lab also established conditional *Nsd1* knockout mice. We observed that hematopoietic (using a *Vav1-iCre* ablator line) inactivation of the *Nsd1* gene in mice led to a fully penetrant leukemia-like disease between 6-25 weeks of age, presenting with anemia, reticulocytosis, thrombocytopenia, splenomegaly and multi-organ infiltrations with blasts on peripheral blood smears. BM and spleen cells from diseased mice expressed modest levels of the transferrin receptor ($CD71^{low/+}$) and variable amounts of Kit. In addition, all diseased $Nsd1^{-/-}$ mice significantly accumulated $CD71^{low}$/$Ter119^-$ cells in BM and spleen, indicating a deterioration in erythroid maturation. The cells formed abnormal BFU-like colonies in erythropoietin (EPO)-containing methylcellulose. Transplantation of BM cells from diseased mice propagated the disease in wild type recipients, alone or in competition with wild type cells. Despite constitutive expression of the erythroid master regulator GATA1 at the protein level, $Nsd1^{-/-}$ FL and BM-derived erythroblast cells failed to achieve terminal erythroid differentiation in an *in vitro* culture assay containing EPO[188], signified by a reduction in hemoglobin production and failure to unregulated erythroid associated genes such *Gypa*. Surprisingly, retroviral overexpression of *Gata1* was able to partially overcome the terminal differentiation block *in vitro*.

## 3.2. Aim of the project

Erythroleukemia can either develop as an acute and rapidly progressive malignancy or as a chronic form of leukemia. Patients with PEL have poor prognosis and despite new treatments options, erythroid neoplasms are still regarded as incurable malignancies due to primary induction failure, relapse, and the toxicity of chemotherapeutic agents. More research is needed to decipher molecular players and targets in these highly fatal neoplasms, in order to develop better, more targeted therapies and improve patients' prognostic outcome. Our recent results indicate a role for NSD1 in erythropoiesis regulation in a conditional knockout model (*Nsd1*$^{fl/fl}$*.iVav-Cre*). In particular, NSD1 seems to be essential for driving the differentiation of immature proerythroblast into mature erythrocytes. Erythroid precursor cells lacking the NSD1 protein accumulate as immature proerythroblasts leading to a leukemia-like disease in mice *in vivo*, and are resistant to terminal differentiation *in vitro* in an erythrocytes culture model. Thus, it would be of great interest to understand the mechanisms by which NSD1 regulates cellular fate and differentiation, specifically during erythropoiesis. NSD1 can interfere with chromatin organization and gene transcription through its histone methyltransferase activity and/or by forming complexes with key factors of erythropoiesis[93, 103].

Mechanistic studies of NSD1 proteins are hindered by several obstacles such as i) the large size of the protein/cDNA, which would significantly reduce gene transfer efficacy, ii) the absence of ChIP-grade antibodies, and iii) the inability of establishing NSD1 overexpressing hematopoietic cells. Murine Erythroleukemia (MEL) cells are an established erythroid progenitor cell lines model that are derived from the spleen of susceptible mice infected with *Friend virus[189]*. These cells are arrested at the proerythroblast stage of development and can be chemically induced (DMSO or HMBA) to differentiate into more mature, hemoglobinized precursors[190]. Due to their high proliferation rate and undemanding culture conditions, MEL cells have been extensively used as model for erythrocytes differentiation *in vitro.* Thus, we decided to use MEL cells as an *in vitro* model to uncover NSD1 regulatory function during erythroid differentiation utilizing chromatin immunoprecipitation-sequencing (ChIP-seq) analysis.

# 3.3. Results

**$Nsd1^{fl/fl}$ but not $Nsd1^{-/-}$ hematopoietic cells cultured in erythroid conditions express both myeloid and erythroid surface markers**

We applied an optimized two-step culture system in which erythroblasts are first enriched and expanded in "maintenance medium" (MM, containing dexamethasone, mSCF and hEPO) followed by induced terminal erythroid maturation in "differentiation medium" (DM, containing mSCF, hEPO) (**Fig.10A**). Using this system, we found that that terminal differentiation of *Nsd1*$^{-/-}$ bone marrow (BM)-derived erythroblasts was significantly impaired compared to wild-type *Nsd1*$^{fl/fl}$ control cells (**Fig.10B & 10C**). Unfortunately, a simple direct comparison of *Nsd1*$^{-/-}$ versus *Nsd1*$^{fl/fl}$ cells is hampered by the observation that while primary *Nsd1*$^{-/-}$ cultures only consisted of CD71$^{+}$ and Kit$^{+}$ erythroblasts, WT *Nsd1*$^{fl/fl}$ erythroblasts cultures always contained cells expressing both myeloid and erythroid markers (CD71, Kit, Mac-1 and Gr-1) in all combinations (**Fig. 10D**). To overcome the problem of unequal *Nsd1*$^{-/-}$ and *Nsd1*$^{fl/fl}$ cell component cultures, we designed a flow cytometry sorting strategy whereby Mac-1$^{+}$ Gr-1$^{+}$ cells were excluded to select for WT type cells expressing only CD71 and Kit (**Fig. 11A**). Upon expansion in maintenance cultures, we observed a comparable markers expression and cellular morphology between *Nsd1*$^{-/-}$ BM-derived erythroblasts with the sorted but not the unsorted *Nsd1*$^{fl/fl}$ cells (**Fig. 11B**). Gene expression analysis showed an upregulation in the erythroid associated genes *Gata1* and *Hb-B* suggesting a ''purer'' erythroid cell culture in the sorted compared to unsorted cells (**Fig. 11C**). However, unfortunately we could not use these cells for further experiments as it was not possible to expand them to sufficient numbers in maintenance culture (**Fig. 11D**) suggesting that the limited expansion capacity of the unsorted *Nsd1*$^{-/-}$ cells is probably driven by these ''multi-lineaged'' cells.

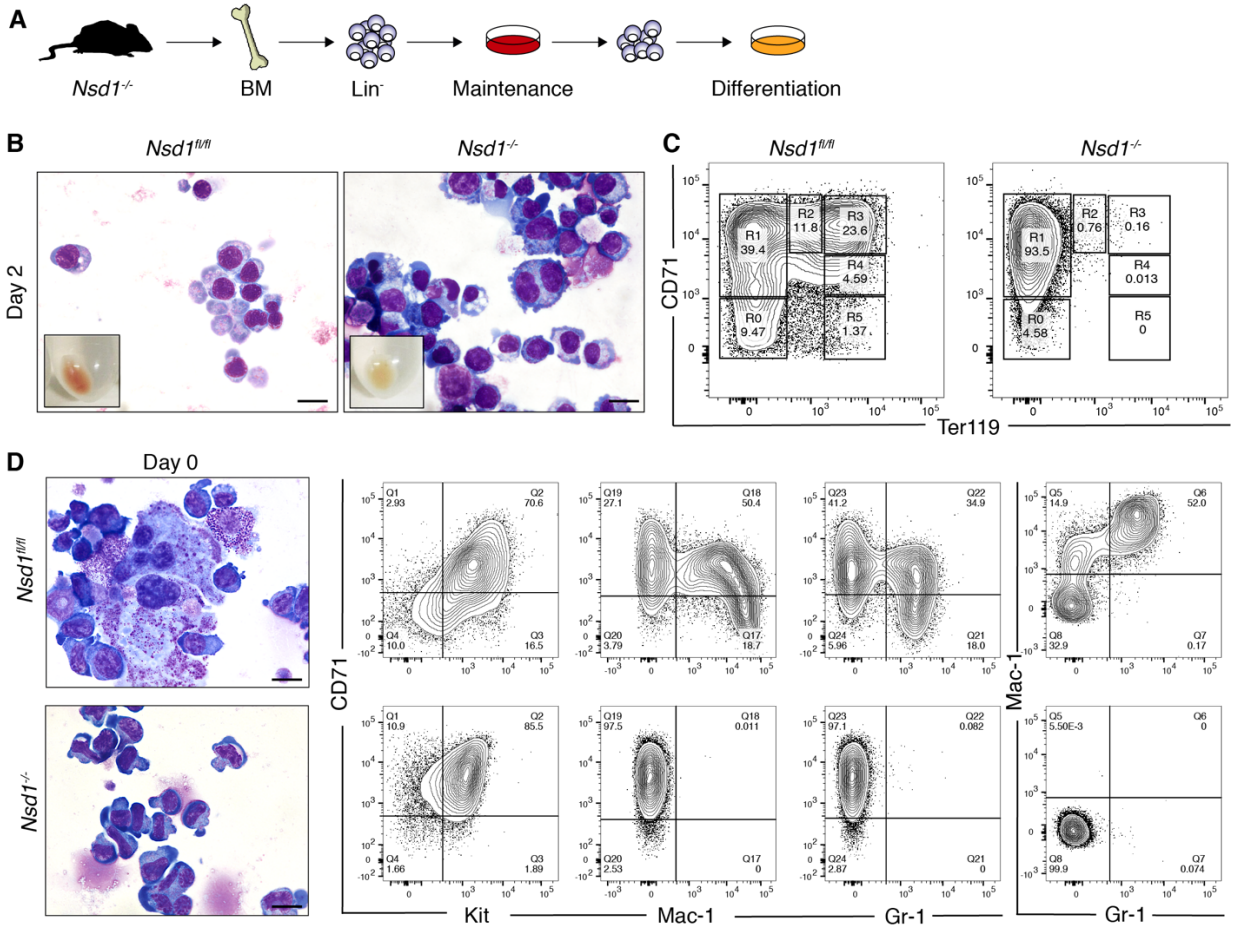



**Fig. 10: *Nsd1*$^{fl/fl}$ cells expressed both myeloid and erythroid surface markers**

**(A)** Schematic illustration of experimental erythroid differentiation assay: lineage marker-depleted (lin$^-$) adult BM cells of *Nsd1*$^{fl/fl}$ and *Nsd1*$^{-/-}$ mice were grown in MM (>1 week, StemSpan$^{TM}$ plus mSCF, hEPO and dexamethasone,) before induction of maturation in DM (4 days, IMEM plus, FCS. hEPO, &mSCF). **(B)** Representative images (1000X) of Wright Giemsa-stained cytospin preparations of lin$^-$ *Nsd1*$^{fl/fl}$ and *Nsd1*$^{-/-}$ BM-derived erythroblasts following 2 days in DM. Small inserts show cell pellets. (1000X, the size bar=10μM). **(C)** Flow cytometric analysis of the transferrin receptor (CD71) and Ter119 marker expression of *Nsd1*$^{fl/fl}$ and *Nsd1*$^{-/-}$ BM-derived erythroblasts after 4 days in DM. The gating strategy distinguishes different stages of erythroid maturation: "R0" fraction (CD71$^{-/low}$, Ter119$^-$), "R1" (CD71$^+$, Ter119$^-$), "R2" fraction (CD71$^+$, Ter119$^{+/low}$), "R3" fraction (CD71$^+$, Ter119$^+$), "R4" fraction (CD71$^{+/low}$, Ter119$^+$), and "R5" fraction (CD71$^-$, Ter119$^+$). **(D)** Representative images (1000X) of Wright Giemsa-stained cytospin preparations of *Nsd1*$^{fl/fl}$ and *Nsd1*$^{-/-}$ BM-derived erythroblasts in MM (day 0- Left panel), and flow cytometric analysis of CD71, Kit, Mac-1 and Gr-1 markers expression of *Nsd1*$^{fl/fl}$ and *Nsd1*$^{-/-}$ BM-derived erythroblasts in MM (day 0- Right panel). This data represents 1 of 4 independent experiments.

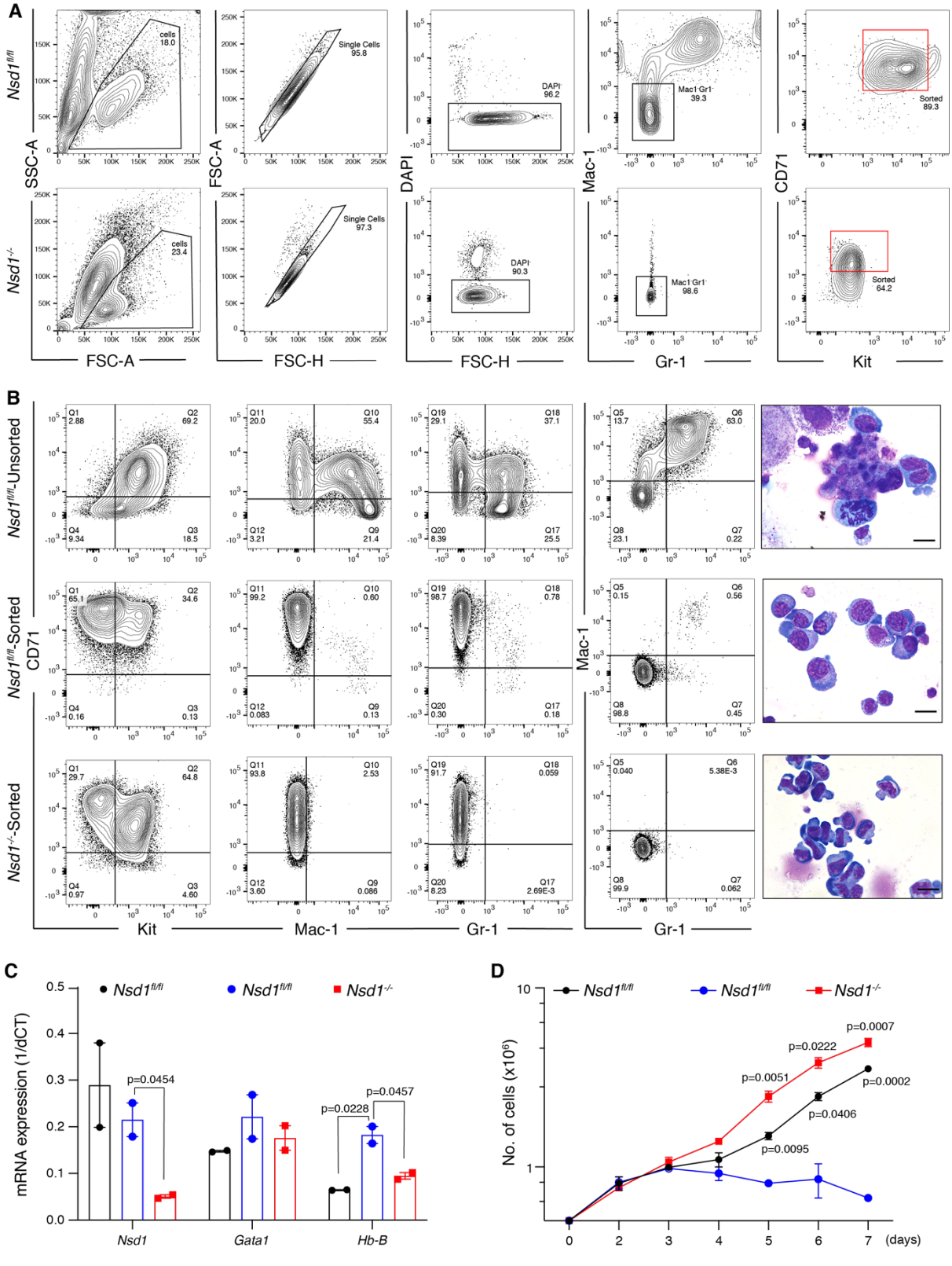



**Fig. 11: CD71⁺ Kit⁺ sorted *Nsd1*$^{fl/fl}$ BM-derived erythroblasts failed to expand in maintenance culture**

**(A)** Flow cytometry example of the gating strategy used to sort CD71⁺ Kit⁺ double positive *Nsd1*$^{fl/fl}$ and *Nsd1*$^{-/-}$ BM-derived erythroblasts. **(B)** Flow cytometric analysis of CD71, Kit, Mac-1 and Gr-1 markers expression (left panel) of unsorted *Nsd1*$^{fl/fl}$ (top) sorted *Nsd1*$^{fl/fl}$ (middle) and sorted *Nsd1*$^{-/-}$ (bottom) BM-derived erythroblasts in MM (day 0). Representative images (1000X) of Wright Giemsa-stained cytospin preparations (right panel) of unsorted *Nsd1*$^{fl/fl}$ (top) sorted *Nsd1*$^{fl/fl}$ (middle) and sorted *Nsd1*$^{-/-}$ (bottom) BM-derived erythroblasts in MM (day 0). This data represents 1 of 3 independent experiments.

**(C)** *Nsd1*, *Gata1* and *Hb-B* relative mRNA expression of unsorted *Nsd1*$^{fl/fl}$ (black bars), sorted *Nsd1*$^{fl/fl}$ (blue bars) and sorted *Nsd1*$^{-/-}$ (red bars) BM-derived erythroblasts in MM (n= 2 per group). Values are shown as relative expression normalized to *Gapdh.* **(D)** Growth of unsorted *Nsd1*$^{fl/fl}$ (black line), sorted *Nsd1*$^{fl/fl}$ (blue line) and sorted *Nsd1*$^{-/-}$ (red line) BM-derived erythroblasts in MM (1-7 days). Nucleated living cells were counted by the Trypan blue exclusion (n= 3 per group).

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 D tested with a paired two-tailed t- test.

**Retroviral expression of *Nsd1* reduced *Nsd1*$^{-/-}$ cells clonogenic potential**

To investigate the role of *Nsd1* in controlling erythroid differentiation, we decided to compare induced terminal differentiation upon viral transduction of *Nsd1*$^{-/-}$ BM-derived erythroblasts with either a WT *Nsd1* transcript or a previously described catalytically-inactive *Nsd1*$^{N1918Q}$ SET-domain mutant[93] (**Fig. 12A**). Unfortunately, establishing stably *Nsd1* expressing cells proved to be very challenging, due to a very low rate of transduction (**Fig. 12B**), and phenotype conversion after prolong expansion *in vitro*. Nevertheless, and despite low transduction rate for the *Nsd1* and *Nsd1*$^{N1918Q}$ compared to empty (*Ctrl*) transduced cells, they proliferated and expanded in MM and we could not detect significant difference in their growth rate **(Fig. 12C)**. *Nsd1* expression reduced clonogenic capacity of the *Nsd1*$^{-/-}$ erythroblasts denoted by the formation of few small reddish colonies in EPO-containing MC and significant reduction in colony and cell numbers (**Fig.12D-F**). *Nsd1* expressing cells harvested from MC showed a slight reduction in *kit* and increase in *Gata1 and Hb-B* mRNA expression compared to empty (*Ctrl*) and *Nsd1*$^{N1918Q}$ expressing cells (**Fig. 12G & H**).

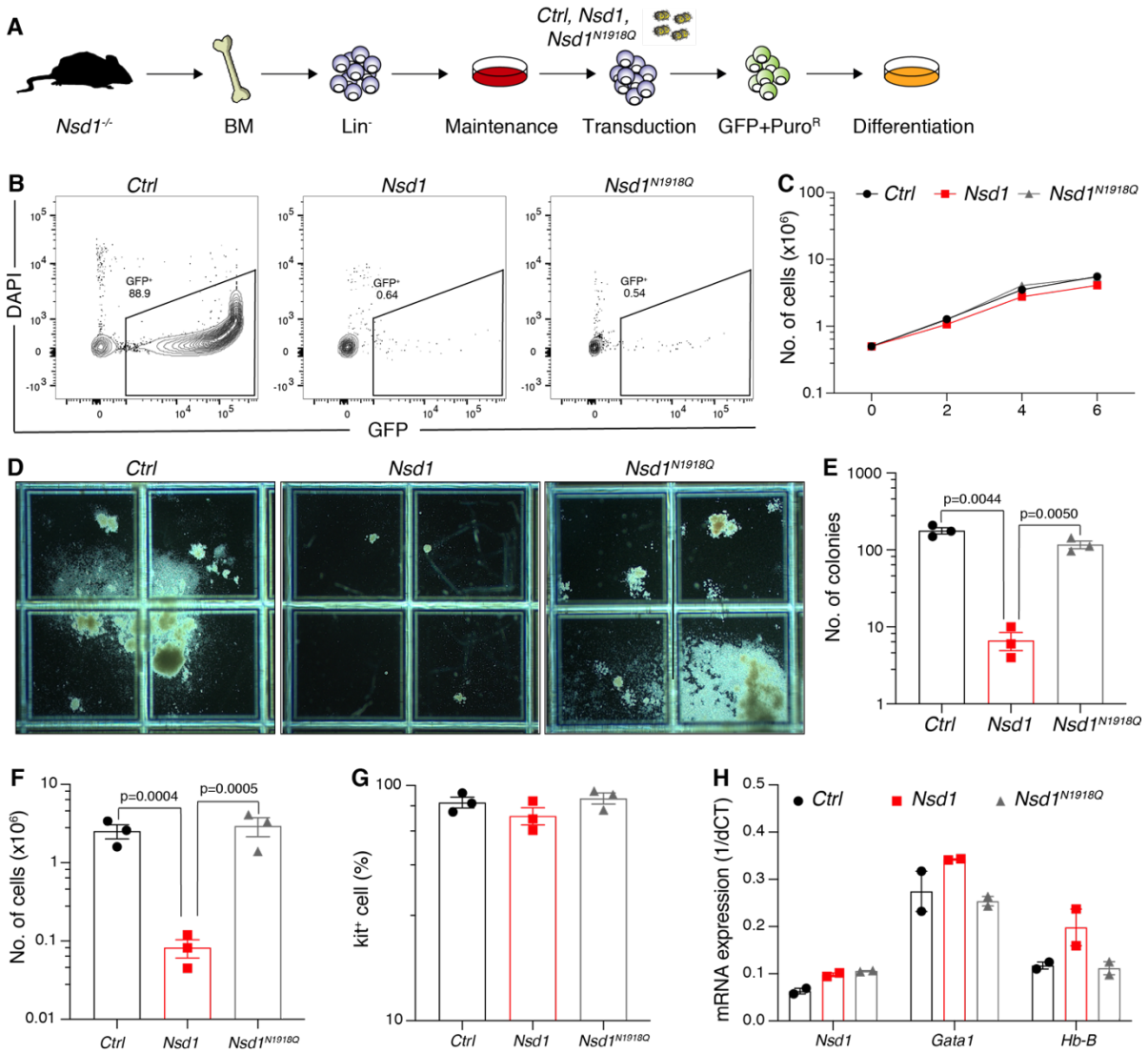



**Fig. 12: Retroviral *Nsd1* expression increased colony formation of *Nsd1*$^{-/-}$ BM-derived erythroblasts**

**(A)** Schematic illustration of the experimental set up: *Nsd1*$^{-/-}$ BM-derived erythroblasts were transduced with either *pMSCV-GFP-Puro (Ctrl), pMSCV-Nsd1-GFP-Puro (Nsd1)* or *pMSCV-Nsd1*$^{N1918Q}$*-GFP-Puro (Nsd1*$^{N1918Q}$*)* in MM, GFP$^{+}$ cells were selected with 1µg Puromycin before expansions and differentiation analysis. **(B)** Flow cytometry example of the gating strategy used to sort GFP$^{+}$ expressing cells in MM. **(C)** Growth of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control *(Ctrl,* black line), *Nsd1* (red line) and *Nsd1*$^{N1918Q}$ (grey line) in MM (1-6 days). Nucleated living cells were counted by the Trypan blue exclusion (n= 3 per group). **(D)** Representative images of colonies formed in MC (H3434) by 1x10$^{4}$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control *(Ctrl,* left*)*, *Nsd1* (middle) or *Nsd1*$^{N1918Q}$ (right) after 11 days. This data represents 1 of 3 independent experiments. **(E)** Number of colonies formed in MC (M3434) by 1x10$^{4}$ *Nsd1*$^{-/-}$ erythroblasts expressing control *(Ctrl,* black bar), *Nsd1* (red bar) and *Nsd1*$^{N1918Q}$ (grey bar) after 11 days (n=3 per group). **(F)** Total number of cells obtained from 1x10$^{4}$ in MC (M3434) by *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control *(Ctrl,* black bar), *Nsd1* (red bar) and *Nsd1*$^{N1918Q}$ (grey bar) after 11 days. Nucleated living cells were counted by the Trypan blue exclusion (n= 3 per group). **(G)** Quantification of the percentage of Kit$^{+}$ cells obtained from 1x10$^{4}$ in MC (M3434) by *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control *(Ctrl,* black bar), *Nsd1* (red bar) and *Nsd1*$^{N1918Q}$ (grey bar) after 11 days (n=3 per group). **(H)** *Nsd1*, *Gata1* and *Hb-B* relative mRNA expression in MC (M3434) by *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control *(Ctrl,* black bar), *Nsd1* (red bar) and *Nsd1*$^{N1918Q}$ (grey bar) after 11 days (n= 2 per group). Values are shown as relative expression normalized to *Gapdh.*

**Integrity of the *Nsd1* SET-domain was essential for *in vitro* erythroblast terminal maturation**

To study the effect of *Nsd1* expression on the cells' erythroid differentiation potential, we switched empty (*Ctrl*), *Nsd1* and *Nsd1*$^{N1918Q}$ transduced cells to liquid DM and assessed their phenotype. We detected significant increase in terminal erythroid maturation in *Nsd1* but not *Ctrl* or *Nsd1*$^{N1918Q}$ cells, illustrated by drastic change in cellular morphology and size, formation of hemoglobinized cell pellet, and reduced proliferation (**Fig.13A-C**). The morphological erythroid maturation of *Nsd1* expressing cells was coupled with a significant shift in Ter119 and CD71 surface expression, along a significant reduction in Kit expression (**Fig. 13D-G**).

Erythroid differentiation is associated with gradual and consistent reduction in the maturing cell's size leading to enucleation and cell cycle arrest[191]. In comparison to *Nsd1*$^{N1918Q}$ expressing cells, *Nsd1* expression significantly reduced the size of *Nsd1*$^{-/-}$ cells during liquid media differentiation as illustrated by the increase of the FSC$^{-}$ population (**Fig. 14A-B**). *Nsd1* expression also resulted in significant reduction of cells in the S- and $G_2$M- phase of the cell cycle by day 2 and 4 in DM (**Fig. 14C-D**).

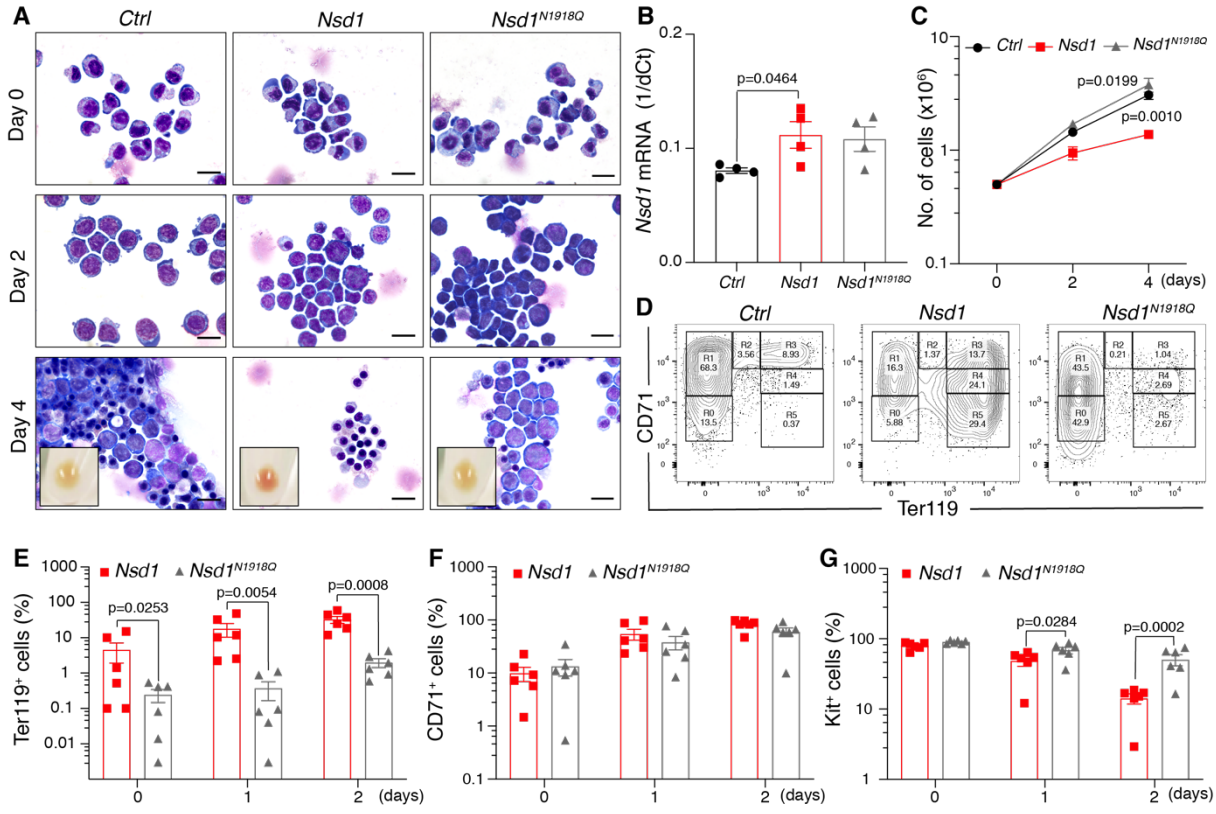



**Fig. 13: The integrity of the *Nsd1* SET domain was essential for *in vitro* erythroblast terminal maturation**
**(A)** Representative pictures (1000X, size bars=10μM) of Wright-Giemsa-stained cytospin preparations of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control *(Ctrl*, left), *Nsd1* (middle) and *Nsd1*$^{N1918Q}$ (right) in MM (day 0, top panel) and after 2 days (day 2, middle panel) and 4 days (Day 4, bottom panel) in DM. Small inserts show cell pellets. **(B)** *Nsd1* mRNA relative expression levels (1/dCt) in *Nsd1*$^{-/-}$ BM-derived MM (n=4

per group). Values are shown as relative expression normalized to *Gapdh*. **(C)** Growth of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control *(Ctrl,* black line), *Nsd1* (red line) and *Nsd1*$^{N1918Q}$ (grey line) in DM (0-4 days). Nucleated living cells were counted by the Trypan blue exclusion (n= 3 per group). **(D)** Flow cytometry analysis of CD71 and Ter119 markers' expression of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control *(Ctrl,* left panel), *Nsd1* (middle panel) and *Nsd1*$^{N1918Q}$ (right panel) after 4 days in DM. The gating strategy distinguishes different stages of erythroid maturation: "R0" fraction (CD71$^{-/low}$, Ter119$^{-}$), "R1" (CD71$^{+}$, Ter119$^{-}$), "R2" fraction (CD71$^{+}$, Ter119$^{+/low}$), "R3" fraction (CD71$^{+}$, Ter119$^{+}$), "R4" fraction (CD71$^{+/low}$, Ter119$^{+}$), and "R5" fraction (CD71$^{-}$, Ter119$^{+}$). This data represents 1 of 6 independent experiments. **(E-G)** Quantification of the percentage of Ter119$^{+}$ **(E)**, CD71$^{+}$ **(F),** and Kit$^{+}$ **(G),** of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* (red bar) and *Nsd1*$^{N1918Q}$ (grey bar) in MM (day 0) and after 1 and 2 days in DM (n=6).

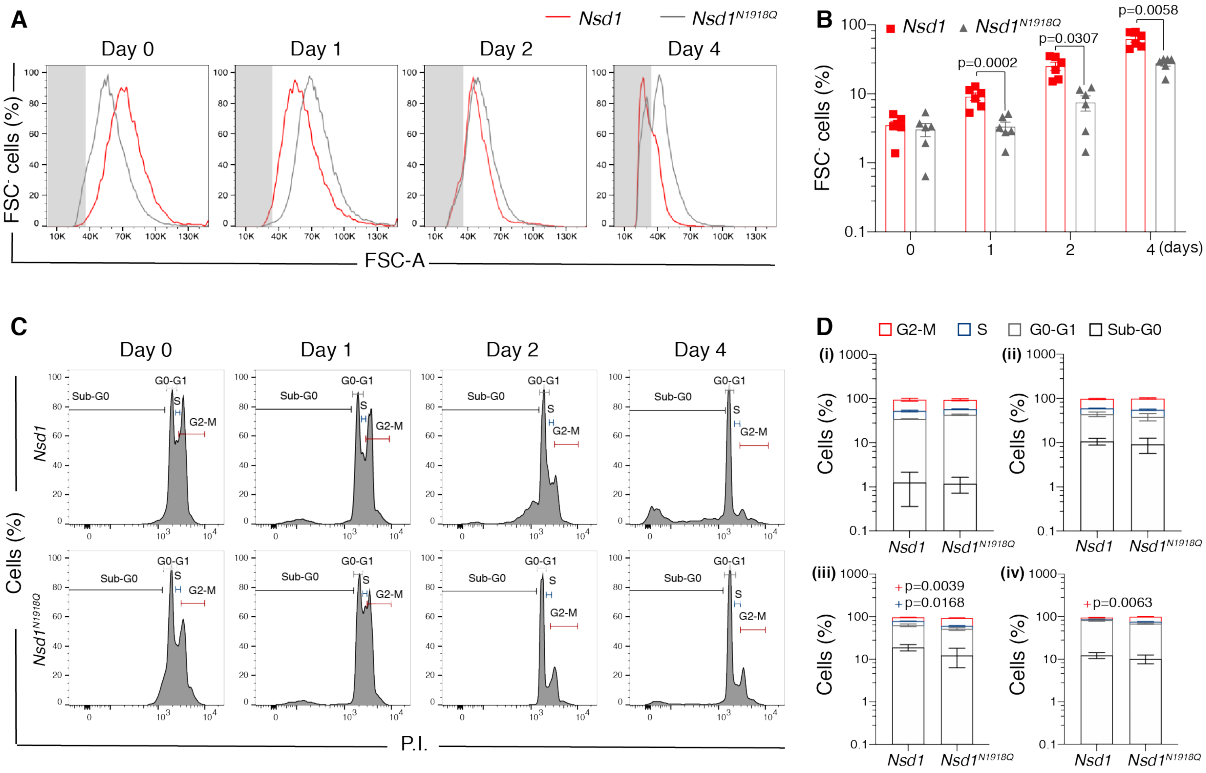



**Fig. 14: *Nsd1*- induced *in vitro* maturation of *Nsd1*$^{-/-}$ erythroblasts is coupled with significant reduction in cell size and a cell cycle arrest**

**(A)** Flow cytometry example of the gating strategy (grey shaded area) used to quantify Forward scatter-negative (FSC$^{-}$) of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* (red histogram) and *Nsd1*$^{N1918Q}$ (grey histogram) in MM (day 0) and after 1, 2 and 4 days in DM. **(B)** Quantification of the percentage of FSC$^{-}$ of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* (red bar) and *Nsd1*$^{N1918Q}$ (grey bar) in MM (day 0) and after 1, 2 and 4 days in DM (n=5). **(C)** Flow cytometry example of the gating strategy (Propidium Iodide, P.I.) used to quantify different cell cycle populations (Sub-G0, Black) (G0-G1, grey) (S-phase, blue) and (G2-M, red) of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* (Top panel) and *Nsd1*$^{N1918Q}$ (bottom panel) in MM (day 0) and after 1, 2 and 4 days in DM. **(D)** Quantification of the percentage of different cell cycle populations of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* and *Nsd1*$^{N1918Q}$ in MM (day 0, **i**) and after 1 (**ii**), 2 (**iii**), and 4 (**iv**) days in DM (n=5). 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 tested with a paired two-tailed t- test.

***Nsd1* expression led to upregulation of erythroid signature genes**

To address the molecular mechanisms of how *Nsd1* controls erythroid differentiation, we performed RNA-sequencing from *Nsd1*$^{-/-}$ erythroblasts expressing *Nsd1* and *Nsd1*$^{N1918Q}$, prior (MM) and upon one day of differentiation induction (24h) (**Fig.15A**). We chose these time points because we wished to explore whether the observed effects of *Nsd1* reintroduction are exclusively confined to differentiation induction, or that subtle changes have already taken place during expansion. When comparing differential gene expression (DEG) between the two differentially transduced cells in expansion culture (MM, 0h), only 53 genes were significantly changed (**Supplementary table S1**). Interestingly, among these 53 genes, we found a significant reduction in the H3K4 Lysine-specific demethylase 1 (LSD1) and growth factor independence 1 (Gfi1) interactor *Rcor2*. *Rcor2* was reported to facilitate LDS1 demethylation activity, and collaboratively repress target genes during erythro-megakaryotic differentiation[192]. Notably, 24h of induced differentiation, caused a significant change of expression of around 2% of genes either significantly ($p<0.05$) increased (270 of 15804) or decreased (318 of 15804) (**Fig.15B, Supplementary table S2**). Remarkably, when considering only the top 100 DEGs, we observed mostly significantly down-regulated genes (**Fig15B, Supplementary table S1**). A similar pattern in DEG has been reported for fetal liver-derived mouse hematopoietic cells during transitioning from Ter119$^{-}$ to Ter119$^{+}$ cells, where this erythroid maturation was predominately associated with gene repression[193]. Among the significantly differentially upregulated genes, we found the cell cycle regulators *Cdc7* and *Cdk2,* the epigenetic regulator *Kmt5a (Setd8)* and the high mobility group protein (*Hmgb2*), all previously reported to be upregulated during erythroid differentiation[173, 175, 194, 195]. Among the differentially downregulated genes, we found the *Setd8* target vimentin (*vim*)[175], the transcription factor *Gata2*, and the RNA-binding protein *Zfp36l2[196, 197]* known for their role of regulating self-renewal of hematopoietic stem and erythroid progenitor cells, and which expression need to be suppressed during erythroid maturation.

Gene set enrichment analysis (GSEA) of DEGs of *Nsd1*$^{-/-}$ cells expressing *Nsd1* after 24h of differentiation (24h vs. 0h) revealed significant ($p<0.0001$) positive enrichment with signatures linked to erythroid development, differentiation, and putative GATA1 target genes, and a significant decline in a negative regulatory differentiation-related signature (**Fig. 15C, Supplementary table S3**). As very few genes were differentially expressed between cells

expressing *Nsd1* or the *Nsd1*$^{N1918Q}$ mutant in MM (0h vs. 0h), GSEA revealed positive yet not significant signatures associated with erythroid development, heme metabolism and cell cycle checkpoint progression, and a decline in MYC targets signature (**Fig. 15D**). However, the same comparison of *Nsd1* and *Nsd1*$^{N1918Q}$ expressing cells after 24h in DM (24h vs. 24h), GSEA revealed significant positive correlations with erythroid lineage development, heme metabolism and putative GATA1 target genes, and significant negative correlation with signatures related to regulation of differentiation (**Fig. 15E, Supplementary table S4).**

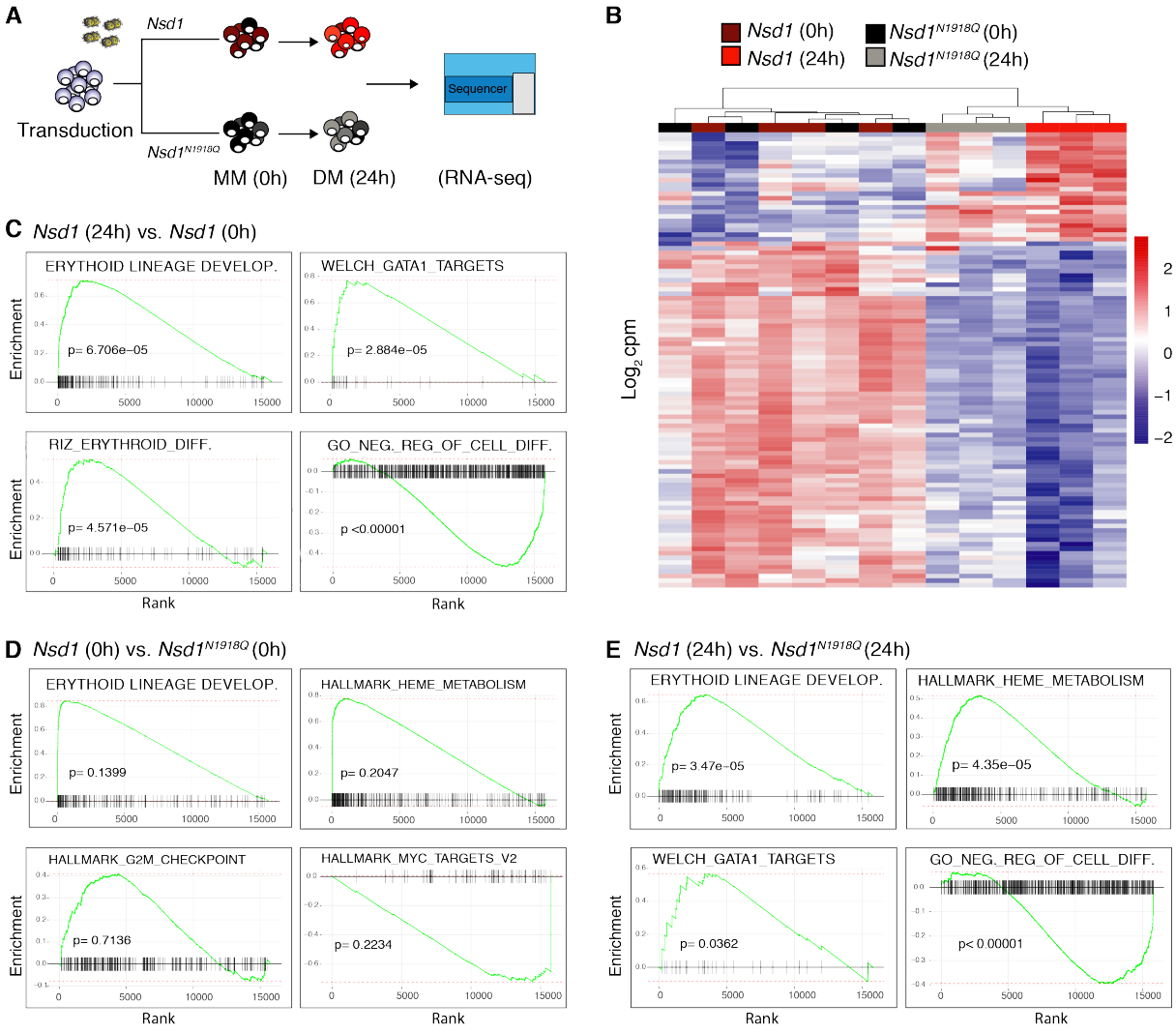


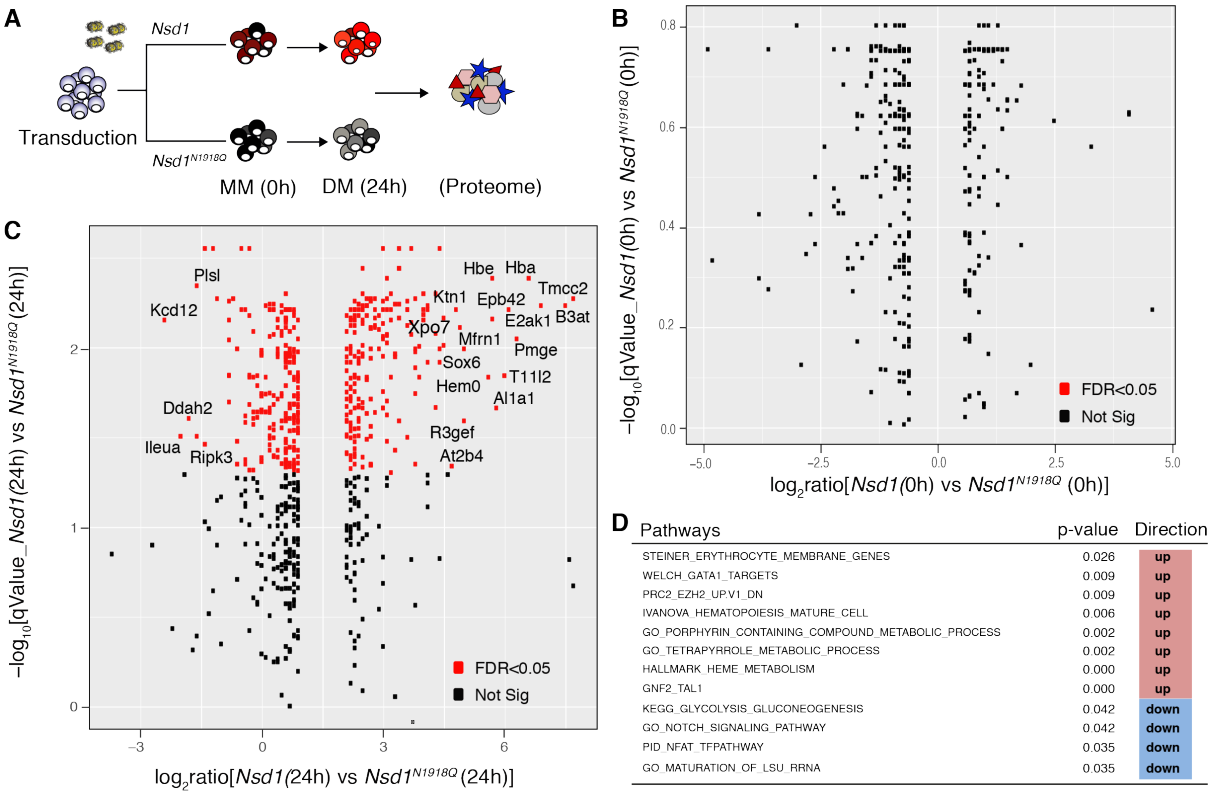

**Fig. 15: *Nsd1* expression led to upregulation in erythroid signature genes**

**(A)** Schematic illustration of experimental set up: *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing either *Nsd1* or *Nsd1*$^{N1918Q}$ were analyzed using total mRNA-sequencing in MM (0h) and after 24h in DM (24h). **(B)** Heatmap of the top 100 DEG (corresponding to FDR < 1.06 *10$^{9}$) of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing either *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (brown and black squares respectively, 0h) after 24h in DM (red and grey squares, 24h). Columns clustering was done by Wards linkage on correlations. **(C)** GSEA of differential expression between *Nsd1*$^{-/-}$ BM-derived erythroblast expressing *Nsd1* in MM(0h) and after 24h in DM (24h). **(D)** GSEA of differential expression between *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing either

*Nsd1* or *Nsd1*$^{N1918Q}$ kept in MM (0h). **(E)** GSEA of differential expression between *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing either *Nsd1* or *Nsd1*$^{N1918Q}$ after 24h in DM (24h).

### *Nsd1* expression increased the expression of erythroid regulatory proteins

To corroborate our transcriptomic data, we compared changes in global protein contents of BM-derived *Nsd1*$^{-/-}$ expressing *Nsd1* or *Nsd1*$^{N1918Q}$ during expansion (MM, 0h) and after 24h of induced differentiation (DM, 24h) (**Fig. 16A**). In accordance with the DEGs RNA seq data, we also found no significant difference in protein expression between *Nsd1*$^{-/-}$ erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (**Fig. 16B**). However, we observed a drastic shift in differential protein expression in *Nsd1* vs. *Nsd1*$^{N1918Q}$ expressing erythroblasts after 24h in DM. *Nsd1* transduced cells expressed significantly (FDR< 0.05) higher protein levels of several proposed GATA1 targets like hemoglobin (HBA, HBB1, HBE) and exportin 7 (XPO7)[198]. We also detected increased expression of some membrane proteins associated with terminal erythroid maturation such as the erythrocyte membrane protein band 4.2 (EPB42) and the coiled-coil domain family2 (TMCC2) (**Fig. 16C, Supplementary table S5**)[199, 200]. Protein pathways analysis showed a significant upregulation for pathways involved in GATA1 target activation, erythroid maturation and metabolism, and a significant downregulation of pathways regulating stemness (**Fig. 16D**). Collectively, differential expression at the mRNA and protein level suggest that the catalytic activity of NSD1 is essential for erythroid maturation and regulation of GATA1 targets.




| Pathways | p-value | Direction |
|---|---|---|
| STEINER_ERYTHROCYTE_MEMBRANE_GENES | 0.026 | up |
| WELCH_GATA1_TARGETS | 0.009 | up |
| PRC2_EZH2_UP.V1_DN | 0.009 | up |
| IVANOVA_HEMATOPOIESIS_MATURE_CELL | 0.006 | up |
| GO_PORPHYRIN_CONTAINING_COMPOUND_METABOLIC_PROCESS | 0.002 | up |
| GO_TETRAPYRROLE_METABOLIC_PROCESS | 0.002 | up |
| HALLMARK_HEME_METABOLISM | 0.000 | up |
| GNF2_TAL1 | 0.000 | up |
| KEGG_GLYCOLYSIS_GLUCONEOGENESIS | 0.042 | down |
| GO_NOTCH_SIGNALING_PATHWAY | 0.042 | down |
| PID_NFAT_TFPATHWAY | 0.035 | down |
| GO_MATURATION_OF_LSU_RRNA | 0.035 | down |

**Fig. 16: *Nsd1* expression increased the expression of erythroid regulatory proteins**
**(A)** Schematic illustration of experimental set up: *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing either *Nsd1* or *Nsd1*$^{N1918Q}$ were analyzed using global proteome in MM (0h) and after 24h in DM (24h). **(B)** Differential protein expression of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (0h, n=3 per group, FDR< 0.05, p-value <0.05). **(C)** Differential protein expression of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ after 24h in DM (24h, n=3 per group, FDR< 0.05, p-value <0.05). Labels are shown for proteins with logFC>3. **(D)** Differentially expressed protein pathways determined by difference in corresponding protein products (measured by mass-spectrometry) in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ measured after 24h in DM. Significance level at FDR<0.05, using Fisher's exact test.

## *Nsd1* expression resulted in gain of H3K36me1/2 and loss of H3K27me3 upon induced erythroid differentiation

Alterations in histones posttranslational modifications have been implicated in many hematologic disorders, ranging from severe anemias, myelodysplastic disorders, to leukemias[201, 202]. Erythroid differentiation in mouse and human cells is associated with dynamic changes in histone methylation, leading to progressive changes in chromatin structure and gene expression[193, 203]. Previous studies have shown that dysregulation of histone modifiers, such as the H4K20 methyltransferase SetD8 and EZH2 targeting H3K27me3, lead to impaired erythroid differentiation and anemia[173, 204, 205]. In order to better study the changes in histone methylation dynamics in upon *Nsd1* expression, particularly with

limited access to ChIP-grade H3K36me antibodies, we used tandem mass spectrometry analysis to determine changes on various marks (H3K36me1, H3K36me2, H3K36me3, H4K20me1, H4K20me2, H4K20me3, H3K27me1, H3K27me2, H3K27me3, H3K9me1, H3K9me2, H3K9me3) on histones extracted from *Nsd1* or *Nsd1*$^{N1918Q}$ transduced cells before (0h) and after (24h) induced differentiation (**Fig. 17, Supplementary table S6**)[206]. Although none of changes reached statistical significance, we could nevertheless detect several changes occurring in histone tails lysine methylation upon *Nsd1* expression. In agreement with NSD1 reported specific role as H3K36me1/2 methyltransferase, we observe an increase in H3K36me1/2 in *Nsd1* expressing cells compared to *Nsd1*$^{N1918Q}$ after 1 day of erythroid induction, while H3K36me3 seemed less affected (**Fig. 17A**). In addition, we also observed an increase in H4K20me1/2 before (0h) and upon differentiation (24h) in *Nsd1* expressing cells compared to the SET-domain mutant (**Fig. 17B**). Surprisingly, we detected two opposite patterns in H3K27 methylation status upon differentiation of the differently transduced cells: an increase in H3K27me1, coupled to a decrease in H3K27me3 mark (**Fig. 17C**).

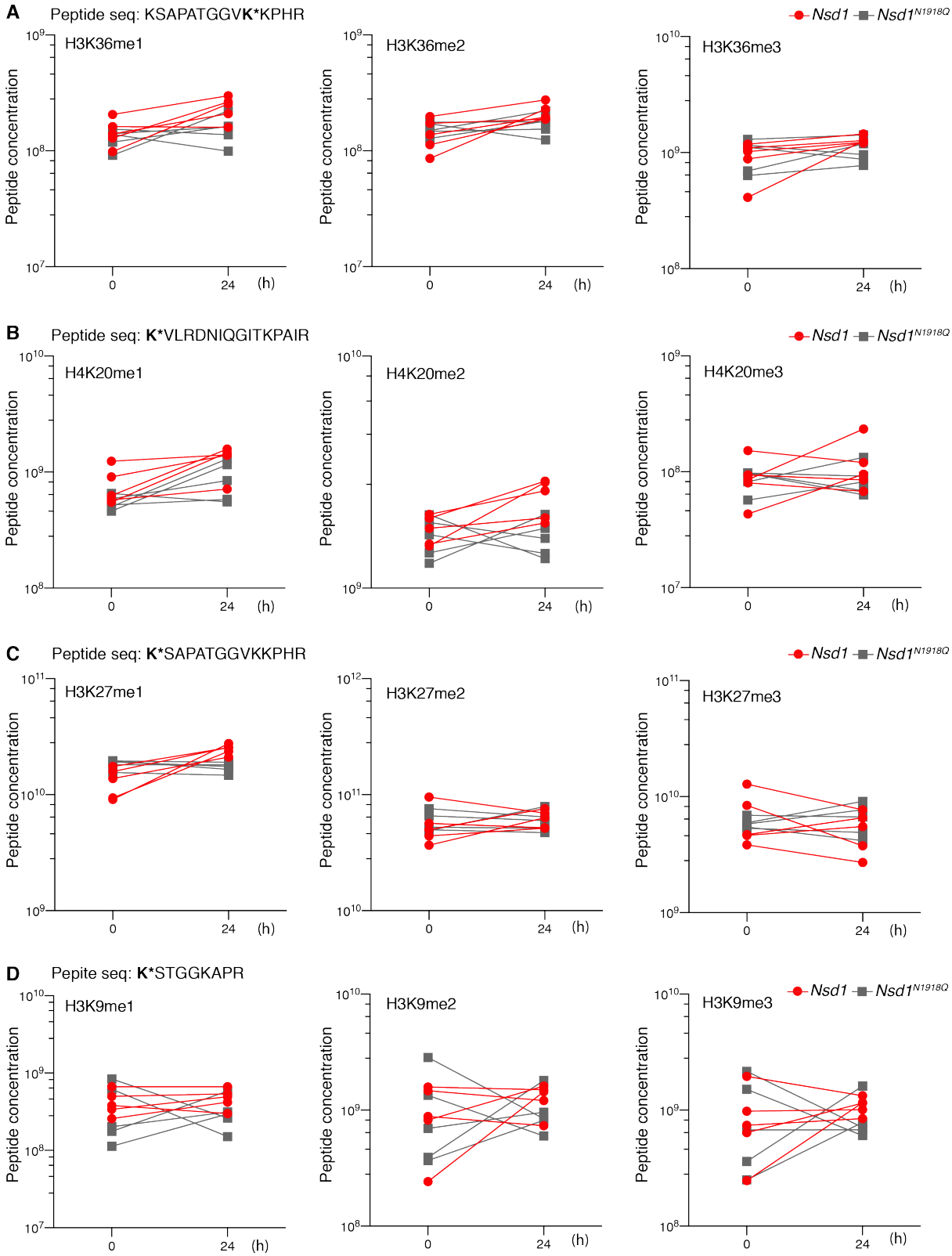



**Fig. 17: *Nsd1* expression resulted in gain of H3K36me1/2 and loss of H3K27me3 upon differentiation**

**(A)** Comparative quantification of Histone 3 lysine 36 mono, di and tri methylation (H3K36me1, H3K36me2 & H3K36me3) peptide concentration, in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (0h) and after 24h in DM (24h). Peptides containing H3K36me1 and H3K36me2 are increased in abundance upon differentiation (24h) in *Nsd1* expressing cells (red circles) compared to *Nsd1*$^{N1918Q}$ expressing cells (grey circles), whereas, peptides containing H3K36me3 are detected at similar levels (n=5). **(B)** Comparative quantification of Histone 4 lysine 20 mono, di and tri methylation (H4K20me1, H4K20me2

& H4K20me3) peptide concentration, in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (0h) and after 24h in DM (24h). Peptides containing H4K20me1 and H4K20me2 are increased in abundance upon differentiation (24h) in *Nsd1* expressing cells (red circles) compared to *Nsd1*$^{N1918Q}$ expressing cells (grey circles), whereas, peptides containing H4K20me3 are detected at similar levels (n=5). **(C)** Comparative quantification of Histone 3 lysine 27 mono, di and tri methylation (H3K27me1, H3K27me2 & H3K27me3) peptide concentration, in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (0h) and after 24h in DM (24h). Peptides containing H3K27me1 are increased in abundance upon differentiation (24h) in *Nsd1* expressing cells (red circles), compared to *Nsd1*$^{N1918Q}$ expressing cells (grey circles). Peptides containing H3K27me3 are decreased upon differentiation (24h) in *Nsd1* expressing cells (red circles), compared to *Nsd1*$^{N1918Q}$ expressing cells (grey circles). Peptides containing H3K27me2 are detected at similar levels (n=5). **(D)** Comparative quantification of Histone 3 lysine 27 mono, di and tri methylation (H3K9me1, H3K9me2 & H3K9me3) peptide concentration, in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (0h) and after 24h in DM (24h). Peptides containing all three modifications are detected at similar levels (n=5). Values are presented as individual points. Statistical significances in A-D tested with a paired two-tailed t- test.

While H3K27me3 is frequently associated with gene expression silencing, H3K27me1 was reported to mark active gene transcription[207]. Lastly, we do not observe any meaningful changes in H3K9 methylation, a mark for silencing and heterochromatin formation, upon *Nsd1* expression and differentiation (**Fig. 17D**). Collectively, MS-based analysis showed that upon erythroid induction, *Nsd1* restoration results in histone marks modifications linked to activated transcription.

### *Nsd1* expression altered histone marks and increases GATA1 chromatin binding

*Nsd1* expression in *Nsd1*$^{-/-}$ BM-derived erythroblast lead to slight but gradual increase in *Gata1* mRNA expression following 1 and 2 days of induced differentiation, however, we did not detect drastic changes in GATA1 protein levels during this time period (**Fig. 18A-B**). Next, we looked for changes in histone 3 lysine 36 (H3K36) methylation marks, the reported major substrate for the NSD1 methyltransferase activity[103, 208, 209]. Upon *Nsd1* expression we observed a clear increase in global H3K36 di-methylation (H3K36me2) mark in MM (day 0), but we could detect only a minor increase in H3K36me1, and almost no difference in H3K36me3 (**Fig. 18C**). Nevertheless, considering all the previous observations, we decided to compare *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* and *Nsd1*$^{N1918Q}$ in GATA1 chromatin binding, changes in H3K27 acetylation (H3K27ac) and H3K36me3 marks by ChIP-seq in MM (0h) and following 24h of differentiation (**Fig. 18D**). Although we did not detect major changes in H3k36me3 by immunoblotting, we still decided to determine changes of this mark, as at that time we were not aware of any well-performing ChIP-grade antibody that is able to dissect H3K36me1 from H3K36me2 and H3k36me3. Preliminary results using ChIP-PCR

analysis showed an increase in the occupancy of GATA1 to its locus and the *Hb-B* locus in *Nsd1* expressing cells after 24h of induced differentiation (**Fig. 18E**). In agreement with our previous RNA-seq and proteome data, ChIP-seq analysis, we did not detect any significant changes in GATA1 binding, H3K27ac or H3K36$^{me3}$ marks in MM between *Nsd1* and *Nsd1*$^{N1918Q}$ mutant expressing cells (0h)(**data not shown**). However, after 24h in DM, we observed increased occupancy of GATA1 protein at over 3000 sites in the genome overlapping with 1362 genes ($p<0.01$) in cells expressing *Nsd1* in comparison to the *Nsd1*$^{N1918Q}$ mutant (**Fig. 18F, Supplementary table S7**). Of genes with significantly increased binding of GATA1, 731 of them had the promotor regions decorated by H3K27ac while H3K36me3 marks overlapped with 1179 gene bodies (**Supplementary tables S8 &S9**). Thus, while the global levels of GATA1 protein remained constant, reintroduction of *Nsd1* resulted in increased DNA binding to available GATA1 sites in promotor regions, similarly reflected in changes in H3K36me3 and H3K27ac at the genomic coordinates. Interestingly, changes in gene expression aligned with H3K27ac around TSS, confirming that these epigenetic marks are directly regulating the down-stream transcriptional programming (**Fig. 18G**). However, we could not detect any gene loci with statistically significant increase of all three GATA1, H3K36me3 and H3K27ac marks, which could be a matter of the order of events occurring along the activation pathway. Nevertheless, *Nsd1*-induced changes of several erythroid regulators were associated with simultaneous changes in GATA1 binding, H3K27ac and H3K36me3 marks. As an example, the *Pklr* gene locus, encoding for the liver-red cell pyruvate kinase that is linked to erythroid differentiation, and *Art4*, encoding for the developmentally regulated Dombrock blood group glycoprotein were both upregulated in *Nsd1*$^{-/-}$ cells expressing *Nsd1* associated with a narrow GATA1 peak in the promotor region within a broader decoration of H3K27ac, followed by gene body-wide H3K36me3 marks (**Fig. 18H i & ii**)[210, 211]. The opposite was observed for the gene encoding for *Fgf*2 (fibroblast growth factor 2) associated with inhibition of efficient erythroid differentiation that appeared upregulated in *Nsd1*$^{N1918Q}$ expressing cells (**Fig. 18H ii**)[212]. Collectively, integrated analysis by RNA-seq, proteomics and ChIP sequencing indicates that *Nsd1* plays an instructive role deriving erythroid differentiation.

