

Dissecting the mechanisms of blocked differentiation in erythroleukemia

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Samantha Tauchmann

Basel, 2022

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von:

Prof. Dr. med. Jürg Schwaller

Prof. Dr. phil. Nat. Markus Affolter

PD Dr. med. Alexandre Theocharides

Basel, den 15.11.2022

Prof. Dr. Marcel Mayor, Dekan

«Wie? In diesem kurzen, eiligen, von einem ungeduldigen Dröhnen begleiteten Leben eine Treppe hinunterlaufen? Das ist unmöglich. Die dir zugemessene Zeit ist so kurz, daß du, wenn du eine Sekunde verlierst, schon dein ganzes Leben verloren hast, denn es ist nicht länger, es ist immer nur so lang, wie die Zeit, die du verlierst.

Hast du also einen Weg begonnen, setze ihn fort, unter allen Umständen, du kannst nur gewinnen, du läufst keine Gefahr, vielleicht wirst du am Ende abstürzen, hättest du aber schon nach den ersten Schritten dich zurückgewendet und wärest die Treppe hinuntergelaufen, wärest du gleich am Anfang abgestürzt und nicht vielleicht sondern ganz gewiß.

Findest du also nichts hier auf den Gängen, öffne die Türen, findest du nichts hinter diesen Türen, gibt es neue Stockwerke, findest du oben nichts, es ist keine Not, schwinge dich neue Treppen hinauf. Solange du nicht zu steigen aufhörst, hören die Stufen nicht auf, unter deinen steigenden Füßen wachsen sie aufwärts.«

“So if you find nothing in the corridors open the doors, and if you find nothing behind these doors there are more floors, and if you find nothing up there, don’t worry, just leap up another flight of stairs. As long as you don’t stop climbing, the stairs won’t end, under your climbing feet they will go on growing upwards.”

Fürsprecher– Franz Kafka - 1936

Table of contents

Abbreviations.....	7
Chapter 1	9
<i>General Introduction.....</i>	9
1.1. Hematopoiesis	10
1.2. Erythropoiesis.....	12
1.2.1. Erythropoiesis is regulated by a crosstalk between signaling molecules, epigenetic regulators, and transcription factors	14
1.2.1.1. Extracellular signals during erythropoiesis.....	15
1.2.1.2. DNA methylation during erythropoiesis.....	15
1.2.1.3. Histone modifications during erythropoiesis	15
1.2.1.4. Transcription factors that regulate erythropoiesis	17
1.2.1.5. GATA1 transcription factor complexes	19
1.3. GATA1 regulation in erythropoiesis	21
1.3.1. Regulation of <i>GATA1</i> mRNA translation	22
1.3.2. Regulation of GATA1 protein levels through posttranslational modifications.....	22
1.3.3. Regulation of GATA1 protein levels through caspase-mediated cleavage	22
1.3.4. Dysregulated GATA1 protein levels can result in hematological malignancies	23
1.4. Acute myeloid leukemia (AML).....	25
1.5. Acute erythroleukemia (AEL).....	28
1.5.1. History, diagnosis and characteristics of AEL	28
1.5.2. Genetic alterations in AEL.....	29
1.5.3. Models to study AEL	30
1.5.4. AEL as an epigenetic disease.....	32
Chapter 2	34
<i>Dissecting molecular mechanisms of impaired differentiation in erythroleukemia ...</i>	34
2.1. Aims and Objectives.....	35
2.2. NSD1 is a critical regulator of erythroid differentiation	35
2.3. Is GATA1 a key player in erythroleukemia?	36

Chapter 3	37
<i>Nuclear interacting SET domain protein 1 inactivation (NSD1) impairs GATA1-regulated erythroid differentiation and causes erythroleukemia.....</i>	37
3.1. Abstract	38
3.2. Experimental contribution	38
3.3. Additional experiments; unpublished findings (<i>not included in the manuscript</i>).....	38
3.3.1. GATA1 interactome in undifferentiated and 4 days differentiated MEL cells.....	38
3.3.2. GATA1 interactome in <i>Nsd1</i> ^{-/-} erythroblasts retrovirally overexpressing <i>mGata1</i>	41
3.3.3. The EZH2 small molecule inhibitor GSK126 depletes H3K27me3, but does not induce differentiation of <i>Nsd1</i> ^{-/-} erythroblasts.....	43
3.3.4. DZNep treatment reduces cell growth and induces erythroid differentiation in <i>Nsd1</i> ^{-/-} erythroblasts	43
3.3.5. Tracing GATA1 expression in <i>Nsd1</i> ^{-/-} mice	45
Chapter 4	48
<i>NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer..</i>	48
4.1. Rationale.....	49
4.2. Experimental approach	50
4.3. Results	50
Chapter 5	52
<i>Dissecting GATA1 protein interactions involved in impaired differentiation of malignant erythroblasts</i>	52
5.1. Introduction	53
5.2. Rationale.....	54
5.3. Experimental approach	54
5.4. Results	55
5.4.1. Proteome analysis separates human AEL cells from human erythroblasts with full differentiation potential.....	55
5.4.2. Exogenous GATA1 expression induces erythroid differentiation	60
5.4.3. Differential GATA1 interactome in normal and leukemic erythroblasts.....	62

5.4.4.	A targeted CRISPR/Cas9 screen identifies GATA1-associated proteins/genes that maintain the differentiation block of K562 cells.....	64
5.4.5.	Characterization of KPNA2, RCOR1, and CCT4 as GATA1 interactors that contribute to the erythroid differentiation block.....	67
5.4.6.	coREST inhibition enhances erythroid differentiation	69
5.5.	Additional experiments (will most likely not be included in the manuscript).....	70
5.5.1.	Characterization of the differentiating proteome of hEBST and HUDEP2 cells.....	70
5.5.2.	Characterization of the differentiating GATA1 interactome of hEBSTs.....	74
5.5.3.	mRNA and protein levels of erythroid master transcription factors do not correlate ..	76
5.5.4.	Effects of exogenous NUP155-FL and deletion mutant expression in hEBST and HUDEP2 cells	77
5.5.5.	NUP155 knockdown in K562 induces erythroid differentiation and enhances erythropoiesis in CD34 ⁺ CB cells.....	79
Chapter 6	82
<i>General Discussion</i>	<i>.....</i>	<i>82</i>
6.2.	Profiling the proteomic landscape in normal and malignant erythropoiesis.....	87
6.3.	Dissecting GATA1-interacting proteins in normal and malignant erythroid cells.....	89
6.4.	CRISPR/Cas9 screening in K562 AEL cells: limitations.....	100
6.5.	Concluding remarks	101
Chapter 7	102
<i>Additional Information</i>	<i>.....</i>	<i>102</i>
7.1.	Material and methods	103
7.2.	Supplementary figures	116
7.3.	Supplementary tables.....	121
7.4.	References.....	139
7.5.	Acknowledgements	164
7.6.	Appendices.....	168

Abbreviations

AEL	acute erythroleukemia	LDB1	LIM domain binding protein 1
AEV	avian erythroblastosis retrovirus	LMO2	rhombotin-2
AGM	aorta-gonad-mesonephros	LRPPRC	leucine rich pentatricopeptide repeat containing
ALB	albumin	M6a	erythroid/myeloid leukemia
AMKL	acute megakaryoblastic leukemia	M6b/PEL	pure erythroid leukemia
AML	acute myeloid leukemia	MBD	methyl-CpG binding domain protein
ANK1	ankyrin	MBIP	MAP3K12 binding inhibitory protein 1
ASH1L	ASH1-like histone lysine methyltransferase	MDS	myelodysplastic syndromes
Baso-EB	basophilic erythroblast	MEF	mouse embryonic fibroblasts
BFU-E	burst-forming unit-erythroid	MEL	murine erythroleukemia cells
BCL11A	BAF chromatin remodeling complex subunit	MEP	megakaryocyte-erythrocyte progenitor
BM	bone marrow	MFI	mean fluorescent intensity
BRD	bromodomain protein	MLF1	myeloid leukemia factor 1
C-ZnF	C-terminal zinc finger	MPN	myeloproliferative neoplasm
CBFA2T3/ETO2/MTG16	CBFA2/RUNX1 partner transcriptional co-repressor 3	MPP	multipotent progenitor
CCT	chaperone containing TCP1	MS	mass spectrometry
CD	cluster of differentiation	MTA	metastasis-associated protein
CD34	hematopoietic progenitor cell antigen	MTA2	metastasis associated 1 family member 2
CD36	platelet glycoprotein 4	MYC	MYC proto-oncogene
CFU-E	colony-forming unit-erythroid	N-TAD	N-terminal transactivation domain
CHD	chromodomain-helicase-DANN-binding protein	N-ZnF	N-terminal zinc finger
ChIP	chromatin immunoprecipitation	NCOR	nuclear receptor corepressor
cKIT/CD117	receptor tyrosine kinase KIT	NFX1	nuclear transcription factor, X-box binding 1
CLP	chronic lymphoid progenitor	NGS	next-generation sequencing
CML	chronic myeloid leukemia	NID	nuclear receptor interacting domain
CMP	chronic myeloid progenitor	NLS	nuclear localization signal
coREST	REST corepressor	NPC	nuclear pore complex
CREBBP/CBP	CREB binding protein	NSD1	nuclear receptor binding SET domain protein 1
CRIP1	cystein rich protein 1	NSE	non-specific esterase
DADS	diallyl disulfide	NUP93	nucleoporin 93
DBA	Diamond-Blackfan anemia	NuRD/Mi-2	nucleosome remodeling and deacetylase
DMSO	dimethyl sulfoxide	Ortho-EB	orthochromatic erythroblast
DNMT	DNA methyltransferase	p53/TP53	tumor protein p53
DOT1L	DOT1-like histone lysine methyltransferase	PARK7/DJ-1	parkinson-associated deglycase
Dox	doxycycline	PHD	plant homeodomain zinc finger
DZNep	3-Deazaneplanocin A	PLT	platelet
E2A/HEB	transcription factor E2-alpha	Poly-EB	polychromatic erythroblast
EED	embryonic ectoderm development	PRC2	polycomb repressive complex 2
EMT	epithelial-mesenchymal transition	Pro-EB	pro-erythroblast
ENO1	enolase 1	PTGES	prostaglandin E synthase
EP300	E1A binding protein P300	PTM	post-translational modifications

EPO	erythropietin	PWWP	proline-tryptophan-tryptophan-proline
EPOR	erythropietin receptor	QPRT	quinolinate phosphoribosyltransferase
ERG	ETS transcription factor ERG	RBBP	retinoblastoma binding protein
Ery	erythrocyte/red blood cell	RBC	red blood cell/erythrocyte
ESRE	extensively self-renewing erythroblast	RCOR1/coREST	REST corepressor 1
EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit	REDS	red cell differentiation signal
FAB	French-American-British	Retic/RTC	reticulocyte
FOG1/ZFPM1	friend of GATA1/zinc finger protein 1	ROS	reactive oxygen species
FUBP1	far Upstream Element binding protein 1	RPS28	ribosomal protein S28
FV	friend virus	RSAD2	radical S-adenosyl methionine domain containing 2
GATA	GATA binding protein	RUNX1	RUNX Family transcription factor 1
GFI-1B	growth factor independent 1b transcriptional repressor	sAML	AML from secondary neoplasm
GMP	granulocyte-monocyte progenitor	SCF	stem cell factor
GRN	gene regulatory networks	SERBP1	serpine binding protein
GSC	goosecoid protein	SET	Su(var)3-9, enhancer-of-zeste, trithorax
GYPA/TER119/CD235a	glycophorin A	SKI	SKI proto-oncogene
HAT	histone acetyltransferase	SLC	solute carrier transporter
HBA	alpha-globin	SLC4A1	solute carrier family 4 member 1
HBB	beta-globin	SMC1A	structural maintenance of chromosomes 1A
HBD	delta-globin	SP1	specificity protein 1
HBE	epsilon-globin	SPI1/PU.1	Spi-1 proto-oncogene
HBG	gamma-globin	SPTA1	spectin-alpha
HBZ	zeta-globin	SPTB	spectin-beta
HDAC	histone deacetylase	STAT5	signal transducer and activator of transcription 5
hEBST	human erythroblasts	SUZ12	SUZ12 polycomb repressive complex 2 subunit
HELLS/SMARCA6	helicase remodeler	t-AML	therapy-related AML
HKMT	histone lysine methyltransferase	TAL1/SCL	T-cell acute lymphocytic leukemia protein 1
HMBA	hexamethylene bisacetamide	TBCA	tubulin-specific chaperone A
HSC	hematopoietic stem cell	TBL1X	transducin beta like 1 x-linked
HSP	heat shock protein	TBL1XR1	transducin beta like 1 x-linked receptor 1
ICAM1	intercellular adhesion molecule 1	TET	ten-eleven translocation family of protein
IKZF1	ikaros family zinc finger 1	TF	transcription factor
IP	immunoprecipitation	TFR1/TFRC/CD71	transferrin receptor
IPO11/RanBP11	importin 11	TMT	tandem mass tag
JAK2	janus kinase 2	TRiC	TCP1 ring complex
KDM1A/LSD1	lysine demethylase 1A	UHRF1BP1/BLTP3A/ICBP90	ubiquitin-like containing PHD and RING finger domains 1-binding protein
KLF1/ELKF	krüppel-like transcription factor	WBC	white blood cells
KMT2C/MLL3	lysine methyltransferase 2C	WHO	world health organization
KMT5A/SETD8	lysine methyltransferase 5A	WT	wildtype
KPNA2	importin-alpha, IPOA1	XPO11	exportin 1
KPNB1	importin-beta, IPOB	ZEB2/SIP1	zinc finger E-box binding homeobox 2