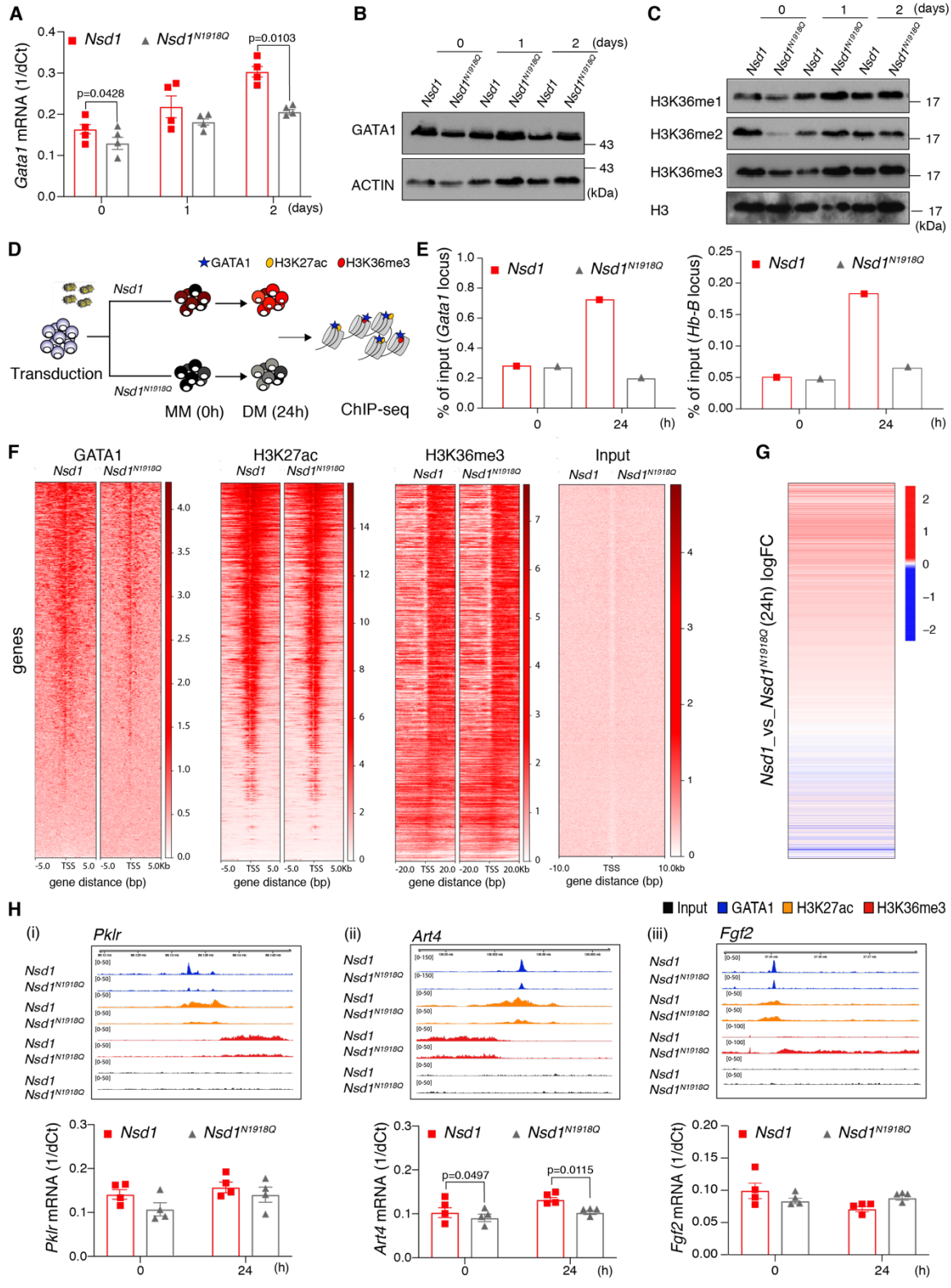
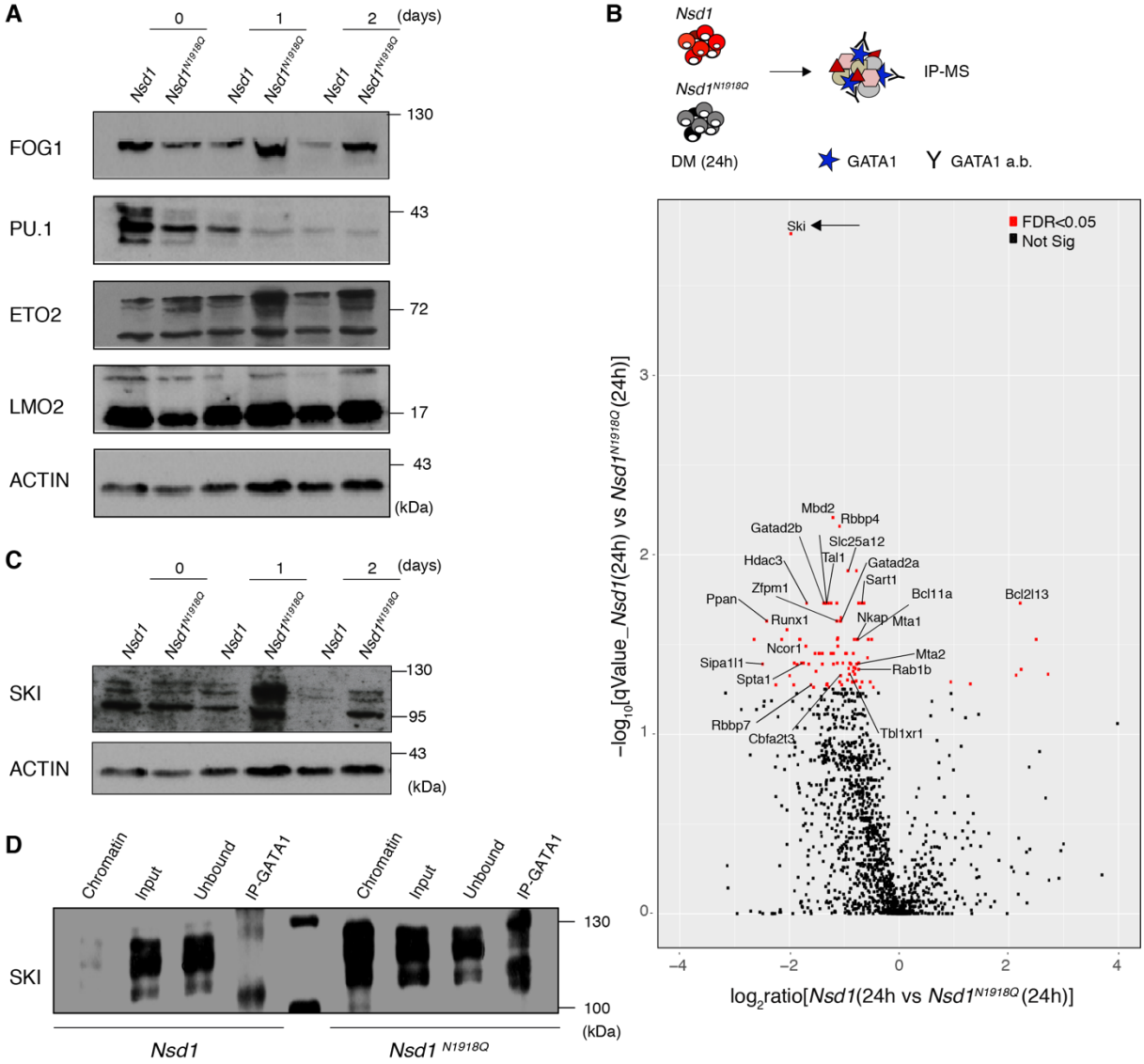



**Fig. 18: *Nsd1* expression altered histones' marks patterns and increases GATA1 chromatin binding**
**(A)** Relative *Gata1* mRNA expression levels (1/dCt) in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* (red bars) or *Nsd1*$^{N1918Q}$ (grey bars) in MM (day 0) and after 1 and 2 days in DM. Values were normalized to *Gapdh* (n=4 per group). **(B)** Western blot showing GATA1 protein expression in 1x10$^{6}$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (day 0), and after 1 and 2 days in DM. Actin was used as loading control. **(C)** Western blot showing H3K36 mono (H3K36me1), Di (H3K36me2) and Tri (H3K36me3) methylation expression in 1x10$^{6}$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ upon

expansion in MM (day 0), and after 1 and 2 days in DM. Histone 3 (H3) was used as loading control. **(D)** Schematic illustration of experimental set up: *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing either *Nsd1* or *Nsd1*$^{N1918Q}$ were analyzed using Chromatin immunoprecipitation-sequencing (ChIP-seq) in MM (0h) and after 24h in DM (24h). **(E)** ChIP-qPCR performed on *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing either *Nsd1* (red bars) or *Nsd1*$^{N1918Q}$ (grey bars) in MM (0h) and after 24h in DM (24h) using a specific GATA1 antibody and a non-specific control IgG. Graph shows enrichment for GATA1 binding to its own locus (*Gata1* locus) and one regulatory target (*Hb-B* locus). Results are expressed relative to input chromatin (n=1). **(F)** Heatmaps of genome-wide ChIP-seq signals in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* (left column) or *Nsd1*$^{N1918Q}$ (right column) after 24h in DM for GATA1, H3K27ac, and H3K36me3. All heatmaps are sorted decreasingly according to read coverage around transcriptional start sites (TSS) of GATA1 (leftmost). Input denotes sheared non-immunoprecipitated DNA (rightmost), serving as visual control. This data represents 1 of 2 independent experiments. **(G)** One-dimensional heatmap of logFC between gene expression of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ after 24 in DM (Figure 15B) sorted according to read coverage around TSS for H3K27ac ChIP (data as shown in panel Figure 18F). Only overlapping genes are displayed). **(H)** Integrated genome viewer (IGV) representation of GATA1, H3K27ac, and H3K36me3 ChIP peaks in the *Pklr* (left panel, **i**), *Art4* gene locus (middle panel, **ii**) and *Fgf2* (right panel, **iii**) from *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ after 24h in DM. Lower panels show *Pklr*, *Art4* and *Fgf2* mRNA relative expression levels (1/dCt) in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (0h) and after 24h in DM (24h). Values are shown as relative expression normalized to *Gapdh* (n=4). ChIP-IP was done by Samantha Tauchmann in collaboration with Cecile Thirant (Gustave Roussy Institute, Paris). 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 A, E and H tested with a paired two-tailed t- test.

### *Nsd1* expression was associated with changes of GATA1 protein interaction partners

To address whether impaired chromatin binding and transactivation of GATA1 in absence of *Nsd1* might be associated with aberrant expression of regulatory GATA1-interacting proteins, we performed Western blot analysis on *Nsd1*$^{-/-}$ BM-derived erythroblasts before (day 0) and 1 and 2 days after differentiation for several candidate proteins (**Fig. 19A**). We observed downregulation of PU.1 during induced differentiation of *Nsd1* and *Nsd1*$^{N1918Q}$ expressing cells. Interestingly *Nsd1*$^{N1918Q}$ expression in *Nsd1*$^{-/-}$ cells did not result in downregulation of FOG1 and ETO2 expression, both proteins are known to act as GATA1 co-repressors[213]. Thus, we hypothesized that the lower GATA1 chromatin binding observed in *Nsd1*$^{N1918Q}$ expressing *Nsd1*$^{-/-}$ cells could be a consequence of its enhanced binding to one or multiple transcriptional co-repressors. To test this hypothesis, we immunoprecipitated GATA1 followed by MS in *Nsd1*$^{-/-}$ BM-derived erythroblasts either expressing *Nsd1* or *Nsd1*$^{N1918Q}$ mutant after 24h induced differentiation (**Fig. 19B**). We identified 413 differentially expressed proteins ($p<0.05$) (**Supplementary table 10**) of which the most significant ones included known interactors of GATA1 such as MBD2, RBBP4, ZFPM1, RUNX1 and TAL1 suggesting functionality of the assay[213] (**Fig.19B**).




E

| Mono-methyl | Histone | Molecular Weight | *Nsd1*$^{N1918Q}$-IgG | *Nsd1*$^{N1918Q}$-GATA1 | *Nsd1*-IgG | *Nsd1*-GATA1 |
|---|---|---|---|---|---|---|
| | H1.2 | 21 kDa | 3 | 4 | 3 | 0 |
| | H4 | 11 kDa | 0 | 1 | 0 | 0 |
| | H3.1 | 15 kDa | 1 | 6 | 0 | 1 |
| **Total** | | | 4 | 11 | 3 | 1 |
| | | | | | | |
| **Di-methyl** | | | | | | |
| | H4 | 11 kDa | 0 | 0 | 1 | 1 |
| | H1.4 | 22 kDa | 2 | 5 | 6 | 4 |
| | H3.1 | 15 kDa | 2 | 6 | 2 | 3 |
| **Total** | | | 4 | 11 | 9 | 8 |
| | | | | | | |
| **No modification** | | | | | | |
| | H1.2 | 21 kDa | 60 | 55 | 70 | 55 |
| | H4 | 11 kDa | 21 | 36 | 23 | 40 |
| | H1.4 | 22 kDa | 47 | 45 | 50 | 48 |
| | H3.1 | 15 kDa | 5 | 4 | 2 | 5 |
| **Total** | | | 73 | 85 | 75 | 93 |

**Fig. 19: *Nsd1* expression resulted in changes of GATA1 protein interaction**

**(A)** Western blot showing FOG1, PU.1, ETO2 and LMO2 expression in 1x10$^6$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ upon expansion in MM (day 0), and after 1 and 2 days in DM, Actin was used as loading control. **(B)** Schematic illustration of experimental set up (Top): *Nsd1*$^{-/-}$ BM-derived erythroblasts

expressing either *Nsd1* or *Nsd1*$^{N1918Q}$ (10x10$^6$) were analyzed by GATA1 immunoprecipitation followed by mass spectrometry (IP-MS) after 24h in DM (24h). Volcano plot of differential protein enrichments by GATA1 immunoprecipitation (GATA1-IP) in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ after 24h in DM normalized to IgG control (n=2) (Bottom). Significant reduction in GATA1-SKI association (indicated by a black arrow) was observed upon expression of *Nsd1* compared to *Nsd1*$^{N1918Q}$ (FDR < 0.05). **(C)** Western blot analysis showing SKI protein expression in 1x10$^6$ BM-derived *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ in MM (day 0), and after 1 and 2 days in DM. Actin was used as a loading control. **(D)** Western blot analysis showing SKI protein expression in GATA1-immunoprecipitates from nuclear extracts of 10x10$^6$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ after 24 hours in DM. **(E)** Table of the number of mono, di methylated and unmodified lysine residues on histones measure by mass-spectrometry in GATA1-IP and IgG-IP control in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *Nsd1* or *Nsd1*$^{N1918Q}$ after 24h. GATA1-IP was done by Samantha Tauchmann.

Surprisingly, we observed that early differentiation of *Nsd1* expressing erythroblasts was coupled with a strikingly significant reduction (logFC=-1.96; p<1.08 x10$^{-7}$) of the transcriptional co-repressor protein SKI ("Sloan-Kettering institute"). SKI protein was previously proposed to interact with and inhibit GATA1 activation, most likely in cooperation with the nuclear corepressor (NCoR) complex[214]. Notably, several members of the NCoR-complex (NCOR1, HDAC3, TBLXR1) co-appeared with SKI, as differentially regulated (**Fig. 19B**). We also observed persistence SKI protein expression, and co-precipitations in *Nsd1*$^{N1918Q}$ compared to *Nsd1* expressing cells after induced differentiation (**Fig. 19C-D**). As NSD1 has been previously shown to methylate non-histone proteins, we wondered whether the MS analysis would be able to detect differentially methylated proteins pulled-down by the GATA1 IP. However, we did not detect any changes in the methylation status of pulled-down proteins, including GATA1 and SKI. Unexpectedly, there were a general reduction in methylation of pulled-down histones, which could be due modification switch (acetylation) or loss of repressive methylation (**Fig. 19E**).

**SKI-knockdown induced terminal erythroid differentiation of *Nsd1*$^{-/-}$ erythroblasts**

To explore whether reducing the expression of SKI is sufficient to trigger terminal maturation of *Nsd1*$^{-/-}$ BM-derived erythroblasts, we applied virally expressed shRNA-mediated *Ski* knockdown (**Fig.20A**). We observed that SKI knockdown resulted in significantly reduced number and size of colonies formed in MC composed of cells expressing reduced levels of the stemness marker, Kit (**Fig. 20B-E**). SKI knockdown also increased *in vitro* terminal erythroid maturation of *Nsd1*$^{-/-}$ erythroblasts, as indicated by cellular morphology, shift in surface markers expression (CD71/Ter119/Kit), and growth (**Fig. 20F-J**) in liquid cultures kept followed over 4 days. Notably, SKI-knockdown did not significantly alter total GATA1 protein levels, as

detected by immunoblotting or immunoprecipitation (**Fig.20K-L**). Nevertheless, prolonged culture of cells in DM was associated with a general reduction of SKI levels suggesting a role for SKI during initiation rather than terminal differentiation. Collectively, this data suggests that in absence of *Nsd1,* terminal erythroid maturation is blocked as a consequence of impaired GATA1 transactivation dependent on its association with the transcriptional repressor SKI.

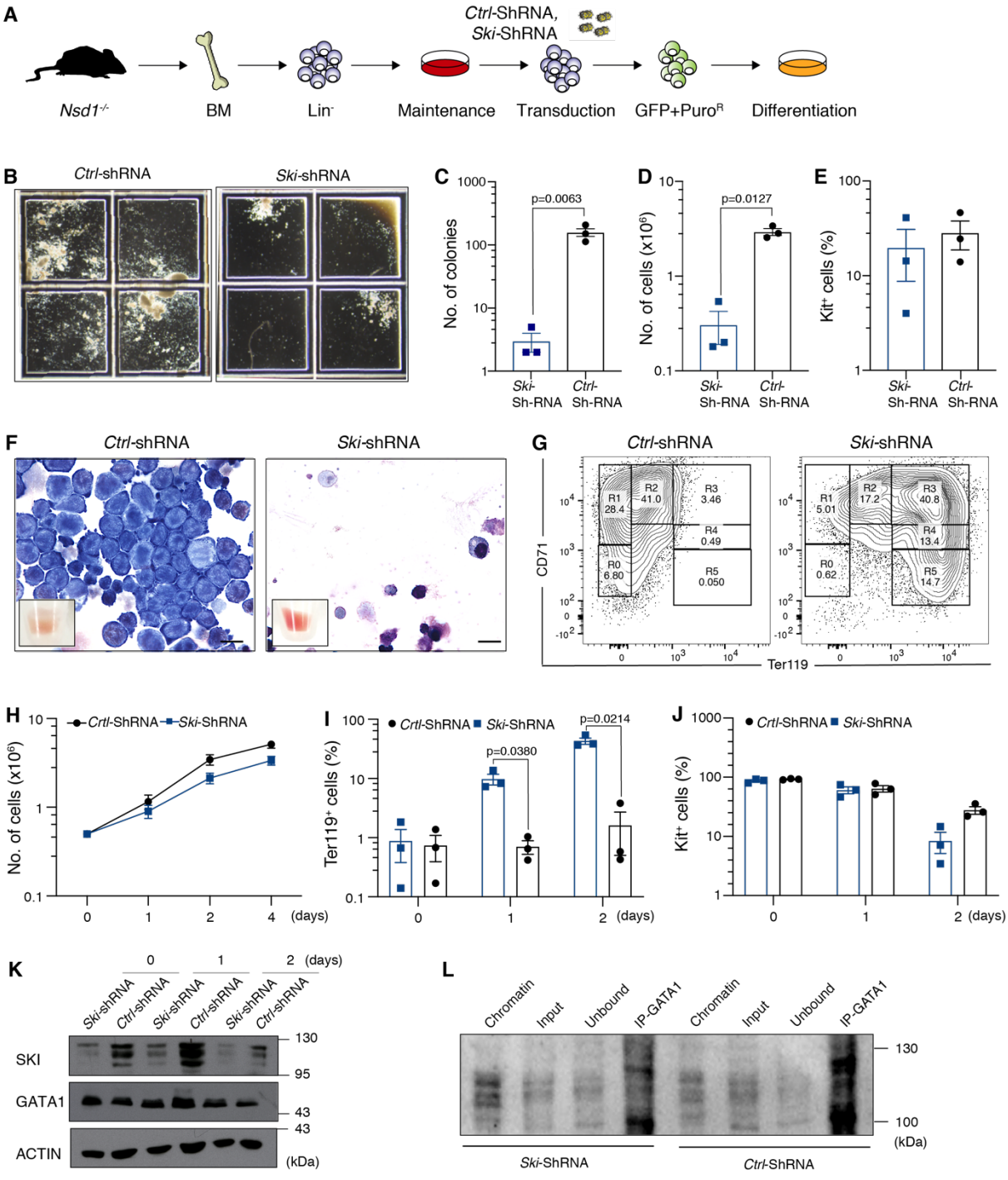


**Fig. 20: Knockdown of SKI expression led to terminal erythroid differentiation of *Nsd1*$^{-/-}$ erythroblasts**
**(A)** Schematic illustration of experimental setup: BM-derived *Nsd1*$^{-/-}$ erythroblasts were transduced with either *pLMP-empty-shRNA-GFP-Puro* (*Ctrl*-shRNA*),* or *pLMP-Ski-shRNA-GFP-Puro* (*Ski*-shRNA*)* in MM, sorted for GFP and selected with Puromycin for 2 days before induced differentiation and analysis. **(B)** Representative images of colonies in MC cultures (M3434) by 1x10$^4$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing either *control* (*Ctrl*-shRNA, left*)* or *Ski* shRNA (*Ski*-shRNA*,* right*)* after 11 days. This data represents 1 of 3 independent experiments. **(C)** Total number of colonies counted at day 11 after plating of 1x10$^4$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *control* (*Ctrl*-shRNA, black bar*)* or *Ski shRNA* (*Ski-shRNA,* blue bar*)* in MC (M3434) after 11 days*.* (n=3 per group). **(D)** Total number of cells obtained from 1x10$^4$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *control* (*Ctrl*-shRNA, black bar*)* or *Ski shRNA* (*Ski-shRNA,* blue bar) after 11 days in MC (M3434) after 11 days. Nucleated living cells were counted by the Trypan blue exclusion (n=3 per group). **(E)** Percentage of Kit$^+$ living cells obtained from 1x10$^4$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing *control* (*Ctrl*-shRNA, black bar*)* or *Ski shRNA* (*Ski-shRNA,* blue bar) in MC (M3434) after 11 days. (n=3 per group). **(F)** Representative images of Wright Giemsa-stained cytospin preparations (x1000, size bar=10μM) of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control (*Ctrl*-shRNA, left panel) and Ski shRNA (*Ski*-shRNA*)*, right panel) after 2 days in DM. The small inserts show the cell pellets before analysis. This data represents 1 of 3 independent experiments. **(G)** Flow cytometry analysis of CD71 and Ter119 markers' expression in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control (*Ctrl*-shRNA) or *Ski* shRNA (*Ski*-shRNA) after 4 days in DM. The gating strategy distinguishes different stages of erythroid maturation: "R0" fraction (CD71$^{-/low}$, Ter119$^-$), "R1" (CD71$^+$, Ter119$^-$), "R2" fraction (CD71$^+$, Ter119$^{+/low}$), "R3" fraction (CD71$^+$, Ter119$^+$), "R4" fraction (CD71$^{+/low}$, Ter119$^+$), and "R5" fraction (CD71$^-$, Ter119$^+$). This data represents 1 of 3 independent experiments. **(H)** Growth of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control shRNA (*Ctrl*-shRNA, black line) or Ski ShRNA (*Ski*-shRNA, blue line) in DM (1-4 days). Nucleated living cells were counted by Trypan blue exclusion (n= 3 per group). **(I)** Quantification of the percentage of Ter119$^+$ of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control shRNA (*Ctrl*-shRNA, black bar) or Ski ShRNA (*Ski*-shRNA, blue bar) in MM (day 0) and after 1 and 2 days in DM. (n=3 per group). **(J)** Quantification of the percentage of kit$^+$ of *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control shRNA (*Ctrl*-shRNA, black bar) or Ski ShRNA (*Ski*-shRNA, blue bar) in MM (day 0) and after 1 and 2 days in DM. (n=3 per group). **(K)** Western blot showing SKI and GATA1 protein expression in *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control shRNA (*Ctrl*-shRNA) or Ski ShRNA (*Ski*-shRNA) in MM and after 1 and 2 days in DM. Actin was used as a loading control. **(L)** Western blot analysis showing SKI protein expression in GATA1-immunoprecipitates from nuclear extracts of 10x10$^6$ *Nsd1*$^{-/-}$ BM-derived erythroblasts expressing control shRNA (*Ctrl*-shRNA) or Ski ShRNA (*Ski*-shRNA) after 24 hours in DM. 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, D, E, H, I and J tested with a paired two-tailed t- test.

**Murine erythroleukemia (MEL) cells: an alternative model to study erythroid differentiation**

Our observations in the conditional KO mouse model suggests that *Nsd1* is a critical regulator of erythroid differentiation. Although we were able to show that NSD1 regulates the activity of the GATA1 transcriptional master regulator, the detailed molecular mechanism still a mystery. Particularly, the limitations in cell number as well as good molecular tools (e.g. lacking ChIP grade anti-NSD1 antibodies) clearly set some limits to our assays.

Murine erythroleukemia (MEL) cell lines are transformed erythroid progenitor cells derived from the spleens of susceptible mice infected with the Friend virus complex[53]. These virally transformed cells are arrested at the proerythroblast stage of development and can be induced to undergo erythroid differentiation by treatment with polar-planar compounds such as DMSO and HMBA[215]. We decided to use MEL cells as a more practical alternative model to primary proerythroblasts to decipher the chromatin occupancy of *Nsd1* during erythroid differentiation. Hereby we aimed to used CRISPR-Cas9-mediated GE to i) add an epitope-tag to endogenously expressed *Nsd1* and ii) determine the impact of *Nsd1* inactivation on induced erythroid differentiation of MEL cells. Before starting GE, we excluded the presence of copy number alterations involving the *Nsd1* locus by performing quantitative PCR analysis using genomic DNA isolated from MEL cells and WT-BM (**Fig. 21A**). A ratio of ≈1 indicates equal allele copy number between the two tested genes (*Nsd1* and the house keeping gene Glucuronidase-Beta, *Gusb*) in the compared cell types (**Fig. 21A**). We also confirmed that our MEL cell line expresses *Nsd1* mRNA to a similar level as in WT-BM cells (**Fig. 21B**).

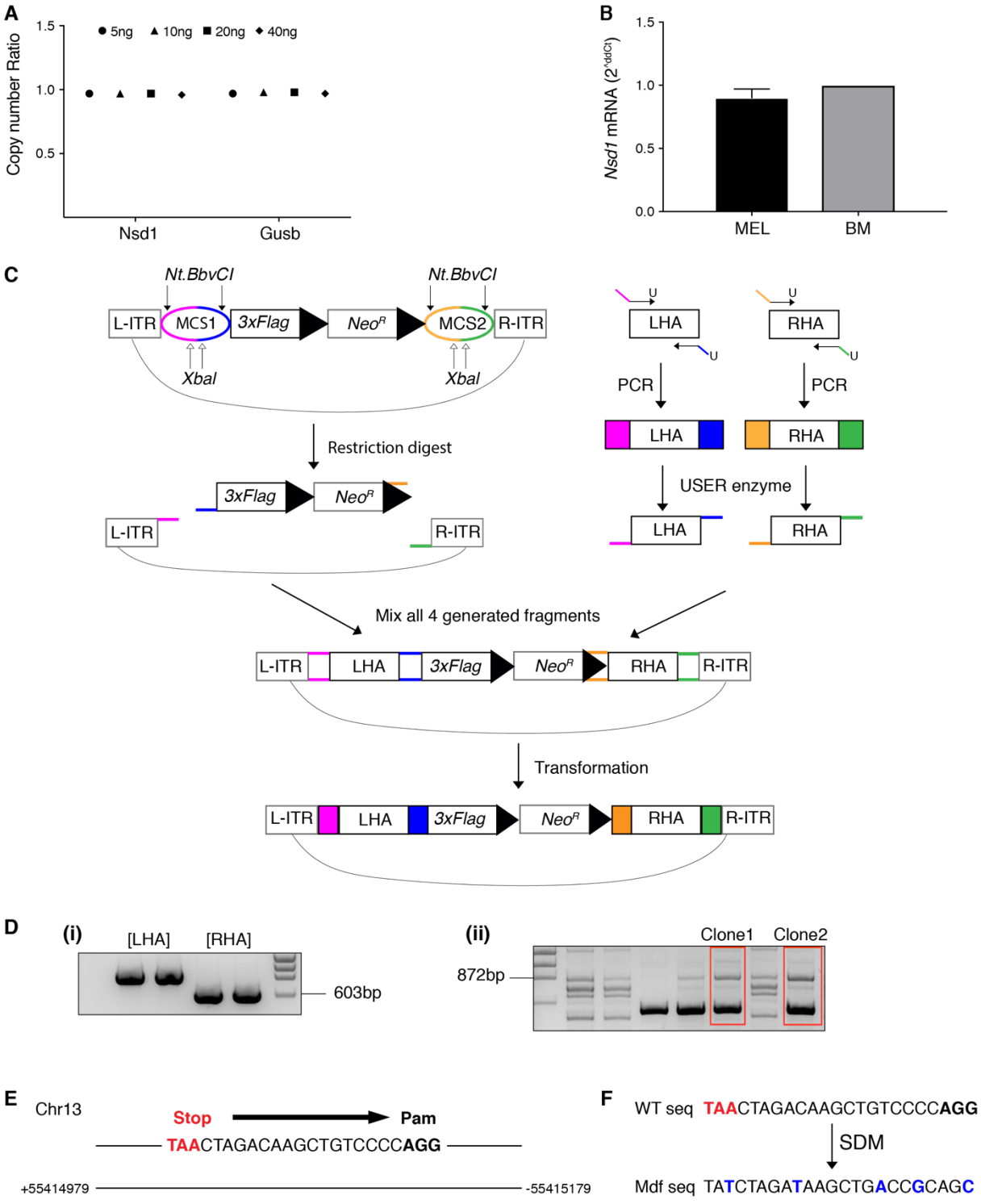



**Fig. 21: Cloning *Nsd1* template and sgRNA plasmids for GE**

**(A)** Copy number ratio of *Nsd1* and *Gusb* genes in MEL cells compared to wild-type bone marrow (wt-BM). Quantitative PCR was performed using increased concentrations of genomic DNA (gDNA) isolated from MEL cells and wt-BM using primers corresponding to *Nsd1* (unknown copy number) and *Gusb* (2 copies). A ratio of 1 indicates identical copy number in both cell types (n=1). **(B)** Fold change of *Nsd1* mRNA expression in MEL and WT-BM cells measured by quantitative PCR using primers corresponding to *Nsd1* (Ex13-14 junction) (n=3 per group). **(C)** Schematic illustration depicting the rapid one-step "USER cloning" of the *rAAV-3XFlag* knock-in vector. The *rAAV*-mediated targeting vector contains a multiple cloning site1 (MCS1) between L-ITR (left inverted terminal repeat) and 3xFlag sequences, and multiple cloning site 2 (MCS2), between the right lox-P site and R-ITR (right inverted terminal repeat) of the AAV-3xFlag knock-in vector to

generate the *pAAV-USER-3xFlag-KI* vector. These fragments contain two inversely oriented nicking endonuclease sites (*Nt.BbvCI*) separated by restriction endonuclease sites (*XbaI*). After treatment with *Nt.BbvCI* and *XbaI* restriction enzymes, the *pAAV-USER-3xFlag-KI* vector was digested into a Tag- *loxP-Neo-loxP* fragment flanked by two 5' single-stranded overhangs (blue and yellow lines) and a vector backbone flanked by two 5' overhangs (pink and green lines). PCR was then used to amplify left and right homologous arms of *Nsd1* from PSG5-*mNsd1* vector[216]. The sequence GGGAAAGU is added to the 5' of the forward left arm primers, and GGAGACAU is added to the reverse left arm primers. The PCR products were then treated with the 1 U of USER enzymes (New England Biolab) to generate single-stranded overhangs. Finally, the left and right arms fragments were mixed with the two vector fragments followed by bacterial transformation. (*Adapted from Zhang et al. 2008*)[217]. **(D)** Agarose gel image showing PCR products of the left homology arm (LHA) (1016bp) and the right homology arm (RHA) (625bp) of *Nsd1* (i). Colony PCR was performed screen for the correct insertion of the LHA and RHA (ii). **(E)** Single guided RNA (sgRNA) sequence spanning the STOP codon of murine *Nsd1* at exon 23, and the corresponding Pam sequence (NGG). **(F)** Template DNA sequence before (top) and after (bottom) base pair substitutions (in blue) by site directed mutagenesis (SDM). Mutating the STOP codon (red, A→T) created an open reading frame (ORF) resulting in the production of a fusion protein of NSD1-3xFLAG. Further base alterations (C→T, G→A, C→G and G→C) in the targeting DNA template were created to avoid undesired the cleavage by the sgRNA.

**Strategy used to integrate a 3x*FLAG* sequence at the C-terminus of *Nsd1* using CRISPR-Cas9 and homologous recombination repair**

As we could not find a commercially validated ChIP grade antibody for pulling down murine NSD1, we decided to knock-in a *Flag* sequence at the C-terminus of the *Nsd1* endogenous locus in MEL cells utilizing CRISPR-Cas9 and homology repair (HR) technology. Using the rapid Uracil excision-based cloning ''USER'' technique[217, 218], we generated a knock-in vector with right and left homology arms (RHA and LHA respectively) corresponding to the C-terminus of the murine *Nsd1* locus flanked by triple *Flag* and neomycin cassette (*3xFlag-Neo*$^{R}$) (**Fig. 21C-D**). Simultaneously, we cloned a CRISPR-Cas9 expression vector with a single guided RNA (sgRNA) spanning the STOP codon of the endogenous *Nsd1* locus (**Fig. 21E**)[219-221]. In order to avoid the cleavage of our cloned template vector with the expressed sgRNA, and to ensure the expression of the Flag sequence in frame (fused) with the *Nsd1* transcript, we created a 5 base pair substitution in our template DNA vector using site directed mutagenesis (**Fig. 21F**).

To Create clonally modified MEL-NSD1-3xFLAG cells, we devised an experimental workflow with two steps of transfection, GFP- single cell sorting and antibiotic resistance screening (**Fig. 22A**). Following the creation of a double stranded break (DSB) by the Cas9 protein, the cells would use the provided the template DNA plasmid to repair the breakage site by homologous recombination. The *Neo*$^{R}$ cassette is then excised by LoxP sites combinations in response to Cre-recombinase expression, resulting in the desired

modification at the *Nsd1* locus (**Fig. 22B**). The change in the size of the *Nsd1* locus caused by the modification was detected by genomic PCR screening using primer sets located at and spanning the *Neo*$^{R}$ cassette (**Fig. 22C-D**).

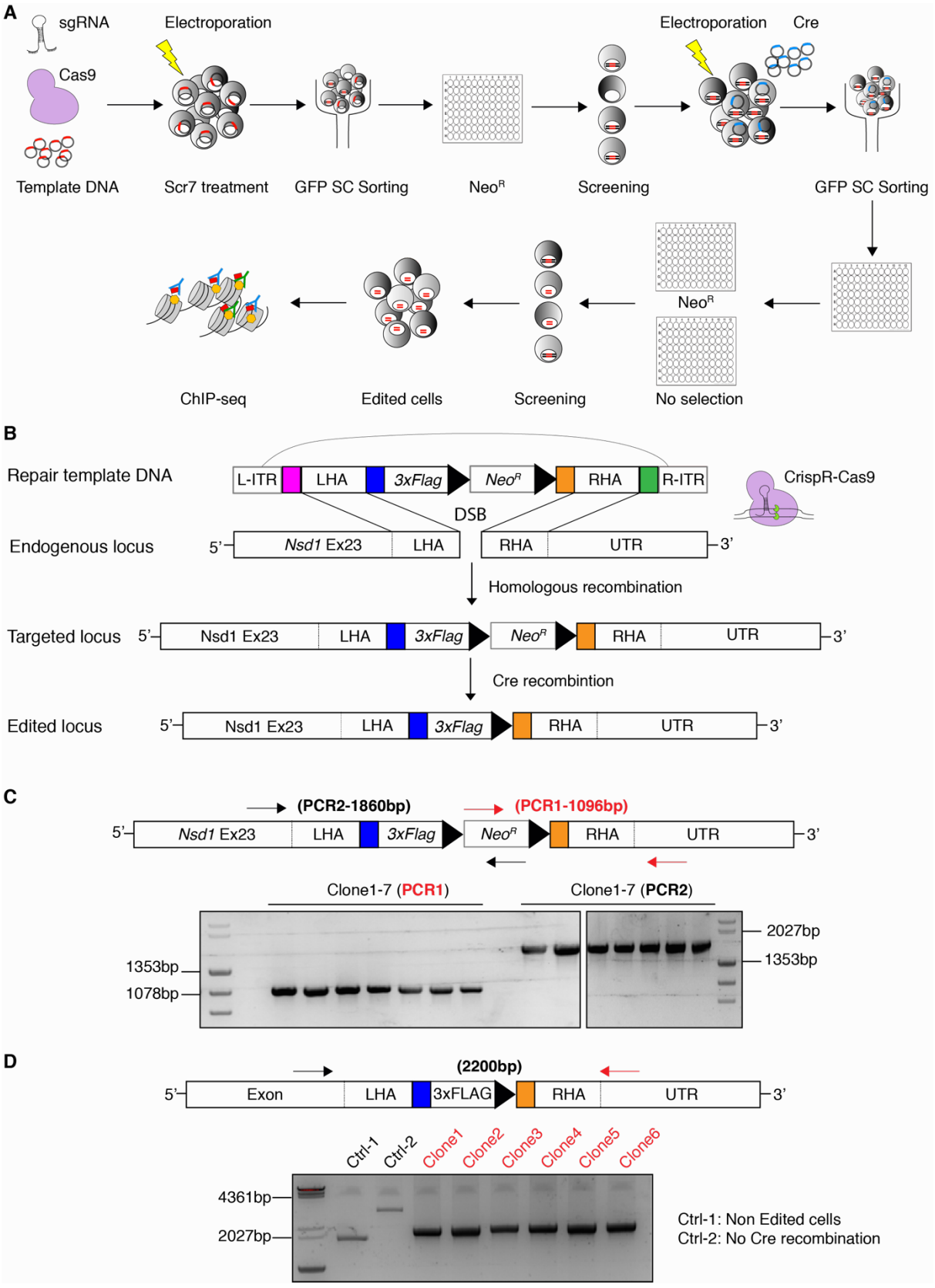


**Fig. 22: Strategy used to integrate a 3x*Flag* sequence at the C-terminus of *Nsd1* using CRISPR-Cas9 and homologous recombination repair**

**(A)** Schematic illustration of the generation of clonally edited MEL cells using CRISPR-Cas9 technology. MEL cells were transfected by electroporation with a *GFP-Cas9-sgRNA* and a repair template targeting plasmid (*pAAV-Nsd1-3xFlag-KI)*. Following treatment with 10uM of the ligase inhibitor Scr7 for 2 days, 2-5% of the highest expressing $GFP^+$ cells were sorted as single cells (GFP-SC) in a 96-well plate and expanded for 1 week. The cells were then treated with 1mg/mL of G418 (Neomycin derivative) for 10 days, and resistance cells were screened by PCR. Correctly edited clones were expanded and then subsequently transfected with a *pCMV-Cre-GFP* expressing plasmid prior to GFP (2-5% highest expressing cells) single cell sorting in 96-well plates. Following expansion for 1 week, the cells were divided into 2x96 well plates where only one plate was treated with 1mg/ml of G418 for 10 days. Sensitive cells were picked from the second untreated plate and screened for *Neomycin* expression cassette excision by PCR. Several clones of edited cells were further expanded, banked and used for Chromatin immunoprecipitation sequencing (ChIP-seq). **(B)** Schematic illustration depicting the molecular steps involved in the generation of the *Nsd1*-3xFlag MEL cells clones. A *3xFlag-LoxP-*$Neo^R$*-loxP* cassette was knocked in at the 3`end of Ex23 of the *Nsd1* locus by Cirspr-Cas9 mediated double strand breakage (DSB) and homologous recombination. Excision of the Neomycin resistance cassette was achieved by LoxP recombination via the expression of Cre-recombinase. **(C)** Schematic illustration of edited N*sd1* locus showing the position and size of expected PCR products for screening the insertion of *3xFlag-LoxP-*$Neo^R$*-loxP* cassette (Top). Agarose gel image of PCR products (Bottom). **(D)** Schematic illustration of edited *Nsd1* locus showing the position and size of expected PCR product for screening the excision of $Neo^R$ cassette (Top). Agarose gel picture of genotyping PCR products (Bottom).

### Knockdown of *Nsd1* impaired erythroid differentiation of NSD1-3xFLAG MEL cells

We consciously selected an sgRNA with low off-target probability for the editing protocol, in order to minimize causing alterations in the physiological behavior of cells. Nevertheless, prior to using the edited cells for further experiments, we first confirmed that the generated clonally-edited cells expressed equivalent levels of *Nsd1* mRNA to unedited controls (**Fig. 23A**), and that they can still undergo erythroid differentiation and hemoglobinization in response to DMSO treatment (**Fig. 23B**). We also validated the expression of the FLAG epitope in edited cells by Western blot analysis (**Fig. 23C**).

Next, we wanted to investigate whether *Nsd1* knockdown would negatively affect the erythroid differentiation potential of MEL cells, so we transduced the cells with five Sh-RNA lenti-viral vectors targeting the *Nsd1* transcript (**Fig. 23D**). Two out the five Sh-RNAs (#379 & 445) showed good knockdown potential and thus were selected for use subsequent experiments and renamed *Nsd1*-ShRNA-1 and *Nsd1*-ShRNA-2 respectively (**Fig. 23D**). We found that *Nsd1* knockdown caused an impairment in the erythroid differentiation of MEL cells as illustrated by benzidine staining, hemoglobinization of cell pellets and the significant reduction in $Ter119^+$ cell populations (**Fig. 23E&F**).

To study the impact of *Nsd1* knockdown on the activity of genes associated with erythroid differentiation, we compared the mRNA expression of several genes prior and following DMSO induced differentiation. We first measured and confirmed the efficiency of *Nsd1* knockdown in transduced cells (**Fig. 24A**). The expression of the erythroid master regulator *Gata1*, as well as several early (*Hb-B* & *Gypa*) and late (*Alas2*, *Ebp4.2* & *Scl2*) erythroid genes was significantly downregulated in *Nsd1*-ShRNA transduced cells compared to controls (**Fig. 23B-G**). *Nsd1* knockdown did not affect the downregulation of *Spi1* (**Fig. 24H**), which suggest that the observed erythroid impairment in these cells most probably follows the same regulatory pathway as the *Nsd1*$^{-/-}$ BM-derived erythroblasts.

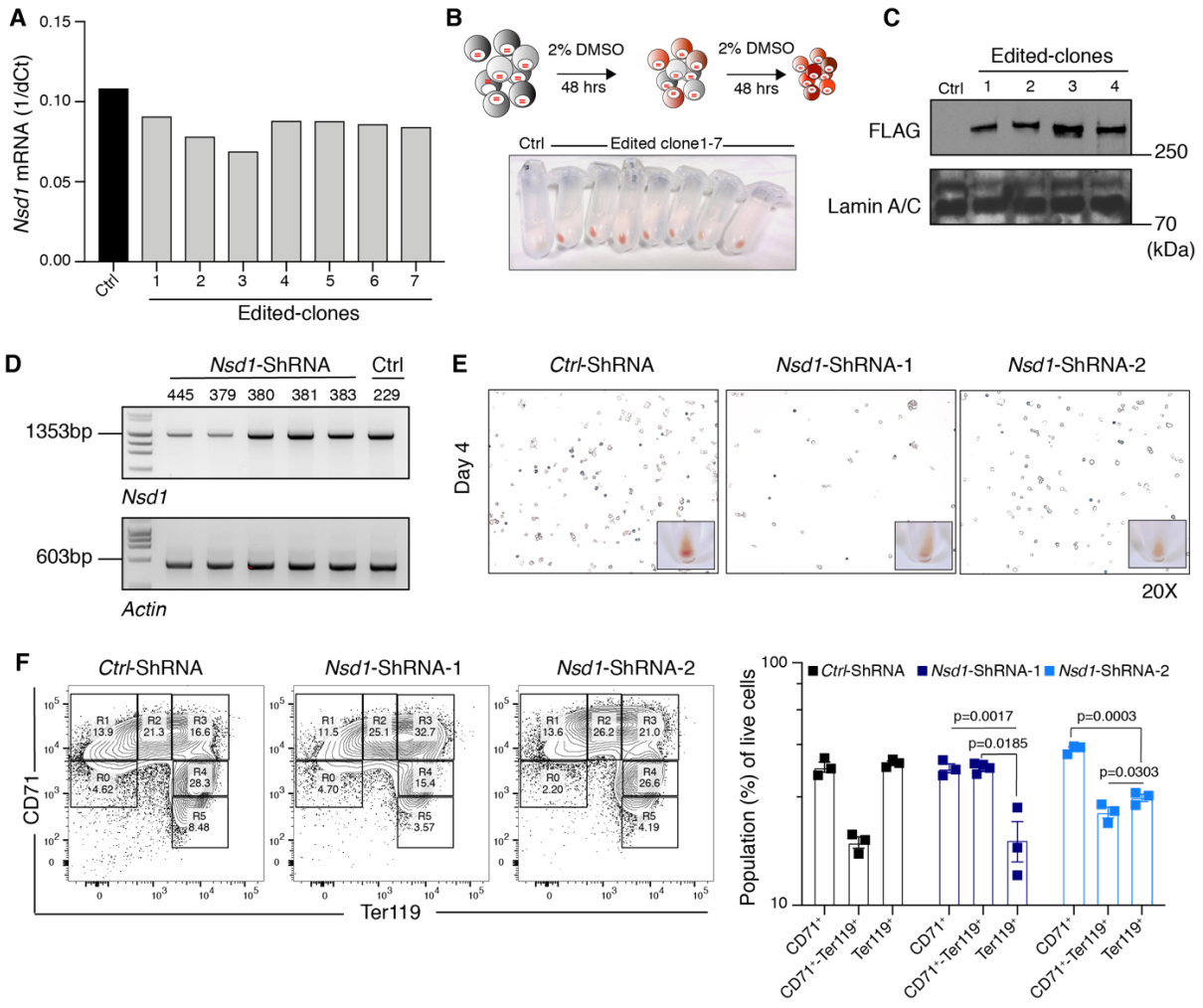



**Fig. 23: Knockdown of *Nsd1* in *Nsd1*-3x*Flag* edited MEL cells led to impaired erythroid differentiation**
**(A)** *Nsd1* mRNA relative expression measured by quantitative PCR in control unedited MEL cells (Ctrl) and *Nsd1*-3x*Flag* edited clones (clone 1-7) normalized to *Gusb* (n=1). **(B)** Schematic illustration of MEL cells 4 days erythroid inducing differentiation assay using 2% DMSO (Top). Picture of control un-induced cells (Ctrl), and DMSO induced *Nsd1*-3x*Flag* edited MEL cell clones (clone 1-7) showing hemaglobinized red cell pellets. **(C)** Western blot analysis using 30ug of nuclear extract proteins isolated from unedited (Ctrl) and *Nsd1*-3x*Flag* edited MEL cell clones (clone 1-4) using FLAG antibody (Cell Signaling). Lamin A/C was used as

loading control. **(D)** Agarose gel image of *Nsd1* PCR product performed on cDNA synthesized using 500ng total RNA extracted from edited MEL cells (Clone 4) transduced with *pLKO.1 Nsd1-ShRNA* (# 445, 379, 380, 381 & 383) and empty control (#229). Actin was used a loading control. **(E)** Representative images of benzidine staining of edited MEL cells (Clone 4) transduced with *pLKO.1 Nsd1*-ShRNA-1 and *Nsd1*-shRNA-2 (# 379 & 445 respectively) and empty control (#229) following 4 days of DMSO induction. Small inserts show cell pellets (20x). **(F)** Representative image of flow cytometry staining for transferrin receptor (CD71) and Ter119 marker of edited MEL cells (Clone 4) transduced with *pLKO.1 Nsd1*-ShRNA-1 & 2 (# 379 & 445 respectively) and empty control (#229) following 4 days of DMSO induction (Left). The gating strategy distinguishes different stages of erythroid maturation: "R0" fraction ($CD71^{-/low}$, $Ter119^{-}$), "R1" ($CD71^{+}$, $Ter119^{-}$), "R2" fraction ($CD71^{+}$, $Ter119^{+/low}$), "R3" fraction ($CD71^{+}$, $Ter119^{+}$), "R4" fraction ($CD71^{+/low}$, $Ter119^{+}$), and "R5" fraction ($CD71^{-}$, $Ter119^{+}$). Bar graph shows the quantification of different cell population following *Nsd1* Knockdown (n=3 per group). 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 A & F, was tested with paired two-tailed t- test.

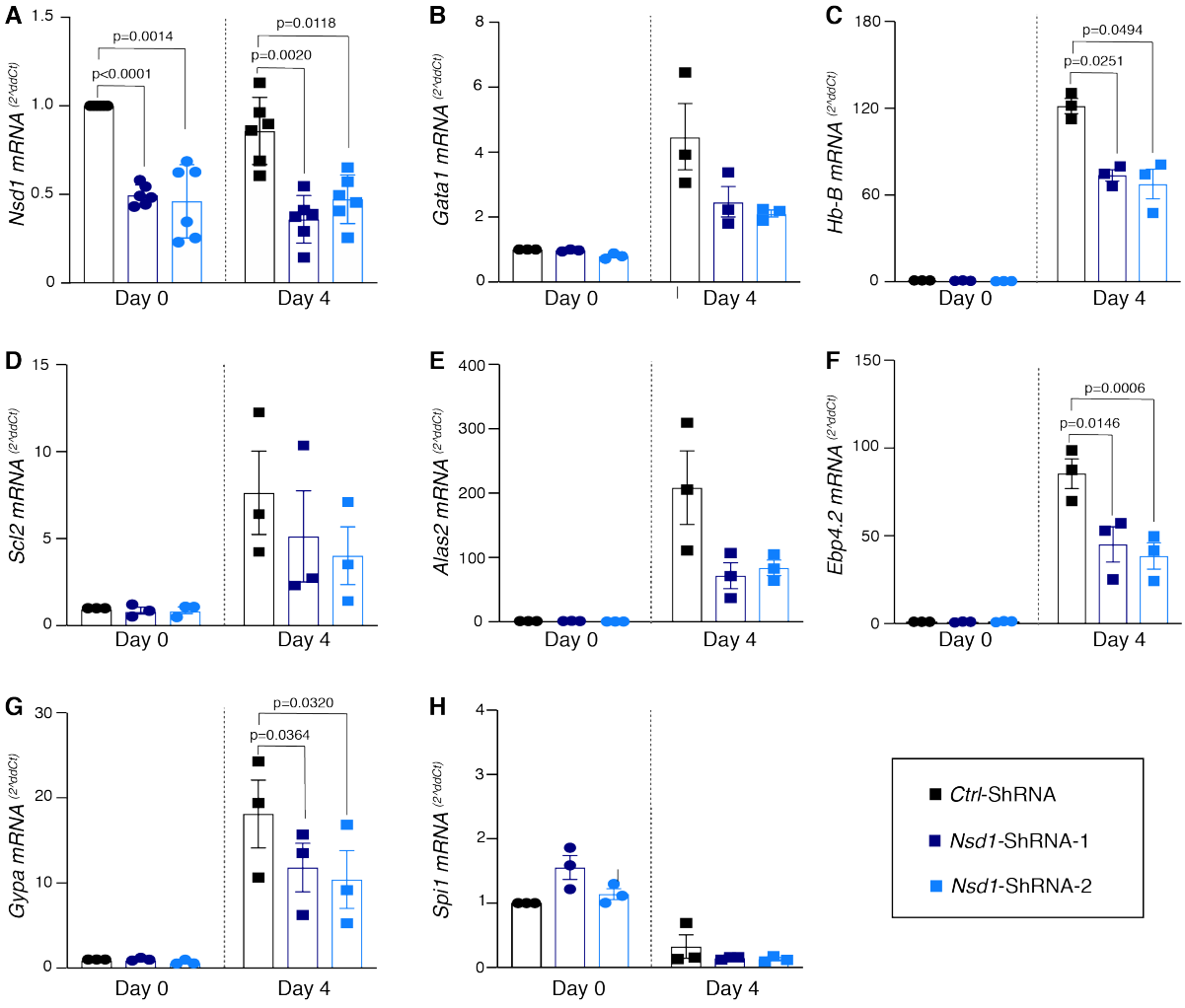



**Fig. 24*: Nsd1* knockdown impaired upregulation of erythroid differentiation associated genes**
**(A)** Fold change expression in *Nsd1* mRNA measured by quantitative PCR of edited MEL cells (Clone 4) transduced with *pLKO.1 Nsd1*-ShRNA-1 & 2 (# 379 & 445 respectively) and empty control (#229) before (Day 0) and following 4 days of DMSO induction (day 4). Data normalized to *Rnh1* (Ribonuclease inhibitor 1) and control cells (#229) at time Day 0 (n=6 per group). **(B)** Fold change expression change in *Gata1* mRNA measured by quantitative PCR of edited MEL cells (Clone 4) transduced with *pLKO.1 Nsd1*-ShRNA-1 & 2 (# 379 & 445 respectively) and empty control (#229) before (Day 0) and following 4 days of DMSO induction

(Day 4). Data normalized to *Rnh1* and control cells (#229) at Day 0 (n=3 per group). **(C-G)** Fold change expression change of erythroid differentiation associated genes mRNA measured by quantitative PCR of edited MEL cells (Clone 4) transduced with *pLKO.1 Nsd1*-ShRNA-1 & 2 (# 379 & 445 respectively) and empty control (#229) before (Day 0) and following 4 days of DMSO induction (Day 4). Data normalized to *Rnh1* and control cells (#229) at Day 0 (n=3 per group). **(H)** Fold change expression in *Spi1* mRNA measured by quantitative PCR of edited MEL cells (Clone 4) transduced with *pLKO.1 Nsd1*-ShRNA-1 & 2 (# 379 & 445 respectively) and empty control (#229) before (Day 0) and following 4 days of DMSO induction (day 4). Data normalized to *Rnh1* and control cells (#229) at Day 0 (n=3 per group). 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 A & F, was tested with paired two-tailed t- test.

**GE-mediated inactivation of *Nsd1* SET domain blocked induced erythroid differentiation of MEL-NSD1-3xFLAG cells**

We designed sgRNAs targeting Ex19 and 20 which together make up the SET domain of *Nsd1* protein to investigate the role of the methyltransferase activity of *Nsd1* in erythroid differentiation (**Fig. 25A**). In a preliminary experiment we sorted *sgRNA*-Ex19 and *sgRNA*-Ex20 transfected cells ($GFP^+$) in bulk and induced them with DMSO. Cells transfect with sgRNA targeting Ex19 or Ex20 were impaired in erythroid differentiation as shown by benzidine staining and cell pellets when compared to control transfected cells (**Fig. 25B**). Gene expression analysis showed decrease expression of *Gata1* and *Hb-B*, however *Nsd1* mRNA expression was not consistent with a knockout phenotype (**Fig. 25C**). We hypothesized that this could be due to the heterogenous mutational nature of the bulk sorted cells, thus, we repeated the experiment with single cell sorting. Hereby we observed a striking change in the transfected cells' size and morphology, such as the presence of abnormal large cells containing multiple nuclei, which were also completely blocked in erythroid differentiation (**Fig. 25D**). Gene expression analysis showed almost a complete loss of *Nsd1* expression at the targeted exons, coupled with significant depression in *Hb-B* expression (**Fig. 25E & F**).

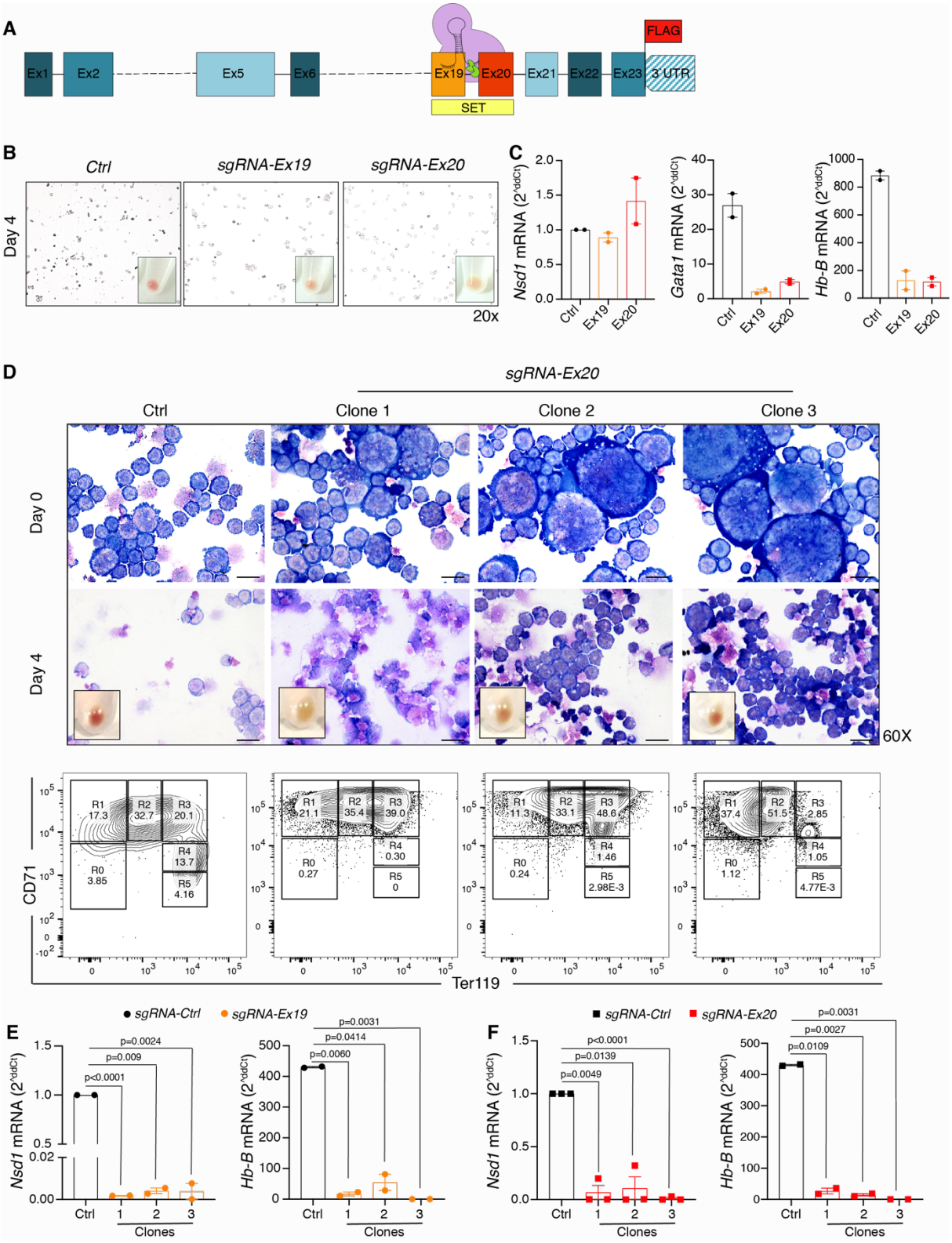



**Fig. 25: GE-mediated mutation of the *Nsd1* SET domain impaired chemically-induced erythroid differentiation of Nsd1-3xFLAG edited MEL cells**

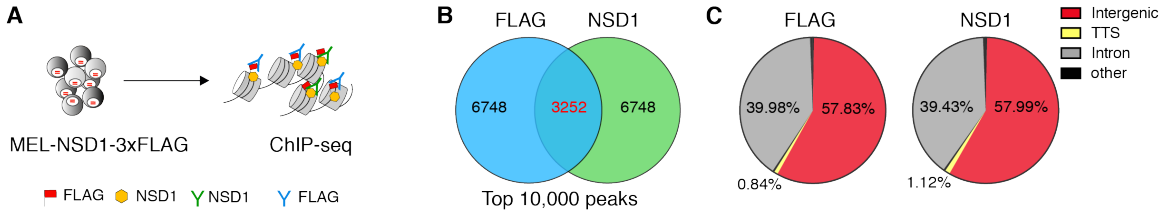
**(A)** Schematic illustration of the mutational strategy using CRISPR-Cas9 technology of edited *Nsd1* locus depicting the position of the targeted SET domain (Ex 19 &20). **(B)** Representative images of benzidine staining of GFP bulk-sorted edited MEL cells (Clone 4) transfected with GFP-sg-empty (Ctrl) or single guided RNA plasmid (sgRNA) targeting Ex 19 and Ex 20 following 4 days of DMSO induction. Small inserts show cell pellets (200x). **(C)** Fold change expression of *Nsd1*, *Gata1* and *Hb-B* mRNA measured by quantitative PCR of

GFP bulk-sorted edited MEL cells (Clone 4) transfected with *GFP-sg-empty* (Ctrl) or single guided RNA plasmid (sgRNA) targeting Ex 19 and EX 20 following 4 days of DMSO induction (n=2 per group). Data normalized to *Rnh1* and control cells (#229) at day 0. **(D)** Representative images (60x) of Wright-Giemsa stained cytospin preparations and cell pellets from GFP single cell-sorted edited MEL cells (Clone 4) transfected with *GFP-sg-empty* (Ctrl) or single guided RNA plasmid (sgRNA) targeting Ex19 and EX 20 before (day 0) and after (Day 4) DMSO-induced differentiation (Top). The lower panels show flow cytometric analysis of CD71 and Ter119 expression of the cells after 4 days in differentiation. The gating strategy distinguishes different stages of erythroid maturation: "R0" fraction ($CD71^{-/low}$, $Ter119^{-}$), "R1" ($CD71^{+}$, $Ter119^{-}$), "R2" fraction ($CD71^{+}$, $Ter119^{+/low}$), "R3" fraction ($CD71^{+}$, $Ter119^{+}$), "R4" fraction ($CD71^{+/low}$, $Ter119^{+}$), and "R5" fraction ($CD71^{-}$, $Ter119^{+}$). **(E)** Fold change expression of *Nsd1* and *Hb-B* mRNA measured by quantitative PCR of GFP single cell-sorted edited MEL cells (Clone 4) transfected with *GFP-sg-empty* (Ctrl) or single guided RNA plasmid (sgRNA) targeting EX19 following 4 days of DMSO induction (n=2 per group). Data normalized to *Rnh1* and control cells (#229) at day 0. **(F)** Fold change expression of *Nsd1* (n=3) and *Hb-B* (n=2) mRNA measured by quantitative PCR of GFP single cell-sorted edited MEL cells (Clone 4) transfected with *GFP-sg-empty* (Ctrl) or single guided RNA plasmid (sgRNA) targeting Ex20 following 4 days of DMSO induction. Data normalized to *Rnh1* and control cells (#229) at day 0. 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 a & F, was tested with paired two-tailed t- test.

**NSD1 occupied intragenic and/or intronic regions of genes associated with cardiac and neuronal function and development**

To better understand the role of NSD1 during erythroid differentiation, it is important to identify the genes and pathways it controls, which in turn could be deduced from its chromatin occupancy. We performed ChIP-seq analysis in un-induced MEL-NSD1-3xFLAG using a FLAG and a non-commercial anti-NSD1 antibody (established by Antoine Peters, FMI, Basel) (**Fig. 26A**). Restricting our analysis to the top 10,000 peaks for both antibodies, we found 3252 common peaks (**Fig. 26B**). NSD1 peaks distribution show extensive association with the intronic and intergenic regions of the MEL genome (**Fig. 26C)**. unlike TF binding profiles, which tend to show a narrow peak at specific genomic loci, we observed long stretches of NSD1 occupancy using both antibodies to regions spanning the Jak2 and Runx2 genes (**Fig. 26D & E**), both of which are targets of activating mutations in myeloproliferative disorders and leukemia[222, 223]. Analysis of the nearest gene transcription start site (TSS) for the common FLAG and NSD1 peaks, identified 450 genes (**Supplementary table S12**), which when tested for pathway analysis seem to mostly belong to pathways involved in cardiac function (**Fig. 26F**), cancer (**Fig. 26G**) and neuronal development (**Fig. 26H) (Supplementary table S13).** Unfortunately, we could not find any significant association with any pathways involved in hematopoiesis or erythroid differentiation. We hypothesized that this is probably due to the fact that we used undifferentiated MEL cells for our experiment. However, the fact that the detected pathways

belong to the most recognizable symptomatic abnormalities associated with Sotos syndrome such as congenital heart defects, mental retardations and increase incident of cancer suggest the functionality of our assay[118]. As these experiments served as proof of concept (n=1) they need to be validated and extended by comparing cells before and after induced erythroid differentiation.




**F**

| Pathways | p-value |
|---|---|
| GO_REGULATION_OF_CARDIAC_MUSCLE_CONTRACTION_BY_CALCIUM_ION_SIGNALING | 0.0050 |
| GO_REGULATION_OF_CARDIAC_MUSCLE_CONTRACTION_BY_REG_OF_THE_RELEASE_OF_SEQUESTERED_CALCIUM_ION | 0.0063 |
| GO_CELL_CELL_SIGNALING_INVOLVED_IN_CARDIAC_CONDUCTION | 0.0065 |
| GO_HIGH_VOLTAGE_GATED_CALCIUM_CHANNEL_ACTIVITY | 0.0065 |
| GO_CARDIAC_MUSCLE_CELL_CONTRACTION | 0.0100 |
| GO_DETECTION_OF_CALCIUM_ION | 0.0100 |
| GO_CELL_COMMUNICATION_INVOLVED_IN_CARDIAC_CONDUCTION | 0.0155 |
| GO_MEMBRANE_DEPOLARIZATION_DURING_CARDIAC_MUSCLE_CELL_ACTION_POTENTIAL | 0.0195 |
| IKEDA_MIR30_TARGETS_DN | 0.0237 |
| GO_VOLTAGE_GATED_CALCIUM_CHANNEL_ACTIVITY | 0.0261 |
| GO_VOLTAGE_GATED_CALCIUM_CHANNEL_COMPLEX | 0.0312 |
| GO_REGULATION_OF_RYANODINE_SENSITIVE_CALCIUM_RELEASE_CHANNEL_ACTIVITY | 0.0367 |

**G**

| Pathways | p-value |
|---|---|
| AGARWAL_AKT_PATHWAY_TARGETS | 0.0011 |
| DING_LUNG_CANCER_BY_MUTATION_RATE | 0.0102 |
| GO_C_TERMINAL_PROTEIN_AMINO_ACID_MODIFICATION | 0.0140 |
| MONTERO_THYROID_CANCER_POOR_SURVIVAL_DN | 0.0181 |
| LI_CISPLATIN_RESISTANCE_UP | 0.0195 |
| PEPPER_CHRONIC_LYMPHOCYTIC_LEUKEMIA_DN | 0.0212 |
| REACTOME_NEGATIVE_REGULATION_OF_THE_PI3K_AKT_NETWORK | 0.0261 |
| VICENT_METASTASIS_UP | 0.0466 |

**H**

| Pathways | p-value |
|---|---|
| GO_RETINA_LAYER_FORMATION | 0.0005 |
| GO_VOCALIZATION_BEHAVIOR | 0.0018 |
| GO_SYNAPSE_ASSEMBLY | 0.0063 |
| REACTOME_DOPAMINE_NEUROTRANSMITTER_RELEASE_CYCLE | 0.0065 |
| GO_NEUROMUSCULAR_JUNCTION_DEVELOPMENT | 0.0065 |
| GO_POSITIVE_REGULATION_OF_EXCITATORY_POSTSYNAPTIC_POTENTIAL | 0.0065 |
| GO_AXON_INITIAL_SEGMENT | 0.0100 |
| GO_PRESYNAPTIC_ACTIVE_ZONE | 0.0100 |
| GO_POSITIVE_REGULATION_OF_SYNAPTIC_TRANSMISSION_GLUTAMATERGIC | 0.0121 |
| GO_MODULATION_OF_EXCITATORY_POSTSYNAPTIC_POTENTIAL | 0.0131 |
| GO_PERIPHERAL_NERVOUS_SYSTEM_NEURON_DIFFERENTIATION | 0.0140 |
| GO_PERIPHERAL_NERVOUS_SYSTEM_NEURON_DEVELOPMENT | 0.0140 |
| GO_AXONAL_GROWTH_CONE | 0.0155 |
| GO_NEURONAL_POSTSYNAPTIC_DENSITY | 0.0234 |
| GO_POSITIVE_REGULATION_OF_SYNAPSE_MATURATION | 0.0261 |
| GO_NEGATIVE_REGULATION_OF_AXON_GUIDANCE | 0.0297 |
| ASTON_MAJOR_DEPRESSIVE_DISORDER_UP | 0.0309 |
| GO_RETINA_MORPHOGENESIS_IN_CAMERA_TYPE_EYE | 0.0376 |
| GO_REGULATION_OF_SYNAPTIC_TRANSMISSION_GLUTAMATERGIC | 0.0377 |
| GO_NEUROMUSCULAR_PROCESS_CONTROLLING_BALANCE | 0.0377 |
| GO_NEURAL_RETINA_DEVELOPMENT | 0.0466 |

*In collaboration with Cecile Thirant*

**Fig. 26: NSD1 occupied intragenic and/or intronic regions of genes associated with cardiac and neuronal function and development.**

**(A)** Schematic illustration of ChIP-seq experiment using edited MEL cells (MEL-Nsd1-FLAG). NSD1 was immunoprecipitated from undifferentiated MEL cells using either a polyclonal NSD1 or a monoclonal anti-FLAG antibody. **(B)** Venn diagram showing the number of peaks overlapping NSD1 and FLAG mapped peaks limited to the top 10,000 peaks for each condition. Further analysis of these 3252 common peaks revealed 450 common genes between the two conditions. **(C)** Venn diagrams showing the distribution of NSD1 and FLAG peaks on the genome. Analysis of binding capacity of NSD1 and FLAG antibodies indicates that NSD1 mostly bind to intragenic and intronic part of the genome, with around only 1% of peaks fall within the transcription termination site (TTS). **(D-E)** Representative image of integrative genomic viewer (IGV) showing the overlap in peaks between NSD1 and FLAG on chromosome 19 (spanning the *Jak2* gene- Left), and on chromosome 17 (spanning the *Runx2* gene- Right). **(F-G)** Significantly associated gene pathways signatures as determined by common peak/gene analysis between NSD1 and FLAG antibodies. Pathways were groups into 3 distinct categories, those that control/affect cardiac development and function (**F**), those that are associated with cancer development and resistance **(G)**, and finally those that regulate neuronal development and function **(H)**. Significance level at $p<0.05$.

## 3.4. Discussion

The motivation for functionally studying the role of NSD1 in erythroid differentiation was based on the fact that genetic inactivation of *Nsd1* during mouse fetal hematopoiesis leads to the development of a fully penetrate erythroleukemia-like disease in adult mice. Hereby we noted that *Nsd1*$^{-/-}$ FL- and BM-derived hematopoietic progenitor cells are impaired in erythroid maturation despite expressing high protein levels of the erythroid transcriptional master regulator GATA1. Unexpectedly, exogenous overexpression of *Gata1* seemed to rescue *in vitro* maturation of *Nsd1*$^{-/-}$ BM-derived erythroblasts.

By performing genetic complementation experiments combined with functional assays, we show that the block in *Nsd1*$^{-/-}$ BM-derived erythroblasts differentiation could be overcome by expression of wild-type *Nsd1*, but not a catalytically inactive SET-domain mutant (*Nsd1*$^{N1918Q}$). Mechanistically, we found that during *in vitro* erythroid differentiation, NSD1 activity affected chromatin binding affinity, and target gene activation of the erythroid master regulator GATA1.

**Genetic complementation experiments to rescue impaired erythroid differentiation of *Nsd1*$^{-/-}$ erythroblasts**

To understand how NSD1 controls erythroid maturation, we initially aimed for simply functionally compare BM-derived *Nsd1*$^{fl/fl}$ to *Nsd1*$^{-/-}$ erythroblasts. This however proved to be unviable due to major shortages: firstly, while *Nsd1*$^{-/-}$ BM-derived erythroblasts showed unlimited exponential expansion capacity in *in vitro* cultures (> 3 months). In contrast, consistent with previous reports[188], we could not expand *Nsd1*$^{fl/fl}$ BM-derived cells beyond 3 weeks. Secondly, unlike the uniformed proerythroblastic nature of the *Nsd1*$^{-/-}$ cells (Ki$t^{+}$, CD71$^{+}$), *Nsd1*$^{fl/fl}$ cells always expressed a combination of erythroid and myeloid (Kit$^{+}$, CD71$^{+}$, Mac-1$^{+}$, Gr1-1$^{+}$) markers, even when cultured for >1 week in the erythroid MM (**Fig. 10**). Sorted CD71$^{+}$ *Nsd1*$^{fl/fl}$ erythroblasts from these cultures spontaneously lost Kit$^{+}$ expression, and failed to expand further in MM. This observation suggested that growth of wildtype erythroblasts in our culture is mediated by the proliferation of these ''multi-lineaged'' cells. These limitations made it impossible to obtain sufficient "truly" erythroid cells to perform further experiments (**Fig. 11)**. Although improved culture conditions supporting erythroid differentiation are constantly reported, most of them were developed to support *ex vivo* erythroid potential of human rather than murine cells[224-227]. Furthermore, switching to a

new culture condition would affect the compatibility between our newly generated data and our previous observations. Therefore, to circumvent these obstacles, we chose a genetic complementation approach by virally expressing WT or a previously reported catalytically-inactive *Nsd1*$^{N1918Q}$ SET-domain mutant[93] .

Epigenetic dysregulation is frequently found in cancer cells and has been also implicated in many hematological malignancies, such as MDS, and leukemia[228]. NSD1 SET-domain is responsible for the catalytic activity of the protein, and is frequently mutated in patients with Sotos syndrome[135, 229]. In leukemia, the transforming potential of the AML-associated NUP98-NSD1 fusion was dependent on the integrity of the NSD1 SET-domain activity. Deletion of the SET-domain of NSD1, or inactivating it with single base pair substitution [replacement asparginine (N) to glutamine (Q) at amino acid position 1918 (*Nsd1*$^{N1918Q}$)], abolished H3K36 methylation, *HoxA7* promoter activations, and immortalization of myeloid progenitors[93]. We therefore reasoned that comparing *Nsd1*$^{-/-}$ BM-derived cells expressing *Nsd1* or *Nsd1*$^{N1918Q}$ SET-domain mutant would be the best approach to link erythroid differentiation to methyltransferase activity of NSD1.

NSD1 is large protein consisting of 2696 aa, coded by around 8300 DNA base pairs. Unfortunately, generation and expansion of stably transduced BM-derived cells expressing WT or SET-domain mutant *Nsd1* was limited by the large size of *Nsd1* cDNA, which also led to low level of transgene expression. In the best-case scenario using concentrated viral particles, we only achieved to transduce between 0.5-2% of primary mouse proerythroblasts. In addition, on several occasions, *Nsd1* transduced cells did not survive the sorting procedure, and all sorted cells died within the first 48 hours. We reason that this is probably due to the fact that on *Nsd1* expression, the cells adapt a more ‘’WT’’ phenotype, making them more fragile, and less disposed for survival. Nevertheless, in total, we managed to generate and expand cells from six independent transductions, isolated from six different mice, which we subsequently used to generate all the discussed data.

We noted that despite rather low levels of exogenous *Nsd1* expression, it appeared sufficient to restore *in vitro* erythroid maturation of *Nsd1*$^{-/-}$ cells. *Nsd1* transduced cells displayed several hallmarks associated with erythroid maturation, with the expected changes in cellular morphology and size, shift in erythroid cell markers expression, and cell cycle arrest (**Figs. 12-14**). Interestingly, *Nsd1*$^{-/-}$ BM-derived erythroblasts transduced with the SET-domain mutant had an indistinguishable phenotype to that of empty transduced cells, which seems to suggest that the rescue of erythroid differentiation is solely dependent on the

methyltransferase activity of *Nsd1*. However, since we did not include other *Nsd1* domain mutants, such NID, PHD or PWWP, in our investigation, we cannot with certainty conclude on their functional requirement during the erythroid differentiation. For the future, it would be interesting to create single and double *Nsd1* domain mutants, and to investigate how they affect erythroid differentiation.

### *Nsd1* expression in *Nsd1*$^{-/-}$ BM-derived erythroblasts enhanced GATA1 activity without modulating its expression

Erythroid differentiation is highly regulated at the level of mRNA transcription with instant modulation of numerous erythroid genes following induction. To gain an insight into the molecular mechanisms driving NSD1 control on erythroid differentiation, we compared the gene expression profiles of WT and SET-domain mutant expressing *Nsd1*$^{-/-}$ BM-derived erythroblasts, prior (0h) and one day (24h) after differentiation. We decided to use the earlier time point of 24h, rather than 48 or 96, in an attempt to decrease the noise in our transcriptomic analysis from highly expressed genes, such Hb-B, and thus, be able to detect the changes in the expression of erythroid regulator genes, rather than responders (**Fig. 15**). Despite not detecting any significant change in gene expression during cell expansion (0h), GSEA pathway correlation indicated that even at this stage, *Nsd1* expressing cells seemed already primed for erythroid differentiation compared to cells expressing the SET-domain mutant. This in turn implies that although, *Nsd1* and *Nsd1*$^{N1918Q}$ expressing cells appear phenotypically compatible during expansion, the underlying changes that allow one to differentiate while the other remain impaired have already occurred.

We found significant changes in gene expression between *Nsd1 and Nsd1*$^{N1918Q}$ expressing cells after 24h of differentiation induction in genes associated with erythroid differentiation, such as *SetD8* (upregulated), and *Gata2* (downregulated). *Gata2* is normally highly expressed in HSCs and some MEPs, and during erythropoiesis, *Gata2* downregulation by GATA1 is a sign that cells are committing to the erythroid lineage[230]. SetD8 is a H4K20 mono methylase enzyme, which was proposed to physically interact with GATA1 and play a role in *Gata2* downregulation during differentiation[175].

GSEA analysis revealed that in addition to positive link with erythroid lineage development and differentiation, more specifically, *Nsd1* expressing cells showed significant positive association with GATA1 targets expression[198]. This was surprising, as *Gata1* expression was not significantly altered between the differently transduced cells (**Fig. 15**). We

obtained similar results when we compared total protein content of *Nsd1* and *Nsd1*$^{N1918Q}$ expressing cells after 24h, where we could detect significant upregulation in proteins associated with erythroid differentiation such as Hb-B, XPO7 and EPB42, while GATA1 protein levels did not significantly change (**Fig. 16**). Thus, we speculated that *Nsd1* expression promotes erythroid differentiation by primarily augmenting GATA1 functional activity and not its expression.

**_Nsd1_ expression altered histone methylation patterns in _Nsd1_$^{-/-}$ BM-derived erythroblasts upon differentiation**

Erythroid differentiation is characterized by extensive remodeling of histones variants and posttranslational modifications leading to gene expression regulation[231]. In addition to marking genomic regulatory elements (promoter & enhancers) for TF binding and gene expression initiation or silencing, histone modification along gene bodies regulate gene expression level by modulating RNA Pol II transcription elongation efficiency[232]. In our model, *Nsd1* expression in *Nsd1*$^{-/-}$ BM-derived erythroblasts caused a change in histone methylation patterns (**Fig 17**), characterized by an increase in marks associated with active transcription (H3K36me1/2, H2K20me1/2, & H3K27me1), together with a decrease in silencing associated modification (H3K27me3)[76].

As stated previously, *Nsd1* catalyzes the mono and di-methylation of H3K36 at promoter regions and gene bodies, thus we concluded that the increase in these two marks is most likely a direct result of NSD1 activity. However, both NSD1 and SetD8 were reported to catalyze H4K20me1, and since we also detect an increase in *Setd8* expression in our transcriptomic data, we cannot conclude whether in this case, H4K20 methylation is a direct consequence of NSD1 activity, or a secondary effect of *SetD8* upregulation. Nevertheless, in addition to marking gene for activation, H4K20me1/2 are also important for cell cycle regulation, chromatin condensation and mitotic progression[233]. A marked accumulation in H4K20me1/2 is essential for effective erythropoiesis, as loss of *Setd8* expression in mice is embryonically lethal due to severe anemia[234]. *SetD8* null erythroblasts maturation is profoundly dysregulated, characterized by defects in transcriptional repression, cell cycle arrest, nuclear condensation, and loss of heterochromatin[234]. Similar to our *Nsd1* inactivation model, *SetD8*$^{-/-}$ erythroblasts impaired erythroid differentiation was also coupled to GATA1 expression level comparable to control cells, suggesting a common mechanism. However, the fact that several studies could not confirm earlier in vitro H4K20me activity of

NSD1[235] suggests that the increase of H4K20me1 results from increased expression of SetD8.

H3K27me2/3 is a catalyzed by the Polycomb repressive complex 2 (PRC2), which is composed of Suz12, Eed, and the SET-domain containing EZH1/2 proteins. Notably, both NSD1 and EZH2 germline mutations are detected in patients with overgrowth syndromes with increased incidents of leukemia[236]. In addition, somatic mutations in *Ezh2* gene causing dysregulation of H3K27 methylation were detection in various hematological malignancies, including MDS and leukemia[237]. Also, similar to NUP98-NSD1 leukemic transformation mechanism, defect in repressing genes associated with self-renewal and proliferation (e.g *HoxA9*) was suggested to be the underlying mechanism of *Ezh2* malignant hematopoiesis[238, 239]. In addition, NSD2 overexpression due to the t(4;14) in multiple myeloma cells was reported to results in wide spread of H3K36me2 and thus modulation of *Ezh2* binding at specific genomic loci, deriving the malignant transformation [240]. A recent study in mESC reported that H3K36me2 co-localize with H3K27me2 and that loss of *Nsd1* results in genome-wide expansion of H3K27me3 at PRC2 controlled genes. Thus, *Nsd1* was proposed as a key modulator of PRC2 function required for regulating the segregation of genome-wide H3K27me2 and H3K27me3 domains in ESCs (**Fig. 27A**)[241]. Previous reports have shown that PU.1 blocks GATA1 activity in MEL cells by creating repressive chromatin structure containing H3K9me3 and heterochromatin protein 1 (HP1)[242]. Notably, we do not detect differences in H3K9 methylation status between *Nsd1* and *Nsd1*$^{N1918Q}$ expressing cells prior or following differentiation (**Fig. 17D**), consolidating our conclusion that *Nsd1* acts via a PU.1 independent mechanism to modulate GATA1 function, leading to erythroid differentiation impairment.

In addition to histone modification cross-talk, another recent study in mESC has shown that *Nsd1*-mediated H3K36me2 is required for maintaining CpG DNA methylation at intergenic regions by DNMT3A[243]. Loss of NSD1 and NSD2 caused a redistribution of DNMT3A to H3K36me3 at gene bodies, and intergenic regions hypomethylation, mediated by its PWWP domain dual recognition function (**Fig. 27B**). Accordingly, blood samples from patients with Sotos syndrome and NSD1-mutated tumors also exhibit hypomethylation at intergenic DNA. We therefore speculate that *Nsd1* regulates gene expression during development and terminal differentiation, by creating a multi-layered context-dependent epigenetic landscapes, through the cross-talk between various histone codes and DNA methylation. In support of this hypothesis, initiation of erythroid differentiation is associated drastic loss of DNA methylation at promoter regions, allowing direct TF binding, and initiation of gene

transcription[244]. However, a recent genome-wide methylation study in human primary erythroid cells reported an inverse association between DNA methylation status at promoter and gene body regions in late stages of erythropoiesis. The study highlights that hypomethylation at promoters was associated with increased gene expression, whereas, hypomethylation at gene bodies was associated with decreased expression of genes[203]. Moreover, analysis of TF binding sites affected by hypomethylation and differential gene expression revealed that a large proportion of these sites were situated in gene bodies, thus suggesting a transcriptional regulatory function for intergenic regions[245].

In order to decipher the cryptic histone code modifications underlying *Nsd1*$^{-/-}$ phenotype, we need to perform more detailed analysis of the genome-wide overlap between all epigenetic histone marks, DNA methylation, and chromatin accessibility, upon the restoration of *Nsd1* the expression, and delineate their combinational effect on gene expression and erythroid differentiation. However, as discussed previously, our current genetic experimental system is hindered by limited cell numbers. Thus, alternative strategies need to be employed, such as *in vitro* conditional manipulation of NSD1 expression via shRNA or CRISPR-interference (CRISPRi)[246], or by generating an *in vivo* conditional rescue mouse model.

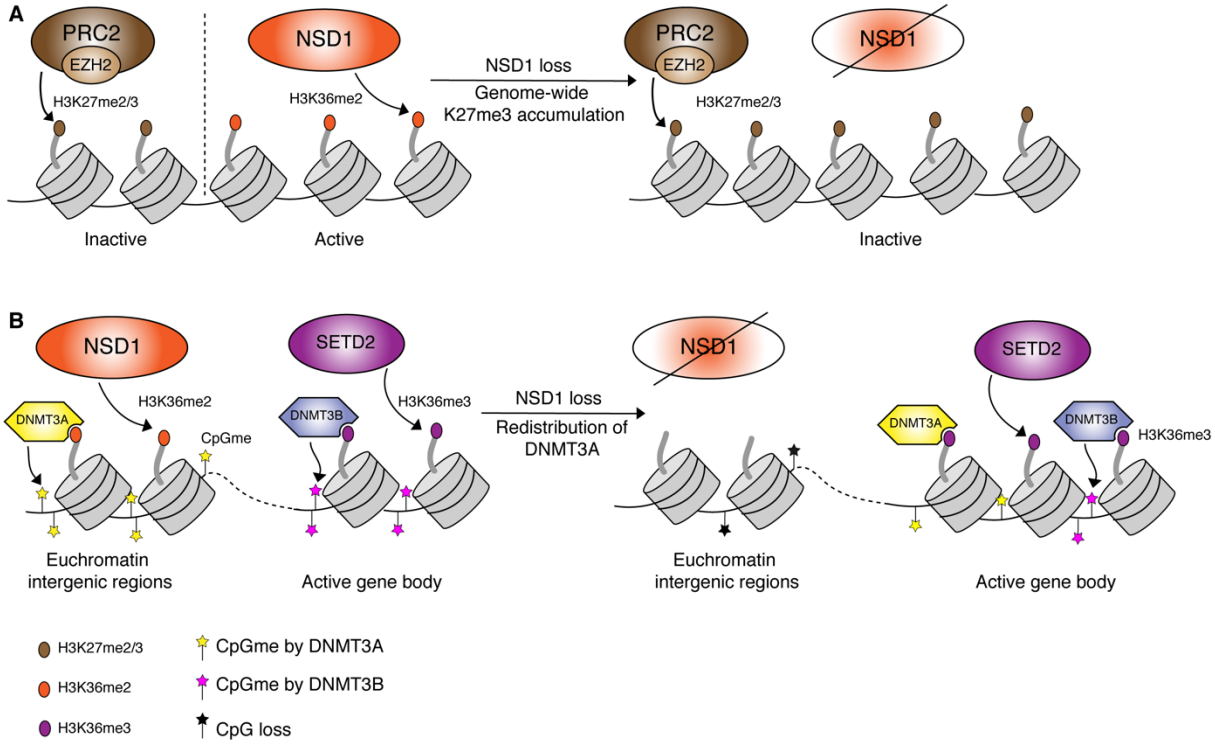


**Fig. 27: Models depicting the changes proposed in chromatin and DNA methylation landscape upon loss of NSD1. (A)** NSD1-mediate H3K36 di-methylation sequesters PRC2-EZH2 mediated H3K27me2/3 repression marks to inactive genomic loci. Loss of NSD1 expression leads genome-wide spreading of H3K27me3 and altered regulation of gene expression**. (B)** DNMT3A and DNMT3B act in parallel to methylate CpG dinucleotides at H3K36me2-enriched intergenic and H3K36me3-enriched genic regions, respectively. Loss of NSD1 activity causes a depletion of intergenic H3K36me2 levels, PWWP-mediated intergenic recruitment of DNMT3A, leading to hypomethylation of intergenic DNA and the redistribution of DNMT3A to H3K36me3- enriched gene bodies (Adapted from [241, 243]).

### *Nsd1* expression increased GATA1 binding to chromatin upon erythroid differentiation

To study the functional activity of GATA1, and to try and understand the differential dynamics of GATA1 transcriptional transactivation between *Nsd1* and *Nsd1*$^{N1918Q}$ expressing cells, we analyzed GATA1 occupancy, and regulatory histone marks (H3K36me3 & H3K27ac) prior, and upon erythroid differentiation. Although NSD1 does not directly catalyze H3K36 tri-methylation, a previous study showed that reduction in NSD1 expression led to global reduction in all three forms of methylated H3K36 within the body of the bone morphogenic protein 4 (BMP4) gene in human colorectal cancer cells[103]. Moreover, H3K36me3 marks along gene bodies are associated with transcription elongation, prevention of cryptic start sites, as well as pre-mRNA splicing and processing[98]. On the other hand, H3K27ac at promoters and distal enhances is associated with gene activation and transcription. Thus, we speculated that differential changes in these two marks between *Nsd1* and *Nsd1*$^{N1918Q}$ expressing cells, should positively correlate with gene expression, and GATA1 transcriptional activity.

Overall, we did not detect significant changes in GATA1, H3K36me3, or H3K27ac during cell expansion (0h). However, GATA1 occupancy to the chromatin showed significant increase in *Nsd1* transduced cells when grown under differentiation-inducing conditions, coupled with changes in H3K27ac and H3K36me3 marks (**Fig.18**). Surprisingly, we did not detect any 'new' binding sites for GATA1 protein upon differentiation, yet GATA1 recruitment became significantly enriched to sites where it was already bound. A recent study described similar DNA binding dynamics for another erythroid regulatory protein, LDB1. Genome wide binding analysis of LDB1 in MEL cells before and after the induction of terminal erythroid differentiation, showed that LDB1 was already recruited to the majority of its controlled genomic loci. Prior to differentiation induction, these genes are expressed at low levels, however, upon differentiation they become strongly induced. Thus, LDB1 complex binding seems to maintain an erythroid-specific gene expression program primed for rapid activation

when differentiation is induced. Mechanistically, the study suggested that in undifferentiated erythroid progenitors, low expression levels of LDB1-activated late erythroid genes is due to LDB1 interaction with the transcriptional repressor complexes ETO2/IRF2BP2 and NCoR1/SMRT. However, once differentiation is induced, LDB1 becomes dynamically associated with coactivators, leading to a boost in erythroid gene expression[247]. Thus, multimeric regulatory protein complexes display a dynamic interplay between activating and repressing components that determines lineage-specific gene expression and cellular differentiation**.** As both GATA1 and LDB1 perform in the same context and were reported to interact within the same complex during erythroid differentiation, we reasoned that GATA1 transcriptional activity might be similarly controlled. i.e. in the absence of *Nsd1*, GATA1 might be functionally trapped within repressing interactors that limit its transactivation potential, which can be overcome by either expression of additional "free" GATA1, or by switching the functional identity of its interaction partners.

**GATA1 is associated with transcriptional repressors**

Our observations suggest that loss of *Nsd1* SET-domain activity has a detrimental effect on the GATA1 transactivation activity during early erythroid differentiation. Comparative proteomic analysis of GATA1 bound proteins in $Nsd1^{-/-}$ BM-derived erythroblast expressing *Nsd1* or $Nsd1^{N1918Q}$, revealed the presence of multiple proteins associated with co-repressor activity, such as ETO2, NcoR1, MBD2 and HDAC3 in SET-domain mutant transduced cells (**Fig. 19**). Remarkably, in addition to these regularly reported GATA1 complex interactors, we also found highly significant reduction in the association of GATA1 with the transcriptional co-repressor SKI in wild-type *Nsd1* expressing cells. SKI was discovered as cellular homologue of the transforming oncogene *v-SKI* found in the genome of multiple acutely transforming avian leukosis retroviruses[248]. SKI expression is mainly localized in the nucleus, and in addition to inducing cell transformation when overexpressed[249], SKI was also found to be overexpressed in certain types of tumors such as melanomas, esophageal carcinomas, and leukemias[250]. In AML, the highest SKI expression was detected in monosomy 7 or deletion 7q (-7/del7q) AML [251], due to loss of *Ski*- targeting *miRNA29a* sequence on ch7q32[252]. *Ski* expression in AML was proposed to repress retinoic acid receptor (RAR) and RUNX1-mediated transcription, thereby leading to a differentiation block and leukemia[253]. Enforced expression of SKI in mouse HSCs and progenitor cells reduced HSC fitness and differentiation,

resulting in myeloproliferative disorder[254]. SKI functions as a transcriptional coregulator, as it only binds to DNA in a complex with other transcriptional factors (**Fig. 28**).

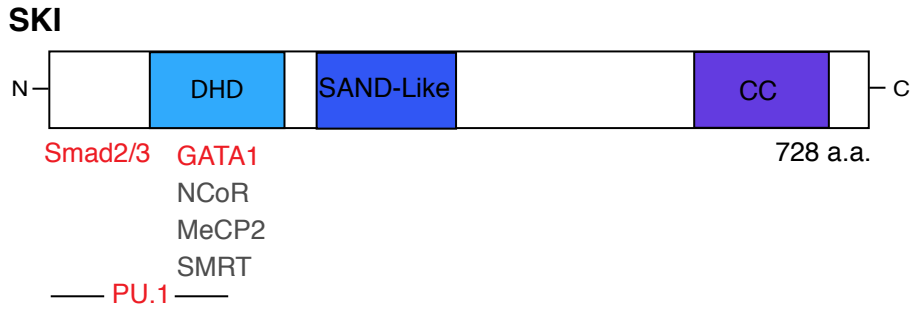



**Fig. 28**: Molecular structure, and protein–protein interactions of the SKI. The domains that define SKI proteins include: Dachshund homology domain (DHD), SAND-like domain, and coiled-coil domain (CC). Known regions of interaction with some partners for SKI are indicated. Transcriptional factors are shown in red (Smad2/3, GATA1 & PU.1), transcription regulators in grey (NCoR, MeCP2 & SMRT. Adapted from ([250]).

Multiple studies have shown that *Ski* negatively regulates the expression of the TGF-β pathway by disrupting the formation of R-SMAD/SMAD4 complexes, as well as by inhibiting SMAD association with the p300/CBP coactivators. In the absence of TGF-β, SKI, and its homologue SnoN, interact with SMAD4 to inhibit the expression of TGF-β target genes (e.g. *Smad7* and *Skil*), by recruiting other corepressors such as HDACs, NcoR1 to their promoters[255]. TGF-β signaling is essential during development, and was reported to play a dual function during erythroid differentiation; where it enhances the maturation of committed cells[256], together with inhibiting the proliferation of immature progenitors[257].

SKI was also reported to bind and co-repress the activity of both PU.1 and GATA1, in an *in vitro* cell models, blocking macrophage and erythroid cells differentiation respectively[258, 259] . However, whereas SKI blocks PU.1-induced transcriptional activation by mediating interaction between PU.1 and HDAC3, it was proposed to block GATA1 function by physically preventing GATA1 binding to DNA. Although we did not detect any significant upregulation in *Ski* gene expression in SET-domain mutant expressing cells, there was an apparent upregulation in SKI protein during induced erythroid differentiation (**Fig. 15B & 19C**). Reducing SKI levels, using *Sh-RNA* mediated silencing, in *Nsd1*$^{-/-}$ BM-derived erythroblast, restored the erythroid maturation, without the requirement for *Nsd1* expression (**Fig. 20**). Thus, we suggest that erythroid maturation of *Nsd1*$^{-/-}$BM-derived erythroblasts with reduced SKI expression, is a consequence of increasing GATA1 chromatin binding, leading to the activation of erythroid transcriptional program. Additionally, we could also speculate that increased TGF-β expression, due reduced SKI expression, plays an additive effect in deriving

effective erythroid maturation. To investigate the validity of this hypothesis, we need to assess the effect of coupling SKI reduction with inhibiting TGF-β signaling on erythroid differentiation of *Nsd1*$^{-/-}$ BM-derived erythroblast. Considering all of our observations together, we propose a model where NSD1 catalytic activity is essential for creating an active genome-wide epigenetic environment facilitating GATA1 binding and transcriptional activation (**Fig 29**).

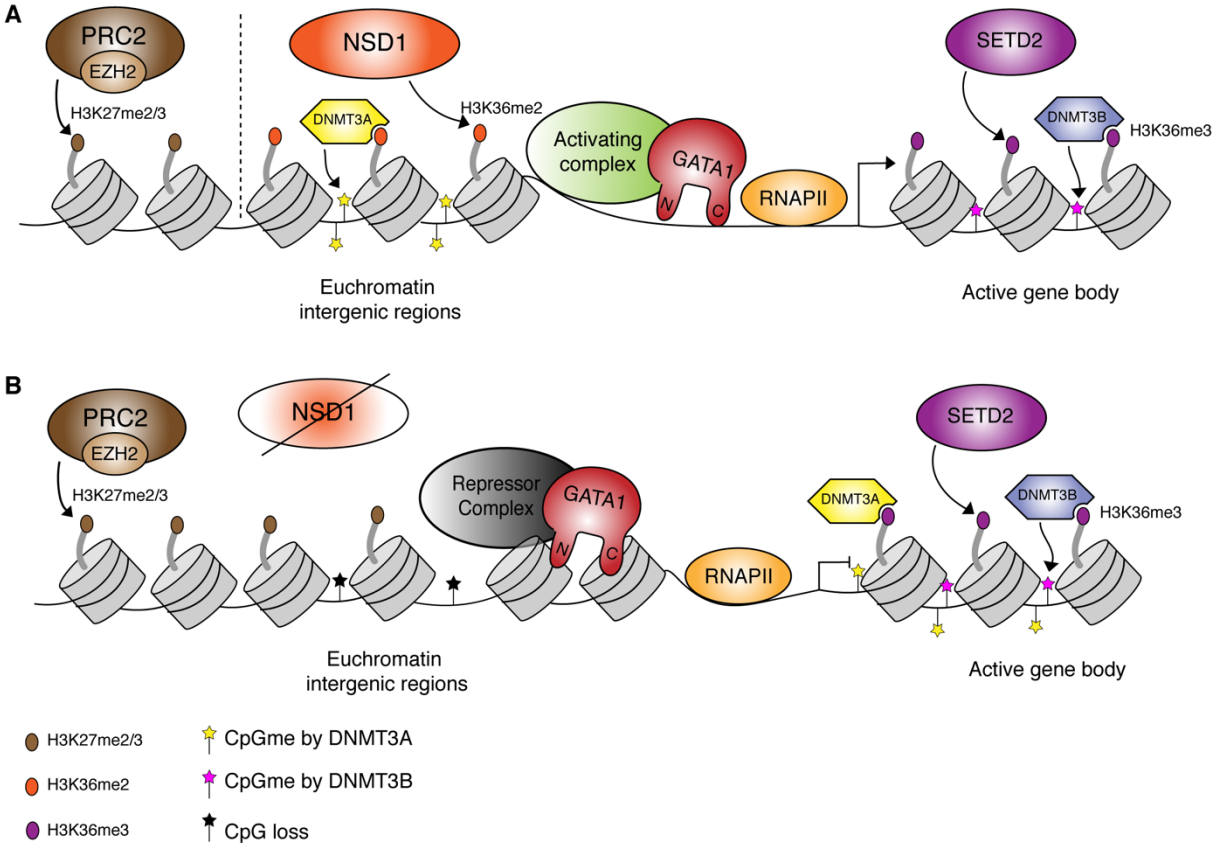



**Fig. 29: Working model of NSD1-mediated methylation in modulating gene transcription during erythroid differentiation.** (**A**) During normal hematopoiesis, NSD1 catalytic activity incites the formation of differentiation promoting histone and DNA methylation patterns, supporting the formation of GATA1 activation complex, and erythroid genes transcription upregulation. (**B**) Complete Loss of NSD1 or its SET domain catalytic activity, causes an improper modulation of chromatin and DNA modification, and tethering of GATA1 transcriptional activity through corepressors interactions.

Throughout performing this study were faced several limitations such as: unequal culture of WT and *Nsd1* KO primary cells, low rate of viral transduction, obtaining adequate cell numbers, and unavailability of a commercial ChIP- or IP grade mouse NSD1 antibody. As an alternative, we decided to genetically engineer MEL cells to express a FLAG tag sequence at the endogenous C-terminus of *Nsd1*, and investigate its genome occupancy by ChIP-seq analysis.

**_Nsd1_ inactivation in MEL cells phenocopied the erythroid differentiation defect of primary _Nsd1_$^{-/-}$ erythroblasts**

MEL is an immortalized erythroid progenitor cell line isolated from the spleen cells of mice infected with the *Friend* virus[53]. MEL cells are a well-established *in vitro* model of erythropoiesis as they represent early erythroid precursors (proerythroblast or CFU-E cells) arrested in development but partial differentiation in the presence of differentiating chemical agents, such as DMSO[190]. Similarly to our *Nsd1*$^{-/-}$ BM-derived erythroblasts model, the erythroid differentiation block of MEL cells is coupled with constitutive abundant levels of GATA1 protein[260]. Decades of extensive investigations resulted in a model where the MEL cell differentiation block is largely attributed to the integration of murine leukemia virus (*F-MULV*) into the *Sfpi1* locus, causing an aberrant activation of the myeloid master regulator PU.1[261]. PU.1 overexpression seems to antagonize GATA1 transcriptional function also by physical interaction, leading to erythroid differentiation arrest[262]. Interestingly, similar to our *Nsd1*$^{-/-}$ cells, activation of exogenous GATA1 in some MEL cell clones have been shown to induce erythroid differentiation [263]. Importantly, similar to *Nsd1*$^{-/-}$ cells, reducing *Nsd1* expression with *ShRNA*-mediated silencing, or inactivating its SET-domain by CRISPR-Cas9, in MEL cells led to marked impairment in their erythroid differentiation, independently from PU.1 (*Spi1*) expression (**Fig. 21-24**).

However, what was more surprising, is that decline/loss of *Nsd1* in MEL cells has somehow led to the generation of large multinucleated cells resembling megakaryocytes (**Fig. 25D**). As we have not performed further experiments to validate the phenotype, such as CD41 marker expression or polyploidy analysis, we cannot be certain of the lineage identity of these cells. Nevertheless, previous studies have shown that MEL cells can differentiate into erythroid and megakaryocytic cells in response to HMBA treatment[264]. Furthermore, both the erythroid and myeloid lineages are controlled by the master regulator GATA1, and that overexpression of GATA1 in G1ME (*Gata1*$^{-/-}$ megakaryocyte-erythroid) cells induced both erythroid and megakaryocytic differentiation[265]. Moreover, in humans, germline *Gata1* mutations are associated with block in erythroid and megakaryocytic maturation resulting in congenital anemia and thrombocytopenia[182]. Somatic mutations in *Gata1* locus resulting in the expression of the short form of the protein (GATA1-S) collaborate with trisomy 21, leading to excess megakaryocytes proliferation and acute megakaryoblastic leukemia[266]. Accordingly, *Nsd1*$^{-/-}$ mice have a marked decrease in platelets count, and are usually presented with thrombocytopenia (not shown). These observations further connect NSD1 catalytic

function in modulating GATA1 activity, and it suggests that this regulation is critical for the proper differentiation of both erythroid and myeloid cell lineages.

**NSD1 chromatin binding in undifferentiated MEL cells corresponded with gene pathways related to Sotos syndrome**

In addition to being altered in multiple types of solid cancers and leukemia, putative loss of function mutations in the *Nsd1* gene are a hallmark for the overgrowth Sotos and Weaver syndromes[118]. NSD1 can act as positive and negative transcriptional co-regulator most likely by binding different nuclear receptors including RAR, RXR, ER, TR and AR[208] .However, little is known about the specific gene/protein NSD1 targets during development and oncogenesis. Using ChIP-seq technology, and a MEL cell line engineered to express an endogenous FLAG epitope tag at the 3' end of the mouse *Nsd1* gene (**Fig. 21 &22**), we identified a number of candidates *Nsd1* regulated pathways (**Fig. 26**). NSD1 genome occupancy was mostly on inter- and intragenic regions of genes (**Fig. 26C & D**), a binding pattern similar to that reported for NSD2 in multiple myeloma[240]. NSD2 overexpression, due to the t(4;14) translocation, led to the aberrant accumulation of H3K36me2 at both the intragenic and intergenic regions of genes[240], suggesting widespread targeting. Furthermore, the broad physiological targeting of NSD proteins was also revealed by examining the molecular consequence of H3K36M mutations in chondroblastoma[267]. Recurrent H3K36M mutations reprogrammed the transcriptome through inhibiting and sequestering H3K36 methyltransferases, resulting in a global reduction in H3K36 methylation. Knockdown of NSD1/2 proteins in mesenchymal progenitor cells (MPCs) mimicked the loss in H3K36me2 along the intragenic and intergenic regions comparable to that in caused by H3K36M[267, 268]. NSD1 knockdown in ESC also reduced H3K36me2 levels throughout the genome, at gene promoters, gene bodies and intergenic regions[241]. Functionally, the broad enrichment of H3K36me2 throughout the genome is thought to play a role in protecting against PRC2 mediated silencing, and inhibiting the spread of H3K27 methylation from active regions[269].

Sotos syndrome is characterized by varying degrees of mental retardation, seizures, advanced bone age, cardiac and renal anomalies, and a slight increased cancer incidence [270]. Our experiments in undifferentiated MEL cells revealed potential NSD1 target genes belonging to pathways regulating cardiac function, neuronal development, and cancer (**Fig. 26F-H**), thus linking NSD1 malfunction to developmental abnormalities diagnosed in Sotos

patients. Unfortunately, we could not find pathways that directly link NSD1 activity to hematopoiesis or leukemia. However, interestingly, NSD1 seems to be involved in modulating calcium ion uptake and cellular concentration, a process of key importance for promoting differentiation and proliferation of erythroid precursors at the stages of BFU-E and CFU-E. Inhibiting $Ca^{2+}$ uptake in erythroid precursor cells, by blocking NMDA receptors leads to cells mortality at the basophilic and polychromatic stages, suggesting their importance for erythropoiesis [271, 272]. In order to validate these preliminary results, and focus the investigation on NSD1 functional activity during erythroid differentiation, we will perform the ChIP-Seq analysis in differentiated MEL cells. In addition, we have also generated a MEL cell clones with a validated 3XHA-Strep-II tag at the 3' of endogenous *Nsd1* locus (not shown). unfortunately, our attempts to immunoprecipitate NSD1 protein using an HA antibody so far failed (not shown). This is most likely due to several factors, such as: NSD1 low expression level, large size, and/or strong chromatin binding of the protein, 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, we could implement several adjustments to our protocol, such as: increasing the number of cells used per experiment, using harsher protein extraction buffers, adding higher units of DNase endonuclease, or generating new MEL clones tagged at the N-terminus of NSD1 instead of the C-terminus.

# 4. Chapter II: A novel inducible mouse model of *MLL-ENL*-driven mixed lineage acute leukemia

Stavropoulou, V., **M. Almosailleakh**, H. Royo, J.-F. Spetz, S. Juge, L. Brault, P. Kopp, M. Iacovino, M. Kyba, A. Tzankov, M. B. Stadler, G. Cazzaniga, A. H. F. M. Peters and J. Schwaller (2018). "A Novel Inducible Mouse Model of MLL-ENL-driven Mixed-lineage Acute Leukemia." HemaSphere.

[273]

# 4.1. Abstract

Previous retroviral and knock-in approaches to model human t(11;19)$^{+}$ acute mixed lineage leukemia in mice resulted in myeloproliferation and AML not fully recapitulating the human disease. We established a doxycycline (DOX)-inducible transgenic mouse model "*iMLL-ENL*" in which induction in long term hematopoietic stem cells (LT-HSC), lymphoid primed multi-potent progenitor cells (LMPP), multi-potent progenitors (MPP4) but not in more committed myeloid granulocyte-macrophage progenitors (GMP) led to a fully reversible acute leukemia expressing myeloid and B-cell markers. *iMLL-ENL* leukemic cells expressed lower *MLL-ENL* mRNA than those obtained after retroviral transduction. Disease induction was associated with *iMLL-ENL* levels exceeding the endogenous *Mll1* at mRNA and protein levels. In leukemic cells from t(11;19)$^{+}$ leukemia patients, *MLL-ENL* mRNA also exceeded the endogenous *MLL1* levels suggesting a critical threshold for transformation. Expression profiling of *iMLL-ENL* acute leukemia revealed gene signatures that segregated t(11;19)$^{+}$ leukemia patients from those without an MLL translocation. Importantly, B220$^{+}$ *iMLL-ENL* leukemic cells showed a higher *in vivo* leukemia initiation potential than co-existing B220$^{-}$ cells. Collectively, characterization of a novel transgenic mouse model indicates that the cell-of-origin and the fusion gene expression level are both critical determinants for *MLL-ENL*-driven acute leukemia (**Appendix 1**).

## 4.2. Results

**Modeling t(11;19)$^{+}$ mixed lineage leukemia by inducible *MLL-ENL* expression in mouse hematopoietic stem and progenitor cells**

The chromosomal translocation t(11;19) leading to expression of an MLL-ENL fusion is a molecular hallmark of acute leukemia that either presents as ALL, MLL or rarely also as pure AML[79]. To model the cellular origin of MLL-ENL driven acute leukemia we generated a transgenic mouse line "*iMLL-ENL*" that express the human *MLL-ENL*[274] oncogenic fusion cDNA under the control of the reverse-type tetracycline-controlled transactivator (*rtTA*) (**Fig. 30A**). We induced the expression of *iMLL-ENL* expression *in vivo* by administrating DOX (0.4mg/ml) to the drinking water of the mice. The mean latency of disease development was 104.3 ± 16.9 days. Strikingly, discontinuation of DOX administration (**Fig. 30B**, arrow) resulted in the survival of all but one mouse. Peripheral blood smears from symptomatic mice that survived after DOX removal showed normal blood cells' morphology except for the occasional focal infiltration cells with pyknotic appearing nuclei (**Fig. 30C**). Flow cytometry analysis in symptomatic mice with highly (>95%) infiltrated BM, showed that the majority of the leukemic cells were of intermediate size (86.8% ± 7.4), with minor fractions (1.7% ±0.9 and 1.4% ±0.3) of cells of large and small size respectively (**Figure 30D**) suggesting the presence of blasts with myeloid and lymphoid differentiation. Notably, only the small sized cells expressed the lymphoid markers B220 (37% ±7.5, n=3) and CD19 (15%±5.3, n=3) (**Fig. 30E**). We also detected a minor population of the B220$^{+}$ cells expressing the myeloid markers Mac-1 (8.1% ±4.4, n=3) or Gr1 (4.4% ±3.3, n=3) (**Fig. 1F**). In contrast, the majority of large and intermediate sized cells showed little were Mac-1 or Mac-1/Gr1 positive, and only had limited expression of B220 or CD19 (**Fig. 30E&G**). PCR analysis suggested that in contrast to normal BM cells, B220$^{+}$ and B220$^{+}$/CD19$^{+}$ *iMLL-ENL* leukemic blasts do not undergo complete *IgH D-J* gene rearrangement[275] as we can detect the presence of a germline (GL) band in these population from diseased *iMLL-ENL* mice (**Fig. 30H**). Collectively, *iMLL-ENL* is a novel robust inducible and reversible transgenic mouse model for acute leukemia expressing myeloid and lymphoid markers.

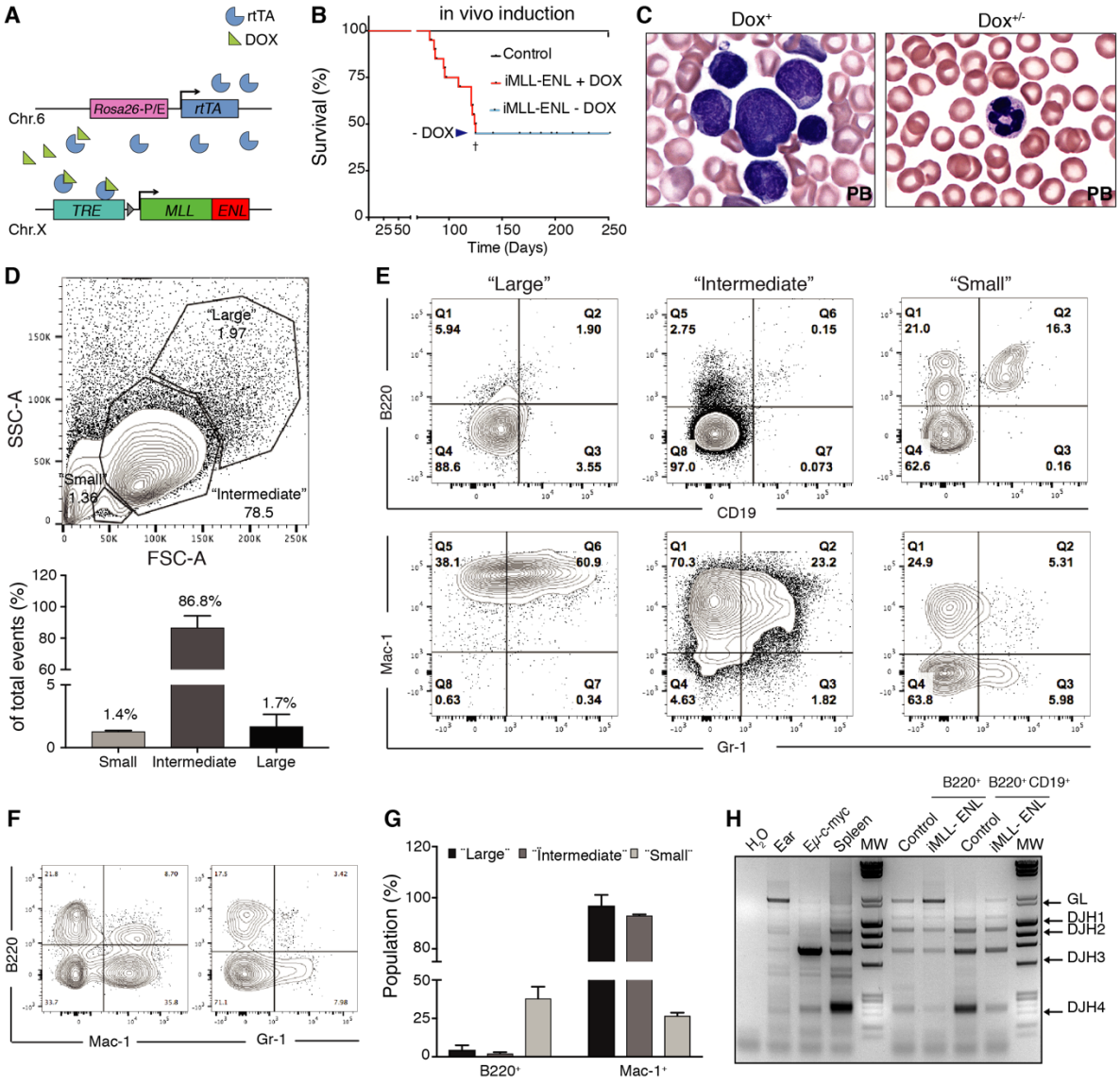



**Fig. 30: Induction of *iMLL-ENL* resulted in reversible mixed lineage acute leukemia**

**(A)** Schematic illustration of the inducible *iMLL-ENL* transgenic mouse model. Expression of the human *MLL-ENL* fusion cDNA is controlled by a Tet-ON-responsive element (TRE). The *reverse tetracycline transactivator* (*rtTA*) cassette was integrated into the *Rosa26* gene locus and the *MLL-ENL* fusion gene was targeted into the *Hprt* locus. Doxycycline (DOX) binds to *rtTA* and causes a conformational change allowing it to bind to the TRE promoter and activate the expression of *MLL-ENL*. **(B)** DOX administration induced acute leukemia in *iMLL-ENL* mice after a median latency of 104.3 ± 16.9 days (n=10). Upon DOX removal (arrow) the remaining animals (n=10) survived up to 300 days without developing the disease, except of one mouse with extensive multi-organ leukemic infiltration (cross). **(C)** Wright-Giemsa-stained blood smears showing leukemic blasts of different sizes upon DOX administration ($DOX^+$). Upon DOX removal ($Dox^{+/-}$) the blood smears were normal. (**D**) Top panel: representative flow cytometry analysis of leukemic blasts isolated from highly infiltrated BM of *iMLL-ENL* mouse grouped according Forward and Side scatter (FSC-A & SSC-A respectively) into three cell size populations ''large'', ''intermediate'' and ''small'' cells. Bottom panel: quantification of the percentage of designated three cell population size ("large", "intermediate" and "small") in diseased mice. **(E)** Detailed flow cytometric analysis of BM cells from a representative diseased *iMLL-ENL* mouse showing the expression for markers representing myeloid (Mac-1, Gr-1) and B-cell lineages (B220, CD19) grouped according to different cell size ("large", "intermediate" and "small"). **(F)** Flow cytometry analysis plot showing the expression of lymphoid marker B220 against the myeloid markers Mac-

1 and Gr-1 in BM cells from 3 diseased iMLL-ENL mice. (**G**) Quantification (%) of cell population expressing lymphoid (B220) versus myeloid marker (Mac-1), grouped according to size. **(H)** PCR analysis of the *IgH D-J* gene rearrangements using DNA isolated from the BM cells of a leukemic mouse (input) of single $B220^+$ and double $B220^+/CD19^+$ flow-sorted cells versus control DNA samples isolated from the ear, the spleen of a wild type (B6) mouse, and from a lymphoma of *Eμ-c-myc* transgenic mice (GL = germline). The bars represent the percentage of mean of expression measured from BM cells of 3 independently analyzed diseased *iMLL-ENL* mice.

**Fusion gene expression levels determine the transformation potential of *iMLL-ENL***

To investigate whether similarly to other MLL -fusions (i.e. MLL-AF9)[276] the expression level of the *iMLL-ENL* fusion correlates with transformation potential, we compared the mRNA levels of the *iMLL-ENL* and endogenous WT *Mll1*. We observed that in the absence of DOX or within 24hr of DOX removal, *iMLL-ENL* mRNA expression levels were below those of the WT-*Mll1* (**Fig. 31A**). In leukemic *iMLL-ENL* cells from diseased mice, the estimated fusion mRNA expression level consistently exceeded *Mll1* (**Figure 31A**). The expression of *MLL-ENL* in retrovirally transduced BM cells (*rMLL-ENL*) were about 5-10 folds higher than *iMLL-ENL* mRNA expression in BM-derived *in vitro* immortalized cell lines or primary leukemic cells from diseased *iMLL-ENL* mice. Comparably, we also observed that the relative fusion transcript levels in leukemic blasts from five $t(11;19)^+$ ALL patients exceeded those of endogenous *MLL1* (**Figure 31B**). In addition, MLL-ENL protein expression clearly exceeded the expression of wild-type MLL in the $MLL\text{-}ENL^+$ KOPN8 cells, and was about equal to WT- MLL in BM and spleen cells of diseased *iMLL-ENL* mice (**Figure 31C**). These data suggest that MLL-ENL protein levels equal to or exceeding that of the non-rearranged MLL is necessary for leukemic transformation by *MLL-ENL* and possibly most other MLL-fusions. Collectively, I characterized the immunophenotyped of the iMLL-ENL mouse model, and provided experimental evidence (on mRNA and protein level) suggesting that the transforming potential of the MLL-ENL fusion is dose-dependent. These observations were critical for publication of the respective manuscript[273].