Chapter 1

General Introduction

1.1. Hematopoiesis

Hematopoiesis is the generation of mature blood cells from a pool of hematopoietic stem cells (HSCs). Blood cells have a limited life span, which results in the need for continuous self-renewal, a supply of nutrients, and oxygenation. 3×10^9 cells are generated daily in the adult human bone marrow (BM), making blood one of the most regenerative and plastic tissue of the human body. The hematopoietic system is maintained by HSCs which are mostly quiescent but can enter the cell cycle and either self-renew or differentiate into multi-potent progenitors (MPP). These cells then give rise to more mature progenitor populations and ultimately differentiate into red blood cells such as erythrocytes and platelets or white blood cells, such as granulocytes (neutrophils, eosinophils, and basophils), monocytes and lymphocytes ⁽¹⁾. Hematopoiesis is a very tightly-regulated and critical process and has led to many controversies about the exact paths of blood cell generation. For decades, hematopoiesis was viewed as a hierarchical tree with HSCs sitting at the top of the pyramid and differentiating into multipotent progenitors (MPP). Here, the first branching point determines whether cells become lymphoid or myeloid via common lymphoid progenitors (CLP) or common myeloid progenitors (CMP). Further downstream branches segregate CMPs into megakaryocyte-erythrocyte progenitors (MEP) from granulocyte-monocyte progenitors (GMP) and ultimately in distinct blood cell types (**Fig. 1A**) ⁽²⁾. However, the classical view of HSC differentiation has been recently challenged by other factors that control cell differentiation and lineage fate decisions, resulting in a complex landscape of heterogeneous populations which are highly flexible to differentiate into most lineages. In the past years, the microenvironment has been characterized as a major determinant of HSC heterogeneity. HSCs localized within distinct niches are exposed to cell-extrinsic signals such as cytokines and growth factors, mediated by cells of the microenvironment which not only maintains survival and proliferation but also influences lineage choice decisions ⁽³⁾. Moreover, cell-intrinsic cues, such as DNA mutations and epigenetic modifications (DNA methylation, histone modifications, transcription factors) also contribute to HSC function in particular in response to stress or DNA damage and thereby determine differentiation towards a specific lineage⁽²⁻⁵⁾. HSC lineage output has been linked to cell cycle and/or metabolic features ^(2,3,6-8). Dormant HSCs frequently show high rates of glycolysis and a metabolic inactive state. When they leave the G_0 phase to enter into the cell cycle, transcription, cell size, and metabolic activity increase. This results in a switch from glycolysis to oxidative phosphorylation and leads to activation and differentiation of HSCs ⁽³⁾.

Single-cell omics analysis and lineage tracing have majorly contributed to the understanding of cell identity and to reconstitute the path of differentiating HSCs, which outcompetes the old model of artificially separating hematopoiesis into discrete stages of differentiation ⁽⁹⁾. It has been postulated that the first branching in lineage potential does not segregate lymphoid and myeloid output, but in

fact lineage-committed progenitors can directly differentiate from HSCs by bypassing hierarchical stages, which shows that the HSC population is more dynamic and integrated into a highly complex network, whose exact regulation still needs to be elucidated (**Fig. 1B**)⁽¹⁰⁾.

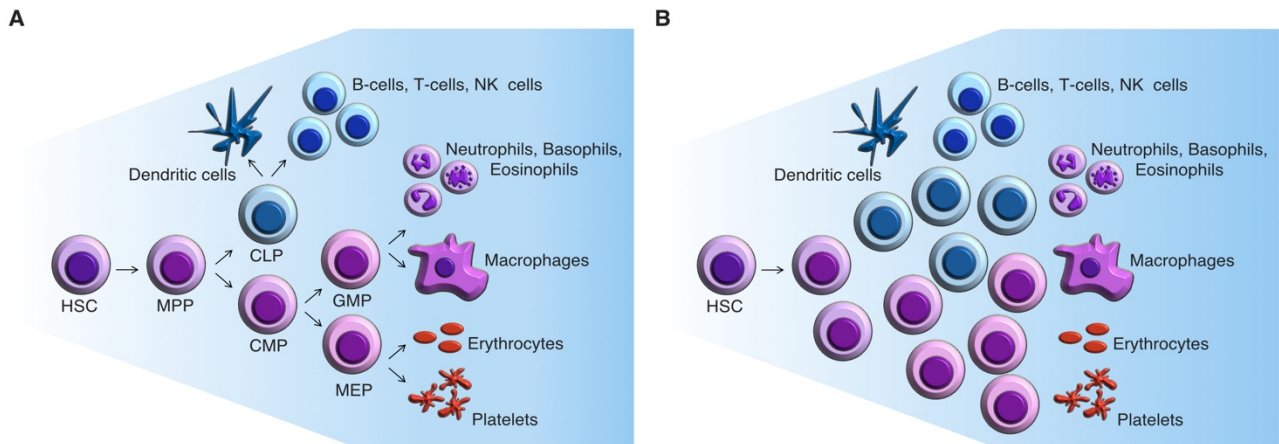


Figure 1. Models of hematopoietic differentiation landscapes. A) In the classical differentiation landscape, HSCs reside at the top of the hierarchic tree and undergo several separate intermediate steps (like MPP, CMP, CLP, GMP, MEP) to produce mature blood cells. **B)** In the continuous differentiation landscape, HSCs at the top can yield all mature blood cells with no obvious boundaries between the different hierarchical stages. Adapted from ⁽⁹⁾.

How blood cells are produced during embryonic development is best understood in the murine hematopoietic system which serves as a model for human hematopoiesis. The first signs of the formation of hematopoietic cells are found right after implantation of the blastocyst into the wall of the uterus. During the development of the embryo, the epiblast undergoes epithelial-mesenchymal transition (EMT) which results in the formation of three germ layers endoderm, ectoderm, and mesoderm. The mesoderm produces a particular cell type called hemangioblasts that migrate to the yolk sac and become committed to endothelial or hematopoietic progenitors. This process is also referred to as “primitive” hematopoiesis and technically occurs before the first heartbeat to ensure the growth and survival of the embryo. The first adult HSCs are generated in the intraembryonic aorta-gonad-mesonephros (AGM) region but quickly migrate to the placenta and fetal liver, which are the major sites of expansion initiating the so-called “definitive” hematopoiesis. Soon after, they also migrate to the spleen and ultimately around birth to the BM, which remain the major site of HSCs after birth ⁽¹¹⁾. The complexity and dynamic of blood generation within different tissues of the embryo make it difficult to assess the course of *de novo* HSC generation. To study HSC engraftment and multilineage differentiation, transplantation is needed. A key feature of adult “definitive” HSCs is their capability to engraft and long-term repopulate the BM to self-renew and differentiate their recipients. Strikingly, embryonic “primitive” HSCs do not provide long-term repopulation after transplantation ⁽¹²⁾.

1.2. Erythropoiesis

Erythropoiesis is the process, whereby billions of red blood cells (RBC) are produced every day. Maturation of erythroblasts is characterized by an expansion in number, hemoglobin accumulation, decrease in cell size, enucleation, and decrease in RNA content. The mature erythrocytes are by far the most abundant cells in the blood comprising about 45% of its volume. Depending on acute or chronic stress, such as trauma, blood loss, or infections, erythrocyte production can increase or decrease significantly. In humans, cells survive for 120 days and are mainly responsible for delivering oxygen to tissues. Aged and damaged erythrocytes are removed by macrophages that reside in the liver and the spleen. The iron that is released from senescent erythrocytes is recycled for heme synthesis and subsequently, hemoglobin molecules are incorporated into newly formed erythrocytes (13–16).

The generation of red blood cells is a highly complex multistep process and occurs in two waves, also called “primitive” and “definitive” erythropoiesis (15). Generally, all cells undergo 3 transitions from erythroid progenitors to erythroblastic precursors and ultimately mature RBC. “Primitive” erythroblasts are generated from mesodermal cells after approximately 18-20 days of gestation. The mesodermal layer is not part of the embryo but belongs to the yolk sac where erythroid cells aggregate in so-called blood islands. “Primitive” erythroblasts are the only circulating cells between 3-6 weeks and nucleated primitive erythroid cells can be found throughout the first trimester in the developing embryo. An important feature of human “primitive” erythroblasts is the expression of embryonic *HBE* (ϵ) and *HBZ* (ζ) globin genes. The “definitive” wave is characterized by “primitive” erythroid cells migrating to the fetal liver, producing “definitive” erythroblasts predominantly expressing human globins *HBA* (α) and *HBG* (γ). This event is marked by the replacement of fetal γ to adult *HBB* (β) globin, whereas α continues to be expressed throughout life. Also, the *HBD* (δ) globin is expressed after birth, although only small amounts are produced from this gene (**Fig. 2A**). In humans two hemoglobin “switches” have to occur, from embryonic to fetal and later from fetal to adult globin (17). Notably, defects in the globin “switching”, are a major cause of hemoglobinopathies (such as thalassemia syndromes and sickle cell disease) (18).

“Definitive” erythropoiesis not only occurs in the fetal liver but also in the postnatal BM where HSCs differentiate into early multipotent progenitors (MPP), which give rise to a variety of progenitor cells including megakaryocyte-erythroid progenitors (MEP). The first erythroid-specific progenitor is the burst-forming unit-erythroid (BFU-E) that gives rise to more mature cells, the so-called colony-forming unit-erythroid (CFU-E). The nomenclature of erythroid progenitors reflects their morphology. The earliest erythroid cell that can be clearly distinguished by morphology is the pro-erythroblast (Pro-EB). Pro-erythroblasts are large cells of 12-20 μ m in diameter with a large nuclear volume that occupies about 80% of the entire cell. Through mitosis basophilic erythroblasts (Baso-EB) are formed which

are smaller and characterized by an increase of the cytoplasm accompanied by condensation of the chromatin. Polychromatic erythroblasts (Poly-EB) are characterized by hemoglobinization of the cytoplasm and irregularly condensed nuclei. Orthochromatic erythroblasts (Ortho-EB) display nearly complete hemoglobinization and pyknotic nuclei. At this stage, cells are prepared for extrusion of the nucleus and become reticulocytes (Retic). The final stage of terminal erythroid differentiation involves extensive shedding of organelles and maturation into erythrocytes (Ery) (**Fig. 2B**). Terminal erythroid differentiation occurs in erythroblastic islands, which consist of a specialized multicellular structure composed of a central macrophage that is surrounded by maturing erythroid cells. This process takes about 4-5 erythroblastic cell divisions and the central macrophage supports this by phagocytosing the expelled nuclei. During this phase, the erythrocytes become bi-concavely shaped through excessive membrane remodeling. As mature erythrocytes only have a diameter of 6-8 μ M, this shape allows them to enter microcapillaries in tissues and enlarges the surface area for gas exchange.