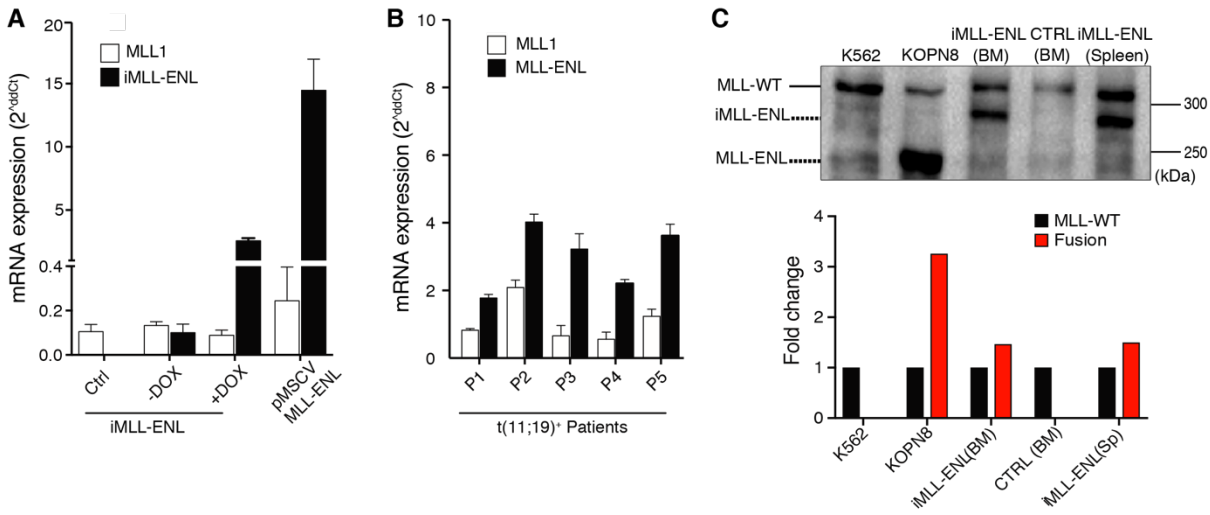



**Fig. 31: *iMLL-ENL* expression levels determine cellular transformation *in vitro* and *in vivo***
**(A)** Relative expression of *Mll1* and *MLL-ENL* mRNA was measured in naïve (off DOX) and leukemic BM cells from iMLL-ENL transgenic mice and BM cells from symptomatic mice transplanted with retrovirally (*pMSCV-MLL-ENL*) transduced cells. Relative mRNA expression levels were normalized to *Gapdh* expression and are expressed as $2^{\wedge ddCt}$. Results are shown as mean values ± S.D. (n=2). **(B)** Relative expression of *MLL1* and *MLL-ENL* mRNA in leukemic cells of human patients diagnosed with ALL and >80% infiltration. Relative mRNA expression levels were normalized to *GAPDH* expression and are expressed as $2^{\wedge ddCt}$. Results are shown as mean values ± S.D. (n=2). **(C)** Western blot analysis of $1x10^6$ total cell lysates from K562 (CML, wildtype MLL), KOPN8 (ALL, t(11;19)$^+$), BM of healthy mice and BM and spleen cells of diseased *iMLL-ENL* mice. The blot was probed with an antibody recognizing the MLL-N-terminus. KOPN8 cells express an MLL-ENL fusion of calculated size of about 170 kDa[12], while the iMLL-ENL fusion results in a protein of 220 kDa [11,12]. Bands were quantified according to intensity, and calculated as fold change normalized to wild-type MLL.

## 4.3. Discussion

Although distinct MLL1 fusion products can be restricted with leukemic subtype, many MLL rearrangements detected in humans appear permissive for AML, ALL or MLL phenotype. Immunophenotyping leukemic blasts positive for the MLL-ENL fusion, revealed simultaneous expression of lymphoid and myeloid markers, suggesting that the disease arises from a multipotent stem or early progenitor cells, retaining a lymphoid and myeloid potential. Surprisingly, modeling MLL-ENL leukemic transformation in mice by viral overexpression, or *in vivo* knockin, resulted in the development of AML or MPD, but never ALL or MLL, which could be due to an inappropriate fusion expression level, or targeted cellular compartment[277-279]. Our *iMLL-ENL* induced acute leukemia was characterized by the coexistence of blasts of various sizes that expressed myeloid and/or lymphoid markers or exclusively lymphoid markers in a small fraction of the cells. Similar to patients with t(11;19)$^{+}$ leukemia, *iMLL-ENL* lymphoid cells from leukemic mice, had immature immunoglobulin rearrangements, suggesting a mixed lineage leukemic phenotype.

A previous study using a similar *MLL-ENL* inducible mouse model, suggested that MLL-ENL can only initiate the leukemic transformation from GMPs, whereas the HSC compartment is inherently protected from transformation[280]. In contrast, in our *iMLL-ENL* model, activation of the oncogene in LT-HSC, MPPs or LMPP but not GMP cells resulted in fully penetrant transplantable and reversible mixed-lineage acute leukemia. Notably, although both models adapt the same inducible system to derived MLL-ENL expression, they differ in the site of fusion insertion. While our MLL-ENL fusion was inserted in the *Hprt* locus, they targeted the fusion to 3' UTR of the murine *Col1a1* gene. This seemingly minor difference, culminated to drastic variation in the expression of the *MLL-ENL* transcript in the two models, specifically in HSCs and MPPs. The expression levels of the *MLL-ENL* fusion transcripts from the *Col1a1* locus in HSCs and MPPs remained clearly below endogenous *Mll1*, while it exceeded *Mll1* in GMPs. Previous reports have already suggested that leukemic transformation of *Mll-r* is dependent on fusion expression exceeding WT *Mll1* in HSC and MPPs[281]. We detected higher fusion protein levels in human t(11;19)$^{+}$ KOPN8 cells, and in leukemic blasts from diseased *iMLL-ENL* mice compared to WT MLL. MLL-fusion proteins were reported to be more stable than WT MLL1. Pharmacologically stabilizing MLL1 protein improved the survival of leukemic mice by displacing the MLL chimera from their oncogenic targets, offering a new concept for the development of therapy [282].

# 5. Chapter III: Generation of a transgenic Erg-YFP reporter mouse line

## 5.1. Results

Experimental and clinical evidence indicated that the ETS-family transcription factor is a critical HSC regulator in normal and malignant hematopoiesis. As we found that ERG expression characterized LT-HSC-derived leukemia cells in our conditional iMLL-AF9 mouse model, we aimed to establish a reporter mouse line that would allow use to track and modify normal and transformed Erg-expressing cells *in vitro* and *in vivo.*

The long-standing tradition of generating mouse transgenic lines using gene targeting in embryonic stem (ES) cells is complex and time-consuming. Nowadays, creating genetically modified mice and rats with site specific genetic modifications has been considerable accelerated and simplified by the direct injection of site-specific nucleases into the one-cell-stage embryo[283, 284]. With the help of the university of Basel transgenic core mouse facility, we adopted this approach to generate mice carrying a *YFP* fluorescence reporter at the C-terminus of the *Erg* (ETS-related) gene (**Fig. 32A**). Due to the extensive A & T base pair repeats at the 3' end of the *Erg* genomic sequence, we needed to increase the specificity and sensitivity of the N1 PCR screening to confidently exclude false positive/negative genotyping. Thus, we designed 3 primers to be used in 2 successive PCR reactions, where the PCR product of the first reaction is used as input to the second reaction, known as ''Nested PCR'' (**Fig. 32B, i**). Out of eight N1 mice, we obtained one positive (#40) female mouse (**Fig. 32B, ii & iii**), which we used for subsequence N2 generation. Although we confirmed the expression of an *Erg-P2A-YFP* mRNA by RT-PCR (**Fig. 32C**), unfortunately, we could not detect a YFP signal in BM cells using flow cytometry analysis (**Fig. 32D**).

It was previously reported that in mice, the expression of *Erg* is upregulated during stress hematopoiesis such as during infections[71]. To mimic infection induced stress hematopoiesis, we treated *Erg-YFP*$^{+}$ and WT (*Erg-YFP*$^{-}$) littermates with either Lipopolysaccharide (LPS) or PBS and then performed flow cytometry analysis on BM cells (**Fig. 33A**). As expected[285], we observed an expansion of the HSC compartment signified by the increase in the kit$^{+}$ Sca-1$^{+}$ cells, however, no YFP signal was detected (**Fig. 33B**). We also tried to enhance the YFP signaling but intracellular staining with a YFP/GFP specific antibody,

nevertheless, again no positive YFP signal was detected in the HSC compartment (**Fig. 33C**). Two hypotheses could explain these results, either that the YFP signal was extremely low, and thus below the detection threshold of flow cytometry, or that during the editing/knockin procedure, the repair DNA template acquired out of frame or missense mutations that lead to deficient YFP protein translation.

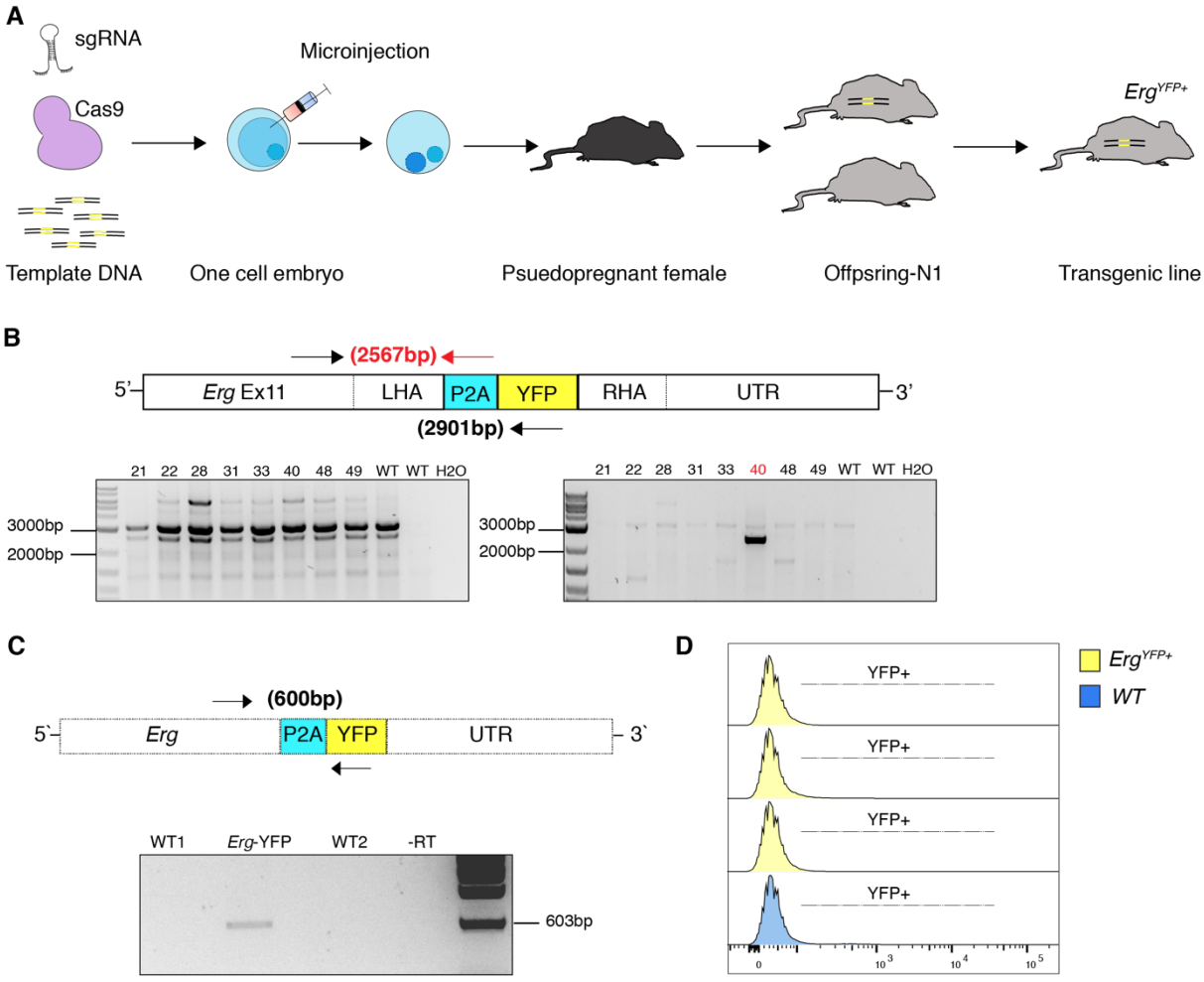



**Fig. 32: GE mediated integration of a *P2A-YFP* expression cassette at the C-terminus of murine *Erg* gene locus**

**(A)** Experimental setup: A mixture of *in vitro* transcribed sgRNA, CAS9 protein and linearized template DNA vector (*Erg-P2A-YFP*) were microinjected into a one cell stage embryo, and then transplanted into psuedopregnant females. N1 offspring were screened for correct editing by PCR, and backcrossed into wild-type mice to generate a transgenic line. DNA template and sgRNA was designed *in situ* and commercially synthesized. **(B)** Schematic illustration of edited *Erg* locus showing the position and size of expected PCR products for screening the insertion of *P2A-YFP* cassette **(i).** Agarose gel images of Nested PCR products, first PCR step **(ii)** and second PCR step **(iii)** of eight N1 offspring (#21, 22, 28, 31, 33, 40, 48, 49) and two non-edited wild-type control (WT). **(C)** Schematic illustration of edited *Erg* mRNA (cDNA) showing the position and size of expected RT-PCR product for detecting the expression of *P2A-YFP* cassette (Top). Agarose gel image RT-PCR product of one Erg-YFP$^{+}$ and two WT (Erg-YFP$^{-}$) littermate. **(D)** Flow cytometry analysis of three *Erg*-YFP$^{+}$ (yellow histogram) and one WT (*Erg*-YFP$^{-}$ , blue histogram) littermate for the expression of YFP fluorescence (YFP+).

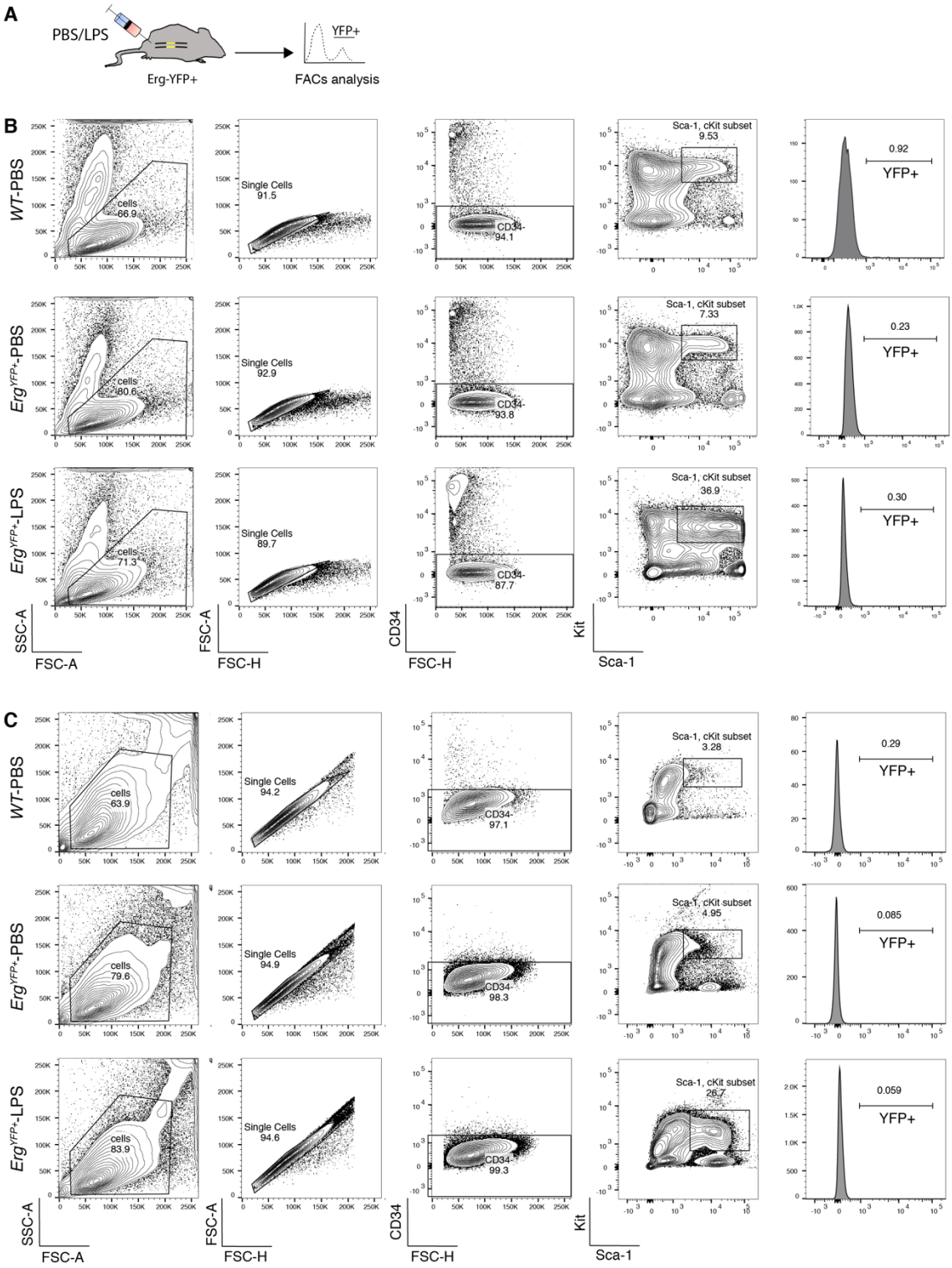



**Fig. 33: LPS treatment lead to HSCP expansion but did not lead to detectable YFP expression in BM of *Erg-YFP*$^{+}$ mice**

**(A)** Schematic illustration of experimental set: *Erg*-YFP$^{+}$ and WT (*Erg*-YFP$^{-}$) littermates were injected (I.P.) with either LPS or PBS 24 hours prior to flow cytometry analysis. **(B)** Representative flow cytometry analysis for CD34, kit, Sca-1 and YFP expression in BM cells from PBS-treated WT (*Erg-YFP*$^{-}$) and *Erg-YFP*$^{+}$ mice compared with LPS-treated *Erg-YFP*$^{+}$ littermates. **(C)** Representative intracellular flow cytometry analysis for CD34, kit, Sca-1 and YFP expression in BM cells from PBS-treated WT (*Erg-YFP*$^{-}$) and *Erg-YFP*$^{+}$ mice compared with LPS-treated *Erg-YFP*$^{+}$ littermates.

# 5.2. Discussion

**Creating a transgenic *Erg-YFP* reporter mouse line**

The visualization of specific HSCs cell populations by transgenic expression of fluorescent proteins driven by cell-type specific promoters has greatly facilitated the analysis of the heterogenous nature of these cells. Hereby researchers either expressed a transgenic fluorescent marker under regulatory elements of an HSC-selective gene, or fused a fluorescence sequence directly, or linked by P2A or IRES to the N- or C-terminus of the respective HSC regulators. One of the best-known examples are *Evi1-GFP*, *Tie2-GFP* or *Hoxb4-mCherry* that were successfully used to track and functionally characterize HSC in mice [67, 286, 287]. Our laboratory has recently identified the HSC transcription factors *Evi1* and *Erg* as markers of AML in mice induced by an inducible mixed lineage leukemia (MLL) fusion. Coupling and comparing the expression levels of *Evi1* and *Erg* into $Evi1^{high}/Erg^{high}$, $Evi1^{low}/Erg^{high}$ and $Evi1^{low}/Erg^{low}$ fractions seems to correspond to different cellular origins and disease outcome in mice as well as in a cohort of human AML patients[288]. Based on the observations of poor prognostic outcomes with higher *Evi1* and *Erg* expression, it is very likely that $Evi1^{high}/Erg^{high}$ expressing cells might represent a particular class of LT-HSC with a particular susceptibility for being immortalized by an MLL fusion, resulting in an aggressive and severe leukemia.

In order to further investigate this observation, we attempted to establish a transgenic *Erg*-reporter mouse (*Erg-YFP*), and then cross it with an established *Evi1-GFP* mice available in the lab [67], generating a double reporter transgenic mouse line. As the intensity of the fluorescence signals would correspond to gene expression level, we planned to sort HSCs into different fractions according to *Evi1* (GFP) and *Erg* (YFP) expression, and perform various *in vitro* (clonogenic assays) and *in vivo* (bone marrow transplantations) experiments. Unfortunately, our attempt at generating transgenic *Erg-YFP* reporter mouse using *in vivo* CRISPR-Cas9 GE technique was unsuccessful. Despite obtaining mice with validated YFP insertion at the 3' of the *Erg* locus, we could not detect a fluoresce signal using flow cytometry in hematopoietic cells (**Fig. 27 &28**).

In comparison to the classical ESC targeting technique, *in vivo* GE using CRISPR-Cas9 has significantly reduced the time needed to generate transgenic animal models. However, the success rate of obtaining locus specific homologous recombination (HR) editing is still very low[289, 290]. In fact, during the course of one year, and after 3 separate attempts, we could

only obtain a single mouse with the desired editing out of 65 (1.5%) generated. Nevertheless, due to the simplicity of the technique, and its potential clinical use in humans for gene therapy[291], extensive research dedicated to increasing the efficiency and accuracy of *in vivo* editing is constantly reported[292-294]. We could thus explore the possibility of using these newly developed and improved protocols to increase the effeciency of creating the transgenic *Erg*-reporter mouse again. In addition, as we suppose that the lack of detectable fluorescence signal could be due to low intensity of the YFP, or low transcription rate. It would be reasonable to improve the design of our repair template, by for example placing the fluorescence signal on the 5' end of the genomic locus, which was reported to improve the transcription rate[295]. We could also replace YFP with a brighter fluorescence signal (e.g. Venus or mOrange), or by adding multiple consecutive fluorescence sequences to improve detection. A recent example on how to improve fluorescence signal was reported for the HSC TF *Hoxb5*[287]. The researchers show that in frame fusion of *Hoxb5* with a single P2A and mCherry sequence, did not results in a detectable fluorescent signal, however, they managed to improve the signal by adding a triple P2A linked mCherry sequence. Notably, although they used Crisper-Cas9 for targeted integration, they still used ES cells and not zygotes, to overcome the limitation of large cassette integration. Their report strongly suggests that increasing the signal strength either by adding multiple copies of the fluorescent marker or using variant with a stronger light emission (e.g. dtTomato instead of mCherry) could be a promising way to finally establish the *Erg*-reporter line.

# 6. Concluding remarks

AML is a rare but severe form of human cancer that results from a limited number of functionally cooperating genetic abnormalities, leading to uncontrolled proliferation and impaired differentiation of hematopoietic stem and progenitor cells. The underlying molecular mechanism of AML is complex, and characterized by deleterious genetic mutations in a variety of genes functional involved in cell fate determination and survival, including epigenetic regulators. Alterations in the function of epigenetic modifiers, such as DNA methylation and histones PTM enzymes, are heritable and highly heterogeneous. Progression of AML undergoes significant clonal evolution, culminating in therapy resistance and eventual relapse. Unfortunately, specific genetic targets linked to hematopoietic transformation following mutations in individual epigenetic modifiers is difficult to interpret, due to multi-regulatory nature of these modifications. Studies in transgenic mice to conditionally activate or delete genes alone, and in combination with other epigenetic modifiers, have already expand our understanding of the impact of these alterations on hematopoiesis. Increased efforts to integrate mutational studies with epigenomic and transcriptomic profiling, and the development more sophisticated tracing techniques, offer promising avenues to delineate genetic targets responsible for malignant transformation. Finally, deciphering the characteristic heterogeneity of AML, and its complex underlying molecular pathogenicity, will hopefully help in improving outcomes for this malignancy, by the optimization and design of novel therapeutic strategies.

# 7. Material and Methods

**Data presentation and statistical analysis**

Bar graphs in the figures represent the mean value of biological replicates. Error bars are standard error of the mean (Mean±SEM). Statistical significance was tested with unpaired (*Nsd1*$^{fl/fl}$ vs. *Nsd1*$^{-/-}$) or paired (viral transduction in *Nsd1*$^{-/-}$ cells and MEL) two-tailed t-test, assuming equal variance, unless otherwise specified. Statistical test was performed in $\log_{10}$ space, or for qPCR kept in $\log_2$ space. Raw gene expression accession number is GSE136811. Supplementary Tables are accessible using the following DOI 10.5281/zenodo.3600185.

**Mice**

Mice carrying a *Nsd1*$^{+/L3}$ allele were previously described[107]. The floxed *pgk-neomycin* selection cassette was removed by viral *Cre* expression in ES cells, leaving two *loxP* sites flanking the largest coding Ex5, here referred as *Nsd1*$^{fl/fl}$. *Nsd1*$^{fl/fl}$ mice were intercrossed with a *Vav1-iCre*$^{tg/+}$ transgenic strain leading to inactivation of the gene in fetal and adult hematopoiesis[296]. All mice in this study were kept under specific pathogen-free conditions. Mice were genotyped using the KAPA Mouse Genotyping Kit Hot Start Kit (Kapa Biosystems, Wilmington, USA, #KK7352) following the manufacturer's instructions. *Erg*$^{YFP+}$ mice were generated in the center for transgenic models (University of Basel, Switzerland) by CRISPR-Cas9 editing technology[297]. sgRNA primers directed at *Erg* were synthesized as gene blocks (IDT), and *in vitro* transcribed using T7 High Yield RNA Synthesis Kit (NEB, #E2040S). Repair template plasmid was designed *in situ* and commercially synthesized (Genescript). Pups were genotyped by nested PCR, followed by verification with Sanger Sequencing. Primers used for genotyping and target repair plasmids can be found in **Tables 5&8** respectively.

**Cell culture**

BM-derived erythroblasts cells were obtained following a previously published protocol[188]. Erythroblast cultures from adult mice were established after lineage depletion of BM cells, Cells were cultured for more than one week in maintenance medium (MM) composed of StemSpan SFEM (StemCell 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 40ng/ml hIGF-1 (Peprotech, London, UK). Cells were split every second day and presence of proerythroblasts was verified by flow cytometry ($DAPI^-$ /$FSC^+$/$Kit^+$/$CD71^+$/$Ter119^-$) and cytospins. Erythroblasts were subjected to terminal maturation in differentiation medium (**DM**) composed of IMDM (Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 1%P/S, 10%FCS, 10%PFHMII (Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 5%hPDS (0.45µM filtered, Blood donation service, University Hospital Basel), monothioglycerol (Sigma Aldrich, Buchs, Switzerland), 100ng/ml mSCF and 2U/ml hEPO.

**Generation of edited single MEL clones**

sgRNA directed against mouse *Nsd1 was* cloned into px458 (Addgene 48138; a gift from F. Zhang). Mouse *Nsd1* right and left homology arms (RHA & LHA respectively) sequences were cloned into pAAv-loxP-Neo-3xFLAG[217] (a gift from Zhenghe Wang, Cleveland-USA) with an N-terminal Flag epitope tag using USER cloning assembly. Site-directed mutagenesis was used to create base pair substitutions in *Nsd1* using the Q5® Site-Directed mutagenesis kit (NEB #E0552S) following the manufacture protocol. To generate single clones with NSD1-3XFLAG site specific insertion lines, MEL cells were transfected with sgRNA-containing px458 and pAAv-loxP-Neo-*Nsd1*-3xFLAG repair template using Amaxa Nucleofector™ (Kit L, Catalog #: VACA-1005, program A-20, Lonza) and incubated for 48 h with SCR7: NHEJ inhibitor, Xcess biosciences Inc (San Diego, CA 92109). Single $GFP^+$ cells were then sorted into 96-well plates, expanded, selected with G418 (1mg/mL) for 10 days, before gDNA isolation. Individual clones were screened for appropriate editing by PCR and individually verified by Sanger sequencing of the target loci. To excise the *Neomycin* cassette correctly edited clones, were then electroporated with pCMV-Cre-GFP, and single $GFP^+$ cells were then sorted into 96-well plates. Following 1 week of expansion, the cells were divided into two 96-well plates, and one plate was grown under G418 selection (1mg/mL) for 10 days. Sensitive clones were verified for *Neomycin* cassette excision using gDNA PCR and Sanger sequencing.

**Colony forming assay**

Approximately $10x10^4$ cells were plated in methylcellulose M3434 (Methocult, StemCell Technologies, Vancouver, Canada), and incubated at 37°C till scored after 8-10 days. Pictures were taken on Olympus IX50 microscope with 2.5 objective magnification. Cells were washed in warm PBS, resuspended and counted with trypan blue exclusion method.

**Benzidine staining**

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

**Cell cycle analysis**

For cell cycle analysis 150,000 cells were harvested at the donated time points, washed with PBS, and fixed with 200 µl of ice-cold 70% EtOH overnight at -20°C. For FACS analysis, fixed cells were washed with PBS, and stained with P.I. buffer (PBS with 100 µg/mL RNase A, 50 µg/mL Propidium Iodide) (Propidium Iodide from Fluka chemicals, cat. No. 81845) for 10 min at 4°C in the dark, and then FACS immediately.

**Flow cytometry**

Cells in suspension were washed with FACS buffer (0.5% BSA, 1mM EDTA in PBS) and incubated with indicated antibodies for 30 min at 4°C, washed and stained with 1mg/ml DAPI (Life Technologies, Paisley, UK) in PBS. Stained cells were analyzed on LSR Fortessa II (BD, New Jersey, USA). Data was analyzed with FlowJo software (version 10.0.8 or 10.2). All antibodies used in this study are indicated in **Table 7**. For differentiation analysis*, in vitro* mouse cells were filtered, washed twice with PBS and stained in 100µL FACS buffer. Populations gating was done according to negative and positive beads staining and unstained negative control.

**Retroviral Gene Transfer**

Full-length cDNAs for murine *Nsd1* (*pSG5*) in was obtained from R. Losson (Strasbourg). Wildtype *(Nsd1*) and a catalytically-inactive (*Nsd1*$^{N1918Q}$*)* mutant ORF were cloned into the murine stem cell virus (*pMSCV*) expression vector and sequence-verified. A retrovirus (*pLMP*) encoding for a SKI-specific Mir-shRNA was a gift from Prof. M. Hayman (Bufallo, NY). Retroviral stocks were produced by transient co-transfection of packaging vectors (*pIPAK6*) and respective plasmids using Jetprime transfection reagent (Life Technologies, Paisley, UK) in HEK293T-LX cells kept in DMEM (Gibco, Lubio, Thermo Fisher Scientific, Reinach, Switzerland) with 10% FSC and 1% P/S. Viral supernatants were harvested 24 and 48 hours after transfection, 10x Vivaspin 20 (Sartorius, Göttingen, Germany) concentrated at 3500rpm for 1.5hrs at 4°C and snap frozen in liquid nitrogen and stored in -80°C until usage. Cells were spin-infected either in MM used for erythroblast culture as described above, in presence of 4µg/ml polybrene (Sigma Aldrich, Buchs, Switzerland) with virus for 90min, 1000rpm at 30°C. 12-18 hours after spin infection, the cells were washed with PBS and plated in MM. Two days after spin infection, the cells were selected with 2µg/ml puromycin (Gibco, Thermo Fisher Scientific, Reinach, Switzerland), or GFP$^{+}$ cells were FACS enriched as previously described. Information regarding the used plasmids can be found in **Table 8.**

**shRNA-mediated knockdown**

shRNAs were expressed from lentiviral vectors (*pLKO.1*). For transduction lentiviral stock was produced by transient co- transfection of packaging vectors (pMD2.G, *pMDL g/pRRE, pRSV-Rev*) and respective lentiviral shRNA plasmid (shRNA Ctrl-#229 and shRNA *NSD1* #445, #379, #380, #381) using lipofectamine 2000 (Invitrogen, Thermo Fisher Scientific, Reinach, Switzerland) in HEK293T-LX cells kept in DMEM (Gibco, Lubio, Thermo Fisher Scientific, Reinach, Switzerland) with 10% FCS and 1% Penicillin/Streptomycin. Viral supernatants were harvested 24, 48 and 72 hours after transfection, ultra-concentrated at 4°C for 2 hours using 24,000 rpm, resuspended in fresh DMEM, snap frozen in liquid nitrogen and stored in -80 °C until usage. MEL were spin-infected in presence of 10 g/ml polybrene (Sigma Aldrich, Buchs, Switzerland) with virus for 90min, 1000rpm at 30°C. 12-18 hours after spin infection, cells were washed with PBS and plated in fresh DMEM medium. Two days after spin infection cells were selected with 2µg/ml puromycin (Gibco, Thermo Fisher Scientific, Reinach, Switzerland). Plasmids used can be found in **Table 8**.

**RT-PCR**

Quantitative RT-PCR: Total RNA was extracted using the RNA Plus extraction kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. cDNA synthesis was carried out using the high capacity cDNA reverse transcription kit (Cat. 4368814, Applied Biosystems, Foster City, USA). Quantitative PCR was performed using SYBR Green reagent (Applied Biosystems, Foster City, USA) and an ABI prism 7500 sequence detection system. Ct values were normalized to housekeeping gene as described in the legends. Expression and relative expression were quantified using 1/dCt or the 2(-ddCt) method[298]. Primers are given in **Table 9.**

**Western Blotting**

For protein detection, total cell extracts were isolated from freshly cultured $1x10^6$ cells using 60µl of Laemmli sample buffer containing 20% SDS. Following 5 min boiling at 100° C, samples were centrifuged at 4°C for 10 min, and supernatant placed in new tube. Nuclear protein lysates were prepared by resuspending cells in hypotonic lysis buffer (10mM HEPES pH7.9, 10mM KCl, 0.1mM EDTA, 0.1mM EGTA, 1mM DTT) for 15min on ice, followed by treatment with 0.1% NP-40 and vortexing. Nuclei were spun down at 14.000rpm for 2min at 4°C, pellets were resuspended in nuclear lysis buffer (20mM HEPES pH7.9, 0.4M NaCl, 1mM EDTA, 1mM EGTA, 1mM DTT). In addition, pellets were sonicated for 3 cycles (30sec sonication, 30sec pause) on a Bioruptor pico sonicator (Diagenode, Seraing, Belgium) and left on ice before spinning down at 14.000rpm for 10 min at 4°C. Lysis buffers were supplemented with Complete Mini protease inhibitors (Cat. 11836153001, Roche). Proteins were quantified by Bradford assay (Biorad, München, Germany) and loading adjusted. Wet transfer was carried out for 3 hours at 4°C. Membranes were blocked in 5-10% non-fatty milk (NFM) in PBS-1% Tween for 2 hours at room temperature. Blots were probed overnight with antibody at 4°C in 2.5%NFM/PBS-1%Tween, washed three times for 15minutes in PBS-1% Tween and probed with a secondary antibody in 2.5%NFM/PBS-1%Tween. Again, blots were washed three times in for 15 minutes in PBS-1%Tween and then probed with Supersignal West Femto Max substrate (Thermo Scientific, Reinach, Switzerland). Carestream Biomax Kodak films were used for development (Sigma, New York, USA). Information regarding the used antibodies can be found in **Table 10**.

**RNA sequencing: RNA isolation and library preparation**

Total RNA from *Nsd1* and *Nsd1*$^{N1918Q}$ transduced cells was isolated using MACHEREY-NAGEL Nucleospin RNA Plus kit (Ref 740984.250). RNA concentration was measure by flurorometry using the QuantiFuor RNA system (cat # E3310, Promega, Mqadison, WI, USA). RNA integrity was measure using the Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA) using the RNA 6000 Nano Chip (Agilent, Cat #5067-1511). Library preparation was performed with 200ng total RNA using the TruSeq Stranded mRNA Library Prep Kit High Throughput (Cat # RS-122-2103, Illumine, San Diego, CA, USA). Libraries were quality-checked on the fragment Analyzer (Advanced Analytical, Ames, IA, USA) using Standard Sensitivity NSG Fragment Analysis kit (Cat # DNF-473, Advanced Analytical). Samples were pooled to equal molarity and 1.4pM was used for clustering on the NextSeq 500 instrument (D-BSSE, ETH Zurich, Basel). Samples were sequenced Single-reads 76 bases using the NextSeq 500 high Output Kit 75-cycles (Illumina, Cat # FC-404-1005). Primary data analysis was performed with the Illumina RTA version 2.4.11 and Basecalling Version bcl2fastq-2.20.0.422.

**RNA-Sequencing analysis**

Quantification of reads was done QuasR by Log2 of counts+1 Per Milion (CPM) was used for plots of expression[299]. Differential expression was performed with [300]. In the addback experiment paired design of donor mouse was accounted for in the following model: ~ 0 + Addback_group + mouse. To test the additive interaction term in the addback experiment the design was: ~ mouse + timepoint + *Nsd1* + *Nsd1*: timepoint, where *Nsd1* is the mutation status in the addback (WT Nsd1, or Set domain mutated *Nsd1*). For log-fold-change-shrinkage apeglm was used[301]; ashr algorithm was used for the Addback experiment, as the design did not allow apeglm[302].

Gene set enrichment analysis was performed using fGSEA [ 10.1101/060012], using parameters nperm=10000 or 100000, minSize = 8, maxSize = 5000. Gene signatures were from MsigDB[https://doi.org/10.1016/j.cels.2015.12.004], Gene Ontology (The Gene Ontology Consortium, 2017) independent studies[198, 303], and erythronDB[304]. The latter “ERYTHROID LINAGE DEVELOPMENT” was obtained using the functionality "Build Gene Lists" within an erythropoietic lineage in a comparison between Proerythroblasts and basophilic erythroblasts. The list can be found in **Supplementary Table 15**.

**Immunoprecipitation-MS analysis of GATA1**

Prior to nuclear extract preparation, all buffers were supplied with Complete Mini proteinase inhibitor (Cat. 11836153001, Roche) to prevent degradation of the target protein. In order to lyse the cell membrane, dry pellets (10 x$10^6$ cells) from *Nsd1* and *Nsd1*$^{N1918Q}$ transduced cells were incubated in hypotonic lysis buffer (10 mM HEPES; pH 7.9, 10 mM KCl; 1.5 mM $MgCl_2$) for 15 minutes on ice. After short centrifugation at 14000 rpm at 4°C the nuclear membrane was dissolved in non-ionic nuclear lysis buffer (10 mM Tris pH 7.5, 100 mM NaCl, 2.5 mM $MgCl_2$, 0.5% Triton X-100) to minimize protein denaturation. In addition, cell pellets were disrupted by benzonase nuclease (Cat. E1014-25KU, Sigma Aldrich, Buchs, Switzerland) and sonicated for 5 cycles (30 sec sonication, 30 sec pause) on a Bioruptor pico sonicator (Cat. B01060001, Diagenode, Seraing, Belgium). Next, cells were incubated on ice for 1.5 hours. After centrifugation for 15 minutes at 4000 rpm at 4°C the supernatant containing the nuclear fraction was pre-cleared for 1h at 4°C to remove non-specific contaminants bound to the Protein G dynabeads (Cat. 1004D, Thermo Scientific, Rheinach, Switzerland). The protein concentration was determined with a colorimetric protein assay, based on the Bradford method (595 nm wavelength). The antibodies against GATA1 N6 (Cat. Sc-265, Santa Cruz) or control IgG2a (Cat. Sc-3883, Santa Cruz) (**Table 11**) were pre-bound to the beads rotating for 15 minutes at room temperature to minimize co-elution of the antibodies. Immunoprecipitations were performed overnight at 4°C. The bead-antibody-antigen complexes were washed in nuclear lysis buffer to remove potential contaminants. Samples were subjected to immunoblotting or trypsin-based (5 µg/ml, Promega) on-bead digestion in 1.6 M urea/ 100 mM ammonium bicarbonate buffer at 27°C for 30 minutes. Cells were subjected to on-bead digestion. Carbamidomethylation of cysteins was performed by addition of 10 mM TCEP and 15 mM Chloroacetamide in the dark at 37 C for 1 hour. Final digestion of proteins was performed by addition of 1 □g Trypsin (Promega) at 37 C overnight. The tryptic digest was acidified by addition of TFA (pH<3) and samples were desalted using C18 reversed phase spin columns (Harvard Apparatus) following the manufacturer instructions. Dried peptides were stored at -20°C until further use. Peptides were dissolved in 0.1% formic acid prior to injection into the mass spectrometer.

**Mass spectrometry analysis and label-free quantification**

Protein samples from were subjected to LC–MS analysis using a dual pressure LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) connected to an electrospray ion source (Thermo Fisher Scientific) as recently described[305]. Peptide separation was carried out on an EASY nLC-1000 system (Thermo Fisher Scientific) equipped with a RP-HPLC column (75 µm × 30 cm) packed in-house with C18 resin (ReproSil-Pur C18–AQ, 1.9 µm resin, Dr. Maisch GmbH). A step-wise gradient from 95% solvent A (0.1% formic acid) and 5% solvent B (80% acetonitrile, 0.1% formic acid) to 50% solvent B over 60 min at a flow rate of 0.2 µl/min was used. Data acquisition mode was set to obtain one high resolution MS scan in the FT part of the mass spectrometer at a resolution of 240,000 full width at half-maximum (at m/z 400) followed by 20 MS/MS scans (TOP20) in the linear ion trap of the most intense ions using rapid scan speed. Unassigned and singly charged ions were excluded from analysis. Dynamic exclusion duration was set to 30 seconds.

MS1-based label-free quantification of MS data was performed using Progenesis QI software (Nonlinear Dynamics (Waters), version 2.0). After MS raw file import the data was analyzed using default parameter settings. MS/MS-data were exported from Progenesis QI (MGF format) and searched with a target/decoy-based strategy against a database containing forward and reverse sequences of the proteome from *Mus musculus* (UniProt, 33984 entries) using MASCOT (version 2.4.1). Search criteria required strict trypsin allowing for three missed cleavages. Carbamidomethyl of cysteine was specified in Mascot as a fixed modification. Oxidation of methionine and acetyl of the n-terminus were specified in Mascot as variable modifications. Mass tolerance was set to 10 ppm for precursor ions and 0.6 Da for-fragment ions. The peptide and protein false discovery rate (FDR) was set to 1%. Results from the database search were imported into Progenesis QI and the resulting peptide measurement list containing peak area values of identified peptides was further used in quantitative analysis. Processing and statistical evaluation of peptide and protein quantities between samples was performed using SafeQuant (PMID: 23017020). Normalized peptide and protein intensities from SafeQuant analysis were used to calculate intensity ratios between experimental conditions for identified proteins. Functional analysis was performed on the corresponding gene annotation, with a one-sided Fishers' exact test for enrichment of signatures in the up- and down- regulated proteins, respectively, using adjustment for multiple hypothesis testing with Benjamini-Hochberg method (FDR).

**Chromatin immunoprecipitation and sequencing (ChIP-seq)**

ChIP protocol was adapted from the EZ-Magna ChIP™ A/G kit protocol (Millipore, Merck KGaA, Darmstadt, Germany). *Nsd1*$^{-/-}$ BM-derived erythroblasts transduced with *Nsd1* or *Nsd1*$^{N1918Q}$ pMSCV virus in MM as well as 24h in DM were fixed with 1% formaldehyde for 10min at RT. Undifferentiated parental (not edited), and edited (NSD1-3XFLAG) MEL cells, were also fixed with 1% formaldehyde for 10min at RT. Fixed cells were lysed with Cell Lysis and then Nuclear Lysis buffers to a concentration of 20X10$^6$ cells per mL, and finally sonicated (20-min cycle on Covaris apparatus; KBioscience). Sheared chromatin was immunoprecipitated overnight (**Table 11**). 1/10 of the sheared chromatin was used as a reference (Input). Immune complex collection was performed using Protein G Sepharose (Sigma-Aldrich, P3296) for 1h30 at 4°C. Rinses were done according to Magna ChIP™ kit protocol with Low salt, High salt and LiCL immune complex wash buffers. Finally, elution was performed according the IPure Kit protocol (Diagenode, Cat.No. C03010012) following manufacturer's instructions. Two independent ChIP-seq experiments were conducted for GATA1, H3K27ac and H3K36me3. A single ChIP-seq experiment was performed for NSD1 and FLAG from undifferentiated MEL cells.

**ChIP-seq data analysis**

Reads were aligned with bowtie2[306](version 2.3.2) to the mouse genome (UCSC version mm10). The output was sorted and indexed with samtools (version 1.7.20) and duplicated reads were marked with picard (version 2.9.2) 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 (bioconductor version 3.6). These window counts were exported in bigWig format using the bioconductor package rtracklayer.

**Peak Calling**

Fragment size was estimated using the correlateReads function from csaw (Lun & Smyth, 2016) (bioconductor version 3.6) using data from chr1 and excluding duplicated reads. The estimated average fragment size (i.e. distance of highest correlation of reads on positive and negative strand) was calculated for each experiment and between 120 and 230. For each group of biological replicates, peaks where called with macs2 (version 2.1.1) (Zhang *et al*. 2008) using the options '-q 0.05 --nomodel --extsize 133 -g 2652783500 --keep-dup all'. Called

peaks were subsequently filtered for a log2[fold-change]>1.5 and an FDR<0.05. The filtered peaks were converted to bigBed format using the UCSC command line tool bedToBigBed. Additionally, peaks were annotated using the refGene transcript annotation (version from December 18, 2015). For the filtered peak set, the fraction of peaks overlapping with annotated promoter regions, genes (intro/exons) and intergenic regions and their respective enrichment assuming a random distribution was calculated. Overlap with differentially expressed genes was calculated from transcription start sites (TSS) +/- 200 base pairs (bp). Differential binding analysis was performed with csaw using two replicates and no input. Parameters were 'max.frag=600, pe="both", discard=repeats' and combining windows less than 1000 bp apart and annotated using window overlap to TxDb.Mmusculus.UCSC.mm10.knownGene. An initial window width of 10 was used for GATA1 and 150 for the histone. Visualization of bigWig files was done in IGV (Robinson *et al.,* 2011).

**MS-based identification of Histone modifications**

A total of $0.51X0^6$ *Nsd1*$^{-/-}$ cells expressing either *Nsd1* or *Nsd1*$^{N1918Q}$ were lysed in 50 ul of 2 M Guanidinium-HCl / 0.2 M HEPES, pH = 8.3 / 5 mM TCEP / 10 mM Chloroacetamide and incubated at 95 C for 10 min. The carbamidomethylation of free cysteines was stopped by addition of 1.25 ul 0.5 M N-acetyl-L-cysteine (Sigma, A7250) for 10 min. Free amino groups (protein n-term and epsilon amino groups of lysine) were blocked by addition of 7.5 ul of 0.15 M N-Acetoxy-d3-succinimide (D3-Ac-NHS, Sigma, 633259) at room temperature for 2 h. Reaction was stopped by addition of 5 ul 1.5M Hydroxylamine for 10 min and samples were further incubated with 32.5 ul of 1 M Sodium phosphate dibasic dihydrate (Sigma, 71643) ph=12 for 20 min to cleave esters also formed during amino group labeling. Sample solution pH was lowered by addition of 25 ul 2M HCl and buffered to pH 7.8 by addition of 70 ul 1 M Ammonium Bicarbonate. Final digestion of proteins was performed by addition of 1 ug Trypsin (Promega) at 37 C overnight. The tryptic digest was acidified by addition of TFA (pH<3) and samples were desalted using C18 reversed phase spin columns (BioPureSPN Midi, The Nest Group) according to the protocol of the manufacturer. Dried peptides were stored at -20 C until further use. Peptides were dissolved in 0.1% formic acid prior to injection into the mass spectrometer.

**Targeted LC-MS analysis of selected histone derived peptides**

Label-free Parallel Reaction-Monitoring (PRM) assays (PMID: 22865924) were used from an earlier in-house study. The PRM assays had been generated from a shotgun data-dependent acquisition (DDA) LC-MS/MS analysis of histone enriched samples on a Q-Exactive HF platform. The setup of the µRPLC-MS system was as described previously (Pubmed-ID: 27345528). Chromatographic separation of peptides was carried out using an EASY nano-LC 1000 system (Thermo Fisher Scientific), equipped with a heated RP-HPLC column (75 µm x 30 cm) packed in-house with 1.9 µm C18 resin (Reprosil-AQ Pur, Dr. Maisch). Peptides were analyzed per LC-MS/MS run using a linear gradient ranging from 95% solvent A (0.15% formic acid, 2% acetonitrile) and 5% solvent B (98% acetonitrile, 2% water, 0.15% formic acid) to 45% solvent B over 60 minutes at a flow rate of 200 nl/min. Mass spectrometry analysis was performed on Q-Exactive HF mass spectrometer equipped with a nanoelectrospray ion source (both Thermo Fisher Scientific). Each MS1 scan was followed by high-collision-dissociation (HCD) of the 10 most abundant precursor ions with dynamic exclusion for 20 seconds. Total cycle time was approximately 1 s. For MS1, 3e6 ions were accumulated in the Orbitrap cell over a maximum time of 100 ms and scanned at a resolution of 120,000 FWHM (at 200 m/z). MS2 scans were acquired at a target setting of 1e5 ions, accumulation time of 50 ms and a resolution of 30,000 FWHM (at 200 m/z). Singly charged ions and ions with unassigned charge state were excluded from triggering MS2 events. The normalized collision energy was set to 27%, the mass isolation window was set to 1.4 m/z and one microscan was acquired for each spectrum.

The acquired raw-files were database searched against a human protein database (UniProt: download date: 07.03.2019, total of 41,592 entries) by the MaxQuant software (Version 1.3.05) using default parameters. Mono-, Di- and Trimethylation of lysine were set as dynamic modifications and d3-acetylation of lysine was set as fixed modification. All peptides identified carrying methylation together with their unmodified (d3-acetlyted) counterpart were selected for PRM assay generation. FDR rates based on decoy database searching were set to 1% for all modification types. The best 6 transitions for each peptide were selected automatically using an in-house software tool and imported to Skyline (version 19.1.0.193). The mass isolation lists containing the modified and unmodified peptides of the selected histone modification sites were exported form Skyline and imported into the QE-HF operating software for PRM analysis using the following settings: The resolution of the orbitrap was set to 60k FWHM (at 200 m/z) and the fill time was set to 110 ms to reach a target value of 3e6

ions for the peptides. Ion isolation window was set to 0.4 Th and the first mass was fixed to 100 Th. A MS1 scan using the same conditions are for DDA was included in each MS cycle. Each condition was analyzed in five biological replicates. All raw-files were imported into Skyline for peptide quantification. To control for variation in injected sample amounts, the total ion chromatogram (only comprising ions with two or more charges) of each sample was determined using LFQ (see below) and used for normalization. Importantly, all PRM data was also database searched using MaxQuant as described above. Only identified peptides were used for quantification.

**Quantitative shotgun LC-MS analysis**

After five PRM LC-MS analysis, a standard DDA LC-MS analysis of the previous sample was carried out using the same gradient and MS-parameters as described above. These samples were also included in the subsequent label-free quantification analysis to increase the number of identified and quantified proteins. The generated raw files were imported into the Progenesis QI software (Nonlinear Dynamics (Waters), Version 2.0) and analyzed using the default parameter settings. MS/MS-data were exported directly from Progenesis QI in mgf format and searched using a target-decoy strategy against a database containing the forward and reverse sequences of the human proteome (human protein database (UniProt, version 07.03.2019, total of 41,592 entries) using MASCOT (version 2.4.1). The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues); 2 missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification; oxidation (M) as variable modification. The mass tolerance was set to 10 ppm for precursor ions and 0.02 Da for-fragment ions. Results from the database search were imported into Progenesis QI and the final peptide measurement list containing the peak areas of all identified peptides, respectively, was exported. This list was further processed and statically analyzed using our in-house developed SafeQuant R script (PubMed-ID: 27345528). The peptide and protein false discovery rate (FDR) was set to 1% using the number of reverse hits in the dataset.

| TARGET | FORWARD PRIMER (5'-3') | REVERSE PRIMER (5'-3') |
|---|---|---|
| ***Vav-iCre*** | CTCTGACAGATGCCAGGACA | TGATTTCAGGGATGGACACA |
| ***Nsd1-LoxP*** | GTCTGCATTAAGTAATTGTGCCCTGAAG | ACTGACTCCTCTTCTGGAGATCTGAGTTC |
| ***Erg-F1*** | GTATGTGCAGGCAGGCAGGTGG | NA |
| ***Erg-F2*** | TGGTAAACCCCAGGCCATCGAGAGAGC | NA |
| ***EYFP-R1*** | TCGAACTTCACCTCGGCGCGGG | NA |
| ***GFP-R2*** | ACCTTGATGCCGTTCTTCTGCTTGTCGGCC | NA |
| ***Nsd1-Neo-RHA*** | CGATGCCTGCTTGCCGAATA | ATG CAT GTT CCA ACC AGG AGA |
| ***Nsd1-Neo-LHA*** | AGT TCC CAC CTT TTA GAT AGG ATC A | ATG CAT GTT CCA ACC AGG AGA |
| ***Neo*** | AGGGGATCCGCTGTAAGTCT | GCACTGACTGCTCATCCAAA |
| ***GFP*** | ATGTCAGCAATTGAGAACATGG | ATCCAAAGGTCCTGAGTTCAAA |

**Table 5: Genotyping primers**

$

| TARGET | FORWARD PRIMER (5'-3') | REVERSE PRIMER (5'-3') |
|---|---|---|
| ***Nsd1-RHA*** | GGGAAAGUGGCCTGAGACTTTCATCACCAG | GGAGACAUTTTCTTTTCTGAATCTGCACAT |
| ***Nsd1-LHA*** | GGTCCCAUTAGGAAACACAGTAAATGTCAC | GGCATAGUCTTGAAAACAGAACTGGAAACC |
| ***Nsd1 Site Directed Mutagenesis*** | CTGACCGCAGCGTACCATTTGAGGAGGGAAATC | CTTATCTAGATATGTGACATTTACATTTACTGTGTTTCC |
| ***Nsd1-SgRNA-Stop*** | CACCGACGAGTTCTTCTGAGGGGAT | AAACATCCCCTCAGAAGAACTCGTC |
| ***Nsd1-SgRNA-Ex19*** | CACCGTGTCGTGCTCTTGTGCGTAA | AAACTTACGCACAAGAGCACGACA |
| ***Nsd1-SgRNA-Ex20*** | CACCGTAGGATCGAATCATTGATGC | AAACGCATCAATGATTCGATCCTAC |

**Table 6: Cloning primers**

| ANTIBODY | CLONE | FLUOROCHROME | CONCENTRATION | SUPPLIER |
|---|---|---|---|---|
| **CD71** | C2 | PE | 1:100 | BD Biosciences |
| **TER119** | TER-119 | APC | 1:100 | BD Biosciences |
| **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 7: Antibodies for Flow cytometry analysis**

| NAME | DESCRIPTION/SOURCE |
|---|---|
| ***Nsd1 (# 508)*** | *pMSCV-mNsd1-pgk-puro-IRES-GFP* |
| ***Nsd1$^{N1918Q}$ (# 553)*** | *pMSCV-mNsd1-Setmut-pgk-puro-IRES-GFP* |
| ***Nsd1-ShRNA*** | pLMP NSD1 ShRNA-<br>Mission ShRNA (TRCN0000441097)- Sigma-Aldrich |
| ***GFP-Puro-Ctrl (#432)*** | *pMSCV-pgk-GFP-puro* |
| ***Ski-ShRNA*** | *pLMP-Ski-shRNA-GFP-Puro* |
| ***Ctrl-ShRNA*** | *pLMP-empty-shRNA-GFP-Puro* |
| ***PAAV-LOXP-Neo-3XFLAG*[217]** | A gift from Zhenghe Wang (Cleveland-USA) |
| ***PSPCAS9(BB)-2A-GFP*[297] (Px458)** | A gift from Feng Zhang (Addgene plasmid #48138) |
| ***PBS185 CMV-CRE*[307]** | A gift from Brian Sauer (Addgene plasmid #11916) |
| ***PUC57-ERG-P2A-YFP*** | Synthesized by Genescript |
| ***SGRNA-Erg.1*** | GCGCGCTAATACGACTCACTATAGGGGTCTCTAGTAGTAGGTGCC<br>GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTA<br>GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTT |
| ***SGRNA-Erg.2*** | GCGCGCTAATACGACTCACTATAGGGTCCATCTGGTCTCTAGTAGT<br>GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTA<br>GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTT |

**Table 8: Plasmids**

| TARGET | FORWARD PRIMER (5'-3') | REVERSE PRIMER (5'-3') |
|---|---|---|
| ***mGaphd*** | ATGACATCAAGAAGGTGGTG | CATACCAGGAAATGAGCTTG |
| ***mRnh1*** | TGCAGGCACTGAAGCACCA | TCCAGTGTGAGCAGCTGAG |
| ***mGusb*** | CCG ACC TCT CGA ACA ACC G | GCT TCC CGT TCA TAC CAC ACC |
| ***mActin*** | GGA CCT GAC AGA CTA CCT CAT GAA | CTG CTT GCT GAT CCA CAT CTG C |
| ***mGata1*** | GTGTCCTCACCATCAGATTCCAC | TCCCTCCATACTGTTGAGCAGTG |
| ***mPklr*** | GACCGCCTCAAGGAGATGAT | CGAATGTTGGCGATGGACTC |
| ***mArt4*** | CGCTGTGGCTTCCAGGAG | TCTGTTCGCTACAGCCTTGG |
| ***mFgf2*** | GGCTGCTGGCTTCTAAGTGT | CAACTGGAGTATTTCCGTGACC |
| ***mHb-B*** | GTCTCTTGCCTGTGGGGAAA | CAACCAGCAGCCTGCCC |
| ***mNsd1 (Ex5-ex6)*** | AAGTCCAGTGTTGGCATGGG | GGCGTTTCTTCTCTGACCGA |
| ***mNsd1 (Ex22-ex24)*** | GGCAAGACCCCAAACCAGA | CAAATGGTACCCTGGGGACAG |
| ***mNsd1 (Ex13-14)*** | TGCTTCTAAAGGTCGTCTGATGC GC | CTAGGGGTGAAGTGATTAGGGCA GA |
| ***mSpi1*** | CGATTCAGAGCTATACCAACGT CC | ACTCGTTTGTTGTGGACATGGTG |
| ***mAlas2*** | TTAGCCACTTTGCCCAGGAG | CCGTCTTTGGTTCGTCCTCA |
| ***mScl2*** | TCACCACCTCCCTGTTCTT | TTTCTTCCTGCCAAACCTG |
| ***mGypa*** | TGAAGTGTCTGCTGCGTT | CCGATAATCCCTGCCATCA |
| ***mEpb4.2*** | GCTTTATCTACCTGGGCAC | TCACTTGCTTGTCCATACTC |

**Table 9: RT-qPCR Primers**

| TARGET PROTEIN | CLONE | DILUTION | SUPPLIER |
|---|---|---|---|
| **NSD1** | 2749 | 1:1000 | A gift from Antoine Peters (FMI, Basel, Switzerland) |
| **GATA1** | D52H6 XP | 1:1000 | Cell Signaling |
| **SKI** | G-8 | 1:1000 | Santa Cruz |
| **ETO2** | C-20/sc-9739 | 1:500 | Santa Cruz |
| **LMO2** | 1A9-1 | 1:500 | Bio-rad |
| **FOG1** | A-6 | 1:500 | Santa Cruz |
| **PU.1** | C-3 | 1:1000 | Santa Cruz |
| **FLAG** | D6W5B | 1:1000 | Cell Signaling |
| **ACTIN** | C-11 | 1:6000 | Santa Cruz |
| **LAMIN-A/C** | E-1/ sc-376248 | 1:5000 | Santa Cruz |
| **h3k36me1** | Lys36-5928S | 1:1000 | Cell Signaling |
| **H3K36 me2** | C75H12-2901S | 1:1000 | Cell Signaling |
| **H3K36 me3** | D5A7-4909P | 1:1000 | Cell Signaling |
| **H3** | D1H2 | 1:1000 | Cell Signaling |

**Table 10: Antibodies used for Western blot analysis**

| ANTIBODY | CLONE | CONCENTRATION | SUPPLIER |
|---|---|---|---|
| **NSD1** | 2749 | 10μg per IP | A gift from Antoine Peters (FMI, Basel, Switzerland) |
| **FLAG** | D6W5B | 10μg per IP | Cell signaling |
| **GATA1- ChIP** | ab11862 | 10μg per IP | Abcam |
| **H3K36me3- ChIP** | ab9050-11 | 10μg per IP | Abcam |
| **H3K27ac-ChIP** | ab4729 | 10μg per IP | Abcam |
| **GATA1-IP** | N6- sc-265 | 2μg per IP | Santa Cruz |
| **IgG2A-IP** | sc-3883 | 2μg per IP | Santa Cruz |

**Table 11: Antibodies used for ChIP-seq and IP-MS analysis**

# 8. References


1. Mikkola, H.K.A. and S.H. Orkin, *The journey of developing hematopoietic stem cells.* Development, 2006. **133**(19): p. 3733.
2. Barminko, J., B. Reinholt, and M.H. Baron, *Development and differentiation of the erythroid lineage in mammals.* Dev Comp Immunol, 2016. **58**: p. 18-29.
3. Palis, J., *Primitive and definitive erythropoiesis in mammals.* Frontiers in physiology, 2014. **5**: p. 3-3.
4. Palis, J., et al., *Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse.* Development, 1999. **126**(22): p. 5073-84.
5. Van Handel, B., et al., *The first trimester human placenta is a site for terminal maturation of primitive erythroid cells.* Blood, 2010. **116**(17): p. 3321-30.
6. Tober, J., et al., *The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis.* Blood, 2007. **109**(4): p. 1433-41.
7. Park, M.A., et al., *Activation of the Arterial Program Drives Development of Definitive Hemogenic Endothelium with Lymphoid Potential.* Cell Rep, 2018. **23**(8): p. 2467-2481.
8. Easterbrook, J., et al., *Analysis of the Spatiotemporal Development of Hematopoietic Stem and Progenitor Cells in the Early Human Embryo.* Stem Cell Reports, 2019. **12**(5): p. 1056-1068.
9. Barminko, J., B. Reinholt, and M.H. Baron, *Development and differentiation of the erythroid lineage in mammals.* Developmental & Comparative Immunology, 2016. **58**: p. 18-29.
10. Eilken, H.M., S. Nishikawa, and T. Schroeder, *Continuous single-cell imaging of blood generation from haemogenic endothelium.* Nature, 2009. **457**(7231): p. 896-900.
11. Swiers, G., et al., *Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level.* Nat Commun, 2013. **4**: p. 2924.
12. Boisset, J.C. and C. Robin, *Imaging the founder of adult hematopoiesis in the mouse embryo aorta.* Cell Cycle, 2010. **9**(13): p. 2489-90.
13. Dzierzak, E. and S. Philipsen, *Erythropoiesis: development and differentiation.* Cold Spring Harb Perspect Med, 2013. **3**(4): p. a011601.
14. Ema, H. and H. Nakauchi, *Expansion of hematopoietic stem cells in the developing liver of a mouse embryo.* Blood, 2000. **95**(7): p. 2284-2288.

15. Morrison, S.J., et al., *The purification and characterization of fetal liver hematopoietic stem cells.* Proc Natl Acad Sci U S A, 1995. **92**(22): p. 10302-6.
16. Harrison, D.E., et al., *Relative to adult marrow, fetal liver repopulates nearly five times more effectively long-term than short-term.* Exp Hematol, 1997. **25**(4): p. 293-7.
17. Kay, M.M., *Mechanism of removal of senescent cells by human macrophages in situ.* Proc Natl Acad Sci U S A, 1975. **72**(9): p. 3521-5.
18. Zhang, J., et al., *Identification of the haematopoietic stem cell niche and control of the niche size.* Nature, 2003. **425**(6960): p. 836-41.
19. Orkin, S.H., *Diversification of haematopoietic stem cells to specific lineages.* Nat Rev Genet, 2000. **1**(1): p. 57-64.
20. Domen, J. and I.L. Weissman, *Self-renewal, differentiation or death: regulation and manipulation of hematopoietic stem cell fate.* Mol Med Today, 1999. **5**(5): p. 201-8.
21. Granick, S. and R.D. Levere, *Heme Synthesis in Erythroid Cells.* Prog Hematol, 1964. **4**: p. 1-47.
22. Gifford, S.C., et al., *A detailed study of time-dependent changes in human red blood cells: from reticulocyte maturation to erythrocyte senescence.* Br J Haematol, 2006. **135**(3): p. 395-404.
23. Bessis, M. and J. Breton-Gorius, *[Ferruginous granules in macrophage cells and erythrocytes in experimental saturnism; electron microscopic examination].* C R Seances Soc Biol Fil, 1957. **151**(2): p. 275-6.
24. Chasis, J.A. and N. Mohandas, *Erythroblastic islands: niches for erythropoiesis.* Blood, 2008. **112**(3): p. 470-478.
25. Hassner, A., et al., *[The monocyte-macrophage system].* Harefuah, 1981. **100**(9): p. 418-22.
26. Ema, H., Y. Morita, and T. Suda, *Heterogeneity and hierarchy of hematopoietic stem cells.* Exp Hematol, 2014. **42**(2): p. 74-82 e2.
27. Wilson, Nicola K., et al., *Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations.* Cell Stem Cell, 2015. **16**(6): p. 712-724.
28. Zhang, P., et al., *Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8705-10.

29. Adachi, I., et al., *First Evidence for cos2beta>0 and Resolution of the Cabibbo-Kobayashi-Maskawa Quark-Mixing Unitarity Triangle Ambiguity.* Phys Rev Lett, 2018. **121**(26): p. 261801.
30. Finch, S.C., *Radiation-induced leukemia: lessons from history.* Best Pract Res Clin Haematol, 2007. **20**(1): p. 109-18.
31. Thys, R.G., et al., *Environmental and chemotherapeutic agents induce breakage at genes involved in leukemia-causing gene rearrangements in human hematopoietic stem/progenitor cells.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2015. **779**: p. 86-95.
32. De Kouchkovsky, I. and M. Abdul-Hay, *'Acute myeloid leukemia: a comprehensive review and 2016 update'.* Blood Cancer Journal, 2016. **6**: p. e441.
33. Papaemmanuil, E., et al., *Genomic Classification and Prognosis in Acute Myeloid Leukemia.* N Engl J Med, 2016. **374**(23): p. 2209-2221.
34. Tricot, G. and A. Broeckaert-Van Orshoven, *8;21 Translocation in acute myeloid leukemia. An ultrastructural study.* Cancer, 1984. **53**(3): p. 453-8.
35. Lavau, C. and A. Dejean, *The t(15;17) translocation in acute promyelocytic leukemia.* Leukemia, 1994. **8**(10): p. 1615-21.
36. Almosailleakh, M. and J. Schwaller, *Murine Models of Acute Myeloid Leukaemia.* Int J Mol Sci, 2019. **20**(2).
37. Lagunas-Rangel, F.A., et al., *Acute Myeloid Leukemia-Genetic Alterations and Their Clinical Prognosis.* Int J Hematol Oncol Stem Cell Res, 2017. **11**(4): p. 328-339.
38. Fisher, J.N., et al., *The Impact of the Cellular Origin in Acute Myeloid Leukemia.* HemaSphere, 2019. **3**(1).
39. Dohner, H., D.J. Weisdorf, and C.D. Bloomfield, *Acute Myeloid Leukemia.* N Engl J Med, 2015. **373**(12): p. 1136-52.
40. Cancer Genome Atlas Research, N., et al., *Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia.* N Engl J Med, 2013. **368**(22): p. 2059-74.
41. Riether, C., C.M. Schurch, and A.F. Ochsenbein, *Regulation of hematopoietic and leukemic stem cells by the immune system.* Cell Death Differ, 2015. **22**(2): p. 187-98.
42. Vardiman, J.W., et al., *The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes.* Blood, 2009. **114**(5): p. 937-51.

43. Arber, D.A., et al., *The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia.* Blood, 2016. **127**(20): p. 2391-405.

44. Leonard, J.P., P. Martin, and G.J. Roboz, *Practical Implications of the 2016 Revision of the World Health Organization Classification of Lymphoid and Myeloid Neoplasms and Acute Leukemia.* J Clin Oncol, 2017. **35**(23): p. 2708-2715.

45. Wang, W., et al., *Pure erythroid leukemia.* Am J Hematol, 2017. **92**(3): p. 292-296.

46. Wang, S.A., et al., *Acute erythroid leukemia with <20% bone marrow blasts is clinically and biologically similar to myelodysplastic syndrome with excess blasts.* Mod Pathol, 2016. **29**(10): p. 1221-31.

47. Zuo, Z., et al., *Acute myeloid leukemia (AML) with erythroid predominance exhibits clinical and molecular characteristics that differ from other types of AML.* PLoS One, 2012. **7**(7): p. e41485.

48. Grossmann, V., et al., *Acute erythroid leukemia (AEL) can be separated into distinct prognostic subsets based on cytogenetic and molecular genetic characteristics.* Leukemia, 2013. **27**(9): p. 1940-3.

49. Valent, P., et al., *Normal and pathological erythropoiesis in adults: from gene regulation to targeted treatment concepts.* Haematologica, 2018. **103**(10): p. 1593-1603.

50. Micci, F., 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): p. 980-2.

51. Panagopoulos, I., et al., *Fusion of ZMYND8 and RELA genes in acute erythroid leukemia.* PLoS One, 2013. **8**(5): p. e63663.

52. Micci, F., et al., *Translocation t(1;16)(p31;q24) rearranging CBFA2T3 is specific for acute erythroid leukemia.* Leukemia, 2011. **25**(9): p. 1510-2.

53. Friend, C., *Cell-free transmission in adult Swiss mice of a disease having the character of a leukemia.* J Exp Med, 1957. **105**(4): p. 307-18.

54. DINAH SINGER, M.C., GEORGE M. MANIATIS, PAUL A. MARKS, AND RICHARD A. RIFKIND, *Erythropoietic Differentiation in Colonies of Cells Transformed by Friend Virus.* Proc Natl Acad Sci U S A, 1974. **71**(7): p. 2668-2670.

55. D'Andrea, P.A.N.a.A.D., *Friend erythroleukemia revisited.* Blood, 2000. **96**: p. 3675-3680.

56. Kumar, C.C., *Genetic abnormalities and challenges in the treatment of acute myeloid leukemia.* Genes Cancer, 2011. **2**(2): p. 95-107.

57. Montalban-Bravo, G., et al., *More than 1 TP53 abnormality is a dominant characteristic of pure erythroid leukemia.* Blood, 2017. **129**(18): p. 2584-2587.

58. Krivtsov, A.V., et al., *Cell of origin determines clinically relevant subtypes of MLL-rearranged AML.* Leukemia, 2013. **27**(4): p. 852-860.

59. Stavropoulou, V., et al., *MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome.* Cancer Cell, 2016. **30**(1): p. 43-58.

60. Visvader, J.E., *Cells of origin in cancer.* Nature, 2011. **469**(7330): p. 314-322.

61. Huntly, B.J., et al., *MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors.* Cancer Cell, 2004. **6**(6): p. 587-96.

62. Chi Wai So, H.K., Emmanuelle Passegue´, Antonio Cozzio, Irving L. Weissman, and Michael L. Cleary, *MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice.* Cancer cell, 2003. **3**: p. 161-171.

63. Zorko, N.A., et al., *Mll partial tandem duplication and Flt3 internal tandem duplication in a double knock-in mouse recapitulates features of counterpart human acute myeloid leukemias.* Blood, 2012. **120**(5): p. 1130-6.

64. Su, G., et al., *Aberrant expression of ecotropic viral integration site-1 in acute myeloid leukemia and acute lymphoblastic leukemia.* Leukemia & lymphoma, 2015. **56**(2): p. 472-9.

65. Goyama, S., et al., *Evi-1 is a critical regulator for hematopoietic stem cells and transformed leukemic cells.* Cell Stem Cell, 2008. **3**(2): p. 207-20.

66. Laricchia-Robbio, L. and G. Nucifora, *Significant increase of self-renewal in hematopoietic cells after forced expression of EVI1.* Blood cells, molecules & diseases, 2008. **40**(2): p. 141-7.

67. Kataoka, K., et al., *Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity*. 2011. p. 2403-2416.

68. Yuasa, H., et al., *Oncogenic transcription factor Evi1 regulates hematopoietic stem cell proliferation through GATA-2 expression.* The EMBO journal, 2005. **24**(11): p. 1976-1987.

69. Eppert, K., et al., *Stem cell gene expression programs influence clinical outcome in human leukemia.* Nature medicine, 2011. **17**(9): p. 1086-93.

70. Yeap, L.-S., K. Hayashi, and M.A. Surani, *ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage.* Epigenetics & chromatin, 2009. **2**(1): p. 12-12.

71. Ng, A.P., et al., *Erg is required for self-renewal of hematopoietic stem cells during stress hematopoiesis in mice.* Blood, 2011. **118**(9): p. 2454-2461.

72. Taoudi, S., et al., *ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification.* Genes and Development, 2011. **25**(3): p. 251-262.

73. Bock, J., et al., *ERG transcriptional networks in primary acute leukemia cells implicate a role for ERG in deregulated kinase signaling.* PLoS One, 2013. **8**(1): p. e52872.

74. Chen, T. and S.Y. Dent, *Chromatin modifiers and remodellers: regulators of cellular differentiation.* Nat Rev Genet, 2014. **15**(2): p. 93-106.

75. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.

76. Linggi, B.E., et al., *Translating the histone code into leukemia.* J Cell Biochem, 2005. **96**(5): p. 938-50.

77. Zhou, V.W., A. Goren, and B.E. Bernstein, *Charting histone modifications and the functional organization of mammalian genomes.* Nat Rev Genet, 2011. **12**(1): p. 7-18.

78. Brien, G.L., D.G. Valerio, and S.A. Armstrong, *Exploiting the Epigenome to Control Cancer-Promoting Gene-Expression Programs.* Cancer Cell, 2016. **29**(4): p. 464-476.

79. Meyer, C., et al., *The MLL recombinome of acute leukemias in 2017.* Leukemia, 2018. **32**(2): p. 273-284.

80. Caligiuri, M.A., et al., *Molecular rearrangement of the ALL-1 gene in acute myeloid leukemia without cytogenetic evidence of 11q23 chromosomal translocations.* Cancer Res, 1994. **54**(2): p. 370-3.

81. Abramovich, C. and R.K. Humphries, *Hox regulation of normal and leukemic hematopoietic stem cells.* Curr Opin Hematol, 2005. **12**(3): p. 210-6.

82. Alharbi, R.A., et al., *The role of HOX genes in normal hematopoiesis and acute leukemia.* Leukemia, 2013. **27**(5): p. 1000-8.

83. Yu, B.D., et al., *Altered Hox expression and segmental identity in Mll-mutant mice.* Nature, 1995. **378**(6556): p. 505-8.

84. Hess, J.L., et al., *Defects in yolk sac hematopoiesis in Mll-null embryos.* Blood, 1997. **90**(5): p. 1799-806.

85. Yagi, H., et al., *Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice.* Blood, 1998. **92**(1): p. 108-17.

86. Ernst, P., et al., *Definitive hematopoiesis requires the mixed-lineage leukemia gene.* Dev Cell, 2004. **6**(3): p. 437-43.

87. Argiropoulos, B. and R.K. Humphries, *Hox genes in hematopoiesis and leukemogenesis.* Oncogene, 2007. **26**(47): p. 6766-76.

88. Biswas, D., et al., *Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes.* Proc Natl Acad Sci U S A, 2011. **108**(38): p. 15751-6.

89. Chen, C.W. and S.A. Armstrong, *Targeting DOT1L and HOX gene expression in MLL-rearranged leukemia and beyond.* Exp Hematol, 2015. **43**(8): p. 673-84.

90. Cerveira, N., et al., *Frequency of NUP98-NSD1 fusion transcript in childhood acute myeloid leukaemia.* Leukemia, 2003. **17**(11): p. 2244-7.

91. Panarello, C., C. Rosanda, and C. Morerio, *Cryptic translocation t(5;11)(q35;p15.5) with involvement of the NSD1 and NUP98 genes without 5q deletion in childhood acute myeloid leukemia.* Genes Chromosomes Cancer, 2002. **35**(3): p. 277-81.

92. Jaju, R.J., et al., *A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia.* Blood, 2001. **98**(4): p. 1264-7.

93. Wang, G.G., et al., *NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis.* Nature cell biology, 2007. **9**(7): p. 804-12.

94. Thanasopoulou, A., A. Tzankov, and J. Schwaller, *Potent co-operation between the NUP98-NSD1 fusion and the FLT3-ITD mutation in acute myeloid leukemia induction.* Haematologica, 2014. **99**(9): p. 1465-71.

95. Ostronoff, F., et al., *NUP98/NSD1 and FLT3/ITD coexpression is more prevalent in younger AML patients and leads to induction failure: a COG and SWOG report.* Blood, 2014. **124**(15): p. 2400-2407.

96. Greenblatt, S., et al., *Knock-in of a FLT3/ITD mutation cooperates with a NUP98-HOXD13 fusion to generate acute myeloid leukemia in a mouse model.* Blood, 2012. **119**(12): p. 2883-94.

97. Kivioja, J.L., et al., *Dasatinib and navitoclax act synergistically to target NUP98-NSD1(+)/FLT3-ITD(+) acute myeloid leukemia.* Leukemia, 2019. **33**(6): p. 1360-1372.

98. Bennett, R.L., et al., *The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer.* Cold Spring Harb Perspect Med, 2017. **7**(6).

99. Huang, N., et al., *Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators.* EMBO J, 1998. **17**(12): p. 3398-412.

100. Kurotaki, N., et al., *Molecular characterization of NSD1, a human homologue of the mouse Nsd1 gene.* Gene, 2001. **279**(2): p. 197-204.

101. Morishita, M., D. Mevius, and E. di Luccio, *In vitro histone lysine methylation by NSD1, NSD2/MMSET/WHSC1 and NSD3/WHSC1L.* BMC Struct Biol, 2014. **14**: p. 25.

102. Pasillas, M.P., M. Shah, and M.P. Kamps, *NSD1 PHD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human sotos syndrome.* Hum Mutat, 2011. **32**(3): p. 292-8.

103. Lucio-Eterovic, A.K., et al., *Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(39): p. 16952-7.

104. Morishita, M. and E. Di Luccio, *Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1.* Biochemical and Biophysical Research Communications, 2011. **412**(2): p. 214-219.

105. Butler, J.S. and S.Y. Dent, *Chromatin 'resetting' during transcription elongation: a central role for methylated H3K36.* Nat Struct Mol Biol, 2012. **19**(9): p. 863-4.

106. Kim, A., C.M. Kiefer, and A. Dean, *Distinctive Signatures of Histone Methylation in Transcribed Coding and Noncoding Human β-Globin Sequences.* Molecular and Cellular Biology, 2007. **27**(4): p. 1271.

107. Rayasam, G.V., et al., *NSD1 is essential for early post-implantation development and has a catalytically active SET domain.* EMBO Journal, 2003. **22**(12): p. 3153-3163.

108. Kivioja, J.L., et al., *Chimeric NUP98-NSD1 transcripts from the cryptic t(5;11)(q35.2;p15.4) in adult de novo acute myeloid leukemia.* Leuk Lymphoma, 2018. **59**(3): p. 725-732.

109. La Starza, R., et al., *Cryptic insertion producing two NUP98/NSD1 chimeric transcripts in adult refractory anemia with an excess of blasts.* Genes Chromosomes Cancer, 2004. **41**(4): p. 395-9.

110. Thol, F., et al., *Analysis of NUP98/NSD1 translocations in adult AML and MDS patients.* Leukemia, 2013. **27**(3): p. 750-4.

111. de Souza, C.F., et al., *Mining gene expression signature for the detection of pre-malignant melanocytes and early melanomas with risk for metastasis.* PLoS One, 2012. **7**(9): p. e44800.

112. Brennan, K., et al., *NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma.* Sci Rep, 2017. **7**(1): p. 17064.

113. Berdasco, M., et al., *Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma.* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21830-5.

114. Bui, N., et al., *Disruption of NSD1 in Head and Neck Cancer Promotes Favorable Chemotherapeutic Responses Linked to Hypomethylation.* Mol Cancer Ther, 2018. **17**(7): p. 1585-1594.

115. Papillon-Cavanagh, S., et al., *Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas.* Nat Genet, 2017. **49**(2): p. 180-185.

116. Quintana, R.M., et al., *A transposon-based analysis of gene mutations related to skin cancer development.* J Invest Dermatol, 2013. **133**(1): p. 239-48.

117. Kurotaki, N., et al., *Haploinsufficiency of NSD1 causes Sotos syndrome.* Nat Genet, 2002. **30**(4): p. 365-6.

118. Douglas, J., et al., *NSD1 mutations are the major cause of Sotos syndrome and occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes.* Am J Hum Genet, 2003. **72**(1): p. 132-43.

119. Jaffe, J.D., et al., *Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia.* Nat Genet, 2013. **45**(11): p. 1386-91.

120. Huang, Z., et al., *NSD2 is recruited through its PHD domain to oncogenic gene loci to drive multiple myeloma.* Cancer Res, 2013. **73**(20): p. 6277-88.

121. Aytes, A., et al., *NSD2 is a conserved driver of metastatic prostate cancer progression.* Nat Commun, 2018. **9**(1): p. 5201.

122. Hudlebusch, H.R., et al., *The histone methyltransferase and putative oncoprotein MMSET is overexpressed in a large variety of human tumors.* Clin Cancer Res, 2011. **17**(9): p. 2919-33.

123. Hudlebusch, H.R., et al., *MMSET is highly expressed and associated with aggressiveness in neuroblastoma.* Cancer Res, 2011. **71**(12): p. 4226-35.

124. Rosati, R., et al., *NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15).* Blood, 2002. **99**(10): p. 3857-60.

125. Kuroda, S., et al., *Cytological Features of a Variant NUT Midline Carcinoma of the Lung Harboring the NSD3-NUT Fusion Gene: A Case Report and Literature Review.* Case Rep Pathol, 2015. **2015**: p. 572951.

126. Suzuki, S., et al., *NSD3-NUT-expressing midline carcinoma of the lung: first characterization of primary cancer tissue.* Pathol Res Pract, 2015. **211**(5): p. 404-8.

127. Angrand, P.O., et al., *NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines.* Genomics, 2001. **74**(1): p. 79-88.

128. Wang, X., et al., *Identification and characterization of a novel androgen receptor coregulator ARA267-alpha in prostate cancer cells.* J Biol Chem, 2001. **276**(44): p. 40417-23.

129. Li, Y., et al., *The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate.* J Biol Chem, 2009. **284**(49): p. 34283-95.

130. Rahman, N., *Mechanisms predisposing to childhood overgrowth and cancer.* Curr Opin Genet Dev, 2005. **15**(3): p. 227-33.

131. Choufani, S., et al., *NSD1 mutations generate a genome-wide DNA methylation signature.* Nat Commun, 2015. **6**.

132. Martin-Herranz, D.E., et al., *Screening for genes that accelerate the epigenetic ageing clock in humans reveals a role for the H3K36 methyltransferase NSD1.* bioRxiv, 2019: p. 545830.

133. Nielsen, A.L., et al., *Nizp1, a novel multitype zinc finger protein that interacts with the NSD1 histone lysine methyltransferase through a unique C2HR motif.* Mol Cell Biol, 2004. **24**(12): p. 5184-96.

134. Lu, T., et al., *Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65.* Proc Natl Acad Sci U S A, 2010. **107**(1): p. 46-51.

135. Tatton-Brown, K., et al., *Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with NSD1 aberrations.* Am J Hum Genet, 2005. **77**(2): p. 193-204.

136. Ghaffari, S., et al., *AKT induces erythroid-cell maturation of JAK2-deficient fetal liver progenitor cells and is required for Epo regulation of erythroid-cell differentiation.* Blood, 2006. **107**(5): p. 1888-91.

137. Wierenga, A.T., E. Vellenga, and J.J. Schuringa, *Down-regulation of GATA1 uncouples STAT5-induced erythroid differentiation from stem/progenitor cell proliferation.* Blood, 2010. **115**(22): p. 4367-76.

138. Moriguchi, T. and M. Yamamoto, *A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation.* Int J Hematol, 2014. **100**(5): p. 417-24.

139. Trainor, C.D., et al., *A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction.* Mol Cell Biol, 1996. **16**(5): p. 2238-47.

140. Tsang, A.P., 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): p. 109-119.

141. Weiss, M.J. and S.H. Orkin, *GATA transcription factors: key regulators of hematopoiesis.* Exp Hematol, 1995. **23**(2): p. 99-107.

142. Zermati, Y., et al., *Caspase activation is required for terminal erythroid differentiation.* The Journal of experimental medicine, 2001. **193**(2): p. 247-254.

143. Martin, D.I., et al., *Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages.* Nature, 1990. **344**(6265): p. 444-7.

144. Pevny, L., et al., *Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1.* Nature, 1991. **349**(6306): p. 257-60.

145. Weiss, M.J., G. Keller, and S.H. Orkin, *Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells.* Genes Dev, 1994. **8**(10): p. 1184-97.

146. Fujiwara, Y., et al., *Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1.* Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12355-8.

147. Weiss, M.J. and S.H. Orkin, *Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis.* Proc Natl Acad Sci U S A, 1995. **92**(21): p. 9623-7.

148. Gutiérrez, L., et al., *Ablation of Gata1 in adult mice results in aplastic crisis, revealing its essential role in steady-state and stress erythropoiesis.* Blood, 2008. **111**(8): p. 4375-4385.

149. Crispino, J.D. and M.S. Horwitz, *GATA factor mutations in hematologic disease.* Blood, 2017. **129**(15): p. 2103-2110.

150. Tsai, F.Y., et al., *An early haematopoietic defect in mice lacking the transcription factor GATA-2.* Nature, 1994. **371**(6494): p. 221-6.

151. de Pater, E., et al., *Gata2 is required for HSC generation and survival.* The Journal of Experimental Medicine, 2013. **210**(13): p. 2843-2850.

152. *Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis.* 1999. **93**(2): p. 488-99.

153. Rodrigues, N.P., et al., *Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis.* Blood, 2005. **106**(2): p. 477-84.

154. Lurie, L.J., et al., *Differential GATA factor stabilities: implications for chromatin occupancy by structurally similar transcription factors.* Biochemistry, 2008. **47**(3): p. 859-69.

155. Kazenwadel, J., et al., *Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature.* Blood, 2012. **119**(5): p. 1283-91.

156. Ting, C.N., et al., *Transcription factor GATA-3 is required for development of the T-cell lineage.* Nature, 1996. **384**(6608): p. 474-8.

157. Garcia-Ojeda, M.E., et al., *GATA-3 promotes T-cell specification by repressing B-cell potential in pro-T cells in mice.* Blood, 2013. **121**(10): p. 1749-59.

158. Lee, G.R., P.E. Fields, and R.A. Flavell, *Regulation of IL-4 Gene Expression by Distal Regulatory Elements and GATA-3 at the Chromatin Level.* Immunity, 2001. **14**(4): p. 447-459.

159. Zaidan, N. and K. Ottersbach, *The multi-faceted role of Gata3 in developmental haematopoiesis.* Open Biol, 2018. **8**(11).

160. Chlon, T.M. and J.D. Crispino, *Combinatorial regulation of tissue specification by GATA and FOG factors.* Development, 2012. **139**(21): p. 3905-16.

161. Tsang, A.P., et al., *Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG.* Genes Dev, 1998. **12**(8): p. 1176-88.

162. Cantor, A.B. and S.H. Orkin, *Coregulation of GATA factors by the Friend of GATA (FOG) family of multitype zinc finger proteins.* Semin Cell Dev Biol, 2005. **16**(1): p. 117-28.

163. Johnson, K.D., et al., *Friend of GATA-1-independent transcriptional repression: a novel mode of GATA-1 function.* Blood, 2007. **109**(12): p. 5230-3.

164. Wadman, I.A., 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): p. 3145-57.

165. Mukhopadhyay, M., et al., *Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation.* Development, 2003. **130**(3): p. 495-505.

166. Yamada, Y., et al., *The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis.* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3890-5.

167. Cross, A.J., et al., *LIM domain binding proteins 1 and 2 have different oligomeric states.* J Mol Biol, 2010. **399**(1): p. 133-44.

168. Xue, Y., et al., *NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities.* Molecular Cell, 1998. **2**(6): p. 851-861.

169. Hong, W., et al., *FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1.* The EMBO journal, 2005. **24**(13): p. 2367-2378.

170. Miccio, A., et al., *NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development.* EMBO J, 2010. **29**(2): p. 442-56.

171. Gregory, G.D., et al., *FOG1 requires NuRD to promote hematopoiesis and maintain lineage fidelity within the megakaryocytic-erythroid compartment.* Blood, 2010. **115**(11): p. 2156-66.

172. Oda, H., et al., *Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development.* Mol Cell Biol, 2009. **29**(8): p. 2278-95.

173. Malik, J., et al., *The Methyltransferase Setd8 Is Essential for Erythroblast Survival and Maturation.* Cell Reports, 2017. **21**(9): p. 2376-2383.

174. Malik, J., M. Getman, and L.A. Steiner, *Histone methyltransferase Setd8 represses Gata2 expression and regulates erythroid maturation.* Mol Cell Biol, 2015. **35**(12): p. 2059-72.

175. DeVilbiss, A.W., et al., *Epigenetic Determinants of Erythropoiesis: Role of the Histone Methyltransferase SetD8 in Promoting Erythroid Cell Maturation and Survival.* Molecular and Cellular Biology, 2015. **35**(12): p. 2073-2087.

176. Hasegawa, A. and R. Shimizu, *GATA1 Activity Governed by Configurations of cis-Acting Elements.* Frontiers in Oncology, 2017. **6**(269).

177. Kerenyi, M.A. and S.H. Orkin, *Networking erythropoiesis.* J Exp Med, 2010. **207**(12): p. 2537-41.

178. Cantor, A.B., S.G. Katz, and S.H. Orkin, *Distinct domains of the GATA-1 cofactor FOG-1 differentially influence erythroid versus megakaryocytic maturation.* Mol Cell Biol, 2002. **22**(12): p. 4268-79.

179. Pevny, L., et al., *Development of hematopoietic cells lacking transcription factor GATA-1.* Development, 1995. **121**(1): p. 163-72.

180. Shimizu, R., et al., *Leukemogenesis Caused by Incapacitated GATA-1 Function.* Molecular and Cellular Biology, 2004. **24**(24): p. 10814-10825.

181. Sankaran, V.G., et al., *Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia.* J Clin Invest, 2012. **122**(7): p. 2439-43.

182. Nichols, K.E., et al., *Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1.* Nat Genet, 2000. **24**(3): p. 266-70.

183. Mehaffey, M.G., et al., *X-linked thrombocytopenia caused by a novel mutation of GATA-1.* Blood, 2001. **98**(9): p. 2681-8.

184. Freson, K., et al., *Different substitutions at residue D218 of the X-linked transcription factor GATA1 lead to altered clinical severity of macrothrombocytopenia and anemia and are associated with variable skewed X inactivation.* Hum Mol Genet, 2002. **11**(2): p. 147-52.

185. Balduini, C.L., et al., *Effects of the R216Q mutation of GATA-1 on erythropoiesis and megakaryocytopoiesis.* Thromb Haemost, 2004. **91**(1): p. 129-40.

186. Yu, C., et al., *X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction.* Blood, 2002. **100**(6): p. 2040-2045.

187. Campbell, A.E., et al., *Analysis of disease-causing GATA1 mutations in murine gene complementation systems.* Blood, 2013. **121**(26): p. 5218-27.

188. England, S.J., et al., *Immature erythroblasts with extensive ex vivo self-renewal capacity emerge from the early mammalian fetus.* 2011. **117**(9): p. 2708-2717.

189. Friend, C., *Cell-free transmission in adult swiss mice of a disease having the character of a leukemia* The Journal of experimental medicine, 1957. **105**: p. 307-326.

190. CHARLOTTE FRIEND, W.S., J. G. HOLLAND, AND TORU 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, 1971. **68**(2): p. 378-382.

191. von Lindern, M., *Cell-cycle control in erythropoiesis.* Blood, 2006. **108**(3): p. 781.

192. Upadhyay, G., et al., *Antagonistic actions of Rcor proteins regulate LSD1 activity and cellular differentiation.* Proc Natl Acad Sci U S A, 2014. **111**(22): p. 8071-6.

193. Wong, P., et al., *Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes.* Blood, 2011. **118**(16): p. e128-e138.

194. Hsieh, F.F., et al., *Cell cycle exit during terminal erythroid differentiation is associated with accumulation of p27 Kip1 and inactivation of cdk2 kinase.* 2019. **96**(8): p. 2746-2755.

195. Laurent, B., et al., *High-mobility group protein HMGB2 regulates human erythroid differentiation through trans-activation of GFI1B transcription.* Blood, 2010. **115**(3): p. 687-95.

196. Liu, J., et al., *ZFP36L2, a novel AML1 target gene, induces AML cells apoptosis and inhibits cell proliferation.* Leuk Res, 2018. **68**: p. 15-21.

197. Zhang, L., et al., *ZFP36L2 is required for self-renewal of early burst-forming unit erythroid progenitors.* Nature, 2013. **499**(7456): p. 92-6.

198. Welch, J.J., et al., *Global regulation of erythroid gene expression by transcription factor GATA-1.* Blood, 2004. **104**(10): p. 3136-3147.

199. Merryweather-Clarke, A.T., et al., *Global gene expression analysis of human erythroid progenitors.* Blood, 2011. **117**(13): p. e96-108.

200. Ebert, B.L., et al., *An erythroid differentiation signature predicts response to lenalidomide in myelodysplastic syndrome.* PLoS Med, 2008. **5**(2): p. e35.

201. Wouters, B.J. and R. Delwel, *Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia.* Blood, 2016. **127**(1): p. 42-52.

202. Ntziachristos, P., O. Abdel-Wahab, and I. Aifantis, *Emerging concepts of epigenetic dysregulation in hematological malignancies.* Nature immunology, 2016. **17**(9): p. 1016-1024.

203. Schulz, V.P., et al., *A Unique Epigenomic Landscape Defines Human Erythropoiesis.* Cell Reports, 2019. **28**(11): p. 2996-3009.e7.

204. Triviai, I., et al., *ASXL1/EZH2 mutations promote clonal expansion of neoplastic HSC and impair erythropoiesis in PMF.* Leukemia, 2019. **33**(1): p. 99-109.

205. Neo, W.H., et al., *Cell-extrinsic hematopoietic impact of Ezh2 inactivation in fetal liver endothelial cells.* Blood, 2018. **131**(20): p. 2223-2234.

206. Garcia, B.A., et al., *Chemical derivatization of histones for facilitated analysis by mass spectrometry.* Nature protocols, 2007. **2**(4): p. 933-938.

207. Ferrari, K.J., et al., *Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity.* Mol Cell, 2014. **53**(1): p. 49-62.

208. Rayasam, G.V., et al., *NSD1 is essential for early post-implantation development and has a catalytically active SET domain.* EMBO J, 2003. **22**(12): p. 3153-63.

209. Qiao, Q., et al., *The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation.* J Biol Chem, 2011. **286**(10): p. 8361-8.

210. Gubin, A.N., et al., *Identification of the dombrock blood group glycoprotein as a polymorphic member of the ADP-ribosyltransferase gene family.* Blood, 2000. **96**(7): p. 2621-7.

211. Aizawa, S., et al., *Ineffective erythropoiesis in mutant mice with deficient pyruvate kinase activity.* Exp Hematol, 2005. **33**(11): p. 1292-8.

212. Bartunek, P., et al., *bFGF signaling and v-Myb cooperate in sustained growth of primitive erythroid progenitors.* Oncogene, 2002. **21**(3): p. 400-10.

213. Rodriguez, P., et al., *GATA-1 forms distinct activating and repressive complexes in erythroid cells.* EMBO J, 2005. **24**(13): p. 2354-66.

214. Ueki, N., L. Zhang, and M.J. Hayman, *Ski Negatively Regulates Erythroid Differentiation through Its Interaction with GATA1.* MOLECULAR AND CELLULAR BIOLOGY, 2004. **24**(23): p. 10118-10125.

215. CHARLOTTE FRIEND, W.S., J. G. HOLLAND, AND TORU SATO, *Hemoglobin Synthesis in Murine Virus Induced Leukemic Cells In Vitro: Stimulation of Erythroid Differentiation by Dimethyl Sulfoxide.* PNAS, 1971. **68**(2): p. 378-382.

216. Ningwu Huang1, E.v.B., Jean-Marie Garnier, Thierry Lerouge, Jean-Luc Vonesch, Yves Lutz, P.Chambon3 and Re ́gine Losson, *Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors.* The EMBO journal, 1998. **17**(12): p. 3398–3412.

217. Zhang, X., et al., *Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies.* Nature methods, 2008. **5**(2): p. 163-165.

218. Kim, J.S., et al., *Epitope tagging of endogenous genes in diverse human cell lines.* Nucleic Acids Research, 2008. **36**(19).

219. Yang, L., et al., *CRISPR/Cas9-Directed Genome Editing of Cultured Cells.* Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.], 2014. **107**: p. 31.1.1-31.1.17.

220. Sakuma, T., et al., *Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system.* Scientific reports, 2014. **4**: p. 5400-5400.

221. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nature protocols, 2013. **8**(11): p. 2281-308.

222. Kuo YH, Z.S., Gornostaeva S, Komori T, Stein GS, Castilla LH. , *Runx2 induces acute myeloid leukemia in cooperation with Cbfß-SMMHC in mice.* Blood, 2009. **113**(14): p. 3323–3332. .

223. Beer, P.A., et al., *Two routes to leukemic transformation after a JAK2 mutation-positive myeloproliferative neoplasm.* Blood, 2010. **115**(14): p. 2891-900.

224. Heshusius, S., et al., *Large-scale in-vitro production of red blood cells from human peripheral blood mononuclear cells.* bioRxiv, 2019: p. 659862.

225. Migliaccio, G., et al., *Humanized Culture Medium for Clinical Expansion of Human Erythroblasts.* Cell Transplantation, 2010. **19**(4): p. 453-469.

226. Poldee, S., et al., *Optimization of an erythroid culture system to reduce the cost of in vitro production of red blood cells.* MethodsX, 2018. **5**: p. 1626-1632.

227. Uchida, N., et al., *Serum-free Erythroid Differentiation for Efficient Genetic Modification and High-Level Adult Hemoglobin Production.* Molecular therapy. Methods & clinical development, 2018. **9**: p. 247-256.

228. Kelly, T.K., D.D. De Carvalho, and P.A. Jones, *Epigenetic modifications as therapeutic targets.* Nat Biotechnol, 2010. **28**(10): p. 1069-78.

229. Saugier-Veber, P., et al., *Heterogeneity of NSD1 alterations in 116 patients with Sotos syndrome.* Hum Mutat, 2007. **28**(11): p. 1098-107.

230. Grass, J.A., et al., *Distinct functions of dispersed GATA factor complexes at an endogenous gene locus.* Mol Cell Biol, 2006. **26**(19): p. 7056-67.

231. Lara-Astiaso, D., et al., *Chromatin state dynamics during blood formation.* Science, 2014. **345**(6199): p. 943-949.

232. Gerber, M. and A. Shilatifard, *Transcriptional elongation by RNA polymerase II and histone methylation.* J Biol Chem, 2003. **278**(29): p. 26303-6.

233. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nat Genet, 2008. **40**(7): p. 897-903.

234. Malik, J., et al., *The Methyltransferase Setd8 Is Essential for Erythroblast Survival and Maturation.* Cell Rep, 2017. **21**(9): p. 2376-2383.

235. Kudithipudi, S., et al., *Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1.* Chemistry and Biology, 2014. **21**(2): p. 226-237.

236. Tatton-Brown, K. and N. Rahman, *The NSD1 and EZH2 overgrowth genes, similarities and differences.* Am J Med Genet C Semin Med Genet, 2013. **163C**(2): p. 86-91.

237. Lund, K., P.D. Adams, and M. Copland, *EZH2 in normal and malignant hematopoiesis.* Leukemia, 2014. **28**(1): p. 44-49.

238. Mochizuki-Kashio, M., et al., *Ezh2 loss in hematopoietic stem cells predisposes mice to develop heterogeneous malignancies in an Ezh1-dependent manner.* Blood, 2015. **126**(10): p. 1172-83.

239. Gao, L., et al., *Higher expression levels of the HOXA9 gene, closely associated with MLL-PTD and EZH2 mutations, predict inferior outcome in acute myeloid leukemia.* Onco Targets Ther, 2016. **9**: p. 711-22.

240. Popovic, R., et al., *Histone methyltransferase MMSET/NSD2 alters EZH2 binding and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation.* PLoS Genet, 2014. **10**(9): p. e1004566.

241. Streubel, G., et al., *The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells.* Mol Cell, 2018. **70**(2): p. 371-379 e5.

242. Stopka, T., et al., *PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure.* EMBO J, 2005. **24**(21): p. 3712-23.

243. Weinberg, D.N., et al., *The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape.* Nature, 2019. **573**(7773): p. 281-286.

244. Shearstone, J.R., et al., *Global DNA demethylation during mouse erythropoiesis in vivo.* Science (New York, N.Y.), 2011. **334**(6057): p. 799-802.

245. Yu, Y., et al., *High resolution methylome analysis reveals widespread functional hypomethylation during adult human erythropoiesis.* J Biol Chem, 2013. **288**(13): p. 8805-14.

246. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.

247. Stadhouders, R., et al., *Control of developmentally primed erythroid genes by combinatorial co-repressor actions.* Nature Communications, 2015. **6**: p. 8893.

248. Li, Y., et al., *Unique sequence, ski, in Sloan-Kettering avian retroviruses with properties of a new cell-derived oncogene.* J Virol, 1986. **57**(3): p. 1065-72.

249. Colmenares, C., et al., *Activation of the c-ski oncogene by overexpression.* J Virol, 1991. **65**(9): p. 4929-35.

250. Tecalco-Cruz, A.C., et al., *Transcriptional cofactors Ski and SnoN are major regulators of the TGF-β/Smad signaling pathway in health and disease.* Signal Transduction and Targeted Therapy, 2018. **3**(1): p. 15.

251. Ritter, M., et al., *Inhibition of retinoic acid receptor signaling by Ski in acute myeloid leukemia.* Leukemia, 2006. **20**(3): p. 437-43.

252. Teichler, S., et al., *MicroRNA29a regulates the expression of the nuclear oncogene Ski.* Blood, 2011. **118**(7): p. 1899-902.

253. Feld, C., et al., *Combined cistrome and transcriptome analysis of SKI in AML cells identifies SKI as a co-repressor for RUNX1.* Nucleic Acids Res, 2018. **46**(7): p. 3412-3428.

254. Muench, D.E., et al., *SKI controls MDS-associated chronic TGF-beta signaling, aberrant splicing, and stem cell fitness.* Blood, 2018. **132**(21): p. e24-e34.

255. Deheuninck, J. and K. Luo, *Ski and SnoN, potent negative regulators of TGF-beta signaling.* Cell Res, 2009. **19**(1): p. 47-57.

256. Krystal, G., et al., *Transforming growth factor beta 1 is an inducer of erythroid differentiation.* J Exp Med, 1994. **180**(3): p. 851-60.

257. Zermati, Y., et al., *Transforming growth factor inhibits erythropoiesis by blocking proliferation and accelerating differentiation of erythroid progenitors.* Exp Hematol, 2000. **28**(8): p. 885-94.

258. Ueki, N., L. Zhang, and M.J. Hayman, *Ski negatively regulates erythroid differentiation through its interaction with GATA1.* Mol Cell Biol, 2004. **24**(23): p. 10118-25.

259. Ueki, N., L. Zhang, and M.J. Hayman, *Ski can negatively regulates macrophage differentiation through its interaction with PU.1.* Oncogene, 2008. **27**(3): p. 300-7.

260. Rekhtman, N., et al., *Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells.* Genes Dev, 1999. **13**(11): p. 1398-411.

261. Cmarik, J. and S. Ruscetti, *Friend Spleen Focus-Forming Virus Activates the Tyrosine Kinase sf-Stk and the Transcription Factor PU.1 to Cause a Multi-Stage Erythroleukemia in Mice.* Viruses, 2010. **2**(10): p. 2235-57.

262. Govinda Rao, N.R., Genhong Cheng1, Tatiana Krasikov and Arthur I Skoultchi, *Deregulated expression of the PU.1 transcription factor blocks murine erythroleukemia cell terminal differentiation.* Oncogene, 1997. **14**: p. 123-131.

263. Choe, K.S., et al., *Reversal of tumorigenicity and the block to differentiation in erythroleukemia cells by GATA-1.* Cancer Res, 2003. **63**(19): p. 6363-9.

264. Hyman, T., et al., *Structural characterization of erythroid and megakaryocytic differentiation in Friend erythroleukemia cells.* Experimental Hematology, 2001. **29**(5): p. 563-571.

265. Chou, S.T., et al., *Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate.* Blood, 2009. **114**(5): p. 983-94.

266. Chou, S.T., et al., *Trisomy 21 enhances human fetal erythro-megakaryocytic development.* Blood, 2008. **112**(12): p. 4503-4506.

267. Fang, D., et al., *The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas.* Science, 2016. **352**(6291): p. 1344-8.

268. Lu, C., et al., *Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape.* Science, 2016. **352**(6287): p. 844-849.

269. Yuan, W., et al., *H3K36 methylation antagonizes PRC2-mediated H3K27 methylation.* J Biol Chem, 2011. **286**(10): p. 7983-9.

270. Ellison, J., *Gene symbol: NSD1. Disease: Sotos syndrome.* Hum Genet, 2008. **124**(3): p. 311.

271. Bogdanova, A., et al., *Calcium in red blood cells-a perilous balance.* Int J Mol Sci, 2013. **14**(5): p. 9848-72.

272. Makhro, A., et al., *N-methyl-D-aspartate receptors in human erythroid precursor cells and in circulating red blood cells contribute to the intracellular calcium regulation.* Am J Physiol Cell Physiol, 2013. **305**(11): p. C1123-38.

273. Stavropoulou, V., et al., *A Novel Inducible Mouse Model of MLL-ENL-driven Mixed-lineage Acute Leukemia.* HemaSphere, 2018.

274. ROBERT K. SLANY, C.L., AND MICHAEL L. CLEARY, *The Oncogenic Capacity of HRX-ENL Requires the Transcriptional Transactivation Activity of ENL and the DNA Binding Motifs of HRX.* MOLECULAR AND CELLULAR BIOLOGY, 1998. **18**(1): p. 122–129.

275. Stam, R.W., et al., *Gene expression profiling-based dissection of MLL translocated and MLL germline acute lymphoblastic leukemia in infants.* Blood, 2010. **115**(14): p. 2835-44.

276. Weili Chen1, A.R.K., Wendy A. Hudson, Quanzhi Li, Baolin Wu, Rodney A. Staggs, Erik A. Lund, Thien N. Sam, and John H. Kersey, *Malignant transformation initiated by Mll-AF9: Gene dosage and critical target cells.* Cancer Cell. **13**(5): p. 432-440.

277. Cano, F., et al., *Leukaemia lineage specification caused by cell-specific Mll-Enl translocations.* Oncogene, 2008. **27**(13): p. 1945-50.

278. Drynan, L.F., et al., *Mll fusions generated by Cre-loxP-mediated de novo translocations can induce lineage reassignment in tumorigenesis.* EMBO J, 2005. **24**(17): p. 3136-46.

279. Forster, A., et al., *Chromosomal translocation engineering to recapitulate primary events of human cancer.* Cold Spring Harb Symp Quant Biol, 2005. **70**: p. 275-82.

280. Ugale, A., et al., *Hematopoietic stem cells are intrinsically protected against MLL-ENL-mediated transformation.* Cell Rep, 2014. **9**(4): p. 1246-55.

281. Yokoyama, A., et al., *A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription.* Cancer Cell, 2010. **17**(2): p. 198-212.

282. Liang, K., et al., *Therapeutic Targeting of MLL Degradation Pathways in MLL-Rearranged Leukemia.* Cell, 2017. **168**(1-2): p. 59-72 e13.

283. Shao, Y., et al., *CRISPR/Cas-mediated genome editing in the rat via direct injection of one-cell embryos.* Nature protocols, 2014. **9**(10): p. 2493-512.

284. Fujii, W., et al., *One-step Generation of Phenotype-expressing Triple-knockout Mice with Heritable Mutated Alleles by the CRISPR/Cas9 System.* The Journal of reproduction and development, 2014. **60**(4): p. 324-7.

285. Nagai, Y., et al., *Toll-like receptors on hematopoietic progenitor cells stimulate innate immune system replenishment.* Immunity, 2006. **24**(6): p. 801-12.

286. Motoike, T., et al., *Universal GFP reporter for the study of vascular development.* Genesis, 2000. **28**(2): p. 75-81.

287. Chen, J.Y., et al., *Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche.* Nature, 2016. **530**: p. 223.

288. Stavropoulou, V., et al., *MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome.* Cancer cell, 2016.

289. Skarnes, W.C., et al., *A conditional knockout resource for the genome-wide study of mouse gene function.* Nature, 2011. **474**(7351): p. 337-42.

290. Wang, H., et al., *One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering.* Cell, 2013. **153**(4): p. 910-918.

291. Skipper, K.A. and J.G. Mikkelsen, *Toward In Vivo Gene Therapy Using CRISPR.* Methods Mol Biol, 2019. **1961**: p. 293-306.

292. Yao, X., et al., *Tild-CRISPR Allows for Efficient and Precise Gene Knockin in Mouse and Human Cells.* Dev Cell, 2018. **45**(4): p. 526-536 e5.

293. Singh, K., et al., *Efficient In Vivo Liver-Directed Gene Editing Using CRISPR/Cas9.* Mol Ther, 2018. **26**(5): p. 1241-1254.

294. Ohtsuka, M., et al., *i-GONAD: a robust method for in situ germline genome engineering using CRISPR nucleases.* Genome Biol, 2018. **19**(1): p. 25.

295. Wang, Y., et al., *2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori.* Scientific reports, 2015. **5**: p. 16273-16273.

296. Georgiades, P., et al., *VavCre transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages.* Genesis, 2002. **34**(4): p. 251-6.

297. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-2308.

298. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.

299. Gaidatzis, D., et al., *QuasR: quantification and annotation of short reads in R.* Bioinformatics, 2015. **31**(7): p. 1130-2.

300. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

301. Zhu, A., J.G. Ibrahim, and M.I. Love, *Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences.* Bioinformatics, 2019. **35**(12): p. 2084-2092.

302. Stevens, W.R., Jr., et al., *Automated event detection algorithm for two squatting protocols.* Gait Posture, 2018. **59**: p. 253-257.

303. Riz, I., et al., *TLX1/HOX11-induced hematopoietic differentiation blockade.* Oncogene, 2007. **26**(28): p. 4115-23.

304. Kingsley, P.D., et al., *Ontogeny of erythroid gene expression.* Blood, 2013. **121**(6): p. e5-e13.

305. Glatter, T., et al., *Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion.* J Proteome Res, 2012. **11**(11): p. 5145-56.

306. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.

307. Sauer, B. and N. Henderson, *Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase.* New Biol, 1990. **2**(5): p. 441-9.

# 9. Acknowledgements

I would like first to say that in order to thank by name everyone who helped, supported and assisted me throughout my PhD, and generally during my time in Basel, I would need a write another 100 pages. As this is not possible, I will just mention the people who made this work possible, and my life in Basel enjoyable. First of all, I would like to thank **Prof. Schwaller, Jürg**, for giving me a second chance to start a PhD, and for all his guidance, mentorships and support. I know it was not easy for him to deal with my spontaneous and spur of the moment nature, nor trying and install in me some Swiss discipline and organization. Nevertheless, the fact that I'm writing this thesis, means that our partnership has somehow worked, and that maybe his efforts have not been in vain. Thank you for the opportunities, for how much you care, for always finding the time to talk and discuss, and for all that I have learnt from you.

The second person I absolutely need to thank is the wonderful **Sabine**. Thank you very much for all that you do for the group, from organizing, ordering and helping with experiments to having to listen to our complains and problems and solving them. You are the force behind the scene and I could not have done it without you. I would like also to thanks all the Childhood leukemia's past, current, visiting and permanent lab members. **Maria** & **Matheus**, thanks a lot for your help and for the most enjoyable times in and out of the lab. **Frederik** and **Samantha**, I am really grateful for all your contributions in making this work possible. **Hugues Etienne & Amber**, I had great pleasure in our discussions, science or none science, and all the ice-creams by the Rhine.

I would also like to extend my thanks to the past and current members of the Stem cells and hematopoiesis, and the experimental hematology groups. It been amazing sharing the lab space with you, as it has led not only to great scientific exchange, but also great friendships. **Anna** and **Jakub,** thank you both for your inputs and for always finding the time to brain storm, troubleshoot, or just chat.

My special thanks go to my PhD Committee members, **Prof. Antoine Peters**, and **Prof. Timm Schroeder**. I am in dept to you both for taking the time to guide me during my PhD and all your inputs.

Lastly, I would like to thank my friends, **Aizhan**, **Jannik**, **Lina**, **Fran & little Omar**. My life has been richer by meeting you, and without your friendship this would not have been possible.

# 10. List of abbreviations

| | |
|---|---|
| **a.a.** | Amino acid |
| **ac** | Acetylation |
| **AF9** | ALL1 Fused gene from chromosome 9 |
| **bp** | Base pairs |
| **CBP-α** | CCAAT/enhancer-binding protein alpha |
| **CBP-β** | CCAAT/enhancer-binding protein beta |
| **Ch** | Chromosome |
| **AEL** | Acute erythroleukemia |
| **AGM** | Aorta-Gonad-Mesonephros region |
| **ALL** | Acute Lymphoid Leukemia |
| **AML** | Acute Myeloid Leukemia |
| **AMKL** | Acute megakaryoblastic leukemia |
| **APL** | Acute promyelocytic Leukemia |
| **Baso-EB** | Basophilic Erythroblast |
| **BFU-E** | Burst-Forming-Unit-Erythroid |
| **BM** | Bone Marrow |
| **Cas9** | CRISPR Associated Protein 9 |
| **CFU-E** | Colony-forming Unit Erythroid |
| **ChIP** | Chromatin immunoprecipitation |
| **ChIP-seq** | Chromatin Immunoprecipitation with high throughput sequencing |
| **Chr** | Chromosome |
| **CLP** | Common Lymphoid Progenitor cell |
| **CML** | Chronic Myeloid Leukemia |
| **CMP** | Common Myeloid Leukemia |
| **CN-AML** | AML with Normal Karyotype |
| **Co-IP** | Co-immunoprecipitation |
| **CRISPR** | Clustered Regularly Interspaced Short Palindromic Repeats |
| **CpGme** | Methylated CG islands |
| **C-ZF** | C-terminal Zinc Finger |
| **DEG** | Differential Gene Expression |
| **DMSO** | Dimethyl sulfoxide |
| **DOX** | Doxycycline |
| **E** | Embryonic day |
| **ENL** | Eleven Nineteen Leukemia |
| **Exon** | Ex |
| **FDR** | False Discovery Rate |
| **FG** | Phenylalanine, Leucine |
| **FL** | Fetal Liver |
| **FLT3** | Fms-Related Tyrosine Kinase 3 |
| **FSC-A** | Forward Scatter Area |
| **GE** | Genome Editing |
| **GMP** | Granulocytes/Macrophage Progenitor cell |
| **GSEA** | Gene Set Enrichment Analysis |
| **H3K27** | Histone 3, Lysine 27 |
| **H3K36** | Histone 3, Lysine 36 |
| **H3K4** | Histone 3, Lysine 4 |

| | |
|---|---|
| **H3K79** | Histone 3, Lysine 79 |
| **H3K9** | Histone 3, Lysine 9 |
| **HA** | Human influenza hemagglutin |
| **HAT** | Histone acetlytransferase |
| **Hb-B** | Hemoglobin Beta (□) |
| **HDAC** | Histone deacetylase |
| **HBMA** | Hexamethylene bisacetamide |
| **HKDM** | histone lysine demethylase |
| **HKMT** | Histone Lysine Methyltransferase |
| **HSC** | Hematopoietic Stem Cells |
| **IDT** | Internal Tandem Duplication |
| **IP** | Immunoprecipitation |
| **ITR** | Inverted Tandem Repeats |
| **MC** | Methylcellulose |
| **MS** | Mass Spectrometry |
| **KO** | Knockout |
| **LHA** | Left Homology Arm |
| **LPS** | Lipopolysaccharides |
| **LSC** | Leukemic Stem Cell |
| **LSK** | Lineage-negative (Lin-), Sca-1+, Kit+ |
| **LT-HSC** | Long-term hematopoietic Stem cells |
| **me1** | Mono-methylation |
| **me2** | Di-methylation |
| **me3** | Tri-methylation |
| **MEL** | Murine erythroleukemia cell line |
| **mESC** | Murine Embryonic Stem Cells |
| **MEP** | Megakaryocyte/erythrocyte Progenitor cell |
| **MSC** | Multiple Cloning Site |
| **MDF** | Modified |
| **MDS** | Myelodysplastic Syndrome |
| **MM** | Multiple Myeloma |
| **MLL** | Mixed Lineage Leukemia |
| **MPN** | Myeloproliferative Neoplasm |
| **MUT** | Mutated |
| **Neo$^{R}$** | Neomycin Resistance |
| **NUP98** | Nucleoporin 98 |
| **MPP** | Multipotent progenitor |
| **NMDA** | N-methyl-D-aspartate |
| **N-ZF** | N-terminal Zinc Finger |
| **Ortho-EB** | Orthochromatic erythroblast |
| **PEL** | Pure Erythroleukemia |
| **P.I.** | Propidium Iodide |
| **Poly-EB** | polychromatophilic Erythroblast |
| **Pro-EB** | Proerythroblast |
| **PTD** | Partial Tandem Duplications |
| **PTM** | Post-translational Modification |
| **RHA** | Right Homology Arm |
| **SC** | Single Cell |
| **SDM** | Site Directed Mutagensis |

| | |
|---|---|
| **sgRNA** | Single-Guided RNA |
| **ST-HSC** | Short-term hematopoietic stem cells |
| **RNAPII** | RNA Polymerase II |
| **RNA-seq** | RNA high-throughput sequencing |
| **RUNX1** | Runt-Related Transcription Factor 1 |
| **rtTA** | Reverse tet-responsive trans activator |
| **SEM** | Standard Error of the Mean |
| **shRNA** | Short hairpin RNA |
| **SSC-A** | Side Scatter Area |
| **TF** | Transcription Factor |
| **TKD** | Tyrosine Kinase domain |
| **UTR** | Untranslated Region |
| **WT** | Wild Type |
| **YS** | Yolk Sac |

# 11. Appendices

11.1. Appendix I (A novel inducible mouse model of *MLL-ENL*-driven mixed lineage acute leukemia

11.2. Appendix II (Mouse models of Acute Myeloid Leukemia)

HemaSphere
Powered by EHA

EHA EUROPEAN HEMATOLOGY ASSOCIATION

Article
OPEN ACCESS

# A Novel Inducible Mouse Model of *MLL-ENL*-driven Mixed-lineage Acute Leukemia

Vaia Stavropoulou[1], Marwa Almosailleakh[1], Hélène Royo[2,3], Jean-François Spetz[2], Sabine Juge[1], Laurent Brault[1], Patrick Kopp[2], Michelina Iacovino[4], Michael Kyba[5,6], Alexandar Tzankov[7], Michael B. Stadler[2,3], Gianni Cazzaniga[8], Antoine H.F.M. Peters[2,9], Juerg Schwaller[1]

**Correspondence:** Juerg Schwaller (e-mail: j.schwaller@unibas.ch); Antoine H.F.M. Peters (e-mail: antoine.peters@fmi.ch).



**Abstract**

Previous retroviral and knock-in approaches to model human t(11;19)$^{+}$ acute mixed-lineage leukemia in mice resulted in myeloproliferation and acute myeloid leukemia not fully recapitulating the human disease. The authors established a doxycycline (DOX)-inducible transgenic mouse model "*iMLL-ENL*" in which induction in long-term hematopoietic stem cells, lymphoid primed multipotent progenitor cells, multipotent progenitors (MPP4) but not in more committed myeloid granulocyte-macrophage progenitors led to a fully reversible acute leukemia expressing myeloid and B-cell markers. *iMLL-ENL* leukemic cells generally expressed lower *MLL-ENL* mRNA than those obtained after retroviral transduction. Disease induction was associated with *iMLL-ENL* levels exceeding the endogenous *Mll1* at mRNA and protein levels. In leukemic cells from t(11;19)$^{+}$ leukemia patients, *MLL-ENL* mRNA also exceeded the endogenous *MLL1* levels suggesting a critical threshold for transformation. Expression profiling of *iMLL-ENL* acute leukemia revealed gene signatures that segregated t(11;19)$^{+}$ leukemia patients from those without an MLL translocation. Importantly, B220$^{+}$*iMLL-ENL* leukemic cells showed a higher in vivo leukemia initiation potential than coexisting B220$^{-}$ cells. Collectively, characterization of a novel transgenic mouse model indicates that the cell-of-origin and the fusion gene expression levels are both critical determinants for *MLL-ENL*-driven acute leukemia.