Although the overall principles are similar between human and murine erythropoiesis, there are nevertheless some significant differences. For example, human erythrocytes have a longer life span (120 days) compared to mouse erythrocytes (42 days) ⁽¹⁹⁾. This also becomes obvious *in vitro*, since terminal erythroid differentiation of mouse cells is much faster than of human erythroblasts (6 vs. 12 days; own observation). Furthermore, fetal hemoglobin is not present in the mouse and the overall globin gene regulation is different ⁽²⁰⁾.

Flow cytometric analysis allows to identify specific cell stages based on surface marker expression in both, human and mouse erythropoiesis. At the beginning of differentiation HSC, MPP, and MEP all express the hematopoietic progenitor cell antigen CD34 and receptor tyrosine kinase KIT (cKit aka CD117). Phenotypically, maturing cells lose CD34 and start expressing platelet glycoprotein 4 (CD36) committing them to the erythroid lineage. Expression of CD36 is then gradually decreased during terminal maturation in which CFU-E start to upregulate the transferrin receptor (TFR1 aka CD71), which remains expressed until the Ortho-EB stage. The expression of the erythroid-specific marker glycophorin A (GYPA) (TER119 in the mouse, CD235a in humans) starts after the pro-erythroblastic stage when cells transition into the basophilic erythroblast stage and gradually become smaller. Upon terminal erythroid differentiation, GYPA expression steadily increases whereas CD71 expression is reduced ^(13,21-24) (**Fig. 2C**).

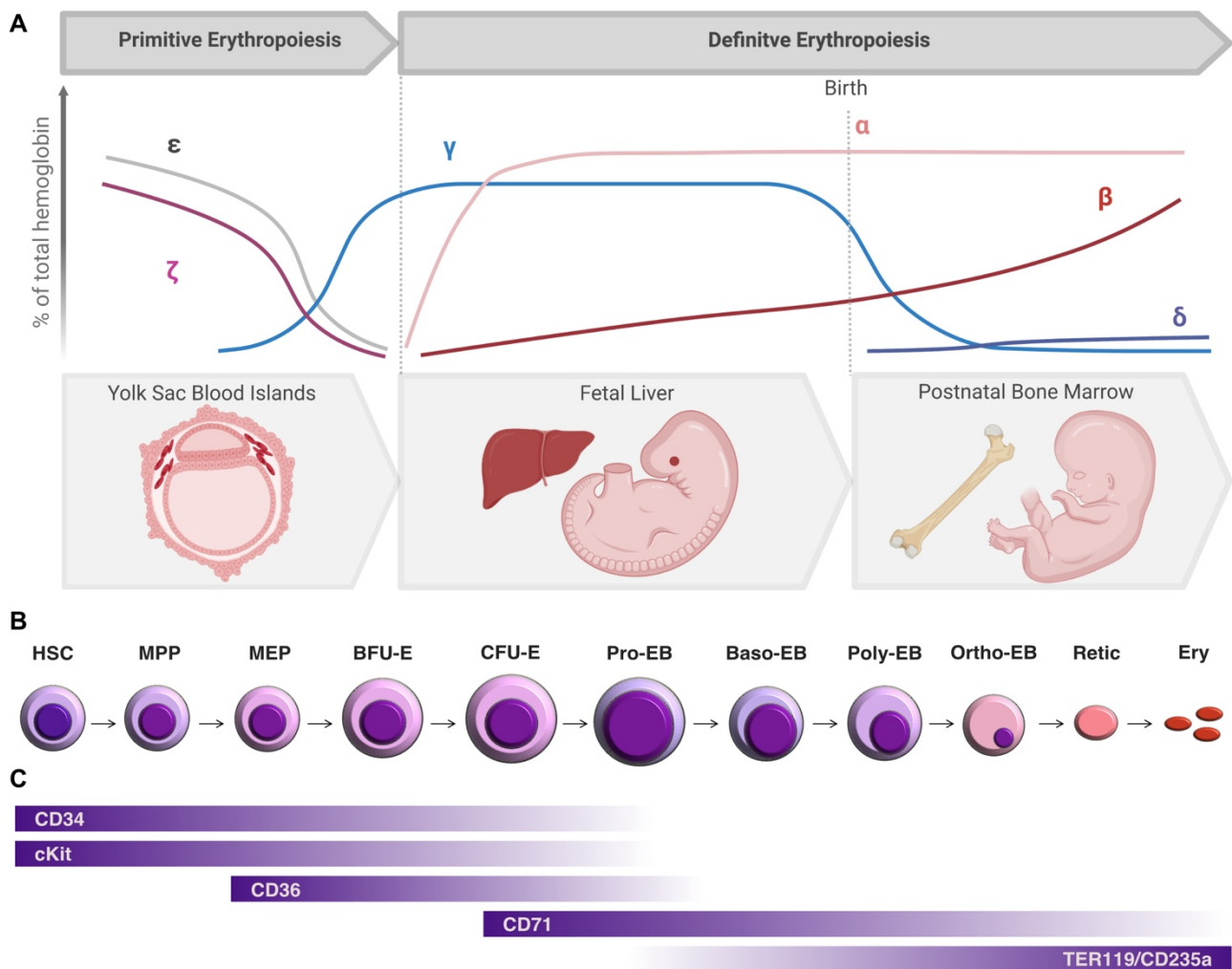


Figure 2. Primitive and definitive human erythropoiesis. **A)** Developmental pattern of hemoglobin expression: embryonic globins (ϵ and ζ) are silenced and fetal globin γ is followed by the expression of adult globin α . Perinatally, a switch from γ to adult globin β occurs. δ globin is expressed after birth at low levels. **(B)** Major anatomical sites of hematopoiesis during development. Erythropoiesis occurs in the yolk sac embryonic blood islands, then in the fetal liver, and finally in the BM (Created with Biorender.com., adapted from ⁽²⁵⁾). **(C)** Adult erythroid maturation into erythrocytes occurs over several stages involving distinct cellular identities which can be identified by **(C)** expression of surface markers (CD34, CD36, KIT, CD71, and TER119/CD235a). Colors on bars indicate high expression (dark) towards low expression (light) (Adapted from ^(13,24)).