## Introduction

The t(11;19)(q23;p13.3) translocation leading to the expression of an MLL-ENL fusion protein is one of the most prevalent alterations affecting the *mixed lineage leukemia 1* (*MLL1*) gene.[1] The ENL fusion partner is a transcriptional elongation factor involved in controlling transcription in lymphoid and myeloid cells.[2] Although the majority of the human MLL-ENL$^{+}$ leukemia cases are diagnosed as B-cell acute lymphoblastic leukemia (ALL), several studies have shown that t(11;19) positive tumor cells simultaneously express lymphoid and myeloid markers characterizing the disease as mixed phenotype or mixed-lineage leukemia.[1,3,4] These observations suggest that the disease initiates by malignant transformation of a hematopoietic stem (HSC) or early progenitor cell with lymphoid and myeloid potential rather than of a lineage-restricted progenitor cell.

*Funding/support: This work was supported by grants from the Swiss National Science Foundation (SNF-31003A_130661 and 31003A_149714/1), Swiss Cancer League (OCS-2357-02-2009, OCS-02778-02-2011, and KFS-3019-08-2012), and SystemsX.ch (Cell plasticity), the Novartis Research Foundation (14B058). HR was supported by grants of the Peter und Traudl Engelhorn foundation and the Marie-Curie program (FP7-PEOPLE-2013-IEF). MBS and AHFMP were supported by the Novartis Research Foundation. JS was supported by the Gertrude Von Meissner Foundation Basel.*
*Disclosure: The authors have indicated they have no potential conflicts of interest to disclose.*
*Marwa Almosailleakh and Hélène Royo have shared second authorship.*
*Antoine H.F.M. Peters and Juerg Schwaller have shared senior authorship.*
*[1]Department of Biomedicine, University Children's Hospital of Basel, Basel, Switzerland*
*[2]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland*
*[3]Swiss Institute of Bioinformatics, Basel, Switzerland*
*[4]Department of Pediatrics, LA Biomedical Research Institute, Torrance, CA, USA*
*[5]Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA*
*[6]Department of Pediatric Oncology/Hematology, University Hospital Basel, Basel, Switzerland*
*[7]Institute for Pathology, University Hospital Basel, Basel, Switzerland*
*[8]Tettamanti Research Center, University of Milano-Bicocca, Ospedale San Gerardo, Monza, Italy*
*[9]Faculty of Sciences, University of Basel, Basel, Switzerland.*

*HemaSphere (2018) 2:4(e51)*
*Received: 23 February 2018 / Received in final form: 18 April 2018 / Accepted: 20 April 2018*
***Citation:*** Stavropoulou V, Almosailleakh M, Royo H, Spetz J-F, Juge S, Brault L, Kopp P, Iacovino M, Kyba M, Tzankov A, Stadler MB, Cazzaniga G, Peters AHFM, Schwaller J. A Novel Inducible Mouse Model of MLL-ENL-driven Mixed-lineage Acute Leukemia. *HemaSphere*, 2018;2:4. http://dx.doi.org/10.1097/HS9.0000000000000051

Previous studies modeling the oncogenic activity of MLL-ENL by retroviral overexpression in mouse bone marrow (BM) cells revealed induction of aberrant self-renewal in vitro, shown by serial replating in methylcellulose (MC) and upon subsequent transplantation into lethally irradiated recipients, the development of acute myeloid leukemia (AML) but not of ALL or mixed-lineage leukemia.[5,6] These myeloid leukemia phenotypes are considerably influenced by viral integration events potentially leading to the activation of cooperating oncogenes or to the loss of tumor suppressors. In addition, in vitro expansion of cells in growth factor-containing medium likely influences cellular lineage fate decisions, and thereby biases the resulting leukemic phenotype. Finally, the oncogenic expression levels driven by retroviral vectors like the murine stem cells virus are beyond those observed in primary patients' cells.[7,8]

To overcome such limitations, several knock-in mouse lines have been established.[9–14] Surprisingly, MLL-ENL expression in most of these models resulted in chronic myeloproliferative disorders rather than in lymphoblastic or mixed-lineage acute leukemia suggesting that the fusion oncogene may not have been expressed in the natural hematopoietic target cell at the appropriate level. One more recent study demonstrated that transgenic doxycycline (DOX) inducible MLL-ENL expression in granulocyte-macrophage progenitors (GMP) resulted in AML, whereas the HSC compartment appeared to be inherently protected from transformation.[15]

To address the impact of the cellular origin of MLL-ENL-driven leukemia, we established a DOX-regulated transgenic mouse model "*iMLL-ENL*," in which activation of the oncogene in long-term hematopoietic stem cells (LT-HSC), multipotent progenitors cells, or lymphoid primed multipotent progenitor cells (LMPP) but not GMP cells resulted in fully penetrant transplantable and reversible mixed-lineage acute leukemia. Interestingly, *iMLL-ENL* expression surpassed endogenous *Mll1* at mRNA and protein levels in tumor cells from diseased mice. Likewise, *MLL-ENL* mRNA levels exceeded those of *MLL1* in 5 $t(11;19)^+$ ALL patients suggesting that the cell-of-origin and oncogene expression levels are both critical determinants for the transforming capacity of MLL-ENL.

## Materials and methods

### Establishment of *iMLL-ENL* transgenic mice

The human *MLL-ENL* cDNA[16] was cloned into *p2Lox*, and electroporated into A2Lox-Cre ES cells.[17,18] The transgene was targeted in a region upstream of the *Hprt* locus. Standard procedures were used to generate the transgenic line that was backcrossed to C57BL/6 for over 10 generations. Double heterozygous female *iMLL-ENL* mice and cells originating from them were used throughout this study. All experiments were done in adherence to Swiss laws for animal welfare and approved by the Swiss Cantonal Veterinary Office of Basel-Stadt.

### Liquid cultures and colony forming assays

Cells were grown in liquid cultures containing RPMI-1640, 10% fetal bovine serum, 1% Pen/Strep supplemented with 10 ng/mL of human interleukin-6 (hIL-6), 6 ng/mL of murine mIL-3, 10 ng/mL of murine mIL-7, 100 ng/mL of murine stem cell factor (mSCF), and 100 ng/mL of mouse Flt3-ligand (mFlt-3L) (PeproTech EC, London, UK) in the presence of 0.5 μg/mL DOX where indicated. Cells were kept in culture for up to 25 days, and the cell number was scored every 2nd day. Total BM and/or progenitors cells were isolated from transgenic mice and plated ($10^4$) in 2 mL MC culture (Methocult M3534 or M3434; StemCell Technologies, Vancouver, BC, Canada), containing IL-3, IL-6, mSCF. The medium was further supplemented with IL-7, and Flt-3L and with 0.5 μg/mL DOX where indicated. Colonies were scored microscopically after 5 to 7 days, then harvested and replated ($10^4$) for up to 5 rounds. Cellular differentiation was examined by Wright-Giemsa staining of cytospin preparations.

### Flow cytometric analysis and sorting of hematopoietic stem and progenitor cells

BM cells were first stained with a lineage cocktail that contains antibodies specific for the following murine BM cell markers: Cd5, Cd11b, CD45R/B220, Ly-6G (Gr-1), and Ter119 (MAGM209, R&D Systems, Minneapolis, MN). Labeled cells were depleted by magnetic separation and the remaining cells were stained with a streptavidin Pacific Blue-conjugated (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA), a biotin-conjugated antimouse IL-7Rα (A7R34, eBioscience, San Diego, CA), a phycoerythrin (PE)-conjugated antimouse FcγRII/III (93, eBioscience) or CD150 (mShad150, eBioscience) or Flt-3L (eBioscience), an Alexa Fluor 647-conjugated antimouse Cd34 (RAM34, BD Biosciences, San Jose, CA), an allophycocyanin-conjugated antimouse c-Kit (2B8, BD Biosciences, Allschwil, Switzerland) and a phycoerythrin-Cy7 (PE-Cy7)-conjugated antimouse Sca-1 (BioLegend, San Diego, CA) monoclonal antibody, a fluorescein isothiocyanate (FITC)-conjugated Cd48 (HM48-1, eBioscience). GMP and common myeloid progenitors (CMPs) were fluorescence-activated cell sorted as IL-7Rα$^-$ Lin$^-$ Sca-1$^-$ c-Kit$^+$ Cd34$^+$ FcγRII/III$^{high}$ and IL-7Rα$^-$ Lin$^-$ Sca-1$^-$ c-Kit$^+$ Cd34$^+$ FcγRII/III$^{low}$, respectively; LT-HSC were sorted as Lin$^-$ Cd34$^-$ Sca-1$^-$ c-Kit$^+$ Cd150$^+$ Cd48$^-$; LSK were sorted as Lin$^-$ Sca-1$^-$ c-Kit$^+$; MPP4s as Lin$^-$ Sca-1$^+$ c-Kit$^+$ Flt-3L$^+$ Cd150$^-$ Cd48$^+$; and LMPPs as Lin$^-$ Sca1$^+$ c-Kit$^+$ Cd34$^-$ Flt-3L$^+$. For BM cell fluorescence activated cell sorting (FACS) analysis, single-cell suspensions from leukemic mice with >95% infiltration were stained with the following monoclonal antibodies: PE-labeled Cd16/32 (FcγRII/III), PE-Cy7-conjugated c-Kit, an allophycocyanin labeled (APC) B220, an allophycocyanin-Cy7-conjugated (APC-Cy7) Gr-1, a hydroxycoumarin (BV510)-conjugated Mac1, and an FITC-labeled Cd34, or Cd19 (all from Bioscience, San Diego, CA).

### RNA sequencing of in vivo mouse samples

RNA was isolated from $3 \times 10^4$ total BM cells (n = 3) or BM cells cultured ex vivo for 48 hours and sorted as B220$^+$ (n = 3) and B220$^-$ (n = 3) from leukemic *iMLL-ENL* mice with over 95% blast infiltration. RNA was also obtained from sorted wt LSK (n = 3) and wt GMP cells (n = 3) using the PicoPure RNA Isolation Kit (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA). The RNA sequencing library was prepared using the NuGen Ovation v2 RNA-Seq kit and sequenced using an Illumina HiSeq 2000 machine. The RNA-sequencing data are available from GEO (GSE84988). For target validation by Q-polymerase chain reaction (PCR) assays we used standard protocols that together with the methodology for the statistical analysis of mouse and human datasets are described in the Supplemental Material section (Supplemental Digital Content 1, http://links.lww.com/HS/A6).

## Results

### A novel reversible transgenic mouse model of MLL-ENL-induced leukemia

We generated "*iMLL-ENL*" transgenic mice in which the expression of the human *MLL-ENL* fusion cDNA[16] is driven by the reverse-type tetracycline-controlled transactivator (rtTA) integrated in the *Rosa26* gene locus (Fig. 1A).[17,18] Induction of *iMLL-ENL* resulted in a rapid expansion of linage marker-depleted BM cells in liquid cultures (Fig. 1B). *iMLL-ENL* cells cultured in growth factor containing MC formed mostly round and dense colonies previously referred as type-I colonies (Fig. 1C).[19,20] DOX removal resulted in rapid cellular differentiation mainly toward monocytes and macrophages but also toward cells expressing high levels of lymphoid markers, as shown by colony morphology (Fig. 1D), Wright-Giemsa staining (Fig. S1A, Supplemental Digital Content 2, http://links.lww.com/HS/A7) and flow cytometry measuring the expression of FcγRII/III, Cd34, Mac-1, and B220 surface markers (Fig. S1B, Supplemental Digital Content 2, http://links.lww.com/HS/A7). Exposure of BM cells to increasing doses of DOX (0–2 μg/mL) resulted in enhanced *iMLL-ENL* mRNA expression returning to background levels 24 hours after DOX removal (Fig. 1E). In the presence of factors supporting myeloid and lymphoid cell growth (mIL-3, hIL-6, mSCF, mFlt-3L, and mIL-7) expression of *iMLL-ENL* resulted in efficient serial replating activity for >4 rounds (Fig. 1F). Collectively, DOX-regulated transgenic *iMLL-ENL* expression reversibly induced aberrant self-renewal, blocked differentiation, and significantly expanded mouse BM-derived hematopoietic stem and progenitor cells (HSPC) ex vivo.

To study the effects of *iMLL-ENL* expression in vivo, DOX (0.4 mg/mL) was administered to the drinking water of the mice. After a mean latency of 104.3 ± 16.9 days, all animals developed signs of disease (Fig. 2A). Importantly, DOX-exposed mice that carried only the *iMLL-ENL* or the *rtTA* allele, or *iMLL-ENL* mice maintained in the absence of DOX never developed any disease. Transplantation of $1 \times 10^6$ total BM cells from naive *iMLL-ENL* mice into lethally irradiated wild type (wt) mice receiving DOX also induced the disease in all recipients (Fig. 2A). Transfer of $1 \times 10^5$ leukemic cells, either originating from directly induced *iMLL-ENL* mice or originating from transplanted mice into sublethally irradiated DOX-pretreated secondary recipients also consistently developed the disease after a shorter latency (Fig. 2A). Upon disease development in 50% of a mouse cohort (Fig. 2B, arrow), discontinuation of DOX administration DOX resulted in the survival of all the remaining mice with the exception of 1 mouse with extensive organ infiltration (not shown). All symptomatic mice presented with elevated white

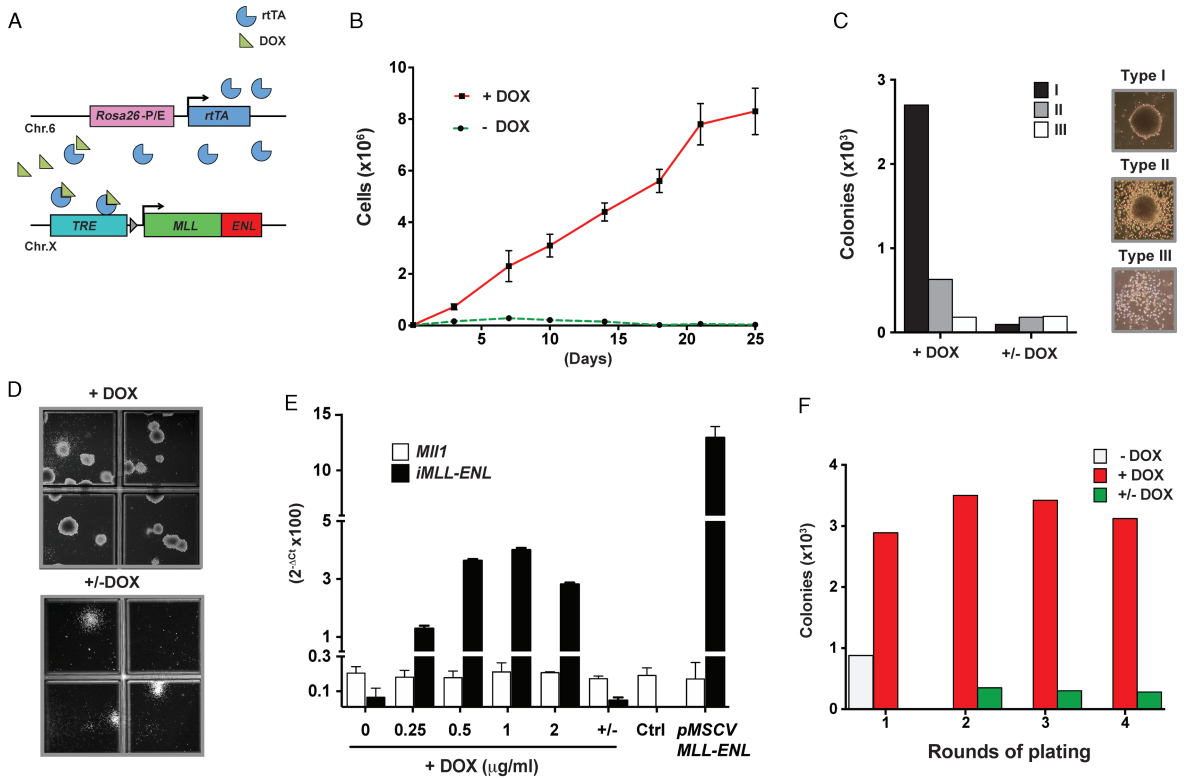



**Figure 1. Ex vivo induction of *iMLL-ENL* provides aberrant self-renewal and blocks differentiation of HSPCs in a fully reversible manner.** (A) *MLL-ENL* expression is controlled by a Tet-responsive element (TRE). The reverse tetracycline transactivator (*rtTA*) cassette was integrated into the *Rosa26* locus and the *MLL-ENL* fusion gene was targeted into the *Hprt* locus. (B) Proliferation of *iMLL-ENL* BM cells grown in liquid cultures with growth factors (10 ng/mL of IL-3, IL-6, and IL-7, 100 ng/mL of SCF and Flt3L) in presence (red) of absence (green) of DOX (1 μg/mL) for up to 25 days. (C) Quantification of colony types formed in the 1st replate in the presence or absence of DOX. (D) Reversibility of transformation, loss of immature compact colony formation capacity, and differentiation of blasts into mature macrophage/monocyte like cells upon removal of DOX as shown by the representative MC cultures. (E) Expression of *iMLL-ENL* mRNA in ex vivo cultured BM cells exposed to increasing amounts of DOX compared with cells immortalized by retroviral MLL-ENL expression, as assessed by quantitative (TaqMan) RT-PCR analysis. (F) Number of colonies of serial replatings of *iMLL-ENL* BM cells in growth factor-containing MC with and without DOX, and upon DOX removal. BM = bone marrow, DOX = doxycycline, MC = methylcellulose.

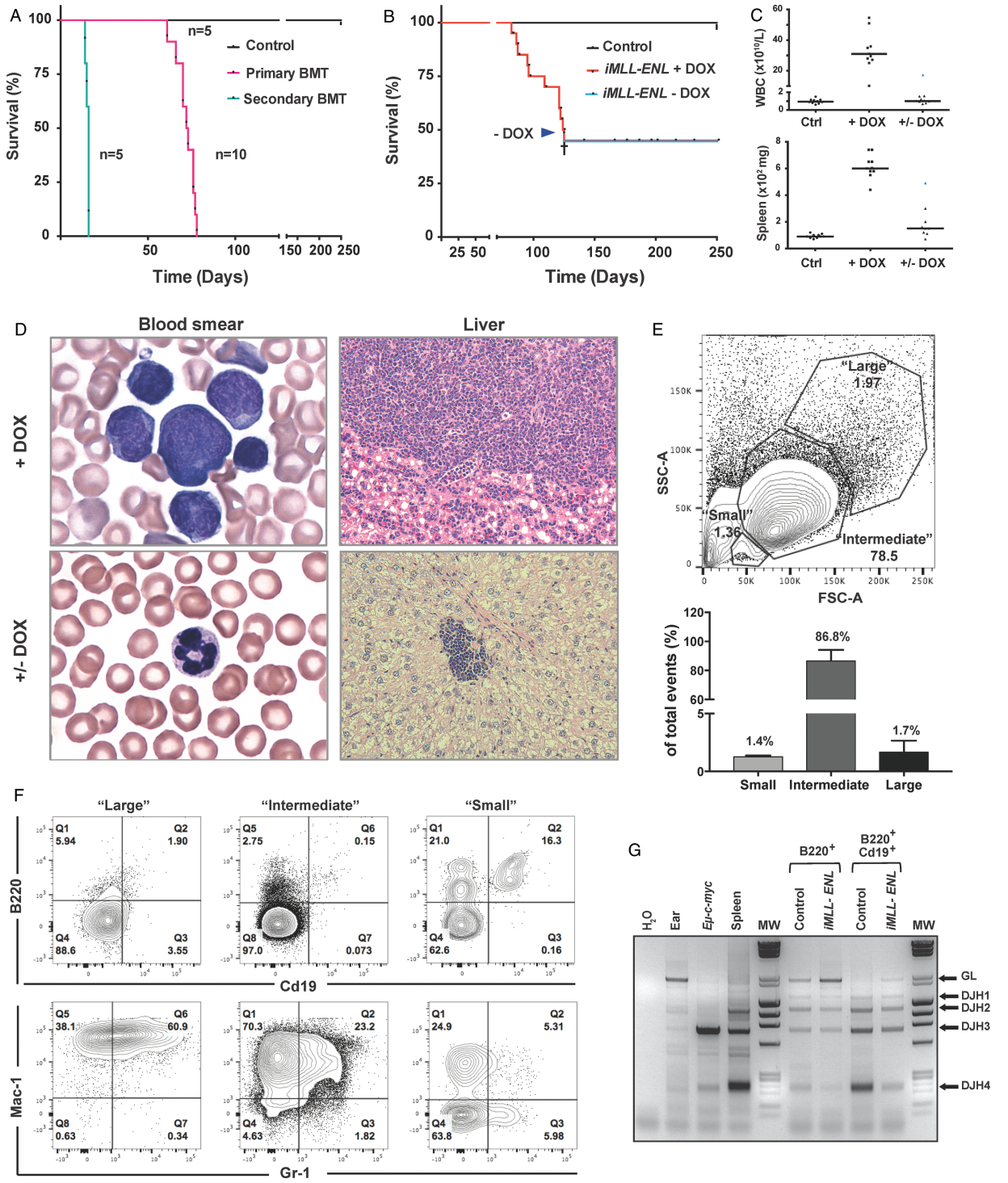



**Figure 2. In vivo induction of *iMLL-ENL* results in a reversible acute leukemia expressing myeloid and lymphoid markers.** (A) Kaplan-Meier survival plot of animals reconstituted with $10^6$ naive BM from *iMLL-ENL* mice (median latency 72.0 ± 5.3 days, n = 10) or transplanted with $10^5$ blasts from leukemic mice (median latency 15.4 ± 0.9 days, n = 5). Wild type recipients of naive BM cells were lethally irradiated ($^{137}$Cs, 9Gy) and DOX-induced 2 weeks after reconstitution. Recipients of leukemic BM were sublethally irradiated (4.5Gy) and received DOX 2 days prior to bone marrow transplantation and onward. (B) DOX administration in the drinking water induced acute leukemia in *iMLL-ENL* (median latency of 104.3 ± 16.9 days, n = 10). Upon DOX removal (arrow) the remaining animals (n = 10) survived up to 300 days without developing the disease, except of 1 mouse with extensive multiorgan leukemic infiltration (cross). (C) White blood cell counts and spleen weight of diseased mice on DOX (+DOX) and after DOX removal (+/− DOX) measured at different time points. (D) Wright-Giemsa-stained blood smears and histopathology of the liver showing infiltration of leukemic blasts of different sizes. Upon DOX removal the blood smears were normal and only small residual infiltrations were found occasionally in the livers. (E) Top panel: representative flow cytometric analysis of leukemic blasts isolated from highly infiltrated BM grouped according to different size ("large," "intermediate," and "small"). Bottom panel: quantification of the representative FACS plot showing the percentage of "large," "intermediate," and "small" cells. The bars represent the percentage of mean of size measured from BM cells of 3 independently analyzed diseased *iMLL-ENL* mice. (F) Detailed flow cytometric analysis of BM cells from a representative diseased *iMLL-ENL* mouse (E), showing the percentage of cell populations expressing lymphoid (B220 and Cd19) versus myeloid markers (Mac-1 and Gr-1), grouped according to size. Cells were costained with antibodies for all the markers. Different populations were gated according to size (SSC-A vs FSC-A), singlet (FSC-A vs FSC-H), live (DAPI$^-$), and expression of the respective markers. (G) Polymerase chain reaction analysis of the *IgH D-J* gene rearrangements using DNA isolated from the BM cells of a leukemic mouse (input) of single B220$^+$ and double B220$^+$/Cd19$^+$ flow-sorted cells versus control DNA samples isolated from the ear, the spleen of a wild type (B6) mouse, and from a lymphoma of *Eμ-c-myc* transgenic mice (GL: 2.1 kb, DJH1: 1.4 kb, DJH2: 1.1 kb, DJH3: 0.7 kb, and DJH4: 0.2 kb). BM = bone marrow, DOX = doxycycline, FACS = fluorescence activated cell sorting, GL = germline.

blood cell counts (WBC), hepatosplenomegaly, lymphadenopathy, and excessive multiorgan infiltration by leukemic blasts (Fig. 2C and D, top right panel, and data not shown). Symptomatic mice that survived after DOX removal had normal WBC, blood smears, spleen, and liver weights measured after 60 to 120 days (Fig. 2C), with occasional focal infiltrations of cells with pyknotic appearing nuclei (Fig. 2D, bottom right panel).

Inspection of blood smears from diseased mice revealed the presence of large, intermediate-sized, and small leukemic blasts (Fig. 2D, top left panel). In symptomatic mice with highly (>95%) infiltrated BM, the majority of the tumor cells were of intermediate size (86.8 ± 7.4%), with minor fractions (1.7 ± 0.9% and 1.4 ± 0.3%) of cells of large and small size, respectively (Fig. 2E). Notably, 37% (±7.5, n = 3) of the small cells expressed the lymphoid marker B220, some also Cd19. A minority of the B220$^+$ cells also expressed the myeloid markers Mac1 (8.1 ± 4.4%, n = 3) or Gr1 (4.4 ± 3.3%, n = 3) (Fig. S1C, Supplemental Digital Content 2, http://links.lww.com/HS/A7). By contrast, the large- and intermediate-sized cells showed little expression of B220 or Cd19, but were mostly Mac1 or Mac-1/Gr1 positive (Figs. 2F and S1D, Supplemental Digital Content 2, http://links.lww.com/HS/A7). PCR analysis showed that a fraction of B220$^+$*iMLL-ENL* leukemic blasts underwent *IgH D-J* gene rearrangement.[20] Whereas complete *IgH D-J* rearrangement was seen in normal controls, incomplete rearrangement, indicated by the presence of a germline (GL) band, was observed in B220$^+$/Cd19$^+$ cells from diseased *iMLL-ENL* mice (Fig. 2G). Collectively, *iMLL-ENL* is a novel robust inducible and reversible transgenic mouse model for acute leukemia expressing myeloid and lymphoid markers.

## Expression signatures of *iMLL-ENL* mouse leukemia separate ALL patients carrying MLL rearrangements

To characterize the *iMLL-ENL* phenotype, we compared the gene expression programs of leukemic cells from diseased mice with FACS-sorted LSK from normal mice. Overall, we identified 1935 differentially expressed genes (absolute logFC > 2, false discovery rate [FDR] < 0.001) of which 780 genes showed increased and 1155 decreased levels in leukemic cells (Fig. 3A and B). Among the genes significantly upregulated in *iMLL-ENL* leukemic cells, we found many previously proposed MLL-fusion targets including *Hoxa9*, *Hoxa10*, *Meis1*, *Eya1*, *Pbx3*, *S100a4*, *S100a8*, and *S100a9*,[19,20] as well as genes implicated in ALL and B-cell maintenance (Table S1, Supplemental Digital Content 3, http://links.lww.com/HS/A8). We performed gene set enrichment analysis (GSEA) to compare the set of genes differentially expressed between leukemic *iMLL-ENL* and LSK cells with previously established expression signatures. We observed significant enrichments with signatures derived from mouse leukemia induced by overexpression of *Hoxa9* and *Meis1* (Fig. 3C).[21] In addition, the signatures also showed similarities to those previously established from late hematopoietic progenitors obtained from normal mice (Fig. 3C). We validated MLL-fusion-dependent expression of *Hoxa9* and *Meis1* in blasts from *iMLL-ENL* and *iMLL-AF9* mice cultured for 24 hours with or without DOX (Fig. 3D). Collectively, leukemic cells from diseased *iMLL-ENL* mice express many known MLL-fusion targets but also genes associated with ALL that were previously not described MLL-ENL$^+$ mouse leukemia models.

We next asked whether the expression signature from *iMLL-ENL* mouse leukemia versus LSK would also be informative in human datasets. We tested whether the signatures would be able to distinguish ALL patients potentially carrying MLL gene rearrangements from a cohort of primary infant ALL expression profiles typified by translocations t(4;11) (n = 29), t(11;19) (n = 21), and t(9;11) (n = 8), or the absence of MLL translocations (n = 14).[22] We found that the *iMLL-ENL* expression signatures fully segregated MLL-ENL$^+$ patients, to some extent also MLL-AF4$^+$ patients and less efficiently MLL-AF9$^+$ patients from patients with no MLL rearrangements (Figs. 3E and S2A and B, Supplemental Digital Content 4, http://links.lww.com/HS/A9). By contrast, expression signatures derived from populations of randomly selected genes with similar expression-level distributions or from all genes did not segregate patients according to their genetic aberrations (Figs. 3E and F and S2A and B, Supplemental Digital Content 4, http://links.lww.com/HS/A9). Thus, this analysis shows that genes differentially expressed in *iMLL-ENL* versus wt LSK are enriched for genes differentially expressed in human MLL-rearranged leukemia. The resulting signature of differentially expressed genes did, however, not allow clustering of human ALLs according to their MLL-status (not shown). In summary, these findings support the notion that *iMLL-ENL* mice develop a disease characterized by a gene expression signature matching to signatures of t(11;19)$^+$ acute leukemia in patients.

## The cellular origin determines the leukemogenic activity of *iMLL-ENL*

Next, we determined the transformation potential of *iMLL-ENL* in different cell populations of the hematopoietic hierarchy. First, we compared in vitro growth of flow-sorted *iMLL-ENL* hematopoietic LSK and GMP in the presence of myeloid and lymphoid growth factors and DOX. We found that LSK cells expanded significantly faster than GMP (Fig. 4A). However, fusion gene expression levels did not significantly differ between LSK- and GMP-derived *iMLL-ENL* cells (Fig. 4B). To address the cellular origin in vivo, we transplanted LT-HSC and GMP from naive *iMLL-ENL* mice into lethally irradiated recipients pretreated with DOX. Hereby we observed that 500 LT-HSC were sufficient to induce the leukemic phenotype after a median latency of 169.6 ± 27.5 days whereas 5000 GMP never resulted in any disease within an observation period of >400 days (Fig. 4C). This difference might not solely be attributed to inherent differences in engraftment capacities of LT-HSC and GMP, as transplantation of the same number of GMP from *iMLL-AF9* mice was sufficient to induce AML.[18] We also transplanted equal numbers ($1 \times 10^4$) of sorted MPP4, LMPP, and CMP progenitor cells and found that, although the disease could be induced by all of these populations, the latency was shorter upon transfer of LMPP (54.4 ± 11.7, n = 5) than of MPP4 (61.0 ± 13.7, n = 5). Notably, only 3 out of 5 mice transplanted with CMP developed the disease within the observation period of 120 days (Fig. 4D). The induced disease was similar to directly induced mice with extensive multiorgan infiltration by leukemic blasts of different sizes expressing myeloid and lymphoid markers (Fig. 4E). Notably, the amount of B220$^+$ tumor cells was significantly higher in *iMLL-ENL* tumors originating from LT-HSC, LMPP, or MPP4 than from CMP (Fig. 4F). Collectively, our data suggest that *iMLL-ENL* preferentially immortalizes hematopoietic stem or early multipotent precursor cells rather than more committed myeloid progenitors.

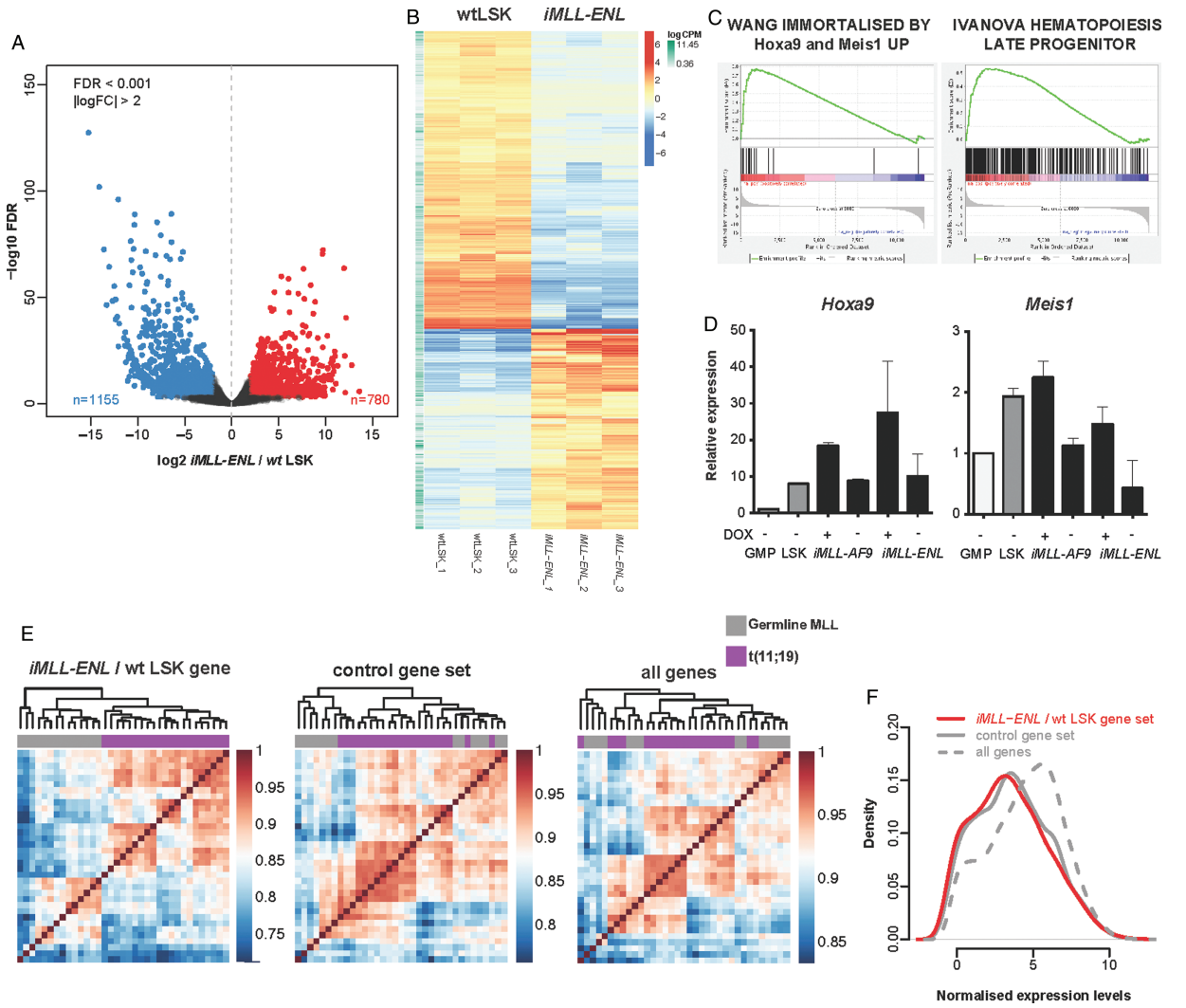
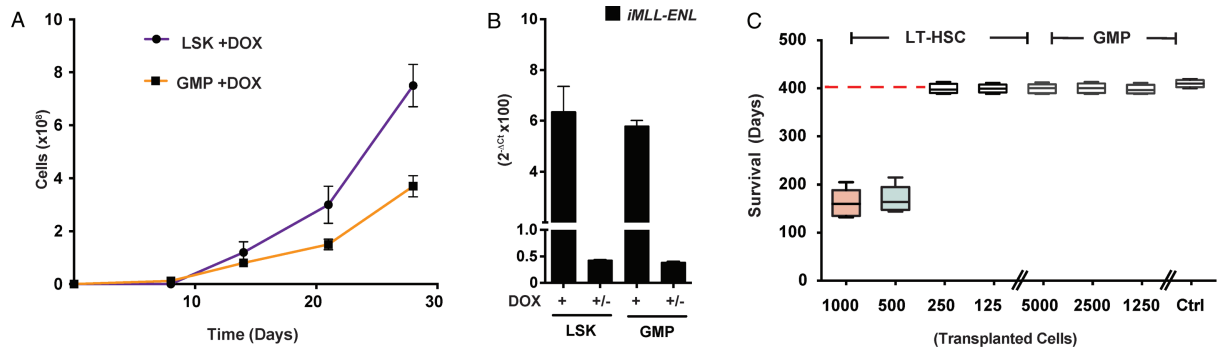



**Figure 3. Cross-species transcriptome analysis of *iMLL-ENL*-derived leukemias revealed common genes that segregate ALL patients with from those without MLL-rearrangements.** (A) Volcano plot showing fold change and statistical significance between wt LSK and *iMLL-ENL*-derived leukemic cells. The 1155 significantly lower expressed genes are plotted in blue and the 780 more highly expressed genes are plotted in red. Cutoffs used for differential expression: FDR < 0.001, absolute log fold change > 2 (same in panels B, C, E, F). (B) Heatmap showing relative expression ($log_2$ level over mean) of 1935 genes differentially expressed between wt LSK and *iMLL-ENL* samples. Average expression is shown on the right as logCPM (green). (C) Gene set enrichment analysis (GSEA) results for 2 signatures of the C2 Molecular Signature Database (MsigDB) enriched in the *iMLL-ENL*/wt LSK signature. Left: "Immortalized by Hoxa9 and Meis1 Up" represents genes upregulated in mouse leukemia by overexpression of *Hoxa9* and *Meis1* (normalized enrichment score [NES]: 2.0880373; nominal *P* value: 0.0; FDR q value: 5.326164E−4; family-wise error rate [FWER] *P* value: 0.006). Right: "Hematopoiesis late progenitor" represents genes upregulated in BM-derived late hematopoietic progenitors (NES: 2.2291334; nominal *P* value: 0.0; FDR q value: 0.0; FWER *P* value: 0.0). (D) *Hoxa9* and *Meis1* mRNA expression measured by qRT-PCR in BM cells isolated from naive wt GMP and wt LSK progenitors and *iMLL-ENL* or *iMLL-AF9* leukemic cell cultured in vitro for 24 hours with DOX or overnight without DOX. Relative expression levels were normalized to *Gapdh* mRNA levels expression and to expression in naive murine GMP. Results are the mean ± standard deviation of triplicates in the indicated AML groups (n = 3 per group). (E) Pair-wise correlation maps and hierarchical clustering of human patients with the t(11;19) (in purple) and germline MLL (in gray) genotypes. Correlations and clustering of patient samples were computed using expression values of human genes that are orthologous to mouse genes being part of the following 3 gene sets. Left panel: genes significantly differentially expressed between *iMLL-ENL* cells and wt LSKs. Middle panel: a control set of genes with the same expression distribution as the genes from the *iMLL-ENL* signature. Right panel: all genes expressed in *iMLL-ENL* and wt LSK samples. (F) Distribution of normalized expression levels ($log_2$ CPM) in the 3 gene sets used in (E). ALL = acute lymphoblastic leukemia, AML = acute myeloid leukemia, DOX = doxycycline, GMP = granulocyte-macrophage progenitors, LSK = $Lin^-$ $Sca\text{-}1^+$ $c\text{-}Kit^+$.

## B220$^+$*iMLL-ENL* leukemic blasts are enriched for leukemia initiating cells

The leukemic disease in *iMLL-ENL* mice was composed by different populations of cells that could be distinguished by the expression level of the B220 surface marker (Fig. 5A, left panel). As B220$^+$ cells were previously proposed in another mouse model to contain leukemic stem cells,[23] we compared the leukemia initiation potential by transplanting $1 \times 10^5$ B220$^+$ or B220$^-$ leukemic blasts sorted from total mBM ex vivo cultures (48 hours) into sublethally irradiated recipients. Transplantation of B220$^+$ leukemic blasts propagated the disease after a significantly shorter latency (Fig. 5A, right panel) than the B220$^-$ cells (15.6 ± 0.5 vs 22.5 ± 1.1 days, $P \leq 0.0028$). Both B220$^+$ and B220$^-$ leukemic cells isolated from terminally diseased mice expressed similar levels of *iMLL-ENL* mRNA




D

| Transplanted Cell Populations | Cell number | Aver. Survival (Days) | Leukemic mice |
|---|---|---|---|
| MPP4 | $10^{+4}$ | 61.0 ± 13.7 | 5/5 |
| LMPP | $10^{+4}$ | 54.4 ± 11.7 | 5/5 |
| CMP | $10^{+4}$ | 120.0 ± 0.0 | 3/5 |

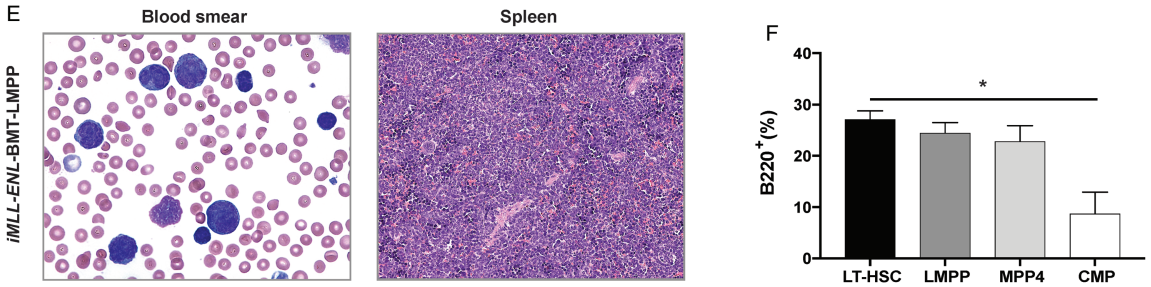



**Figure 4. *iMLL-ENL* preferentially transforms hematopoietic stem cells and pluripotent rather than more committed myeloid progenitor cells.** (A) Ex vivo expansion of naive *iMLL-ENL* LSK- and GMP-derived cells in liquid cultures containing hIL-6 (10 ng/mL), mIL-3 (6 ng/mL), mIL-7 (10 ng/mL), mSCF (100 ng/mL), and mFlt-3L (100 ng/mL) in the presence of DOX. (B) Expression of *iMLL-ENL* mRNA of sorted LSK and GMP cells cultured ex vivo in MC for 3 rounds with growth factors in the presence or absence of DOX. Relative mRNA expression levels were normalized to *Gapdh* expression and are expressed as $2^{-\Delta Ct}$. Results the mean values ± standard deviation of duplicates. (C) Average survival of lethally irradiated mice transplanted with the indicated numbers (1000, 500, 250, and 125) of naive *iMLL-ENL* LT-HSCs or GMPs (5000, 2500, and 1250) into wild type recipients on DOX. Mice receiving 500 to 1000 LT-HSCs developed acute leukemia (median latency 169.6 ± 27.3 days, n = 10) whereas mice receiving 125 to 250 LT-HSCs (n = 10) or GMPs (n = 15) did not develop the disease within 400 days. Mice transplanted with the indicated cells but kept off DOX served as controls (n = 10). (D) Average survival of lethally irradiated mice transplanted with $10^4$ of transgenic MPP4s or LMPPs or CMPs on DOX. Mice receiving MPP4s and LMPPs developed acute leukemia with a median latency of 61.0 ± 13.7 days and 54.4 ± 11.7, respectively (n = 5 per group), whereas 3 out of 5 mice receiving CMPs (n = 5) develop the disease within 120 days. (E) Wright-Giemsa-stained blood smear showing presence of leukemic blasts of different size; and histopathology of the spleen showing infiltration. (F) Quantification of flow cytometry analysis of "small" cell fraction showing the percentage of $B220^+$ in BM of leukemic mice transplanted with different hematopoietic progenitor populations. Notably LT-HSC-derived disease was associated with a significantly higher fraction of $B220^+$ BM cells than CMP-derived leukemia ($P = 0.002$, n = 2, unpaired *t* test). BM = bone marrow, CMP = common myeloid progenitor, DOX = doxycycline, GMP = granulocyte-macrophage progenitors, hIL-6 = human interleukin-6, LMPP = lymphoid primed multipotent progenitor cells, LSK = $Lin^-$ $Sca\text{-}1^+$ $c\text{-}Kit^+$, LT-HSC = long-term hematopoietic stem cells, MC = methylcellulose, MPP4 = multipotent progenitors, mSCF = murine stem cell factor.

and remained fully DOX-dependent (Fig. S3A, Supplemental Digital Content 5, http://links.lww.com/HS/A10). Leukemic mice transplanted with $B220^+$ cells showed more extensive cell infiltrations in the BM, spleen, liver, intestines, kidney, and lungs as compared to the recipients of $B220^-$ cells. All mice, however, ultimately succumb to the same disease as directly induced *iMLL-ENL* mice, characterized by leukemic blasts of different size expressing myeloid and lymphoid surface markers (Fig. 5B and not shown). Similar to primary induced *iMLL-ENL* mice, the tumor cells of diseased mice expressed myeloid and lymphoid surface markers. Interestingly, tumor cells of mice transplanted with $B220^+$ cells again expressed higher B220 levels than those transplanted with $B220^-$ cells (Fig. S3B, Supplemental Digital Content 5, http://links.lww.com/HS/A10 and data not shown). These findings suggest that $B220^+$ cells have higher leukemia initiation potential than $B220^-$ *iMLL-ENL* cells.

To better characterize the biological differences between *iMLL-ENL* leukemia induced by $B220^+$ or $B220^-$ cells, we compared their gene expression signatures. We found 335 differentially expressed genes of which 313 genes were expressed at significantly higher levels (absolute logFC > 1; FDR < 0.01) and only 22 genes at significantly lower levels in $B220^+$ compared with $B220^-$ cells (Fig. S3C and Table S2, Supplemental Digital Contents 5 and 6, http://links.lww.com/HS/A10 and http://links.lww.com/HS/A11, respectively). Gene ontology (GO) analysis revealed terms including "regulation of cell motility," "positive regulation of cellular component movement," "positive regulation of locomotion," "granulocyte migration," and "regulation of cell activation" ($P < 10^{-15}$) (Fig. S3D, Supplemental Digital Content 5, http://links.lww.com/HS/A10). GSEA indicated significant overlaps between the genes more highly expressed in $B220^+$ versus by $B220^-$ cells with genes up-regulated in solid

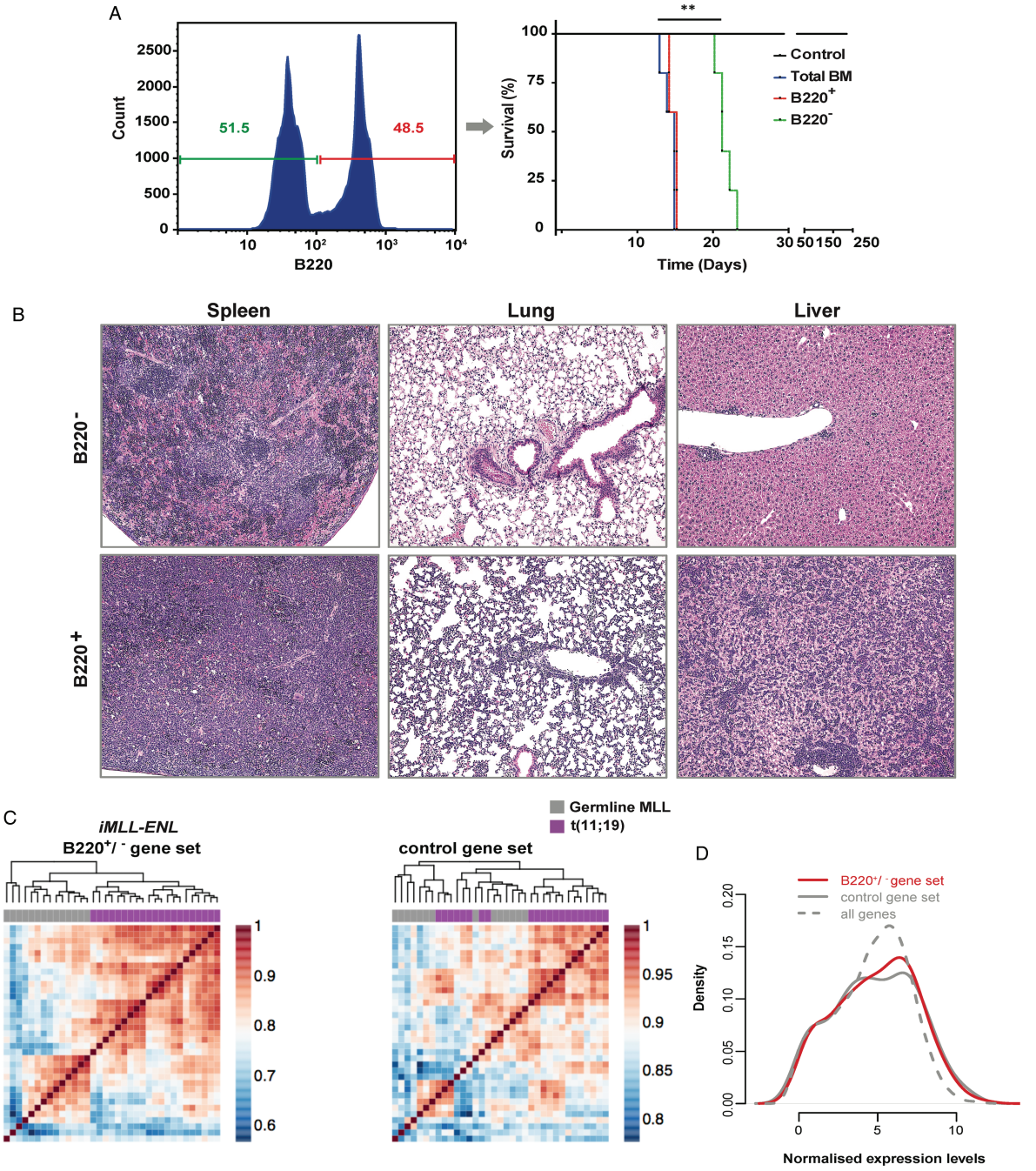



**Figure 5. B220$^+$*iMLL-ENL* leukemic blasts are enriched for leukemia initiating cells.** (A) BM-derived leukemic blasts were cultured overnight in medium with hIL-6 (10 ng/mL), mIL-3 (6 ng/mL), mIL-7 (10 ng/mL), mSCF (100 ng/mL), and mFlt-3L (100 ng/mL) and DOX, then sorted (left panel) according to B220$^+$ versus B220$^-$ expression and transferred into secondary recipients. Kaplan-Meier plot (right panel): the difference of the latency periods was assessed by a log-rank test ($P < 0.0028$). (B) Histopathology of spleen, lungs, and liver revealed massive infiltration of leukemic blasts in mice transplanted with $1 \times 10^5$ *iMLL-ENL* B220$^-$ versus B220$^+$ ex vivo cultured mBM cells with complete loss of normal organ architecture in the latter. (C) Pair-wise correlation maps and hierarchical clustering of human patients with the t(11;19) (in purple) and germline MLL (in gray) genotypes. Correlations and clustering of patient samples were computed using expression values of human genes that are orthologous to mouse genes being part of the following 3 gene sets. Left panel: genes significantly differentially expressed between *iMLL-ENL* B220$^+$ and *iMLL-ENL* B220$^-$ samples. Right panel: a control set of genes with the same expression distribution as the genes from the *iMLL-ENL* B220$^+$/B220$^-$ signature. (D) Distribution of normalized expression levels ($\log_2$ CPM) in the 2 gene sets used in (C). The distribution of normalized expression values of all genes expressed in both samples is given as a comparison. BM = bone marrow, DOX = doxycycline, hIL-6 = human interleukin-6, mSCF = murine stem cell factor.

cancers harboring activating mutations in the *K-RAS* oncogene or the STK33 serine-threonine kinase, previously implied in RAS-mediated transformation, and with genes upregulated in pathways involved in K-RAS and p53 signaling (Fig. S3E, Supplemental Digital Content 5, http://links.lww.com/HS/A10).[24,25]

Finally, we investigated whether the B220$^+$ versus B220$^-$*iMLL-ENL* expression signature would enable identification of ALL patients carrying MLL gene rearrangements within a cohort of primary infant ALL expression profiles.[22] The genes significantly differentially expressed in B220$^+$ versus B220$^-$

enabled clustering of MLL-ENL$^+$ patients (n=21) from those without *MLL1* gene rearrangements (n=14) (Fig. 5C, left panel), contrary to a control gene set or to all genes (Fig. 5D). By contrast, the signature was less effective in separating MLL-AF4$^+$ ALL patients (n=29) and not effective in segregating MLL-AF9$^+$ ALL patients (n=9) from patients without MLL rearrangements (Fig. S3F and G, Supplemental Digital Content 5, http://links.lww.com/HS/A10). Thus, the B220$^+$*iMLL-ENL* signature is enriched for genes that are differentially expressed in MLL-ENL$^+$ as compared to leukemia patients without any MLL-gene rearrangements.

### MLL-ENL protein levels and leukemic transformation

To address whether the expression levels correlate with transformation activity we compared transgenic *iMLL-ENL* mRNA levels with the expression level of endogenous *Mll1*. In the absence of DOX or within 24 hours after DOX removal, *iMLL-ENL* mRNA levels were below *Mll1* expression (Fig. 1E), a condition in which we observed rapid cell differentiation (Fig. S1A and B, Supplemental Digital Content 2, http://links.lww.com/HS/A7). In leukemic *iMLL-ENL* cells from diseased mice, the estimated fusion mRNA expression level consistently exceeded *Mll1* (Fig. 6A). Notably, expression of *MLL-ENL* in retrovirally transduced BM cells (*rMLL-ENL*) appeared 5- to 10-fold higher than *iMLL-ENL* mRNA expression in BM derived in vitro immortalized cell lines or in leukemic cells from diseased *iMLL-ENL* mice. In similar, we found that the relative fusion transcript levels exceeded those of endogenous *MLL1* in leukemic blasts from 5 t(11;19)+ ALL patients (Fig. 6B). In addition, MLL-ENL protein expression clearly exceeded the expression of wild type MLL in the MLL-ENL$^+$ KOPN8 cells, and about equal to wild type MLL in BM and spleen cells of diseased *iMLL-ENL* mice (Fig. 6C). These data suggest that MLL-ENL protein levels equal to or exceeding that of the nonrearranged MLL is necessary for leukemic transformation by the *MLL-ENL* fusion.

## Discussion

We established an inducible transgenic mouse for MLL-ENL-driven acute mixed lymphomyeloid leukemia. Transplanting naive LT-HSCs, MPP4, or LMPP progenitors into DOX-treated recipients efficiently induced the disease phenotype. By contrast, transplantation of CMPs resulted in disease only in some animals and transfer of even more committed GMPs failed to induce the disease. These results contrast strongly to a comparable inducible transgenic mouse model of MLL-AF9-driven AML: here the leukemic phenotype was induced by transplanting *iMLL-AF9* expressing LT-HSCs, short-term HSCs, CMPs, or GMPs, which supports a concept for an MLL fusion-specific transformation potential of particular cellular compartments.[18]

Intriguingly, a recent study claimed that HSCs are intrinsically protected against MLL-ENL-mediated transformation.[15] This conclusion emerged from experiments with a different DOX inducible transgenic mouse strain in which expression of the identical human *MLL-ENL* fusion cDNA,[16] integrated in a DOX responsive operon in the 3′ UTR of the murine *Col1a1* gene, is controlled by the M2 reverse tetracycline transactivator (M2-rtTA), the latter constitutively expressed from the *Rosa26* locus. In contrast to *iMLL-ENL*, transplantation of *Col1a1*-tetO-MLL/ENL LT-HSCs never induced a disease in DOX-treated recipients. Likewise, leukemia induction by transferring MPPs was very poor, whereas GMPs efficiently induced the disease. Interestingly, the expression levels of the *MLL-ENL* fusion transcripts from the *Col1a1*-tetO-MLL/ENL locus in HSCs and MPPs remained clearly below endogenous *Mll1*, while it exceeded *Mll1* in GMPs. In contrast to our *iMLL-ENL* model, the induced disease was consistently pure AML.[15]

Although both models (*iMLL-ENL* and *Col1a1*-tetO-MLL/ENL) are limited in pinpointing the disease origin to specific cell

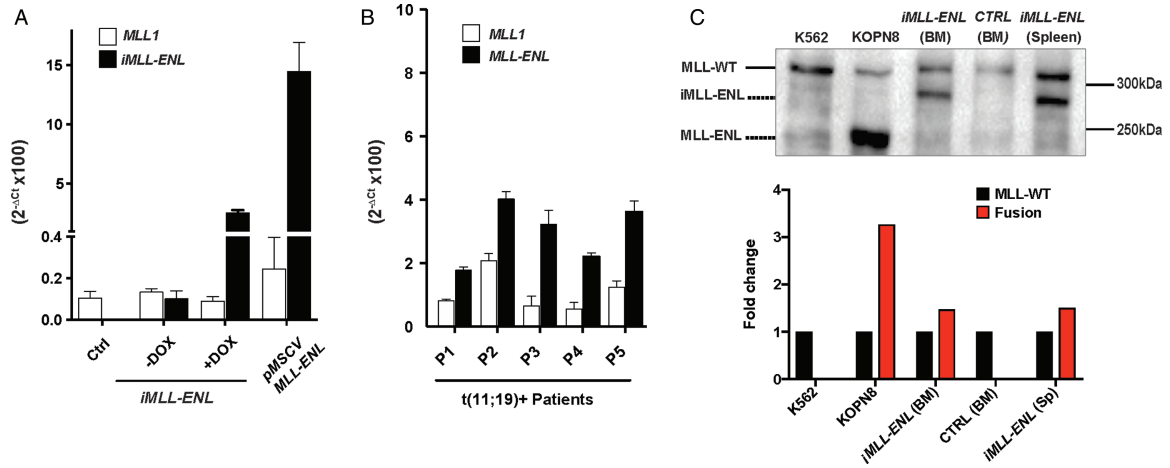



**Figure 6. Expression of *iMLL-ENL* levels determine cellular transformation in vitro and in vivo.** (A) The expression of *Mll1* and *MLL-ENL* mRNA was measured in naive and leukemic BM cells from transgenic mice and BM cells from symptomatic mice transplanted with retrovirally (*pMSCV-MLL-ENL*) transduced cells. Relative mRNA expression levels were normalized to *Gapdh* expression and are expressed as $2^{-\Delta Ct}$. Results are shown as mean values ± standard deviation of duplicates. (B) Expression of *MLL1* and *MLL-ENL* mRNA in leukemic cells of human patients diagnosed with ALL and >80% infiltration. Relative mRNA expression levels were normalized to *Gapdh* expression and are expressed as $2^{-\Delta Ct}$. Results are shown as mean values ± standard deviation of duplicates. (C) Western blot analysis of total cell lysates of K562 (CML, wild type MLL), KOPN8 (ALL, t(11;19)$^+$), BM of healthy mice, and BM and spleen cells of diseased *iMLL-ENL* mice. The blot was probed with an antibody recognizing the MLL-N-terminus. KOPN8 cells express an MLL-ENL fusion of calculated size of about 170 kDa,[29] while the *iMLL-ENL* fusion results in a protein of 220 kDa.[16] Bands were quantified according to intensity, and calculated as fold change normalized to wild type MLL. ALL=acute lymphoblastic leukemia, BM=bone marrow, CML=chronic myelogenous leukemia.

types, our work suggests that expression of *MLL-ENL* mRNA exceeding *Mll1* might be important to potently induce LT-HSCs or MPP cell-derived disease. Low *iMLL-ENL* mRNA expression (comparable or lower than endogenous *Mll1*) in unexposed naive *iMLL-ENL* mice (due to the inherent leakiness of the *rtTA* transcriptional transactivator) never resulted in disease during a normal lifespan, suggesting that a certain threshold level is necessary for the leukemogenic activity in vivo. Importantly, akin to the *iMLL-ENL* mouse model, we also found that *MLL-ENL* fusion transcripts were generally higher than endogenous *MLL1* levels in 5 MLL-ENL$^+$ leukemia patients. Importantly, we observed abundant amounts of the fusion protein in human t (11;19)$^+$ KOPN8 cells but also in leukemic blasts from diseased *iMLL-ENL* mice. This finding extends previous work by others in other human MLL-ENL$^+$ cell lines suggesting that fusion protein expression indeed exceeds MLL1 in leukemic blasts.[26] While this observation warrants further validation in larger patients cohorts, it is intriguing to postulate that the relative expression levels of MLL fusions versus endogenous *MLL1* might be a critical factor for the transforming activity in different cells of the hematopoietic hierarchy. Recent work indeed suggested that stabilization of the wild type MLL1 protein displaces MLL fusions from critical targets and impairs transformation, which suggests competition between wild type and rearranged MLL1.[27]

GSEA identified some similarities with gene expression profiles derived from previous studies on *MLL* fusions verifying that *iMLL-ENL* leukemic cells also express a common signature, similar to what we observed in the *iMLL-AF9* model.[18] Besides the known MLL fusion targets, gene expression profiling of leukemic cells from diseased *iMLL-ENL* mice identified multiple putative target genes never described in previous mouse models for this leukemia. This difference may result from the fact that *iMLL-ENL* model causes mixed-lineage leukemia composed of blasts expressing myeloid and lymphoid surface markers mostly likely originating from multipotent HSPCs.

Cross-species comparison of *iMLL-ENL* expression signatures with those from ALL patients was able to segregate patients with and without *MLL1* gene rearrangements.[22] Interestingly, clustering of ALL patients according to their MLL rearrangement status could only be achieved by signatures generated from the comparison of the gene expression of the leukemic cells with normal LSK but not with normal GMP. As activation of the fusion in stem cells but not in GMP initiated the disease, this observation supports furthermore that the cell of origin for the *iMLL-ENL* leukemia is within the stem cell and/or early hematopoietic progenitor rather than in more committed myeloid progenitor compartment (not shown).

Patients with t(11;19)$^+$ leukemia often display immature immunoglobulin rearrangements and present with a pro-B cell lymphoblastic or mixed-lineage leukemia phenotype.[3,4] *iMLL-ENL* induced acute leukemia was characterized by the coexistence of blasts of various sizes that expressed myeloid and/or lymphoid markers or exclusively lymphoid markers in a small fraction of the cells. This intriguing mixed phenotype differs from previous mouse models expressing MLL-ENL by retroviral transduction, interchromosomal recombination or from conditional transgenes that resulted in MPN and/or AML. However, retrovirally *MLL-ENL* expressing mouse BM cells cultured under B-cell lineage favoring conditions resulted in the outgrowth of cells expressing B-cell markers Cd19 and B220. Transplanting immature Cd19$^-$/B220$^+$ cells resulted in a leukemic phenotype in which the blasts still maintained a monocytoid morphology suggesting that *MLL-ENL* transformed a lymphoid/myeloid precursor leading to a bi-phenotypic-like leukemia.[28] By transplanting leukemic blasts from diseased *iMLL-ENL* mice we found that B220$^+$ leukemic cells were more potent to propagate leukemia than B220$^-$ cells. This situation is somehow similar to mice that developed acute leukemia by transplanting BM retrovirally expressing the *CALM-AF10* fusion gene with leukemia propagating cells expressing B220 with *IgH D-J* gene rearrangements.[23]

## Acknowledgments

The authors thank R. Slany (Erlangen, Germany) for providing the *pMSCV-MLL-ENL* expression plasmid and the KOPN8 cell line; D. Labes, E. Traunecker, and P. Tspaogas for FACS assistance; U. Schneider and his team for animal husbandry; B. Kuchemann for his assistance in establishing the transgenic line; T. Roloff and his team for transcriptional profiling (FMI functional genomics); P. Matthias and Vincent Pionel (FMI, Basel) for providing material; and R.W. Stam (Rotterdam) and J. De Boer (London) for practical technical support.

## Author contributions

Vaia Stavropoulou established the transgenic mouse line, designed and performed experiments, and wrote the manuscript.

Marwa Almosailleakh: designed and performed experiments and wrote the manuscript.

Hélène Royo: designed and performed bioinformatic data analysis.

Jean-François Spetz and Patrick Kopp: established the transgenic mouse line.

Sabine Juge and Laurent Brault: designed and performed experiments.

Michelina Iacovino and Michael Kyba: provided essential material.

Alexandar Tzankov: supervised histopathology analysis.

Michael B. Stadler: supervised bioinformatic data analysis.

Gianni Cazzaniga: provided patient-derived materials.

Antoine H.F.M. Peters: designed, supervised experiments and data analysis, and wrote the manuscript.

Juerg Schwaller: designed, performed, supervised experiments and data analysis, and wrote the manuscript.

## References

1. Meyer C, Burmeister T, Groger D, et al. The MLL recombinome of acute leukemias in 2017. *Leukemia* 2018; 23:273–284.
2. Rubnitz JE, Morrissey J, Savage PA, et al. ENL, the gene fused with HRX in t(11;19) leukemias, encodes a nuclear protein with transcriptional activation potential in lymphoid and myeloid cells. *Blood* 1994; 84:1747–1752.
3. Hayashi Y, Kobayashi Y, Hirai H, et al. Immunoglobulin heavy chain gene rearrangements and mixed lineage characteristics in acute leukemias with the 11;19 translocation. *Cancer* 1988; 61:712–720.
4. Hudson MM, Raimondi SC, Behm FG, et al. Childhood acute leukemia with t(11;19) (q23;p13). *Leukemia* 1991; 5:1064–1068.
5. Lavau C, Szilvassy SJ, Slany R, et al. Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL. *EMBO J* 1997; 16:4226–4237.
6. Cozzio A, Passegue E, Ayton PM, et al. Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes Dev* 2003; 17:3029–3035.
7. Ren R. Modeling the dosage effect of oncogenes in leukemogenesis. *Curr Opin Hematol* 2004; 11:25–34.
8. Haviernik P, Bunting KD. Safety concerns related to hematopoietic stem cell gene transfer using retroviral vectors. *Curr Gene Ther* 2004; 4:263–276.

9. Cano F, Drynan LF, Pannell R, et al. Leukaemia lineage specification caused by cell-specific Mll-Enl translocations. *Oncogene* 2008; 27: 1945–1950.
10. Chambers JS, Tanaka T, Brend T, et al. Sequential gene targeting to make chimeric tumor models with de novo chromosomal abnormalities. *Cancer Res* 2014; 74:1588–1597.
11. Drynan LF, Pannell R, Forster A, et al. Mll fusions generated by Cre-loxP-mediated de novo translocations can induce lineage reassignment in tumorigenesis. *EMBO J* 2005; 24:3136–3146.
12. Forster A, Pannell R, Drynan LF, et al. Engineering de novo reciprocal chromosomal translocations associated with Mll to replicate primary events of human cancer. *Cancer Cell* 2003; 3:449–458.
13. Ono R, Masuya M, Nakajima H, et al. Plzf drives MLL-fusion-mediated leukemogenesis specifically in long-term hematopoietic stem cells. *Blood* 2013; 122:1271–1283.
14. Takacova S, Slany R, Bartkova J, et al. DNA damage response and inflammatory signaling limit the MLL-ENL-induced leukemogenesis in vivo. *Cancer Cell* 2012; 21:517–531.
15. Ugale A, Norddahl GL, Wahlestedt M, et al. Hematopoietic stem cells are intrinsically protected against MLL-ENL-mediated transformation. *Cell Rep* 2014; 9:1246–1255.
16. Slany RK, Lavau C, Cleary ML. The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX. *Mol Cell Biol* 1998; 18:122–129.
17. Iacovino M, Bosnakovski D, Fey H, et al. Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cells* 2011; 29:1580–1588.
18. Stavropoulou V, Kaspar S, Brault L, et al. MLL-AF9 expression in hematopoietic stem cells drives a highly invasive AML expressing EMT-related genes linked to poor outcome. *Cancer Cell* 2016; 30:43–58.
19. Somervaille TC, Cleary ML. Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. *Cancer Cell* 2006; 10:257–268.
20. Krivtsov AV, Twomey D, Feng Z, et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* 2006; 442:818–822.
21. Wang QF, Wu G, Mi S, et al. MLL fusion proteins preferentially regulate a subset of wild-type MLL target genes in the leukemic genome. *Blood* 2011; 117:6895–6905.
22. Stam RW, Schneider P, Hagelstein JA, et al. Gene expression profiling-based dissection of MLL translocated and MLL germline acute lymphoblastic leukemia in infants. *Blood* 2010; 115:2835–2844.
23. Deshpande AJ, Cusan M, Rawat VP, et al. Acute myeloid leukemia is propagated by a leukemic stem cell with lymphoid characteristics in a mouse model of CALM/AF10-positive leukemia. *Cancer Cell* 2006; 10:363–374.
24. Azoitei N, Hoffmann CM, Ellegast JM, et al. Targeting of KRAS mutant tumors by HSP90 inhibitors involves degradation of STK33. *J Exp Med* 2012; 209:697–711.
25. Scholl C, Frohling S, Dunn IF, et al. Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. *Cell* 2009; 137:821–834.
26. Yokoyama A, Lin M, Naresh A, et al. A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer Cell* 2010; 17:198–212.
27. Liang K, Volk AG, Haug JS, et al. Therapeutic targeting of MLL degradation pathways in MLL-rearranged leukemia. *Cell* 2017; 168: 59–72.
28. Zeisig BB, Garcia-Cuellar MP, Winkler TH, et al. The oncoprotein MLL-ENL disturbs hematopoietic lineage determination and transforms a biphenotypic lymphoid/myeloid cell. *Oncogene* 2003; 22:1629–1637.
29. Wilkinson AC, Ballabio E, Geng H, et al. RUNX1 is a key target in t(4;11) leukemias that contributes to gene activation through an AF4-MLL complex interactions. *Cell Rep* 2013; 3:116–127.

International Journal of
*Molecular Sciences*

MDPI

*Review*

# Murine Models of Acute Myeloid Leukaemia

**Marwa Almosailleakh and Juerg Schwaller ***

Department of Biomedicine, University Children's Hospital beider Basel (UKBB), University of Basel, 4031 Basel, Switzerland; m.almosailleakh@unibas.ch

* Correspondence: J.Schwaller@unibas.ch; Tel.: +41-61-265-3504; Fax: +41-61-265-2350

Received: 31 October 2018; Accepted: 11 January 2019; Published: 21 January 2019


**Abstract:** Acute myeloid leukaemia (AML) is a rare but severe form of human cancer that results from a limited number of functionally cooperating genetic abnormalities leading to uncontrolled proliferation and impaired differentiation of hematopoietic stem and progenitor cells. Before the identification of genetic driver lesions, chemically, irradiation or viral infection-induced mouse leukaemia models provided platforms to test novel chemotherapeutics. Later, transgenic mouse models were established to test the in vivo transforming potential of newly cloned fusion genes and genetic aberrations detected in patients' genomes. Hereby researchers constitutively or conditionally expressed the respective gene in the germline of the mouse or reconstituted the hematopoietic system of lethally irradiated mice with bone marrow virally expressing the mutation of interest. More recently, immune deficient mice have been explored to study patient-derived human AML cells in vivo. Unfortunately, although complementary to each other, none of the currently available strategies faithfully model the initiation and progression of the human disease. Nevertheless, fast advances in the fields of next generation sequencing, molecular technology and bioengineering are continuously contributing to the generation of better mouse models. Here we review the most important AML mouse models of each category, briefly describe their advantages and limitations and show how they have contributed to our understanding of the biology and to the development of novel therapies.

**Keywords:** acute myeloid leukaemia; AML; mouse models; transgenic mice; bone marrow reconstitution; genome editing; patient-derived xenografts; PDX

## 1. Introduction

Acute myeloid leukaemia (AML) is a disease of an uncontrolled clonal proliferation of abnormal myeloid stem and progenitor cells in the hematopoietic tissue. The transformed myeloid cells or 'leukemic blasts' exhibit aberrant differentiation and accumulate in the bone marrow (BM). This process diminishes normal haematopoiesis, often leading to thrombocytopenia and anaemia, hematopoietic failure and mortality [1]. The genomic landscape of AML has been extensively studied since the 1970s, starting by the examination of chromosomal karyotypes of patients' leukemic cells [1,2]. Several prevalent balanced chromosomal rearrangements, including t(8;21)(q21;q22), inv(16)(p13q22) and t(15;17)(q22;q12) were identified in tumour cells from AML patients and molecularly characterized [3,4]. All three of these rearrangements share the remarkable feature of generating chimeric fusion proteins, in which at least one of the fusion partners is a gene encoding for a transcriptional regulator that is required for normal haematopoiesis. The advent of higher resolution next generation sequencing (NGS) has led to the identification of additional recurring and singleton alterations including cytogenetically-silent translocations, point mutations in metabolic regulators and small copy number changes [5]. NGS studies also revealed that despite the detection of recurrent genomic aberration, the majority of the genomes of de novo diagnosed AML contain fewer

number of mutations compared to most solid tumours [6]. A comprehensive NGS landmark study by Papaemmanuil and colleagues has identified 5234 driver mutations in 76 genes from a cohort of 1540 AML patients [1]. Recurrent AML-associated mutations can be grouped into different categories according to their functional consequences: those which are involved in epigenetic regulation such as DNA methylation (e.g., *DNMT3A, TET1, TET2, IDH1, IDH2*) and chromatin modification (e.g., *EZH2, ASXL1, KMT2A/MLL*), cellular signalling pathways of proliferation and survival (e.g., *FLT3, N-RAS, K-RAS*), key transcriptional regulators of haematopoiesis (e.g., *CEBPA, RUNX1, GATA2*), tumour suppressor genes (e.g., *TP53, WT1, PHF6*), RNA splicing (e.g., *SRSF2*, *U2AF1*, *SF3B1*) and formation of cohesions complex and chromatin architecture (e.g., *SMC1A, SMC3, STAG2*) [1,7]. In fact, recent functional studies revealed that in significant number of patients without a detectable cytogenetic aberration, AML emerges from functional cooperation of multiple alterations (e.g., *DNMT3A, TET2, IDH*, spliceosome mutations) that are often identified as molecular markers of potential pre-leukemic states such as clonal haematopoiesis of indeterminate potential (CHIP) and myelodysplastic syndromes (MDS) [7].

Although improved modern technologies have simplified the detection of genetic alterations in AML cells, the challenge remain in validating their function during initiation and development of the disease. These alterations are categorized as either potential driver mutations necessary for disease induction and/or maintenance or neutral passenger mutations that may not be part of disease aetiology. Despite improvement in ex vivo cell culture systems, significantly expanding primary AML blasts while preserving their naïve characters over a long period remains a challenging task [8]. In addition, significant patient-to-patient cell heterogeneity complicates studying common mechanisms that control AML biology. Thus, comprehensive functional characterization of many pathogenic phenomena could only be addressed using in vivo animal models, in particular, in genetically modified mouse strains. Unfortunately, despite collective efforts from many laboratories around the world, none of the existing models ideally recapitulate all aspects of the human disease. Nevertheless, the fast development of molecular and genetic engineering approaches has led to considerable progress. Faithfully modelling the complex heterogeneity of human AML in vivo will ultimately result in a better understanding of the molecular pathogenesis of the disease, identify genetic markers with predictive and prognostic value and develop novel personalized and efficient treatments strategies. Currently, mouse leukaemia models range from carcinogen-induced tumours, to transgenic animals expressing AML-associated proto-oncogenes and xenograft models based on transplantation of primary patient cells into immune-compromised mice (Figure 1).

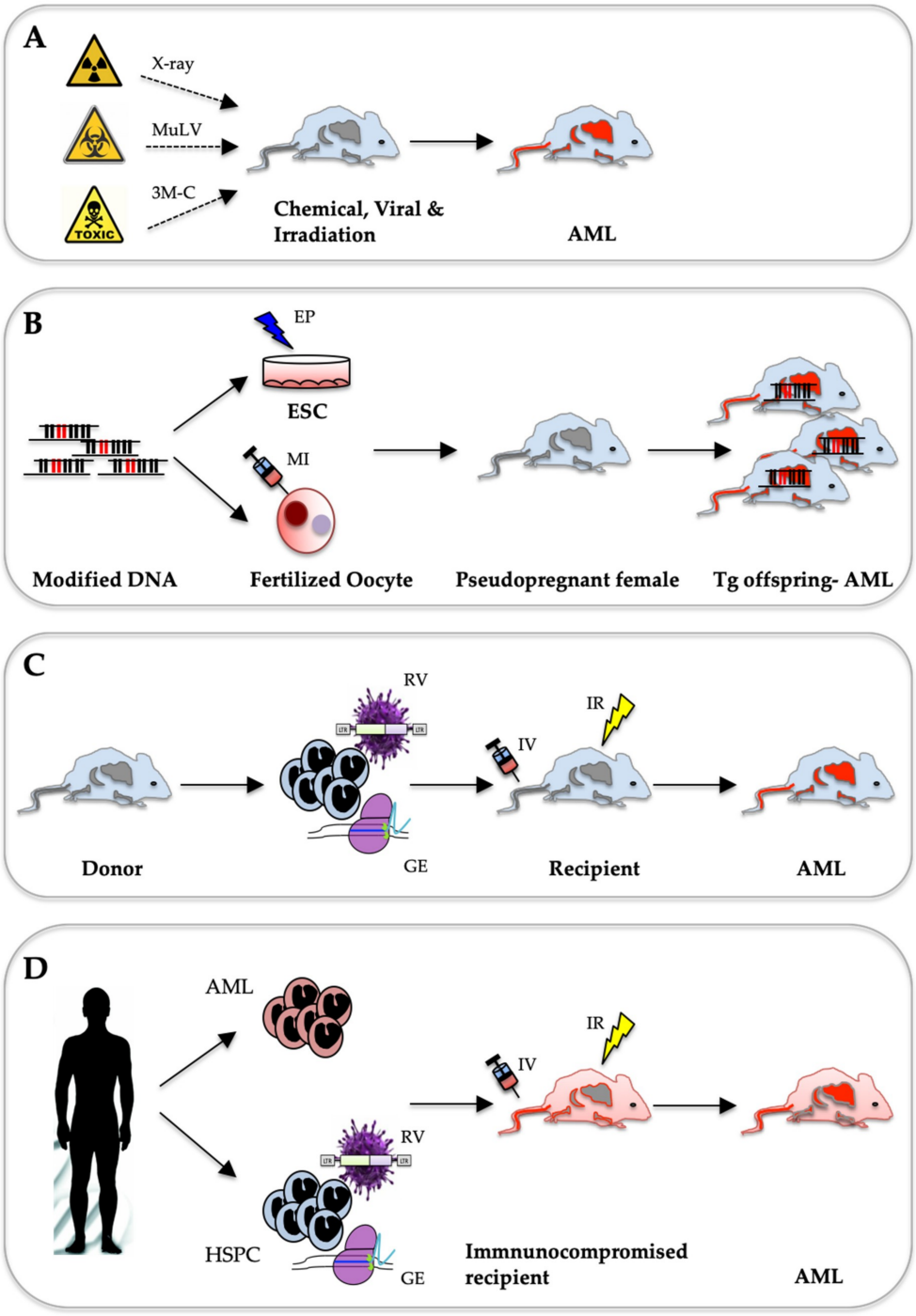



**Figure 1.** Schematics of different strategies for remodelling AML in mice. (**A**) Spontaneous AML development upon exposure to carcinogens like chemicals (e.g., 3-methylcholantrene; 3M-C), biologicals (e.g., murine leukaemia virus, *MuLV*) or radiation (X-rays). (**B**) Conventional transgenic approach: Transgenic (Tg) mouse lines are generated by DNA insertion into the genome, either randomly by pronuclear microinjections (MI) into fertilized Oocytes, or targeted by electroporation (EP) and homologous recombination in embryonic stem cells (ESC). (**C**) Adaptive transfer method of in vitro modified murine HSPC cells using either retroviral transduction (RV) or genome editing (GE) techniques followed by tail intravenous (IV) transplantation in irradiated (IR) recipients. (**D**) Xenotransplantation of either leukemic blasts or in vitro modified HSPC into immuno-compromised mice intravenously (IV) injected into irradiated (IR) recipients.

## 2. AML Mouse Models Induced by Chemicals, Viral Infection or Irradiation

A large number of studies that date from the last century have shown that AML can be spontaneously triggered in mice by chemical compounds, irradiation or particular viral infections (Figure 1A). Notably, modelling chemically and irradiation-induced AML also accounts for the effects of the environment, which is mostly disregarded in genetically engineered transgenic or xenograft leukaemia models.

### 2.1. Chemically Induced Leukaemia Models

One of the first reported and widely used leukaemia model is the L1210 cell line (L stands for Lloyd Law) isolated from DBA/2 mice exposed to the carcinogen 3-methylcholantrene [9,10]. The cells can be propagated in vitro and give rise to secondary leukaemia when transplanted. This model permitted the study of disease initiation, kinetics and effectiveness of newly developed leukaemia therapies [11]. The majority of chemotherapeutic agents, such as the widely used cytarabine, were selected for AML therapy during the late 1960s based on the in vivo efficacy against leukemic L1210 and other similar leukaemia models (P388, P1534 and L5178Y) [12]. However, the use of such cell line-based models had significantly diminished in the last decades due to several limitations. First, the pathology of the leukaemia induced by these cells does not fully phenocopy human AML, as mice often develop a lymphoid disease. Secondly, only a small number of animals develop the disease after a long latency on exposure to the inducing carcinogen [13]. Therefore, the study of AML development in individuals as a consequence of chemical exposure (e.g., benzene) [14–16] or as results of chemotherapy (e.g., alkylating agents and topoisomerase II inhibitors) [17,18] have relied mainly on epidemiological and direct analysis of primary patient-derived material, rather than on the use of mouse models [19].

### 2.2. Radiation-Induced Leukaemia Models

Leukaemia was one of the first malignancies reported as a radiation-induced cancer. Leukaemia incidents were significantly higher among X-ray workers and scientists working in close proximity to particle accelerators, especially before the introduction of safety measures [20]. Survivors of the atomic bomb explosion in Hiroshima and Nagasaki were exposed to high doses of irradiation of high energy radiation which resulted in rather rapid increased risk for developing haematological malignancies in particular chronic myeloid leukemia (CML), acute lymphoblastic leukemia (ALL) and AML [21]. The Chernobyl accident on the other hand, resulted in exposure to isotopes of lower energy which significantly increased the risk to develop thyroid cancers but was associated with a much lower risk of developing hematologic malignancies [22]. Multiple murine strains develop leukaemia on exposure to high and low-grade radiation, including the RFM, CBA, C3H and SJL/J [23]. Single high dose or prolonged low-grade full body irradiation such as gamma radiation, X-rays and neutrons has reported to induce leukaemia or mixed leukaemia/lymphoma development in mice.

Radiation-induced AML (RI-AML) in the RFM mouse line correlates with human data with comparable time of exposure to leukaemia development latency [24,25]. The clinical presentation of RI-leukaemia in the SJL/J mouse closely resembles that of secondary human AML, occurring after radiation therapy of patients with Hodgkin's disease [26]. One of the most interesting findings was that the incidence of RI-AML in SJL/J mice increased upon co-administration of corticosteroids and colony-stimulating factor-1 (CSF-1). This correlated with human findings where higher expression of CSF-1 could be associated with poor outcome in AML [27,28]. Strikingly, the most common cytogenetic feature detected in RI-AML models was recurrent deletions of chromosome 2. The identification that the minimal deleted region contained the *Sfpi1* gene encoding for the PU.1 transcriptional master regulator of myeloid differentiation shed some light on the underlying mechanism of disease initiation [25]. Later studies have shown that loss of one *Sfpi1* (PU.1) allele is not sufficient to induce a myeloid malignancy, despite the cells having a growth advantage [29]. A "second-hit" in these cells, in the form of a point mutation in the second *Sfpi1* (PU.1) allele in its DNA binding

domain (R235), is believed to transform these cells leading to clonal expansion and cancer [30,31]. Sequencing of AML samples from survivors of the Chernobyl accident showed similar mutational pattern with large chromosomal deletions and loss-of-heterozygosity (LOH) in multiple locations in the genome [32]. Experimental irradiation was also shown to accelerate the development of leukaemia in engineered mouse models, for example, such as the acute lymphoblastic leukaemia (ALL) associated with t(12;21)(p13;q22) leading to a *TEL-AML1* (aka *ETV6-RUNX1*) fusion, coupled with a loss of the *CDKN2A* cell cycle regulator gene [33]. This finding supported a model in which environmental low-grade radiation exposure may induce cooperating mutations to existing initiation lesions resulting in the expansion of pre-leukemic clones. Understanding the underlying molecular pathogenic mechanisms leading to RI-AML would help for radiation mitigation and to develop better radio-protective agents to reduce the incidence of secondary malignancies.

### 2.3. Virally Induced Leukaemia Models

Murine leukaemia viruses (*MuLV*) have been widely used to model the disease in susceptible mouse strains [34]. Pioneering work in the 1950s demonstrated that leukaemia could be induced and serially transmitted by injecting cell-free filterable *MuLV* supernatants into new-born mice [35]. Historically, murine leukaemia viruses were named after the scientist who originally characterized them, such as *Gross-MuLV*, *Friend-MuLV*, *Moloney-MuLV*, *Graffi-MuLV* and *Rauscher-MuLV*. Each of these virus strains results in recognizable and predictable patterns of disease in particularly susceptible mice strains, such as NIH/Swiss, DBA/2, AKXD, BXH-2 and C57BL-6 [36]. The original Friend virus preparation contains two retroviruses, a defective spleen focus forming virus (*SFFV*) and a replication competent murine leukaemia virus (*F-MuLV*) [37,38]. Two different *SFFV* strains have been identified: ($SFFV_P$) which reproducibly lead to polycythaemia and ($SFFV_A$) which results in anaemia. The target cell in which both *SFFV* express their pathogenic effect is an erythropoietin (EPO)-responsive progenitor cell identified as a late erythroid burst forming unit (BFU-E) or colony-forming unit (CFU-E). The envelope protein encoded by *SFFV* interacts with and activates the EPO receptor and sf-Stk (a truncated form of the Stk/RON receptor tyrosine kinase) causing EPO-independent proliferation, differentiation and survival. In the second stage, *F-MuLV* integration into the *Sfpi1* locus activates the myeloid transcription factor PU.1, blocking erythroid cell differentiation. Cells from diseased mice can be serially transplanted in vivo and propagated as permanent cell lines in vitro known as murine erythroleukemia (MEL) cells [39]. Subsequent studies suggested that aberrant PU.1 expression leading to functional inhibition of the GATA1 major erythroid transcriptional regulator is the causal event for blocked terminal differentiation [40,41].

Virally induced AML was also studied in the AKXD (recombinant inbred strain derived from AKR/J expressing two endogenous MuLV, *Akv-1* and *Akv-2*, and DBA/2J.) inbred mouse strain to identify putative leukaemia-inducing oncogenes through insertional mutagenesis [42]. Notably, the *ecotropic virus integration site-1* (*EVI1*) gene on 3q26 in the human genome, today a well characterized molecular marker in aggressive AML, was identified by analysis of retroviral integration site in *MuLV*-infected diseased AKXD mice [42]. Although a conclusive link between viral infection and AML induction in humans was never established, the use of these models has been instrumental for the identification and function of many AML-associated proto-oncogenes and the development of anti-leukemic therapeutic strategies. Improved molecular tools such as NGS for genome-wide viral integration site definition and the subsequent development of more sophisticated viral strains [43] resulted in several high-throughput insertional mutagenesis screens using both virus- or transposon-based (e.g., *Sleeping Beauty*) systems [44–46]. This approach of forward genetics was critical in identifying many cooperating proto-oncogenes that accelerate leukaemia development and eventually confer drug resistance [47].

## 3. Genetically Engineered Mouse Models

The molecular revolution in biological methods in the 70–80s of the last century allowed researchers to transfer foreign genetic elements into the germline of mice to create homogenous transgenic lines. To study the activity of putative proto-oncogenes, researchers integrated expression cassettes and mini-genes with promoter/enhancer elements, open reading frames (ORF) and transcript stabilization elements. A linearized copy of the engineered DNA is introduced by either direct injection into mouse oocytes or by electroporation into mouse embryonic stem (ES) cells (Figure 1B). Driven by the success of modelling B-cell leukaemia/lymphoma in transgenic mice expressing the *c-myc* oncogene under the control of immunoglobulin heavy chain (*IgH*) gene promoter [48,49], several conventional transgenic AML mouse lines were generated (Table 1). Further refinement of the technology rapidly increased the number of AML models driven by proto-oncogenes controlled either by their endogenous promoter or by inducible expression from a heterologous promoter.

### 3.1. Conventional Transgenic AML Models

The "classical" approach to establish a transgenic AML mouse model is based on direct injection of DNA fragments containing an ORF of the desired genetic and regulatory sequences into the pro-nucleus of fertilized oocytes. The zygotes are then transplanted into pseudo-pregnant foster mother mice. This results in a random integration of the transgene and founder animals are typically identified by either restriction enzyme digests and Southern blotting or genomic PCR assays [50]. Several groups explored this strategy to establish transgenic models for the *PML-RARA* fusion gene resulting from the t(15;17)(q24;q21) chromosomal translocation present in the vast majority of patients with acute promyelocytic leukaemia (APL). Hereby different regulatory elements directing transgene expression towards the myeloid lineage derived from human/mouse *cathepsin G (CG)* [51], *CD11b* [52] or *MRP8* [53] (*S100A9*) genes were used. The rather wide spectrum of the resulting phenotypes illustrates the complexity and limitations of this classical transgenic approach. Whereas expression of *hCG* or *MRP8* controlled *PML-RARA* expression was able to induce AML or APL-like phenotypes with incomplete penetrance after long latency [51,53], *CD11b/PML-RARA* transgenic mice did not develop any leukaemia [52]. Nevertheless, classical transgenic mice were instrumental to show that *PML-RARA* indeed is the genetic driver of APL and to study the underlying molecular mechanisms leading to the first (and so far only) really efficient targeted AML therapy based PML-RARA degradation by all-trans-retinoic acid (ATRA) and/or arsenic trioxide [54]. Notably, another classical transgenic model for APL associated with a *PLZF-RARA* fusion gene derived from t(11;17)(q23;q21) revealed that ATRA was unable to induce remission in mice, faithfully recapitulating the clinical response in the respective patients [55]. We list some of the most important classical transgenic AML models in Table 1 and refer to respective review articles [56–58]. Unfortunately, the classical transgenic mouse approach is inefficient, technically challenging, time- and cost-consuming and as illustrated by the APL mouse model and others, unable to recapitulate the desired phenotypes [59]. Therefore, it is most likely that a large number of classical transgenic mouse lines expressing leukaemia-associated proto-oncogenes that were not able to phenocopy human disease remained unpublished.

**Table 1.** Transgenic mouse lines modelling AML.

| Year | Transgene | Strategy | Promoter | Inducer | Cellular Target | Phenotype | Ref. |
|---|---|---|---|---|---|---|---|
| 1996 | *PML-RARA* | Conventional | *CD11b* | | Myeloid lineage (BM, periphery) | Abnormal myelopoiesis. No APL | [52] |
| 1997 | *PML-RARA* | Conventional | *hCG* | | Myeloid lineage (BM, periphery) | Myeloid cells expansion in BM and spleen. AML-like with 30% penetrance after long (> 100 days) latency | [51] |
| 1997 | *PML-RARA* | Conventional | *hMRP8* | | Myeloid lineage (BM, periphery) | APL-like disease (median 174 days) | [53] |
| 2000 | *RUNX1-ETO* | Conditional | *Tet* | *tTA* | BM | Abnormal haematopoiesis. No AML | [60] |
| 2001 | *RUNX1-ETO* | Conventional | *hMRP8* | | Myeloid (neutrophils & monocytes) | AML-Only upon new-born treatment with ENU | [61] |
| 2006 | *Cbfb-MYH11* | Conditional | *Cbfβ* | *Mx-iCre* | BM (LSK) | AML-Aberrant myeloid progenitors, blocked megakaryotic differentiation. | [62] |
| 2008 | *Mll-AF9* | Knock-in (*Mll1; Mllex8-AF9* cDNA) | *Mll* | | | AML-Higher *MLL-AF9* expression in HSCs than GMPs. | [63] |
| 2014 | *MLL-ENL* | Conditional | *TRE (Col1a)* | *rtTA* | LT-HS, pMeg/E, HSC, MPP, GMLP, CLP | AML- no leukaemia from HSC | [64] |
| 2016 | *MLL-AF9* | Conditional | *TRE (Hprt)* | *rtTA* | LT-HSC, ST-HSC, CMP, GMP | AML-dependent on DOX dose and cellular origin | [65] |
| 2018 | *MLL-ENL* | Conditional | *TRE (Hprt)* | *rtTA* | LT-HSC, LMPP, CMP | AML-MLL-dependent on DOX dose and cellular target | [66] |

Conventional (DNA injection into Oocytes), Knock-in (homologous DNA recombination in ES cells), Conditional (regulated expression), LSK (lineage marker negative, $Sca1^+$, $cKit^+$), MPP (multipotent progenitors), GMLP (granulocyte-macrophage-lymphoid progenitors), CLP (common lymphoid progenitor), ST-HSC (short term hematopoietic stem cells), DOX (doxycycline).

### 3.2. Transgenic AML Models by Homologous Recombination in ES Cells

The site of integration of the cloned DNA can be specifically directed toward the desired gene locus thanks to the development of targeted homologous recombination (HR) in murine embryonic stem (ES) cells [67]. This technology was explored early on to establish AML mouse models and a positive proof of concept was provided by the first transgenic model for the *MLL-AF9* fusion gene associated with myelomonocytic leukaemia [68]. Physiologically, the mixed lineage leukaemia (*MLL*; aka *KMT2A*) gene encodes for a regulator of self-renewal and differentiation of hematopoietic stem cells (HSC) and is target of recurrent chromosomal translocations that lead to fusions of its amino-terminus to the carboxy-terminus of one of > 60 different partner loci [68]. MLL fusion genes are the molecular hallmark of more than 70% of infant acute leukaemia, 5–10% of adult de novo AML and an increasing number of secondary and therapy-related AML [69]. The most prevalent translocations comprise t(9;11)(p22;q23), t(11;19)(q23;p13) and t(4;11)(q21,q23) leading to the *MLL-AF9*, *MLL-ENL* and *MLL-AF4* fusion genes, respectively [69]. To express the *MLL-AF9* fusion from its native regulatory elements, Rabbitts and co-workers successfully integrated a short *MLL exon8-AF9* cDNA-poly-A fragment into the mouse *Mll1* locus by homologous recombination [70]. Interestingly, despite the widespread activity of the *Mll1* promoter, chimeric mice only developed AML. Notably, joining *Mll1* exon8 with a bacterial *lacZ* gene was sufficient to induce leukaemia in some chimeric mice after prolonged latency [71]. Subsequent studies with the *Mll-AF9* knock-in mouse line demonstrated pre- and postnatal stepwise progression of the disease [72], the role of the HOXA9 homeobox transcription factor as downstream effector [73] and gene dosage effects as well as putative cellular targets of MLL-AF9 to initiate AML [63].

The success of modelling AML by constitutive integration of a driver fusion oncogene into its natural locus encouraged researchers to model the function of other alterations. The fusion genes of the core-binding factor (CBF), a heterodimeric essential HSC regulator composed of RUNX1 (AML1) bound to CBFβ, is involved in balanced chromosomal rearrangements found in 20–30% of human AML [74]. RUNX1, was initially identified as a target of the t(8;21)(q21;q22) chromosomal translocation which results in expression of a fusion protein that contains the N terminus of RUNX1 fused to a nearly full-length ETO (Eight-Twenty-One, aka RUNXT1 or MTG8) protein. Knocking-in the *RUNX1-ETO* fusion gene into the murine *Runx1* promoter lead to embryonic lethality and a lack of definitive haematopoiesis in the foetal liver, very similar to those seen in *Runx1*$^{-/-}$ knockout mice [75]. CBF is also target of inv(16)(p13q22) leading to expression of a *CBFβ-MYH11* fusion gene. The resulting fusion protein was shown to interact with RUNX1 and to outcompete binding of wild-type CBFβ in a dominant-negative fashion. Not surprisingly, expression of a single copy of *CBFβ-MYH11* from the *Cbfb* promoter (*Cbfb*$^{+/MYH11}$) resulted in a similar lethal phenotype as *Runx*$^{-/-}$ mice [76]. Thus, to be able to study the role of these CBF fusions for leukemogenesis in vivo, it was essential to express them in a spatially and temporally-controlled manner.

### 3.3. Conditional Transgenic AML Mouse Models

#### 3.3.1. Modelling AML-Associated Fusions

Conditional gain-of-function models of AML-driving (fusion)-oncogenes are mostly generated by inserting a strong translational and transcriptional termination (STOP) sequence flanked by *LoxP* or *Flp* recombinase recognition target site (FRT) cleavage sites between the promoter sequence and the ORF of interest. In presence of a C- (Cre) or FLP recombinase, the STOP cassette is removed, allowing the expression of the transgene. The same approach can also be used to ablate essential parts of a gene of interest by deleting regions flanked by *LoxP* sites. Utilizing this setting, transgenic mice carrying the floxed genes are usually crossed with transgenic lines that express the Cre recombinase in the hematopoietic tissue. Cre expression is typically driven under the control of spatially and/or temporally controllable promoters such as the hematopoietic-restricted *Vav1* (*Vav1-iCre*) promoter, the interferon-inducible *Mx1* promoter (*Mx1-iCre*) or as a fusion to a mutated oestrogen receptor (ER) ligand binding domain (*Cre-ER*), which can be activated by tamoxifen (TAM) [77].

An additional layer of complexity can be added by engineering transgenic cassettes controlled by minimal promoters that are sensitive to chemical inducers like tetracycline (Tet) [78], or one of its derivative such as doxycycline (DOX). The Tet system can be designed to inhibit (Tet-off) or induce (Tet-on) the expression of a transgene, by either coupling it with a Tet-sensitive transcriptional repressor (*tTA*) or a reverse-tTA (*rtTA*) transcriptional activator respectively [79]. A Tet-off expression system was applied to bypass the embryonic lethality associated with constitutive *RUNX1-ETO* expression [60]. Another study established Cre-responsive conditional *RUNX1-ETO* knock-in mice [80]. However, despite robust expression of the *RUNX1-ETO* fusion transgenes in the BM upon tetracycline withdrawal in the first model and efficient excision of the *floxed STOP* codon in the second model, no leukaemia developed. In subsequent studies, researchers showed that mice conditionally expressing *RUNX1-ETO* developed leukaemia only upon treatment with genotoxic agents such as N-ethyl-N-nitrosourea (ENU) [61]. ENU is a strong carcinogenic mutagen which transfers its ethyl group to oxygen or nitrogen radicals into DNA, resulting in miss-pairing and base pair substitutions which translates to the production of proteins with missense mutations and aberrant splicing events [81]. A *RUNX1-ETO* leukaemia model that allows for conditional and reversible controlled mosaic expression of the fusion in hematopoietic progenitors was established by transplanting whole BM carrying a *ROSA26-iM2-tetO* DOX inducible promoter and the fusion cDNA (*ROSA26-iM2-tetOGFP/TgPtet-AML1-ETO*) into lethally irradiated mice [82]. Hereby the researchers were able to recapitulate the slow disease evolution and mosaic expression found in human RUNX1-ETO$^{+}$ AML. Transcriptional analysis from different hematopoietic populations during disease progression demonstrated that the fusion alters the transcriptional expression of HSC and committed progenitors. However, despite showing signs of a myeloproliferative leukaemia-like disease, all the mice survived. This finding is consistent with the idea that *RUNX1-ETO* expression is necessary but not sufficient to induce a fully penetrant AML. Indeed shRNA-targeted degradation of the fusion significantly reduced proliferation and survival of *RUNX1-ETO*-expression AML cells [83]. Later studies found functional cooperation of *RUNX1-ETO* with mutations in tyrosine kinases such as *c-KIT, FLT3-ITD* or the *TEL-PDGFβR* fusion in different mouse models [84,85].

A similar conditional mouse model was developed for the *CBFβ-MYH11* fusion gene, called *Cbfb*$^{+/56M}$ [62]. Wild-type *Cbfb* cDNA (exon 5 and 6 and a polyadenylation signal) flanked by *LoxP1* sites was inserted into intron 4 of the previously generated transgenic *Cbfb*$^{+/MYH11}$ knock-in allele. Hereby the wild-type *Cbfb* transcript is temporarily expressed from the ''floxed'' *Cbfb*$^{56M}$ allele. However, in presence of Cre, the knock-in allele is restored and a *Cbfβ-MYH11* fusion is expressed. Strikingly following injection of polyinosinic:polycytidylic acid (poly(I:C)) activating *Mx1-iCre*, 90% of the mice developed AML after a median latency of 5 months demonstrating that the fusion is indeed a driver of AML [62].

We have established a series of Tet-regulated transgenic mice to model acute leukaemia driven by the most prevalent MLL fusion genes [65,66]. We were particularly interested to study the role of the cellular origin on AML onset and progression. Using this model, we were able to show that conditional expression of the MLL-AF9 fusion in long-term HSC (LT-HSC) resulted in a more aggressive phenotype than activation in the committed granulocyte-macrophage (GMP) or common myeloid progenitors (CMP) [65]. Notably, in a subset of mice, activation of MLL-AF9 led to a particularly invasive and drug-resistant phenotype characterized by expression of genes previously associated with epithelial-mesenchymal transformation (EMT) observed in solid cancers. Cross-species comparative gene expression profiling suggested that similar to MLL-AF9 driven AML in mice, some AML patients (not only those carrying MLL-fusions) expressed similar EMT-related genes associated with poor outcome [65]. In contrast to *MLL-AF9*, conditional expression of the *MLL-ENL* fusion using the same conditional Tet-on system was not able to transform GMP but induced a rather mixed myeloid-lymphoid leukaemia when activated in HSC, lymphoid-myeloid progenitor population (LMPP) or CMP [66]. Comparison with another Tet-regulated MLL-ENL transgenic mouse model

suggested that the leukemic phenotypes might be influenced by the expression level of the transgene in cells of a particular stage of the hematopoietic hierarchy [64].

#### 3.3.2. Modelling AML-Associated Mutations and Aberrantly Expressed Genes

Several transgenic mouse models have been generated to model AML carrying NPM1 mutations [86]. Conventional transgenes in which expression of mutated *NPM1* was regulated by the human *MRP8* promoter [87] and a knock-in model mimicking the human mutation in the mouse *Npm1* [88] developed myeloproliferative disease only but no AML. Conditional ex vivo activation of a human *NPM1* mutant cDNA integrated in the *Hprt* locus followed by transplantation into irradiated WT mice induced a late-onset AML-like disease in about 30% of the recipients [89]. Conditional expression of a humanized *NPM1c* knock-in allele in the hematopoietic system (mediated by *Mx1-iCre*) resulted in the development of late onset AML in about 30% of the mice, however this percentage increased to 80% following the activation of cooperating proto-oncogenes through the use of the *Sleeping Beauty* insertional mutagenesis system [46]. Collectively, these models indicated that an NPM1 mutant is not sufficient to induce clinical AML.

Transgenic mouse models have also been established to model the role of the Flt3 (Fms-related tyrosine kinase 3) internal tandem repeats (FLT3-ITD) mutation found in > 20% of human AML [90]. Two independently established knock-in mouse lines carrying an ITD mutation in the juxta-membrane domain of murine *Flt3* slowly developed a myeloproliferative disease but no acute leukaemia [91,92]. However, it is important to note that $Flt3^{ITD}$ models were instrumental to demonstrate the impact of the gene dosage, loss of the wild-type allele and FLT3 ligand on phenotype development [93,94].

Very similar to FLT3 mutations, activation of conditional transgenic knock-in alleles of AML-associated $K\text{-}RAS^{G12D}$ and $N\text{-}RAS^{G12D}$ mutations resulted a highly penetrant myeloproliferative phenotype but was not sufficient to induce AML [95]. Thus, to be able to study cooperation between co-occurring mutations in AML, an increasing number of compound transgenic/knock-in mouse lines are generated (Table 2). For example, crossing the *NPM1c* with $Flt3^{ITD}$ knock-in strains revealed a powerful molecular synergy with the development of highly penetrant acute leukaemia [96]. Transgenic $Flt3^{ITD}$ expression was also shown to cooperate with $Mll^{PTD}$ [97], the *NUP98-HOXD13* fusion [98], the $Wt1^{R394W}$ [99] mutation or with *Dnmt3a* [100] haploinsufficiency to cause AML. Potent in vivo oncogenic cooperation was also demonstrated by crossing the $N\text{-}Ras^{G12D}$ knock-in strain with transgenics expression of the *MLL-AF9* fusion gene [101], the anti-apoptotic regulator *BCL2* [102] or the *Cbfβ-SMMHC* fusion [103]. Transgenic expression of $K\text{-}RAS^{G12D}$ increased the penetrance of the APL-like phenotype in cathepsin-G driven *PML-RARA* transgenic mice [104]. Collectively, mouse models have shown that AML-associated NPM1c, FLT3-ITD and N-/K-RAS mutations are *per se* not sufficient to induce the disease but act as potent cooperating lesions.

**Table 2.** Compound transgenic mouse AML models.

| Year | Co-Op Mutations | Activity | Promoter | Inducer | Cellular Target | Phenotype | Ref. |
|---|---|---|---|---|---|---|---|
| 2007 | *NRAS12D* + *BCL2* | Const.<br>Cond. | *hMPP8*<br>*Tet* | *rtTA* | Myeloid lineage (BM, periphery) | MDS/AML | [102] |
| 2012 | *MLL-PTD* + *FLT3-ITD* | Const. | *Mll* + *Flt3* | | *Mll* and *Flt3* expressing cells | AML with 100% penetrance | [97] |
| 2012 | *NUP98-HOXD13* + *FLT3-ITD* | Conv. (*FLT3-ITD*)<br>Conv. (*NUP98-HOXD12*) | *Flt3*<br>*Vav* | | Hematopoietic lineage cells (FL, BM) | AML with 100% penetrance | [98] |
| 2012 | *KRAS-G12D* + *PML-RARA* | Cond. (*Kras-G12D*)<br>Const. (*PML-RARA*) | <br>*hCG* | *Mx-iCre* | Myeloid lineage (BM, periphery) | APL-like Disease with 69% penetrance, remaining mice developed MDS | [104] |
| 2013 | *NPM1c* + *FLT3-ITD* | Cond. (*NPM1c*)<br>Const. (*Flt3-ITD*) | *Mx1* | *Mx-iCre* | Hematopoietic lineage cells (BM) | AML after short latency (median 49 days) | [96] |
| 2014 | *NRAS-G12D* + *CBFβ-SMMHC* | Cond. | *Mx1* | *Mx-iCre* | Hematopoietic lineage cells (BM) | AML after short latency (median 13.7 weeks) and full penetrance | [103] |
| 2017 | *NPM1c* + *NRAS-G12D*<br>*NPM1c* + *FLT3-ITD* | Cond. | *Mx1* | *Mx.iCre* | Hematopoietic lineage cells (BM) | AML with 95% penetrance, some mice develop MPN<br>AML with 100% penetrance | [46] |
| 2018 | *WT1-R394W* + *FLT3-ITD* | Const. | *Wt1* and *Flt3* | | Wt1 and Flt3 expressing cells | MPN-like disease or T-ALL after short latency-AML associated with LOH of *Flt3* | [99] |

Const. (constitutive expression), Cond. (regulated expression), Conv. (conventional), FL (fetal liver), MPN (myeloproliferative neoplasms), T-ALL (T-cell acute lymphoblastic leukemia).

A transgenic mouse line remodelling the aberrant expression of the *EVI1* gene mediated by 3q21-3q26 chromosomal translocations or inversions leading a hallmark of particularly aggressive AML was recently established [105,106]. All the breakpoints detected in patients cluster within an approximately 25kb region, which in the mouse maps to -77 kb upstream of the *Gata2* gene. To test whether this region possess enhancer activity, researchers established a transgenic mouse line with a fluorescent reporter cloned 186 kb downstream of 5′ sequences flanking the *Gata2* gene [106]. Strong reporter signal was detected in HSPC and the sequence was thus designated as the *Gata2* distal hematopoietic enhancer (G2DHE). The same researchers then established a bacterial artificial chromosome (BAC) transgenic mouse that allowed the induction of *EVI1* expression with or without *G2DHE* region. All mice with an intact *G2DHE* developed leukaemia in accordance with transgene copy number, where two copies gave rise to B-cell, three copies resulted in myeloid and four copies led to mixed lineage leukaemia within 200 days. However, mice lacking the *G2DHE* region did not show sign of disease during the 400 days of observation indicating that the GATA2 enhancer plays a critical role [106]. This study confirmed and extended the observations that genomic excision of a distal *GATA2* enhancer led to *EVI1* silencing, growth inhibition and differentiation of human AML cells with inv(3)(q21q26) or t(3;3)(q21;q26) [107]. More recently, a transgenic mouse line was established in which *EVI1* expression is under the control of a Tet inducible (tet-on, "TO") promoter ($Evi1^{TO/+}/Rosa26^{rtTA}$) [108]. To recapitulate the clinical presentation of EVI1 overexpressing leukaemia, researchers performed competitive 1:1 transplantation with $Evi1^{TO/TO}/Rosa26^{rtTA}$ with WT BM cells. Using this approach all mice developed symptomatic AML within 90–119 days, clearly demonstrating its oncogenic activity [108].

Transgenic mouse models of leukaemia have been vital for our understanding of the role of genetic aberration in the induction and maintenance of the leukemic condition. However, one of the main shortcomings of these models is their inability to reliably reproduce the leukemic phenotype observed in patients carrying the genetic lesion. Several factors could attribute for that; such as the evolutionary difference between the human and mouse haematopoiesis systems, the effect of unaccounted genetic variability in the human genome (e.g., SNP) and generating transgenic mouse lines with the cDNA sequence only. This might subsequently lead to the potential loss of essential regulatory elements located in the intronic regions of mutated genes, causing an alteration in the dynamic expression of the genetic lesion in targeted cells and thus phenotypic differences. For example, the dynamic expression of the *Gata1* gene in erythroid cells versus HSC was shown to be depended on untranslated regulatory elements located at its 5′ region [109,110].

## 4. Mouse Models Based on Adaptive Transfer of Hematopoietic Cells Virally Expressing an AML-Associated Proto-Oncogene

Technologies developed during last two decades of the 20th century allowed to transfer the cloned leukaemia-associated genetic aberrations into hematopoietic cells to explore their transforming potential in vitro and in vivo. Production of replication-incompetent high titre retrovirus expressing a gene of interest was critical to develop the widely used adoptive transfer protocol to model the effects of leukaemia-associated genetic lesions in hematopoietic stem and progenitor cells (HSPC) of the mouse. Hereby, virally transduced cells are transplanted into lethally or sub-lethally irradiated syngeneic recipients, resulting in chimeric animals in which the donor-derived transformed HSPC may outcompete the host haematopoiesis ultimately leading to leukaemia (Figure 1C). The power of this strategy became first evident by studies of the Baltimore laboratory that modelled the effect of the chronic myeloid leukaemia (CML)-associated *BCR-ABL* fusion gene [111]. Transplantation of BM cells transduced with a retrovirus carrying the *BCR-ABL* fusion cDNA induced hematologic malignancies in about half of the recipients: either a CML-like myeloproliferative syndrome, acute lymphoblastic leukaemia (ALL) or tumours containing macrophage-like cells occurring after mean latencies of 9, 14 and 16.5 weeks respectively. Notably they were able to transfer the disease phenotype by transplanting tumour cells into irradiated secondary recipients.

Following this landmark study, this approach, often referred to as the transduction-transplantation model, was further refined and successfully used to model the transforming activity of a large number of AML-associated genetic alterations [112]. BM reconstitution with HSPC expressing a gene or mutation of interest was not only instrumental to demonstrate the transforming potential but also to validate functional cooperation of different mutation classes necessary to induce a leukemic phenotype, such as transcription factor fusion genes involving *CBF*, *RARA* or *NUP98* cooperating with *FLT3* or *N-/K-RAS* mutations [113]. Some of the most important models that were established by this approach are listed in Table 3. Many of these studies suggested that most AML-associated mutations are not sufficient to induce the disease. The versatility of the system allowed researchers to define many critical downstream effectors of AML driver mutations. In addition, such studies also suggested that the cellular origin might be an important nominator of transforming potential of AML-associated mutations. Transduction of enriched hematopoietic stem and distinct progenitor cells showed that in contrast to *BCR-ABL*, AML-associated *MOZ-TIF2* [114], *MLL-AF9* [115,116], *MLL-ENL* [117,118], *AML1-ETO* [119] and *MLL-GAS7* [120] fusion genes were able to transform committed progenitor cells. In addition, selective expression in different myeloid progenitor cell populations (CMP vs. GMP) revealed a differential transforming activity of the of the *meningioma 1* (*MN1*) gene, often overexpressed in aggressive AML [121].

In the majority of these studies, researchers used replication-deficient *murine stem cell virus* (*MSCV*)-based expression vectors, which allow efficient transduction and stable transgene expression in hematopoietic progenitor cells [122]. However, it is worth noting that viral integration events, potentially non-physiological expression level, batch to batch transduction and transplantation variability and the inherent transduction bias for early multi-potent HSPC may influence the disease phenotype. Nevertheless, in general the AML disease arising in these mice share a common histopathological and immunophenotypic features, best illustrated by the MLL-AF9 fusion. Independ whether the fusion is expressed retrovirally or as a knock-in transgene, the resultant disease is characterized by extensive infiltration of the BM and other organs by myeloid progenitors and monoblasts expressing high levels of Gr1, Mac1 and c-Kit surface markers [63,65,115,116]. The adaptive transfer model is still the prime experimental method to investigate the in vivo transforming potential of AML-associated genetic aberration. It provides a relatively rapid and robust methodology to explore the function of one or more AML-associated mutations or overexpressed genes in cells of the hematopoietic system.

**Table 3.** AML mouse models based on viral transduction and transplantation.

| Year | Transgene | Viral Vector | Cellular Target | Phenotype | Ref. |
|---|---|---|---|---|---|
| 1990 | *BCR-ABL* | *pMSCV-pgk-neo* | Total BM | Myeloproliferative malignancy, ALL and CML-like | [111] |
| 1997 | *MLL-ENL* | *pMSCV-IRES-GFP* | *Thy-1$^{lo}$Sca-1$^{+Hi}$-2K$^{hi}$, 5-FU treated BM* | Self-renewal in vitro & AML in vivo | [117] |
| 2002 | *RUNX1-ETO* | *pMSCV-IRES-GFP* | *HSC c-Kit$^{+}$Sca-1$^{-}$Lin$^{-}$* | Myeloid developmental abnormality but no AML | [118] |
| 2003 | *MLL-GAS7* | *pMSCV-pgk-neo* | *HSPC* | Mixed lineage leukaemia phenotype | [120] |
| 2004 | *MOZ-TIF2, BCR-ABL* | *pMSCV-IRES-GFP* | *CMP, GMP* | MOZ-TIF2 but not BCR-ABL resulted in transplantable AML in vivo | [114] |
| 2006 | *MLL-AF9* | *pMSCV-IRES-GFP* | *GMP* | Transplantation of transduced cells propagated in MC resulted in AML in vivo | [115] |
| 2011 | *MN1* | *pMSCV-pgk-neo* | *CMP, GMP* | CMP are susceptible for MN1 transformation, GMP required co-expression of MEIS1 for AML induction | [121] |
| 2012 | *MLL-AF9* | *pMSCV-pgk-puro* | *Evi1$^{+/-}$* MLL-AF9 transduced cells | Knockdown of *Evi1* delayed leukaemia induction in vivo | [116] |

FU (Fluorouracil), MC (methylcellulose).

## 5. Modelling AML by Transferring Patient-Derived Cells into Immune-Compromised Mice

Ex vivo maintenance and expansion of even the most clinically aggressive patient-derived leukemic blasts remains a technical challenge. Even very sophisticated culture systems cannot fully replace the complex interactions between leukaemia cells and the BM microenvironment. To overcome these limitations researchers explored transplantation of human primary AML cells into immune compromised mice (Figure 1D). Several immunodeficient mouse strains were developed for patient-derived cell xenotransplants (PDX) including *nude* (*nu*), severe combined immunodeficient (SCID), non-obese diabetic (NOD), NOD-SCID and NOD-SCID-IL2r$\gamma^{null}$ (NSG) strains [123,124].

In one of the first PDX experiments, researchers transplanted primary AML cells into *nude* mice that are athymic due to a homozygous *nude* mutation (encoding for a forkhead box transcription factor (FoxN1), resulting in lack of functional T cells. However, due to an intact B cell and NK cell function, grafting of normal as well as leukemic cells remained poor and was often associated with formation of extramedullary granulocytic tumours [125]. Even in mice carrying triple homozygous mutations in *nude*, *beige* (affecting the lysosomal trafficking regulator; *Lyst*) and *Xid* (X-linked immunodeficiency gene, Bruton's tyrosine kinase; *Btk*) xenografting of human AML cells remained inconsistent and unreliable [126]. The development of severe combine immunodeficient (SCID) mice was an important step forward for the development of humanized AML mouse models. SCID mice carry inactivating mutations in the *protein kinase DNA-activated catalytic polypeptide* (*Prkdc*) gene, which protein product is involved in DNA repair pathways. This leads to improper immunoglobulin *V-D-J* gene recombination, subsequently resulting in mice lacking functional mature T and B cells, however retaining NK function [125]. Although primary AML injected intraperitoneally or implanted under the kidney capsules showed improved engraftment rates, intravenous injection remained poor [127]. To overcome these limitations, researchers began to transfer the cells directly into the recipients' BM by intrafemoral injection. To further improve engraftment rates, mouse models with more severe immunodeficiency were developed by combining the SCID background with the non-obese diabetic (NOD) strain. Combined non-obese diabetic NOD-SCID mice have no functional B or T cells and reduced NK cell and macrophage activity [128]. They showed superior engraftment rate compared to SCID mice even when injecting fewer primary AML cells [129]. Moreover, the morphologic, phenotypic and genetic characteristics of the expanded AML specimens seemed mostly preserved [129]. The ability to initiate the AML from few number of phenotypically stable cells allowed researchers to propose the existence of an AML-cell hierarchy with leukemic stem cells (or SCID-Leukaemia initiating cells; SL-ICs) enriched in the lineage marker-negative $CD34^+/CD38^-$ compartment [130,131]. Crossing of NOD-SCID mice with *IL2R*$\gamma^{-/-}$ mice resulted in an even more immune compromised (NOD/LtSz-*scid* with *IL2*$\gamma_c{}^{null}$; NSG) strain [132]. Deletions in the interleukin-2 receptor gamma chain (*IL2R*$\gamma$) led to an almost complete absence of the murine immune system and improved AML engraftment [133]. To further humanize the hematopoietic system, the NSG strain was crossed with knock-in mice expressing genes of three human cytokines (*hIL3*, *hGM-CSF* and *hSCF*) (NOD/LtSz-*scid IL2*$\gamma_c{}^{null}$–SGM3; or NSG-S) [134,135]. NSG-S mice showed significantly improved expansion of normal human myeloid cells and enhanced engraftment rates of primary patient AML cells [134,136]. NSG strains carrying null alleles for major histocompatibility complex class I and class II beta2-microglobulin (β2m) called NSG-β2m were developed to minimize reactivity of human immune cells against host tissue and thus specifically reduce graft versus host disease (GVHD) [137,138]. Notably, increased in engraftment rate for AML cell lines and primary paediatric patient samples in these mice without the need for irradiation was reported [139]. More recently, $NSG^{W/V}$ and $NSG^{W41}$ mouse strains were obtained by breeding NSG with strains carrying *c-kit* loss-of-function alleles (*Kit*$^{WV/WV}$ and/or *Kit*$^{W41/W41}$) [140]. Loss of *c-Kit* impairs HSCs of the host and thereby creating empty BM niches leading to a competitive advantage for transplanted human HSPC. These strains supported engraftment of human $CD34^+$ cord blood cells (CBCs) without prior host irradiation. They also showed greater engraftment and appropriate differentiation of human cells of the erythroid and megakaryocytic lineages [141]. In addition to the severity of immunodeficiency of the host, expression of human engraftment-enhancing cytokines and

creating empty niches in the BM, the mutational status of the AML cells and the observation time seem also key determinants for successful expansion in PDX [142–144].