1.2.1. Erythropoiesis is regulated by a crosstalk between signaling molecules, epigenetic regulators, and transcription factors

At each cell stage during lineage commitment and erythroid differentiation, specific epigenetic programs are executed. Intracellular signal transduction proteins and transcription factors (TFs) can interact with DNA-binding proteins to modify chromatin and establish a unique stage-specific epigenetic landscape ⁽²⁶⁾. In contrast to well-known exogenous factors like cytokines and hormones including stem cell factor (SCF), erythropoietin (EPO), thyroid hormones, androgens, estrogens, corticosteroids, vitamins, iron, and other growth factors, regulation of erythropoiesis by endogenous

epigenetic regulators remains less understood. The expression of genes that regulate erythropoiesis is not controlled by solely one modification, but rather a complex interplay between a variety of extracellular signals, chromatin modifiers (DNA methylation, histone modifications), and TFs that control each stage of erythropoiesis (**Fig. 3**).

1.2.1.1. Extracellular signals during erythropoiesis

Immature BFU-Es start expressing EPO receptor (EPOR), however, the erythroid-stimulating hormone EPO, is not required at this stage yet. BFU-Es then differentiate into CFU-Es that are fully EPO-dependent ⁽²⁷⁾. EPOR expression is lost when erythroblasts undergo terminal erythroid differentiation. EPO is however crucial for erythropoiesis and promotes survival. Its binding to EPOR triggers several signaling pathways, such as the Janus kinase 2/signal transducer and activator of transcription 5 (JAK2/STAT5), phosphoinositide-3 kinase/AKT, and Shc/Ras/mitogen-activated kinase (MAPK) pathways (**Fig. 3A**). Deletion of EPOR or JAK2 is embryonically lethal due to ineffective erythropoiesis. In particular disruption of the JAK2/STAT5 pathway has deleterious effects on RBC production. STAT5 activation can overcome defects in erythropoiesis in the absence of EPOR and JAK2 ⁽²⁸⁾. However, erythropoiesis cannot be rescued in STAT5a^{-/-}5b^{-/-} mouse embryos that accumulate erythroblasts due to impaired terminal differentiation resulting in severe anemia ⁽²⁹⁾.

1.2.1.2. DNA methylation during erythropoiesis

At the onset of terminal maturation, erythroblasts undergo dramatic chromatin remodeling ⁽²⁶⁾. Generally, DNA methylation is mediated by DNA methyltransferases (DNMTs). Four DNMTs have been found in mammals: DNMT1, DNMT2, DNMT3A, and DNMT3B. DNMT1 is characterized as a maintenance methylase, whereas DNMT3A and 3B are *de novo* methyltransferases ⁽³⁰⁾. In particular, DNMT3A and 3B methylate CpG dinucleotides (CpG islands) at the 5` position of cytosine residues, which is generally associated with gene repression ⁽³¹⁾. This silencing is also supported by interaction with transcription factors at specific loci resulting in site-specific DNA methylation at promoters. DNA methyl-CpG binding domain proteins, such as MBD1, MBD2, and MBD3 can be recruited to gene promoters, mainly acting as transcriptional repressors ^(31,32). Erythroid differentiation is associated with global DNA demethylation which is mediated by the ten-eleven translocation (TET) family of proteins, TET2 and TET3 (**Fig. 3B**)^(26,33).

1.2.1.3. Histone modifications during erythropoiesis

Posttranslational modifications of histone tails are key regulatory mechanisms that determine gene expression controlling erythroid maturation and nuclear condensation ⁽²⁶⁾.

Tail acetylation and deacetylation mainly occur at histones H3 and H4 and are a reversible process catalyzed by histone acetyltransferases (HATs) and histone deacetylases (HDACs). During acetylation, chromatin becomes accessible for transcription through the neutralization of the positive charge of the histone tail, which loosens the affinity to the negatively charged DNA. The best-known HAT co-activator are CREB binding protein (CREBBP aka CBP) and E1A binding protein P300 (EP300), which has been shown to interact with GATA binding protein 1 (GATA1) or Krüppel-like transcription factor (ELKF aka KLF1) to facilitate target gene activation ⁽³⁴⁾. Although, unlike global DNA demethylation, a global decrease in histone acetylation is not observed during erythroid differentiation ⁽²⁶⁾. However, a selection of HDACs are likely to play an important role during the process by removing histone marks and thereby repressing transcription ⁽³⁵⁾. Particularly in maturing erythroid cells, HDACs 2, 3, and 5 are highly expressed and play important roles in chromatin condensation during terminal erythroid maturation ⁽²⁶⁾. Knockdown of HDAC2 was shown to block nuclear condensation and enucleation ⁽³⁶⁾. Importantly, most HDACs do not contain DNA binding activities and need TFs to recruit them to specific loci. HDACs have in general a low substrate specificity and are therefore part of multiprotein complexes, such as the nucleosome remodeling and deacetylase (NuRD/Mi-2) or REST corepressor 1 (coREST) complex to execute their function ^(37,38).

Erythropoiesis is not only regulated by HATs and HDACs but also through dynamic histone lysine methylation. Histone lysine methyltransferases (HKMTs) that methylate H3K4, H3K36, and H3K79 are activating transcription, whereas H3K9, H3K27, and H4K20 are implicated in repressing transcription. HKMTs are very specific in their action and each methyltransferase usually only has one histone that it can methylate ^(35,39). The DOT1-like histone lysine methyltransferase (DOT1L), a H3K79 HKMTs is especially important during primitive embryonic erythropoiesis and controls normal expression of GATA2 and Spi-1 proto-oncogene (SPI1 aka PU.1) TFs ⁽⁴⁰⁾. The ASH1-like histone lysine methyltransferase (ASH1L), which methylates H3K4 and H3K36, has been proposed to correlate with differentiation of human erythroid progenitor cells and shRNA-mediated knockdown showed a reduction in α , β and γ -globin expression ⁽⁴¹⁾. Two other HKMTs, nuclear receptor binding SET domain protein 1 (NSD1) and lysine methyltransferase 5A (KMT5A aka SETD8), have been described to play critical roles during erythroid differentiation and gene regulation. NSD1 is essential for active transcription of erythroid genes by mono- and di-methylation of H3K36 whereas SETD8 monomethylates H4K20, which is important for chromatin condensation in late-stage erythroblasts prior to enucleation ⁽⁴²⁻⁴⁴⁾. Silencing is mediated by the H3K27 methyltransferase enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2) which is part of the polycomb repressive complex 2 (PRC2) recruited by GATA1 ⁽⁴⁵⁾. A well-known described histone lysine demethylase is lysine demethylase 1A (KDM1A aka LSD1), which demethylates H3K4 and silences transcription (**Fig 3B**) ⁽⁴⁶⁾.

1.2.1.4. Transcription factors that regulate erythropoiesis

Erythropoiesis is not only intimately correlated with chromatin modifications but also includes transcriptional regulators to establish an appropriate chromatin landscape. During erythroid commitment, the lymphoid and myeloid branches have to be suppressed to allow restriction towards erythroid cells. TFs at these branch points such as PU.1 and GATA family of TFs have been studied extensively. PU.1 mediates differentiation into lymphoid and myeloid lineages restricting differentiation into erythroid cells, whereas the interplay between GATA1 and GATA2 promotes differentiation along the myeloid-erythroid branch (**Fig. 3C**)^(47–49).

The GATA TF family consists of 6 (GATA1-GATA6) structurally and evolutionarily conserved members, all of which contain two characteristic zinc finger domains that enable specific binding to the DNA motif, (A/T)GATA(A/G) (WGATAR)⁽⁵⁰⁾. The binding to DNA depends on the C-terminal zinc finger (C-ZnF), while the N-terminal zinc finger (N-ZnF) is involved in protein-protein interactions to strengthen the binding of the C-ZnF on chromatin^(51–53). Amongst all GATA TFs the zinc finger domains are more than 70% conserved, whereas sequences at the N-terminus and C-terminus show less homology^(54,55). Both zinc fingers are important for GATA1 self-association in the presence or absence of DNA. Whereas the N-ZnF does not homodimerize itself, it can interact with the C-ZnF to mediate the self-association of GATA1⁽⁵⁶⁾. Single mutations of the tails of both zinc fingers showed reduced GATA1 self-association. The most severe effects, however, were obtained by the double mutation, which severely impaired self-association and transcriptional activation⁽⁵⁶⁾. It is possible that self-association does not only involve dimerization but that GATA1 also accumulates and forms multimers, but so far this has only been shown for GATA4⁽⁵⁷⁾.

GATA1 is known as the “master regulator” of erythroid differentiation. It is expressed during primitive and definitive erythropoiesis and interacts extensively with epigenetic regulators and proteins to shape the chromatin architecture. GATA1 was first discovered by its binding to the DNA regulatory sequences of globin genes highlighting the necessity of GATA1 for erythropoietic development⁽⁵⁸⁾. Next to the erythroid lineage, GATA1 expression was also detected in megakaryocytes, eosinophils, mast cells, and in Sertoli cells of the testis^(53,59,60). Loss of GATA1 results in arrested development of mouse embryonic red cell progenitors⁽⁶¹⁾. The *GATA1* gene is located on the X chromosome in mice and humans and encodes for a protein that contains 413 amino acids (aa) resulting in a molecular weight of approximately 48kDa^(62,63).

In contrast, GATA2 is important in mast and megakaryocytic development and is highly expressed in HSCs, CMP, and MEPs and its expression gradually decreases at the onset of erythroid commitment at the proerythroblast stage. GATA2-null mice lack definitive HSCs and show severe anemia with embryonic lethality at E10-11. Already heterozygous *Gata2*^{+/-} knockout mice are defective in the expansion and function of HSCs^(64,65).

A “TF-switch” has been proposed in which PU.1 functionally antagonizes GATA1. Elevation of PU.1 levels inhibited erythroid differentiation by physically interacting with GATA1 ⁽⁶⁶⁾. However, new TF-reporter mouse lines and computational analysis that trace lineage choices suggest that TFs only reinforce lineage choices, but decisions are already made. For example, the expression of PU.1 seems not to change during, but long after the predicted fate decision, opposing a putative “PU.1/GATA1 switch” ^(67,68). Another example is the so-called “GATA switch”. It was long thought that GATA1 and GATA2 compete on the same target loci and that abundance determines differentiation. However, even in HSCs GATA2 binds despite an excess of GATA1 and the downregulation of GATA2 is likely not to be a stochastic event ⁽⁶⁹⁾. It becomes obvious that finer molecular tools will possibly provide a new twist to how erythropoiesis is regulated.