Many compounds that showed significant anticancer effects in vitro and in transgenic mouse models failed to show efficacy in clinical trials, most likely due to the unaccounted complexity of the mutational load of human AML and effect of the microenvironment [145]. To circumvent this limitation, the PDX model has been suggested as a good system to evaluate the efficacy of chemotherapeutic agents on human AML cells in vivo [146,147]. Combination therapy of cytarabine and doxorubicin on freshly transplanted human MLL-AF9$^{+}$ leukaemia in NSG mice resulted in a reduction in residual disease burden [147]. Doxorubicin treatment had a profound effect on AML cells compared to mouse BM cells, in contrast to cytarabine which had a greater toxic effect on mouse BM cells. Transplanted primary samples showed variable sensitivity to chemotherapy, correlating with patients' clinical outcome [146]. In another study, transplantation of Ara-C-resistant primary human AML cells into NSG mice revealed a role for mitochondria and elevated oxidative metabolism in leukemic cells' chemo-resistance [148]. Thus, the PDX system seem to provide an experimental platform to test the efficacy of novel therapeutic compounds against primary human AML cells and to study the mechanisms of chemo-resistance.

Although AML xenotransplantation into immunodeficient murine models is a valuable tool for the expansion and study of some aspects of the biology of human AML, these models are still limited by their inability to address the interplay of leukemic blasts with different cells of the immune system and to dissect the cell autonomous from cell non-cell autonomous aspect of the disease as they tend to develop other spontaneous malignancies. To overcome these limitations, scientists took advantage of new advances in the fields of bioengineering and synthetic material development to create biological inserts or scaffolds [149]. The function of these scaffolds is to create humanized microenvironment in the mouse that is efficient in supporting implanted cells expansion and differentiation without altering their character and function. Successful primary AML cells engraftment was achieved using polyurethane scaffolds coated with freshly isolated human BM-derived mesenchymal stem cells (hMSC) in NOD-SCID mice [150]. The subcutaneously implanted scaffold remodelled the architecture of human BM niche (with de novo vascularization and osteoclast and adipocyte development) at the site of implantation and supported the initial expansion and spreading (BM, liver and kidney) of pre-implanted and retro-orbitally injected AML cells. In another study, hMSC coated ceramic scaffolds were able to support the engraftment of favourable non-engrafting AML samples when implanted subcutaneously in NSG mice [151]. The implanted insert supported cellular proliferation and maintained clonal heterogeneity and leukemic stem cells' (LSC) self-renewal capacity in methylcellulose cultures. In a different approach, researchers also used freshly collected human BM biopsies from hip replacement patients and directly transplanted them subcutaneously into the flanks of NSG mice, using Matrigel as carrier [152]. The human BM tissue showed vascularization and rapid engraftment of intravenously injected MOLM-13 human AML cells.

In an attempt to study the heterogenous sub-clones detected in human patients [153] and follow the stage specific transformation of MDS to AML [154], scientists took advantage of induced pluripotent stem cells (iPSC) technique to establish AML- and MDS-iPSC respectively. Despite the challenges in deriving myeloid malignant IPS cells due to their inherent resistance to reprogramming and apoptotic priming in ex vivo culture, both studies showed that in a pluripotent state, AML-iPSC do not possess a transformed phenotype. However, these cells appear to retain their leukemic potential upon induced hematopoietic differentiation and are able to induce disease in NSG mice. Interestingly, when stimulated to differentiate towards non-hematopoietic lineage, AML-derived IPSC are able to form non-malignant cells from all three embryonic germ layers [153].

These novel approaches in leukemic remodelling using the classical PDX system have provided proof of concept solutions on how to overcome some of challenges associated with the system, such as the difference between the mouse and human BM microenvironment and AML samples the *intra* and *inter* heterogeneity.

## 6. AML Mouse Models Generated with Genome-Editing Techniques

The recent development of gene editing tools such as the clustered regularly interspaced short palindromic repeats (CRISPR) and the transcription activator-like effector nucleases (TALEN) offer novel tools to study the biology of AML by engineering disease-associated mutations in primary cells (Figure 1C,D). Pioneering work by the Ebert lab explored disease modelling by genome-editing of AML-associated mutations and inactivation of multiple tumour suppressor genes using a double lentiviral expression system [155]. Hereby, one vector delivered *Cas9* and a green fluorescent marker (eGFP), while the other carried the guide RNA (sgRNA) targeting the *Tet2*, *Runx1*, *Dnmt3a*, *Nf1*, *Ezh2* and *Smc3* genes in conjunction with another fluorescent marker (RFP-567). Viral transduction of lineage marker-depleted Sca1$^+$; cKit$^+$ (LSK) cells from C57B1/6 wild-type and *Flt3-ITD* knock-in mice with pooled sgRNA virus followed by transplantation into lethally irradiated recipients caused significant myeloid skewing of haematopoiesis and development of splenomegaly and leukaemia in some mice. Sequencing of genomic DNA from clonal leukemic cells revealed mutations in *Tet2*, *Dnmt3a*, *Runx1*, *Nf1* and *Ezh2* in single cells, thus indicating clonal outgrowth and transformation. This strategy was further refined to model mutations associated with CHIP [156]. Several CHIP and AML mutated genes (*Dnmt3a, Ezh2, Nf1, Runx1, Ascl1, Smc3* and/or Ep300) were edited simultaneously in murine HSPC, followed by transplantation into lethally irradiated recipients [156]. Genomic sequencing following long term observation and malignancy development showed single clonal expansion, especially from those harbouring *Dnmt3a* mutations. The mice showed a general increase in myeloid chimerism and clonal expansion reminiscence of CHIP. Some mice died of severe anaemia, while others progressively developed hematopoietic failure and AML. Genomic DNA sequencing detected deleterious mutations in all targeted genes except *Ep300*, leading to block in differentiation and activation of RAS-MAPK pathway [156]. A similar approach was used by another group to edit commonly mutated AML genes such as *TET2*, *ASXL1*, *DNMT3A*, *RUNX1*, *TP53*, *NF1*, *STAG2* and *SMC3* in human umbilical cord blood (UCB) and adult CD34$^+$ cells, by introducing a pool of 11-targeted sgRNAs [157]. Consistent with patients' data, in vitro generated colonies derived from single edited UBC carried bi-allelic loss-of-function (LOF) mutations in *TET2*, *DNMT3*, *EZH2*, *TP53* and *NF1* but only single allele mutations of *SMC3*, *ASXL1* and *RUNX1* were detected. To employ the multiplex genome editing in vivo, the researchers edited human adult CD34$^+$ cells with the same pool of sgRNAs together with *FLT3-ITD* and mutated *NPM*. Transplantation of edited cells into immune-compromised NSG-S mice resulted in the development of CHIP and MDS. Genomic DNA screening of in vivo expanded clones showed mutagenic pattern similar to the in vitro experiments, with overrepresentation of clones carrying LOF mutation (mostly deletions leading to frame shift) in *TET2*, *DMNT3A* and *ASXL1*. Despite these promising observations, none of the mice developed AML. The authors suggested that differences between the human and murine BM microenvironment might impair in vivo leukemogenesis [157].

Another study used CRISPR/Cas9-based genome editing to model the recurrent *7q* deletion associated with MDS and AML [158]. The commonly deleted region contains the *mixed lineage leukaemia 3* (*MLL3*) gene but mutations and deletions of *MLL3* have been only detected on one allele, suggesting that *MLL3* functions as a haplo-insufficient tumour suppressor. To prove this hypothesis, researchers transduced tumour-prone ($p53^{-/-}$ with reduced expression of the tumour suppressor *Nf1*) HSPC with sgRNA targeting *Mll3* followed by transplantation into sub-lethally irradiated *C57Bl/6* mice. Compared to controls, targeting *Mll3* significantly accelerated leukaemia development. Subsequent gDNA sequencing of individual *Mll3* edited AML clones revealed both wild type and mutant alleles in the majority of samples. This suggested that leukemogenesis selects for partial but not complete *Mll3* inactivation, providing compelling evidence that *Mll3* is a haplo-insufficient tumour suppressor gene in AML [158].

Genome editing using TALEN effector nucleases was used to specifically generate reciprocal chromosomal translocations of the *MLL* and *AF9* genes (*MLL-AF9* and *AF9-MLL*) in primary human CD34$^+$ UCB-derived HSC, to recapitulate MLL rearrangements in patients' cells [159,160]. In the first

study, edited HSPC showed heterogeneous response to the fusion whereby only some cells showed a clear proliferative advantage. The cells were not sufficiently transformed and could not be significantly expanded in culture and had a limited replating capacity in methylcellulose culture [159]. In contrast, in the second study, researchers were able to induce a leukemic phenotype by transplanting in vitro expanded monoclonal and immortalized cells into NSG mice. Notably, no secondary pathogenic mutations were found by targeted exome and RNA-sequencing, suggesting that this MLL fusion might be sufficient to initiate the disease [160]. Using engineered lentiviral vectors carrying Cas9 and two sgRNA sequences targeting the MLL and ENL locus researchers were able to generate the reciprocal t(11;19) translocation leading to expression of the MLL-ENL fusion in human $CD34^+$ UCB cell [161]. Unfortunately, similar to the first study done with MLL-AF9 fusion, the cells did not display enhanced self-renewal capacity in vitro when cultured in methylcellulose media. However, when injected in sub-lethally irradiated NSG-S mice, they were able to produce leukaemia with monocytic features. Future work will show whether Crispr/Cas9 genome editing will be suitable for the generation of animal models carrying multiple functionally cooperating genetic lesions ultimately progressing into clinical AML.

## 7. Conclusions

None of the currently used mouse AML models faithfully recapitulate the complex biology, cell to microenvironment interactions and dynamic progression of AML. Nevertheless, they have been instrumental in deciphering the underlying pathology of the disease and advancing AML research. Historically, the chemical, irradiation and viral models set the field of AML modelling in mice and were used to develop many AML drugs. Transgenic mouse lines harbouring AML associated mutations have enabled researchers to directly link genetic aberrations to AML initiation and progression. The creation of immunocompromised mouse strain has allowed for the expansion and study of human primary AML cells and the discovery of a hierarchy led by leukemic stem cell. In the future, advancement in genome editing technologies and collaboration between multidisciplinary fields would lead to the generation of more humanized mouse strains, which will ultimately help scientists to accurately model the complex biology of AML in mice.

**Author Contributions:** Conceptualization, Writing-Original Draft Preparation, Writing-Review & Editing: M.A. & J.S.

**Funding:** This research was funded by by the Swiss National Science Foundation (31003A-173224/1), the Swiss Cancer League (KFS-4258-08-2017), the San Salvatore Foundation (201525), the Wilhelm Sander Foundation (2017.035.1), the Novartis Foundation for Biomedical Research (14B058).

**Acknowledgments:** We would like to thank Lina Schukur, Jannik Vollmer and Aysha Rashid for their help and scientific input during the preparation of this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Papaemmanuil, E.; Gerstung, M.; Bullinger, L.; Gaidzik, V.I.; Paschka, P.; Roberts, N.D.; Potter, N.E.; Heuser, M.; Thol, F.; Bolli, N.; et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* **2016**, *374*, 2209–2221. [CrossRef]
2. Wang, M.L.; Bailey, N.G. Acute Myeloid Leukemia Genetics: Risk Stratification and Implications for Therapy. *Arch. Pathol. Lab. Med.* **2015**, *139*, 1215–1223. [CrossRef] [PubMed]
3. Mitelman, F.; Johansson, B.; Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **2007**, *7*, 233–245. [CrossRef] [PubMed]
4. Rowley, J.D. Chromosome translocations: Dangerous liaisons revisited. *Nat. Rev. Cancer* **2001**, *1*, 245–250. [CrossRef]
5. Welch, J.; Ley, T.; Link, D.; Miller, C. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **2012**, *150*, 264–278. [CrossRef] [PubMed]

6. Ley, T.; Miller, C.; Ding, L.; Raphael, B.J.; Mungall, A.J.; Robertson, A.; Hoadley, K.; Triche, T.J., Jr.; Laird, P.W.; Baty, J.D.; et al. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* **2013**, *368*, 2059–2074. [PubMed]
7. Meyer, S.C.; Levine, R.L. Translational implications of somatic genomics in acute myeloid leukaemia. *Lancet Oncol.* **2014**, *15*, e382–e394. [CrossRef]
8. Ito, S.; Barrett, A.J.; Dutra, A.; Pak, E.; Miner, S.; Keyvanfar, K.; Hensel, N.F.; Rezvani, K.; Muranski, P.; Liu, P.; et al. Long term maintenance of myeloid leukemic stem cells cultured with unrelated human mesenchymal stromal cells. *Stem Cell Res.* **2015**, *14*, 95–104. [CrossRef]
9. Alfred, L.J.; Wojdani, A.; Nieto, M.; Perez, R.; Yoshida, G. A chemical carcinogen, 3-methylcholanthrene, alters T-cell function and induces T-suppressor cells in a mouse model system. *Immunology* **1983**, 207–212.
10. Law, L.W.; Taormina, V.; Boyle, P.J. Response of Acute Lymphocytic Leukemias To the Purine Antagonist 6-Mercaptopurine. *Ann. N. Y. Acad. Sci.* **1954**, *60*, 244–250. [CrossRef]
11. Skipper, H.E.; Perry, S. Kinetics Of Normal And Leukemic Leukocyte Populations And Relevance To Chemotherapy. *Cancer Res.* **1970**, *30*, 1883–1897. [PubMed]
12. Chu, M.Y.; Fischer, G.A. A proposed mechanism of action of 1-β-d-arabinofuranosyl-cytosine as an inhibitor of the growth of leukemic cells. *Biochem. Pharmacol.* **1962**, *11*, 423–430. [CrossRef]
13. Kawasaki, Y.; Hirabayashi, Y.; Kaneko, T.; Kanno, J.; Kodama, Y.; Matsushima, Y.; Ogawa, Y.; Saitoh, M.; Sekita, K.; Uchida, O.; et al. Benzene-Induced Hematopoietic Neoplasms Including Myeloid Leukemia in Trp 53-Deficient C57BL/6 and C3H/He Mice. *Toxicol. Sci.* **2009**, *110*, 293–306. [CrossRef] [PubMed]
14. Robert, S. Leukemia and Benzene. *Int. J. Environ. Res. Public Health* **2012**, *9*, 2875–2893.
15. Khalade, A.; Jaakkola, M.S.; Pukkala, E.; Jaakkola, J.J.K. Exposure to benzene at work and the risk of leukemia: A systematic review and meta-analysis. *Environ. Heal. A Glob. Access Sci. Source* **2010**, *9*, 1–8. [CrossRef] [PubMed]
16. Mchale, C.M.; Zhang, L.; Smith, M.T. Current understanding of the mechanism of benzene-induced leukemia in humans: Implications for risk assessment. *Carcinogenesis* **2012**, *33*, 240–252. [CrossRef] [PubMed]
17. Christiansen, D.H.; Andersen, M.K.; Pedersen-Bjergaard, J. Mutations of AML1 are common in therapy-related myelodysplasia following therapy with alkylating agents and are significantly associated with deletion or loss of chromosome arm 7q and with subsequent leukemic transformation. *Blood* **2004**, *104*, 1474–1481. [CrossRef]
18. Thys, R.G.; Lehman, C.E.; Pierce, L.C.T.; Wang, Y.H. Environmental and chemotherapeutic agents induce breakage at genes involved in leukemia-causing gene rearrangements in human hematopoietic stem/progenitor cells. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **2015**, *779*, 86–95. [CrossRef]
19. Poynter, J.N.; Richardson, M.; Roesler, M.; Blair, C.K.; Hirsch, B.; Nguyen, P.; Cioc, A.; Cerhan, J.R.; Warlick, E. Chemical exposures and risk of acute myeloid leukemia and myelodysplastic syndromes in a population-based study. *Int. J. Cancer* **2017**, *140*. [CrossRef]
20. Board, R.S.; Studies, L. *Analysis of Cancer Risks in Populations Near Nuclear Facilities: Phase I*; National Academy of Sciences: Washington, DC, USA, 2012.
21. Finch, S.C. Radiation-induced leukemia: Lessons from history. *Best Pract. Res. Clin. Haematol.* **2007**, *20*, 109–118. [CrossRef]
22. Balonov, M. Third annual Warren K. Sinclair keynote address: Retrospective analysis of impacts of the chernobyl accident. *Health Phys.* **2007**, *93*, 383–409. [CrossRef] [PubMed]
23. Rivina, L.; Davoren, M.; Schiestl, R.H. Radiation-induced myeloid leukemia in murine models. *Hum. Genom.* **2014**, *8*, 13. [CrossRef] [PubMed]
24. Noshchenko, A.G.; Bondar, O.Y.; Drozdova, V.D. Radiation-induced leukemia among children aged 0–5 years at the time of the Chernobyl accident. *Int. J. Cancer* **2010**, *127*, 412–426. [CrossRef] [PubMed]
25. Dekkers, F.; Bijwaard, H.; Bouffler, S.; Ellender, M.; Huiskamp, R.; Kowalczuk, C.; Meijne, E.; Sutmuller, M. A two-mutation model of radiation-induced acute myeloid leukemia using historical mouse data. *Radiat. Environ. Biophys.* **2011**, *50*, 37–45. [CrossRef] [PubMed]
26. Cadman, E.C.; Capizzi, R.L.; Bertino, J.R. Acute nonlymphocytic leukemia. A delayed complication of Hodgkin's disease therapy: Analysis of 109 cases. *Cancer* **1977**, *40*, 1280–1296. [CrossRef]
27. Janowska-Wieczorek, A.; Belch, A.R.; Jacobs, A.; Bowen, D.; Padua, R.A.; Paietta, E.; Stanley, E.R. Increased circulating colony-stimulating factor-1 in patients with preleukemia, leukemia, and lymphoid malignancies. *Blood* **1991**, *77*, 1796–1803. [PubMed]

28. Haran-Ghera, N.; Krautghamer, R.; Lapidot, T.; Peled, A.; Dominguez, M.G.; Stanley, E.R. Increased circulating colony-stimulating factor-1 (CSF-1) in SJL/J mice with radiation-induced acute myeloid leukemia (AML) is associated with autocrine regulation of AML cells by CSF-1. *Blood* **1997**, *89*, 2537–2545.
29. Peng, Y.; Borak, T.B.; Bouffler, S.D.; Ullrich, R.L.; Weil, M.M.; Bedford, J.S. Radiation Leukemogenesis in Mice: Loss of PU.1 on Chromosome 2 in CBA and C57BL/6 Mice after Irradiation with 1 GeV/nucleon 56 Fe Ions, X Rays or γ Rays. Part I. Experimental Observation. *Radiat. Res.* **2009**, *171*, 484–493. [CrossRef]
30. Olme, C.H.; Finnon, R.; Brown, N.; Kabacik, S.; Bouffler, S.D.; Badie, C. Live cell detection of chromosome 2 deletion and Sfpi1/PU1 loss in radiation-induced mouse acute myeloid leukaemia. *Leuk. Res.* **2013**, *37*, 1374–1382. [CrossRef]
31. Verbiest, T.; Finnon, R.; Brown, N.; Cruz-Garcia, L.; Finnon, P.; O'Brien, G.; Ross, E.; Bouffler, S.; Scudamore, C.L.; Badie, C. Tracking preleukemic cells in vivo to reveal the sequence of molecular events in radiation leukemogenesis. *Leukemia* **2018**, *32*, 1435–1444. [CrossRef]
32. Klymenko, S.V.; Smida, J.; Atkinson, M.J.; Bebeshko, V.G.; Nathrath, M.; Rosemann, M. Allelic imbalances in radiation-associated acute myeloid leukemia. *Genes* **2011**, *2*, 384–393. [CrossRef] [PubMed]
33. Li, M.; Jones, L.; Gaillard, C.; Binnewies, M.; Ochoa, R.; Garcia, E.; Lam, V.; Wei, G.; Yang, W.; Lobe, C.; et al. Initially disadvantaged, TEL-AML1 cells expand and initiate leukemia in response to irradiation and cooperating mutations. *Leukemia* **2013**, *27*, 1570–1573. [CrossRef] [PubMed]
34. Rein, A. Murine leukemia viruses: Objects and organisms. *Adv. Virol.* **2011**, *2011*. [CrossRef]
35. Friend, C. Cell-free transmission in adult swiss mice of a disease having the character of leukemia. *J. Exp. Med.* **1957**, *105*, 307–318. [CrossRef] [PubMed]
36. Largaespada, D.A. Genetic heterogeneity in acute myeloid leukemia: Maximizing information flow from MuLV mutagenesis studies. *Leukemia* **2000**, *14*, 1174–1184. [CrossRef] [PubMed]
37. Cmarik, J.; Ruscetti, S. Friend spleen focus-forming virus activates the tyrosine kinase sf-Stk and the transcription factor PU.1 to cause a multi-stage erythroleukemia in mice. *Viruses* **2010**, *2*, 2235–2257. [CrossRef]
38. Moreau-Gachelin, F. Multi-stage Friend murine erythroleukemia: Molecular insights into oncogenic cooperation. *Retrovirology* **2008**, *5*, 1–11. [CrossRef]
39. Singer, D.; Cooper, M.; Maniatis, G.M.; Marks, P.A.; Rifkind, R.A. Erythropoietic differentiation in colonies of cells transformed by Friend virus. *Proc. Natl. Acad. Sci. USA.* **1974**, *71*, 2668–2670. [CrossRef]
40. Rao, G.; Rekhtman, N.; Cheng, G.; Krasikov, T.; Skoultchi, A.I. Deregulated expression of the PU.1 transcription factor blocks murine erythroleukemia cell terminal differentiation. *Oncogene* **1997**, *14*, 123–131. [CrossRef]
41. Yamada, T.; Kondoh, N.; Matsumoto, M.; Yoshida, M.; Maekawa, A.; Oikawa, T. Overexpression of PU.1 induces growth and differentiation inhibition and apoptotic cell death in murine erythroleukemia cells. *Blood* **1997**, *89*, 1383–1393.
42. Mucenski, M.L.; Taylor, B.A.; Ihle, J.N.; Hartley, J.W.; Morse, H.C.; Jenkins, N.A.; Copeland, N.G. Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors. *Mol. Cell. Biol.* **1988**, *8*, 301–308. [CrossRef] [PubMed]
43. Wolff, L.; Koller, R.; Hu, X.; Anver, M.R. A Moloney murine leukemia virus-based retrovirus with 4070A long terminal repeat sequences induces a high incidence of myeloid as well as lymphoid neoplasms. *J. Virol.* **2003**, *77*, 4965–4971. [CrossRef] [PubMed]
44. Kool, J.; Berns, A. High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. *Nat. Rev. Cancer* **2009**, *9*, 389–399. [CrossRef] [PubMed]
45. Ranzani, M.; Annunziato, S.; Adams, D.J.; Montini, E. Cancer Gene Discovery: Exploiting Insertional Mutagenesis. *Mol. Cancer Res.* **2013**, *11*, 1141–1158. [CrossRef] [PubMed]
46. Vassiliou, G.S.; Cooper, J.L.; Rad, R.; Li, J.; Rice, S.; Uren, A.; Rad, L.; Ellis, P.; Andrews, R.; Banerjee, R.; et al. Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. *Nat. Genet.* **2011**, *43*, 470–476. [CrossRef] [PubMed]
47. Lauchle, J.O.; Kim, D.; Le, D.T.; Akagi, K.; Crone, M.; Krisman, K.; Warner, K.; Bonifas, J.M.; Li, Q.; Coakley, K.M.; et al. Response and resistance to MEK inhibition in leukaemias initiated by hyperactive Ras. *Nature* **2009**, *461*, 411–414. [CrossRef] [PubMed]

48. Adams, J.; Harris, W.; Pinkert, C.; Corcoran, L.; Alexander, W.; Cory, S.; Palmiter, R.; Brinster, R. The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* **1985**, *318*, 533–538. [CrossRef] [PubMed]
49. Schmidt, E.V.; Pattengale, P.K.; Weir, L.; Leder, P. Transgenic mice bearing the human c-myc gene activated by an immunoglobulin enhancer: A pre-B-cell lymphoma model. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6047–6051. [CrossRef] [PubMed]
50. Würtele, H.; Little, K.C.E.; Chartrand, P. Illegitimate DNA integration in mammalian cells. *Gene Ther.* **2003**, *10*, 1791–1799. [CrossRef]
51. Grisolano, J.L.; Wesselschmidt, R.L.; Pelicci, P.G.; Ley, T.J. Altered Myeloid Development and Acute Leukemia in Transgenic Mice Expressing PML-RARα Under Control of Cathepsin G Regulatory Sequences. *Blood* **1997**, *89*, 376–387.
52. Early, E.; Moore, M.A.; Kakizuka, A.; Nason-Burchenal, K.; Martin, P.; Evans, R.M.; Dmitrovsky, E. Transgenic expression of PML/RARA impairs myelopoiesis. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 7900–7904. [CrossRef] [PubMed]
53. Brown, D.; Kogan, S.; Lagasse, E.; Weissman, I.; Alcalay, M.; Pelicci, P.G.; Atwater, S.; Bishop, J.M. A PMLRAR transgene initiates murine acute promyelocytic leukemia. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2551–2556. [CrossRef] [PubMed]
54. De Thé, H.; Chen, Z. Acute promyelocytic leukaemia: Novel insights into the mechanisms of cure. *Nat. Rev. Cancer* **2010**, *10*, 775–783. [CrossRef] [PubMed]
55. Cheng, G.X.; Zhu, X.H.; Men, X.Q.; Wang, L.; Huang, Q.H.; Jin, X.L.; Xiong, S.M.; Zhu, J.; Guo, W.M.; Chen, J.Q.; et al. Distinct leukemia phenotypes in transgenic mice and different corepressor interactions generated by promyelocytic leukemia variant fusion genes PLZF-RARA and NPM-RARA. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6318–6323. [CrossRef] [PubMed]
56. Ablain, J.; Nasr, R.; Zhu, J.; Bazarbachi, A.; Lallemand-Breittenbach, V.; de Thé, H. How animal models of leukaemias have already benefited patients. *Mol. Oncol.* **2013**, *7*, 224–231. [CrossRef] [PubMed]
57. Bernardi, R.; Grisendi, S.; Pandolfi, P.P. Modelling haematopoietic malignancies in the mouse and therapeutical implications. *Oncogene* **2002**, *21*, 3445–3458. [CrossRef] [PubMed]
58. McCormack, E.; Bruserud, O.; Gjertsen, B.T. Review: Genetic models of acute myeloid leukaemia. *Oncogene* **2008**, *27*, 3765–3779. [CrossRef]
59. Tomasson, M.H.; Williams, I.R.; Hasserjian, R.; Udomsakdi, C.; McGrath, S.M.; Schwaller, J.; Druker, B.; Gilliland, D.G. TEL/PDGFbetaR induces hematologic malignancies in mice that respond to a specific tyrosine kinase inhibitor. *Blood* **1999**, *93*, 1707–1714.
60. Rhoades, K.L.; Hetherington, C.J.; Harakawa, N.; Yergeau, D.A.; Zhou, L.; Liu, L.Q.; Little, M.T.; Tenen, D.G.; Zhang, D.E. Analysis of the role of AML1-ETO in leukemogenesis, using an inducible transgenic mouse model. *Blood* **2000**, *96*, 2108–2115.
61. Yuan, Y.; Zhou, L.; Miyamoto, T.; Iwasaki, H.; Harakawa, N.; Hetherington, C.J.; Burel, S.A.; Lagasse, E.; Weissman, I.L.; Akashi, K.; et al. AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10398–10403. [CrossRef]
62. Kuo, Y.H.; Landrette, S.F.; Heilman, S.A.; Perrat, P.N.; Garrett, L.; Liu, P.P.; Le Beau, M.M.; Kogan, S.C.; Castilla, L.H. Cbfβ-SMMHC induces distinct abnormal myeloid progenitors able to develop acute myeloid leukemia. *Cancer Cell* **2006**, *9*, 57–68. [CrossRef] [PubMed]
63. Chen, W.; Kumar, A.R.; Hudson, W.A.; Li, Q.; Wu, B.; Staggs, R.A.; Lund, E.A.; Sam, T.N.; Kersey, J.H. Malignant Transformation Initiated by Mll-AF9: Gene Dosage and Critical Target Cells. *Cancer Cell* **2008**, *13*, 432–440. [CrossRef] [PubMed]
64. Ugale, A.; Norddahl, G.L.; Wahlestedt, M.; Säwén, P.; Jaako, P.; Pronk, C.J.; Soneji, S.; Cammenga, J.; Bryder, D. Hematopoietic Stem Cells Are Intrinsically Protected against MLL-ENL-Mediated Transformation. *Cell Rep.* **2014**, *9*, 1246–1255. [CrossRef] [PubMed]
65. Stavropoulou, V.; Kaspar, S.; Brault, L.; Sanders, M.A.; Juge, S.; Morettini, S.; Tzankov, A.; Iacovino, M.; Lau, I.J.; Milne, T.A.; et al. MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome. *Cancer Cell* **2016**, *30*, 43–58. [CrossRef] [PubMed]

66. Stavropoulou, V.; Almosailleakh, M.; Royo, H.; Spetz, J.; Juge, S.; Brault, L.; Kopp, P.; Iacovino, M.; Kyba, M.; Tzankov, A.; et al. A Novel Inducible Mouse Model of MLL-ENL-driven Mixed-lineage Acute Leukemia. *HemaSphere* **2018**, *4*, 1–11. [CrossRef]
67. Thomas, K.R.; Capecchi, M.R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **1987**, *51*, 503–512. [CrossRef]
68. Corral, J.; Lavenir, I.; Impey, H.; Warren, A.J.; Forster, A.; Larson, T.A.; Bell, S.; McKenzie, A.N.J.; King, G.; Rabbitts, T.H. An MII-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: A method to create fusion oncogenes. *Cell* **1996**, *85*, 853–861. [CrossRef]
69. Meyer, C.; Burmeister, T.; Gröger, D.; Tsaur, G.; Fechina, L.; Renneville, A.; Sutton, R.; Venn, N.C.; Emerenciano, M.; Pombo-de-oliveira, M.S. The MLL recombinome of acute leukemias in 2017. *Leukemia* **2018**, 273–284. [CrossRef]
70. Collins, E.C.; Pannell, R.; Simpson, E.M.; Forster, A.; Rabbitts, T.H. Inter-chromosomal recombination of Mll and Af9 genes mediated by cre-loxP in mouse development. *Sci. Rep.* **2000**, *1*, 127–132.
71. Dobson, C.L.; Warren, A.J.; Pannell, R.; Forster, A.; Rabbitts, T.H. Tumorigenesis in mice with a fusion of the leukaemia oncogene MII and the bacterial lacZ gene. *EMBO J.* **2000**, *19*, 843–851. [CrossRef]
72. Johnson, J.J.; Chen, W.; Hudson, W.; Yao, Q.; Taylor, M.; Rabbitts, T.H.; Kersey, J.H. Prenatal and postnatal myeloid cells demonstrate stepwise progression in the pathogenesis of MLL fusion gene leukemia. *Blood* **2003**, *101*, 3229–3235. [CrossRef] [PubMed]
73. Ayton, P.M.; Cleary, M.L. Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. *Genes Dev.* **2003**, 2298–2307. [CrossRef] [PubMed]
74. Sinha, C.; Cunningham, L.C.; Liu, P.P. Core Binding Factor Acute Myeloid Leukemia: New Prognostic Categories and Therapeutic Opportunities. *Semin. Hematol.* **2015**, *52*, 215–222. [CrossRef] [PubMed]
75. Okuda, T.; Cai, Z.; Yang, S.; Lenny, N.; Lyu, C.J.; van Deursen, J.M.; Harada, H.; Downing, J.R. Expression of a knocked-in AML1-ETO leukemia gene inhibits the establishment of normal definitive hematopoiesis and directly generates dysplastic hematopoietic progenitors. *Blood* **1998**, *91*, 3134–3143. [PubMed]
76. Okuda, T.; van Deursen, J.; Hiebert, S.W.; Grosveld, G.; Downing, J.R.; Bae, S.; Yamaguchi-Iwai, Y.; Ogawa, E.; Maruyama, M.; Inuzuka, M.; et al. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* **1996**, *84*, 321–330. [CrossRef]
77. Speck, N.A.; Iruela-Arispe, M.L. Conditional Cre/LoxP strategies for the study of hematopoietic stem cell formation. *Blood Cells Mol. Dis.* **2009**, *43*, 6–11. [CrossRef] [PubMed]
78. Furth, P.A.; Onget, L.S.; B6gert, H.; Grusst, P.; Gossen, M.; Kistner, A.; Bujard, H.; Hennighausentl, L. Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. *Biochemistry* **1994**, *91*, 9302–9306. [CrossRef]
79. Gossen, M.; Freundlieb, S.; Bender, G.; Muller, G.; Hillen, W.; Bujardt, H. Transcriptional Activation by Tetracyclines in Mammalian Cells. *Science* **1995**, *268*, 1766–1770. [CrossRef]
80. Higuchi, M.; O'Brien, D.; Kumaravelu, P.; Lenny, N.; Yeoh, E.-J.; Downing, J.R. 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* **2002**, *1*, 63–74. [CrossRef]
81. Claij, N.; Wal, A. Van Der; Dekker, M. DNA Mismatch Repair Deficiency Stimulates N-Ethyl-N-nitrosourea-induced Mutagenesis and Lymphomagenesis. *Cancer Res.* **2003**, *63*, 2062–2066.
82. Cabezas-Wallscheid, N.; Eichwald, V.; de Graaf, J.; Löwer, M.; Lehr, H.A.; Kreft, A.; Eshkind, L.; Hildebrandt, A.; Abassi, Y.; Heck, R.; et al. Instruction of haematopoietic lineage choices, evolution of transcriptional landscapes and cancer stem cell hierarchies derived from an AML1-ETO mouse model. *EMBO Mol. Med.* **2013**, *5*, 1804–1820. [CrossRef] [PubMed]
83. Heidenreich, O.; Riehle, H.; Hadwiger, P.; John, M.; Heil, G.; Vornlocher, H.; Nordheim, A. AML1/MTG8 oncogene suppression by small interfering RNAs supports myeloid differentiation of t (8; 21)-positive leukemic cells. *Blood* **2003**, *101*, 3157–3163. [CrossRef] [PubMed]
84. Nick, H.J.; Kim, H.; Chang, C.; Harris, K.W.; Reddy, V.; Klug, A.; Dc, W.; Klug, C.A. Distinct classes of c-Kit–activating mutations differ in their ability to promote RUNX1-ETO–associated acute myeloid leukemia. *Blood* **2012**, *119*, 1522–1531. [CrossRef] [PubMed]
85. Schessl, C.; Rawat, V.P.S.; Cusan, M.; Deshpande, A.; Kohl, T.M.; Rosten, P.M.; Spiekermann, K.; Humphries, R.K.; Schnittger, S.; Kern, W.; et al. The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. *J. Clin. Investig.* **2005**, *115*, 2159–2168. [CrossRef] [PubMed]

86. Sportoletti, P.; Varasano, E.; Rossi, R.; Mupo, A.; Tiacci, E.; Vassiliou, G.; Martelli, M.P.; Falini, B. Mouse models of NPM1-mutated acute myeloid leukemia: Biological and clinical implications. *Leukemia* **2015**, *29*, 269–278. [CrossRef]
87. Cheng, K.; Sportoletti, P.; Ito, K.; Clohessy, J.G.; Teruya-feldstein, J.; Kutok, J.L.; Pandolfi, P.P. Brief report The cytoplasmic NPM mutant induces myeloproliferation in a transgenic mouse model. *Blood* **2018**, *115*, 3341–3346. [CrossRef]
88. Chou, S.H.; Ko, B.S.; Chiou, J.S.; Hsu, Y.C.; Tsai, M.H.; Chiu, Y.C.; Yu, I.S.; Lin, S.W.; Hou, H.A.; Kuo, Y.Y.; et al. A Knock-In Npm1 Mutation in Mice Results in Myeloproliferation and Implies a Perturbation in Hematopoietic Microenvironment. *PLoS ONE* **2012**, *7*, 1–10. [CrossRef]
89. Mallardo, M.; Caronno, A.; Pruneri, G.; Raviele, P.R.; Viale, A.; Pelicci, P.G.; Colombo, E. NPMc+ and FLT3_ITD mutations cooperate in inducing acute leukaemia in a novel mouse model. *Leukemia* **2013**, *27*, 2248–2251. [CrossRef]
90. Garg, M.; Nagata, Y.; Kanojia, D.; Mayakonda, A.; Yoshida, K.; Keloth, S.H.; Zang, Z.J.; Okuno, Y.; Shiraishi, Y.; Chiba, K.; et al. Profiling of somatic mutations in acute myeloid leukemia with FLT3-ITD at diagnosis and relapse. *Blood* **2015**, *126*, 2491–2502. [CrossRef]
91. Lee, B.H.; Tothova, Z.; Levine, R.L.; Anderson, K.; Buza-Vidas, N.; Cullen, D.E.E.; McDowell, E.P.; Adelsperger, J.; Fröhling, S.; Huntly, B.J.P.; et al. FLT3 Mutations Confer Enhanced Proliferation and Survival Properties to Multipotent Progenitors in a Murine Model of Chronic Myelomonocytic Leukemia. *Cancer Cell* **2007**, *12*, 367–380. [CrossRef]
92. Lee, B.H.; Williams, I.R.; Anastasiadou, E.; Boulton, C.L.; Joseph, S.W.; Amaral, S.M.; Curley, D.P.; Duclos, N.; Huntly, B.J.P.; Fabbro, D.; et al. FLT3 internal tandem duplication mutations induce myeloproliferative or lymphoid disease in a transgenic mouse model. *Oncogene* **2005**, *24*, 7882–7892. [CrossRef] [PubMed]
93. Kharazi, S.; Mead, A.J.; Mansour, A.; Hultquist, A.; Böiers, C.; Luc, S.; Buza-Vidas, N.; Ma, Z.; Ferry, H.; Atkinson, D.; et al. Impact of gene dosage, loss of wild-type allele, and FLT3 ligand on Flt3-ITD-induced myeloproliferation. *Blood* **2011**, *118*, 3613–3621. [CrossRef] [PubMed]
94. Li, L.; Bailey, E.; Greenblatt, S.; Huso, D.; Small, D. Loss of the wild-type allele contributes to myeloid expansion and disease aggressiveness in FLT3/ITD knockin mice. *Blood* **2011**, *118*, 4935–4945. [CrossRef] [PubMed]
95. Dovey, O.M.; Cooper, J.L.; Mupo, A.; Grove, C.S.; Lynn, C.; Conte, N.; Andrews, R.M.; Pacharne, S.; Tzelepis, K.; Vijayabaskar, M.S.; et al. Molecular synergy underlies the co-occurrence patterns and phenotype of NPM1-mutant acute myeloid leukemia. *Blood* **2017**, *130*, 1911–1922. [CrossRef]
96. Mupo, A.; Celani, L.; Dovey, O.; Cooper, J.L.; Grove, C.; Rad, R.; Sportoletti, P.; Falini, B.; Bradley, A.; Vassiliou, G.S. A powerful molecular synergy between mutant Nucleophosmin and Flt3-ITD drives acute myeloid leukemia in mice. *Leukemia* **2013**, *27*, 1917–1920. [CrossRef] [PubMed]
97. Zorko, N.A.; Bernot, K.M.; Whitman, S.P.; Siebenaler, R.F.; Ahmed, E.H.; Marcucci, G.G.G.G.; Yanes, D.A.; McConnell, K.K.; Mao, C.; Kalu, C.; et al. Mll partial tandem duplication and Flt3 internal tandem duplication in a double knock-in mouse recapitulates features of counterpart human acute myeloid leukemias. *Blood* **2012**, *120*, 1130–1136. [CrossRef] [PubMed]
98. Greenblatt, S.; Li, L.; Slape, C.; Nguyen, B.; Novak, R.; Duffield, A.; Huso, D.; Desiderio, S.; Borowitz, M.J.; Aplan, P.; et al. Knock-in of a FLT3/ITD mutation cooperates with a NUP98-HOXD13 fusion to generate acute myeloid leukemia in a mouse model. *Blood* **2012**, *119*, 2883–2894. [CrossRef] [PubMed]
99. Annesley, C.E.; Rabik, C.; Duffield, A.S.; Rau, R.E.; Li, L.; Huff, V.; Small, D.; Loeb, D.M.; Brown, P. Knock-in of the Wt1 R394W mutation causes MDS and cooperates with Flt3/ITD to drive aggressive myeloid neoplasms in mice. *Oncotarget* **2018**, *9*, 35313–35326. [CrossRef]
100. Meyer, S.E.; Qin, T.; Muench, D.E.; Masuda, K.; Venkatasubramanian, M.; Orr, E.; Suarez, L.; Gore, S.D.; Delwel, R.; Paietta, E.; et al. Dnmt3a haploinsufficiency transforms Flt3-ITD myeloproliferative disease into a rapid, spontaneous, and fully-penetrant acute myeloid leukemia. *Cancer Discov.* **2016**, *6*, 501–515. [CrossRef]
101. Shi, J.; Wang, E.; Zuber, J.; Rappaport, A.; Taylor, M.; Johns, C.; Lowe, S.W.; Vakoc, C.R. The Polycomb complex PRC2 supports aberrant self-renewal in a mouse model of MLL-AF9;Nras-G12D acute myeloid leukemia J. *Oncogene* **2013**, *32*, 930–938. [CrossRef]
102. Omidvar, N.; Kogan, S.; Beurlet, S.; Le Pogam, C.; Janin, A.; West, R.; Noguera, M.E.; Reboul, M.; Soulie, A.; Leboeuf, C.; et al. BCL-2 and mutant NRAS interact physically and functionally in a mouse model of progressive myelodysplasia. *Cancer Res.* **2007**, *67*, 11657–11667. [CrossRef] [PubMed]

103. Xue, L.; Pulikkan, J.A.; Valk, P.J.M.; Castilla, L.H. Nras G12D oncoprotein inhibits apoptosis of preleukemic cells expressing Cbfb -SMMHC via activation of MEK/ERK axis. *Blood* **2014**, *124*, 426–437. [CrossRef] [PubMed]
104. Chan, I.T.; Kutok, J.L.; Williams, I.R.; Cohen, S.; Moore, S.; Shigematsu, H.; Ley, T.J.; Akashi, K.; Beau, M.M. Le; Gilliland, D.G.; et al. Oncogenic K-ras cooperates with PML-RAR to induce an acute promyelocytic leukemia–like disease. *Blood* **2012**, *108*, 1708–1715. [CrossRef] [PubMed]
105. Hinai, A.A.; Valk, P.J.M. Review: Aberrant EVI1 expression in acute myeloid leukaemia. *Br. J. Haematol.* **2016**, *172*, 870–878. [CrossRef] [PubMed]
106. Yamazaki, H.; Suzuki, M.; Otsuki, A.; Shimizu, R.; Bresnick, E.H.; Engel, J.D.; Yamamoto, M. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. *Cancer Cell* **2014**, *25*, 415–427. [CrossRef] [PubMed]
107. Gröschel, S.; Sanders, M.A.; Hoogenboezem, R.; De Wit, E.; Bouwman, B.A.M.; Erpelinck, C.; Van Der Velden, V.H.J.; Havermans, M.; Avellino, R.; Van Lom, K.; et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in Leukemia. *Cell* **2014**, *157*, 369–381. [CrossRef] [PubMed]
108. Ayoub, E.; Wilson, M.P.; McGrath, K.E.; Li, A.J.; Frisch, B.J.; Palis, J.; Calvi, L.M.; Zhang, Y.; Perkins, A.S. EVI1 overexpression reprograms hematopoiesis via upregulation of Spi1 transcription. *Nat. Commun.* **2018**, *9*, 1–12. [CrossRef]
109. Whitelaw, C.B.A.; Archibald, A.L.; Harris, S.; McClenaghan, M.; Simons, J.P.; Clark, A.J. Targeting expression to the mammary gland: Intronic sequences can enhance the efficiency of gene expression in transgenic mice. *Transgenic Res.* **1991**, *1*, 3–13. [CrossRef]
110. Takai, J.; Moriguchi, T.; Suzuki, M.; Yu, L.; Ohneda, K.; Yamamoto, M. The Gata1 5′ region harbors distinct cis-regulatory modules that direct gene activation in erythroid cells and gene inactivation in HSCs. *Blood* **2013**, *112*, 3450–3460. [CrossRef]
111. Daley, G.Q.; Van Etten, R.A.; Baltimore, D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Am. Assoc. Adv. Sci.* **1990**, *247*, 824–830. [CrossRef]
112. Mc Cormack, E.; Bruserud, Ø.; Gjertsen, B. Animal models of acute myelogenous leukaemia—Development, application and future perspectives. *Leukemia* **2005**, *19*, 687–706. [CrossRef] [PubMed]
113. Grove, C.S.; Vassiliou, G.S. Acute myeloid leukaemia: a paradigm for the clonal evolution of cancer? *Dis. Model. Mech.* **2014**, *7*, 941–951. [CrossRef]
114. Huntly, B.J.P.; Shigematsu, H.; Deguchi, K.; Lee, B.H.; Mizuno, S.; Duclos, N.; Rowan, R.; Amaral, S.; Curley, D.; Williams, I.R.; et al. MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell* **2004**, *6*, 587–596. [CrossRef]
115. Krivtsov, A.V.; Twomey, D.; Feng, Z.; Stubbs, M.C.; Wang, Y.; Faber, J.; Levine, J.E.; Wang, J.; Hahn, W.C.; Gilliland, D.G.; et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **2006**, *442*, 818–822. [CrossRef]
116. Bindels, E.M.J.; Havermans, M.; Lugthart, S.; Erpelinck, C.; Wocjtowicz, E.; Krivtsov, A.V.; Rombouts, E.; Armstrong, S.A.; Taskesen, E.; Haanstra, J.R.; et al. EVI1 is critical for the pathogenesis of a subset of MLL-AF9-rearranged AMLs. *Blood* **2012**, *119*, 5838–5849. [CrossRef] [PubMed]
117. Lavau, C.; Szilvassy, S.J.; Slany, R.; Cleary, M.L. Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL. *EMBO J.* **1997**, *16*, 4226–4237. [CrossRef] [PubMed]
118. Cozzio, A.; Passegué, E.; Ayton, P.M.; Karsunky, H.; Cleary, M.L.; Weissman, I.L. Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes Dev.* **2003**, *17*, 3029–3035. [CrossRef]
119. de Guzman, C.G.; Warren, A.J.; Zhang, Z.; Gartland, L.; Erickson, P.; Drabkin, H.; Hiebert, S.W.; Klug, C.A. Hematopoietic Stem Cell Expansion and Distinct Myeloid Developmental Abnormalities in a Murine Model of the AML1-ETO Translocation. *Mol. Cell. Biol.* **2002**, *22*, 5506–5517. [CrossRef]
120. So, C.W.; Karsunky, H.; Passegué, E.; Cozzio, A.; Weissman, I.L.; Cleary, M.L. MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. *Cancer Cell* **2003**, *3*, 161–171. [CrossRef]
121. Heuser, M.; Yun, H.; Berg, T.; Yung, E.; Argiropoulos, B.; Kuchenbauer, F.; Park, G.; Hamwi, I.; Palmqvist, L.; Lai, C.K.; et al. Cell of Origin in AML: Susceptibility to MN1-Induced Transformation Is Regulated by the MEIS1/AbdB-like HOX Protein Complex. *Cancer Cell* **2011**, *20*, 39–52. [CrossRef]

122. Chang, A.H.; Sadelain, M. The genetic engineering of hematopoietic stem cells: The rise of lentiviral vectors, the conundrum of the LTR, and the promise of lineage-restricted vectors. *Mol. Ther.* **2007**, *15*, 445–456. [CrossRef] [PubMed]
123. Shultz, L.D.; Brehm, M.A.; Victor Garcia-Martinez, J.; Greiner, D.L. Humanized mice for immune system investigation: Progress, promise and challenges. *Nat. Rev. Immunol.* **2012**, *12*, 786–798. [CrossRef] [PubMed]
124. Theocharides, A.P.A.; Rongvaux, A.; Fritsch, K.; Flavell, R.A.; Manz, M.G. Humanized hemato-lymphoid system mice. *Haematologica* **2016**, *101*, 5–19. [CrossRef] [PubMed]
125. Nara, N.; Miyamoto, T. Direct and serial transplantation of human acute myeloid leukaemia into nude mice. *Br. J. Cancer* **1982**, *45*, 778–782. [CrossRef] [PubMed]
126. Gelebart, P.; Popa, M.; McCormack, E. Xenograft models of primary acute myeloid leukemia for the development of imaging strategies and evaluation of novel targeted therapies. *Curr. Pharm. Biotechnol.* **2016**, *17*, 42–51. [CrossRef] [PubMed]
127. Sawyers, C.L.; Gishizky, M.L.; Quan, S.; Golde, D.W.; Witte, O.N. Propagation of human blastic myeloid leukemias in the SCID mouse. *Blood* **1992**, *79*, 2089–2098. [PubMed]
128. Cao, X.; Shores, E.W.; Hu-Li, J.; Anver, M.R.; Kelsail, B.L.; Russell, S.M.; Drago, J.; Noguchi, M.; Grinberg, A.; Bloom, E.T.; et al. Defective lymphoid development in mice lacking expression of the common cytokine receptor γ chain. *Immunity* **1995**, *2*, 223–238. [CrossRef]
129. Bonnet, D.; Dick, J.E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **1997**, *3*, 730–737. [CrossRef]
130. Lapidot, T.; Sirard, C.; Vormoor, J.; Murdoch, B.; Hoang, T.; Caceres-Cortes, J.; Minden, M.; Paterson, B.; Caligiuri, M.A.; Dick, J.E. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **1994**, *367*, 645–648. [CrossRef]
131. Ailles, L.E.; Gerhard, B.; Kawagoe, H.; Hogge, D.E. Growth characteristics of acute myelogenous leukemia progenitors that initiate malignant hematopoiesis in nonobese diabetic/severe combined immunodeficient mice. *Blood* **1999**, *94*, 1761–1772.
132. Ito, M.; Hiramatsu, H.; Kobayashi, K.; Suzue, K.; Kawahata, M.; Hioki, K.; Ueyama, Y.; Koyanagi, Y.; Sugamura, K.; Tsuji, K.; et al. NOD/SCID/ gamma c null mouse: An excellent recipient mouse model for engraftment of human cells. *Bone* **2002**, *100*, 3175–3182.
133. Shultz, L.D.; Lyons, B.L.; Burzenski, L.M.; Gott, B.; Chen, X.; Chaleff, S.; Kotb, M.; Gillies, S.D.; King, M.; Mangada, J.; et al. Human Lymphoid and Myeloid Cell Development in NOD/LtSz-scid IL2R null Mice Engrafted with Mobilized Human Hemopoietic Stem Cells. *J. Immunol.* **2005**, *174*, 6477–6489. [CrossRef] [PubMed]
134. Wunderlich, M.; Chou, F.S.; Link, K.A.; Mizukawa, B.; Perry, R.L.; Carroll, M.; Mulloy, J.C. AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3. *Leukemia* **2010**, *24*, 1785–1788. [CrossRef] [PubMed]
135. Billerbeck, E.; Barry, W.T.; Mu, K.; Dorner, M.; Rice, C.M.; Ploss, A. Development of human CD4+FoxP3+ regulatory T cells in human stem cell factor-, granulocyte-macrophage colony-stimulating factor-, and interleukin-3-expressing NOD-SCID IL2Rγ(null) humanized mice. *Blood* **2011**, *117*, 3076–3086. [CrossRef]
136. Feuring-Buske, M.; Gerhard, B.; Cashman, J.; Humphries, R.K.; Eaves, C.J.; Hogge, D.E. Improved engraftment of human acute myeloid leukemia progenitor cells in beta 2-microglobulin-deficient NOD/SCID mice and in NOD/SCID mice transgenic for human growth factors. *Leukemia* **2003**, *17*, 760–763. [CrossRef] [PubMed]
137. King, M.A.; Covassin, L.; Brehm, M.A.; Racki, W.; Pearson, T.; Leif, J.; Laning, J.; Fodor, W.; Foreman, O.; Burzenski, L.; et al. Human peripheral blood leucocyte non-obese diabetic-severe combined immunodeficiency interleukin-2 receptor gamma chain gene mouse model of xenogeneic graft-versus-host-like disease and the role of host major histocompatibility complex. *Clin. Exp. Immunol.* **2009**, *157*, 104–118. [CrossRef]

138. Covassin, L.; Laning, J.; Abdi, R.; Langevin, D.L.; Phillips, N.E.; Shultz, L.D.; Brehm, M.A. Human peripheral blood CD4 T cell-engrafted non-obese diabetic-scid IL2rgnull H2-Ab1tm1Gru Tg (human leucocyte antigen D-related 4) mice: A mouse model of human allogeneic graft-versus-host disease. *Clin. Exp. Immunol.* **2011**, *166*, 269–280. [CrossRef]
139. Gopalakrishnapillai, A.; Kolb, E.A.; Dhanan, P.; Bojja, A.S.; Mason, R.W.; Corao, D.; Barwe, S.P. Generation of Pediatric Leukemia Xenograft Models in NSG-B2m Mice: Comparison with NOD/SCID Mice. *Front. Oncol.* **2016**, *6*, 1–11. [CrossRef]
140. Cosgun, K.N.; Rahmig, S.; Mende, N.; Reinke, S.; Hauber, I.; Schäfer, C.; Petzold, A.; Weisbach, H.; Heidkamp, G.; Purbojo, A.; et al. Kit regulates HSC engraftment across the human-mouse species barrier. *Cell Stem Cell* **2014**, *15*, 227–238. [CrossRef]
141. Morikawa, M.; Koinuma, D.; Mizutani, A.; Kawasaki, N.; Holmborn, K.; Sundqvist, A.; Tsutsumi, S.; Watabe, T.; Aburatani, H.; Heldin, C. Improved human erythropoiesis and platelet formation in humanized NSGW41 mice. *Cell Rep.* **2016**, *6*, 171–180.
142. Culen, M.; Kosarova, Z.; Jeziskova, I.; Folta, A.; Chovancova, J.; Loja, T.; Tom, N.; Bystry, V.; Janeckova, V.; Dvorakova, D.; et al. The influence of mutational status and biological characteristics of acute myeloid leukemia on xenotransplantation outcomes in NOD SCID gamma mice. *J. Cancer Res. Clin. Oncol.* **2018**, *144*, 1239–1251. [CrossRef] [PubMed]
143. Paczulla, A.M.; Dirnhofer, S.; Konantz, M.; Medinger, M.; Salih, H.R.; Rothfelder, K.; Tsakiris, D.A.; Passweg, J.R.; Lundberg, P.; Lengerke, C. Long-term observation reveals high-frequency engraftment of human acute myeloid leukemia in immunodeficient mice. *Haematologica* **2017**, *102*, 854–864. [CrossRef] [PubMed]
144. Griessinger, E.; Vargaftig, J.; Horswell, S.; Taussig, D.C.; Gribben, J.; Bonnet, D. Acute myeloid leukemia xenograft success prediction: Saving time. *Exp. Hematol.* **2018**, *59*, 66–71. [CrossRef] [PubMed]
145. Hogenesch, H.; Nikitin, A.Y. Challenges in pre-clinical testing of anti-cancer drugs in cell culture and in animal models. *J. Control. Release* **2012**, *164*, 183–186. [CrossRef] [PubMed]
146. Saland, E.; Boutzen, H.; Castellano, R.; Pouyet, L.; Griessinger, E.; Larrue, C.; De Toni, F.; Scotland, S.; David, M.; Danet-Desnoyers, G.; et al. A robust and rapid xenograft model to assess efficacy of chemotherapeutic agents for human acute myeloid leukemia. *Blood Cancer J.* **2015**, *5*, 1475–1486. [CrossRef] [PubMed]
147. Wunderlich, M.; Mizukawa, B.; Chou, F.S.; Sexton, C.; Shrestha, M.; Saunthararajah, Y.; Mulloy, J.C. AML cells are differentially sensitive to chemotherapy treatment in a human xenograft model. *Blood* **2013**, *121*, e90–e97. [CrossRef] [PubMed]
148. Farge, T.; Saland, E.; de Toni, F.; Aroua, N.; Hosseini, M.; Perry, R.; Bosc, C.; Sugita, M.; Stuani, L.; Fraisse, M.; et al. Chemotherapy-resistant human acute myeloid leukemia cells are not enriched for leukemic stem cells but require oxidative metabolism. *Cancer Discov.* **2017**, *7*, 716–735. [CrossRef]
149. Abarrategi, A.; Mian, S.A.; Passaro, D.; Rouault-Pierre, K.; Grey, W.; Bonnet, D. Modeling the human bone marrow niche in mice: From host bone marrow engraftment to bioengineering approaches. *J. Exp. Med.* **2018**, *215*, 729–743. [CrossRef]
150. Vaiselbuh, S.R.; Edelman, M.; Lipton, J.M.; Liu, J.M. Ectopic Human Mesenchymal Stem Cell-Coated Scaffolds in NOD/SCID Mice: An In Vivo Model of the Leukemia Niche. *Tissue Eng. Part C Methods* **2010**, *16*, 1523–1531. [CrossRef]
151. Antonelli, A.; Noort, W.A.; Jaques, J.; de Boer, B.; de Jong-Korlaar, R.; Brouwers-Vos, A.Z.; Lubbers-Aalders, L.; van Velzen, J.F.; Bloem, A.C.; Schuringa, J.J. Establishing human leukemia xenograft mouse models by implanting human bone marrow–like scaffold-based niches. *Blood* **2016**, *128*, 2949–2960. [CrossRef]
152. Battula, V.L.; Le, P.M.; Sun, J.C.; Nguyen, K.; Yuan, B.; Zhou, X.; Sonnylal, S.; McQueen, T.; Ruvolo, V.; Michel, K.A.; et al. AML-induced osteogenic differentiation in mesenchymal stromal cells supports leukemia growth. *JCI Insight* **2017**, *2*, 1–18. [CrossRef] [PubMed]
153. Chao, M.P.; Gentles, A.J.; Chatterjee, S.; Lan, F.; Reinisch, A.; Corces, M.R.; Xavy, S.; Shen, J.; Haag, D.; Chanda, S.; et al. Human AML-iPSCs Reacquire Leukemic Properties after Differentiation and Model Clonal Variation of Disease. *Cell Stem Cell* **2017**, *20*, 329–344. [CrossRef] [PubMed]
154. Kotini, A.G.; Chang, C.J.; Chow, A.; Yuan, H.; Ho, T.C.; Wang, T.; Vora, S.; Solovyov, A.; Husser, C.; Olszewska, M.; et al. Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia. *Cell Stem Cell* **2017**, *20*, 315–328. [CrossRef] [PubMed]

155. Heckl, D.; Kowalczyk, M.S.; Yudovich, D.; Belizaire, R.; Puram, R.V.; McConkey, M.E.; Thielke, A.; Aster, J.C.; Regev, A.; Ebert, B.L. Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **2014**, *32*, 941–946. [CrossRef] [PubMed]
156. Shi, X.; Kitano, A.; Jiang, Y.; Luu, V.; Hoegenauer, K.A.; Nakada, D. Clonal expansion and myeloid leukemia progression modeled by multiplex gene editing of murine hematopoietic progenitor cells. *Exp. Hematol.* **2018**, *64*, 33–44. [CrossRef]
157. Tothova, Z.; Krill-Burger, J.M.; Popova, K.D.; Landers, C.C.; Sievers, Q.L.; Yudovich, D.; Belizaire, R.; Aster, J.C.; Morgan, E.A.; Tsherniak, A.; et al. Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. *Cell Stem Cell* **2017**, *21*, 547–555. [CrossRef]
158. Chen, C.; Liu, Y.; Rappaport, A.R.; Kitzing, T.; Schultz, N.; Zhao, Z.; Shroff, A.S.; Dickins, R.A.; Vakoc, C.R.; Bradner, J.E.; et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell* **2014**, *25*, 652–665. [CrossRef]
159. Breese, E.H.; Buechele, C.; Dawson, C.; Cleary, M.L.; Porteus, M.H. Use of genome engineering to create patient specific MLL translocations in primary human hematopoietic stem and progenitor cells. *PLoS ONE* **2015**, *10*, 1–16. [CrossRef]
160. Schneidawind, C.; Jeong, J.; Schneidawind, D.; Kim, I.-S.; Duque-Afonso, J.; Wong, S.H.K.; Iwasaki, M.; Breese, E.H.; Zehnder, J.L.; Porteus, M.; et al. *MLL* leukemia induction by t(9;11) chromosomal translocation in human hematopoietic stem cells using genome editing. *Blood Adv.* **2018**, *2*, 832–845. [CrossRef]
161. Reimer, J.; Knöß, S.; Labuhn, M.; Charpentier, E.M.; Göhring, G.; Schlegelberger, B.; Klusmann, J.H.; Heckl, D. CRISPR-Cas9-induced t(11;19)/MLL-ENL translocations initiate leukemia in human hematopoietic progenitor cells in vivo. *Haematologica* **2017**, *102*, 1558–1566. [CrossRef]