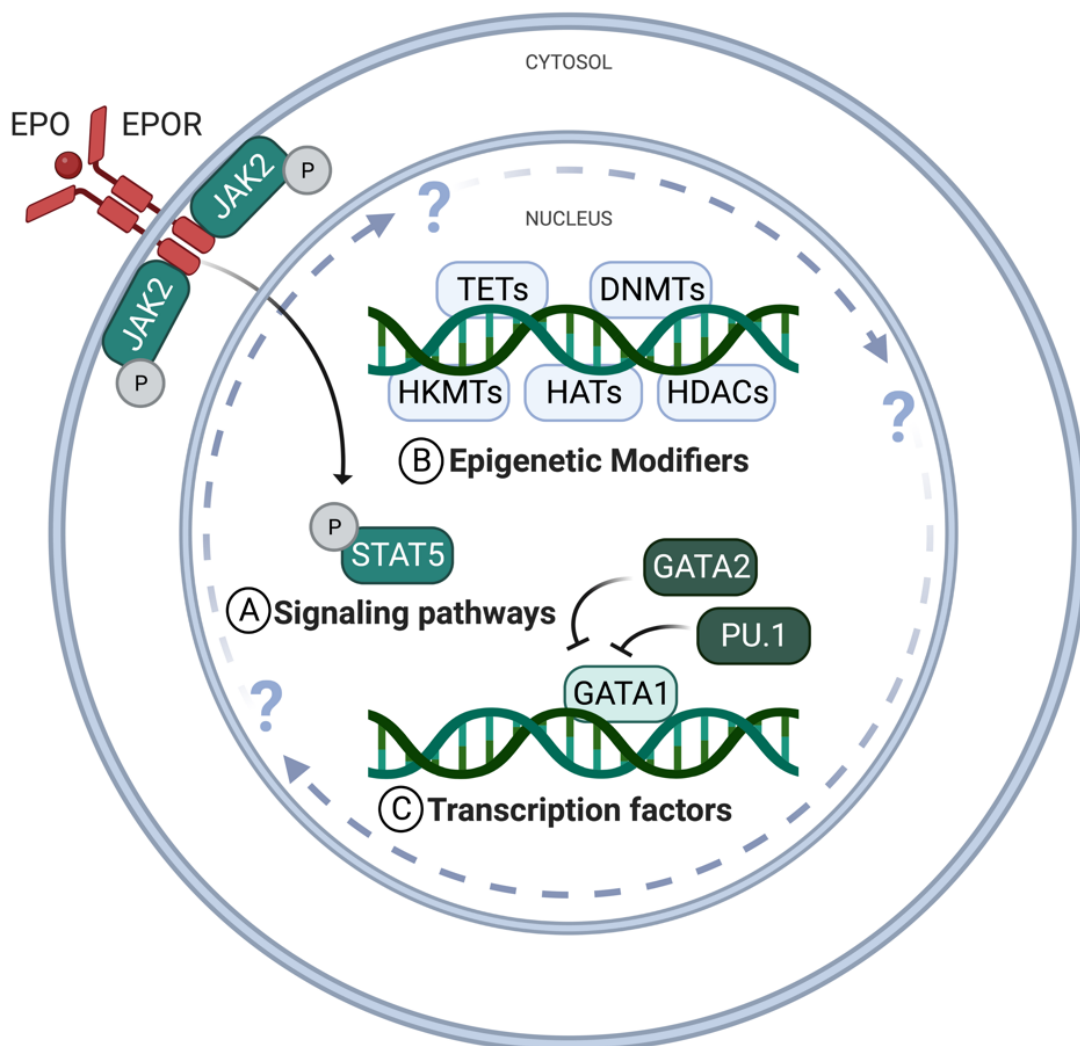


Figure 3. Crosstalk between signaling molecules, epigenetic regulators, and transcription factors controlling erythropoiesis. The expression of genes that regulate erythropoiesis is not controlled by solely one modification, but rather by an incompletely understood interplay between **A)** extracellular signals (EPO/JAK2/STAT5 signaling), **B)** chromatin modifiers (DNA methylation, histone modifications), and **C)** transcription factors that control erythroid commitment and maturation (P; phosphorylation, created with Biorender.com).

1.2.1.5. GATA1 transcription factor complexes

Efforts to understand how GATA1 activity is regulated have revealed interactions with multiple proteins, including other TFs, transcriptional co-regulators, and chromatin-remodeling and -modifying proteins. Depending on the partners, protein complexes result in either transcriptional activation or repression of GATA1 target genes (**Fig. 4**)⁽⁷⁰⁾.

One of the best-studied GATA1 protein interactions is binding to “Friend of GATA1” (FOG1 aka ZFPM1), which was identified as a cofactor modulating GATA1 activity on particular gene promoters⁽⁷¹⁾. Studies with GATA1 and FOG1 mutants, preventing their assembly, revealed that GATA1 activates and represses genes through FOG1-dependent mechanisms (**Fig. 4A-B**)^(72,73). GATA1/FOG1-mediated gene transcription involves NuRD/Mi-2 complex. It includes two core components with enzymatic functions, the chromodomain-helicase-DNA-binding proteins CHD3 (Mi-2 α) and CHD4 (Mi-2 β). In addition, the core complex also contains HDAC1 and HDAC2 that catalyze deacetylation. Furthermore, non-enzymatic subunits include the MBD2 and MBD3 which recruit the complex to methylated DNA, metastasis-associated proteins (MTA) MTA1-3, and retinoblastoma binding proteins (RBBP) RBBP4 and RBBP7. Structural components that directly assemble with histone tails are the transcriptional repressors P66 α /GATAD2a and P66 β /GATAD2b^(70,74). Initially, the NurD complex was thought to be solely required for GATA1/FOG1-mediated repression, however more recent studies have shown the dual function and that gene activation can be achieved through abrogation of HDAC1 activity by CBP/EP300-mediated acetylation (**Fig. 4C, D**)^(73,75–77).

Another well-described GATA1 interaction involves binding to the T-cell acute lymphocytic leukemia protein 1 (TAL1 aka SCL). Together, GATA1 and TAL1 can form a pentameric complex with LIM domain binding protein 1 (LDB1), transcription factor E2-alpha (E2A/HEB), and rhombotin-2 (LMO2) which are associated with the activation of erythroid genes. In erythroid cells, GATA1 directly associates with LMO2 but not with TAL1. These findings resulted in a model in which GATA1 binds to the GATA binding site whereas TAL1 and E2A bind to the adjacent E-Box motif (CANNTG). LMO2 which binds to LDB serves as a bridge to connect GATA1 and TAL1 (**Fig. 4E**)^(78–80).

Further studies in erythroid cells identified CBFA2/RUNX1 partner transcriptional co-repressor 3 (CBFA2T3 aka ETO2 or MTG16) as a component of the pentameric complex. CBFA2T3 associates with TAL1, E2A, and LDB1 in immunoprecipitation assays with the purpose to primarily suppress the activator function of the pentameric complex (**Fig. 4F**)⁽⁸¹⁾. Interestingly, so far an interaction between CBFA2T3 and GATA1 was only shown in megakaryocytic cells⁽⁷⁷⁾.

Rodriguez *et al.* showed that GATA1 moderately interacted with the growth factor independent 1b transcriptional repressor (GFI-1B)⁽⁷⁰⁾. Gene inactivation studies showed that GFI-1B is essential for regulating erythroid development and was found to bind to the coREST complex together with the

REST corepressor 1 (coREST aka RCOR1), LSD1, HDAC1, and HDAC2 ⁽⁸²⁾. Conditional loss of *Rcor1* in mice resulted in a complete block of myelo-erythroid differentiation ⁽⁸³⁾.

Despite efforts to understand the exact transcriptional consequences of GATA1-multiprotein complexes, differences in pull-down strategies, cell type, and developmental stage of cells impact the overall output. Over the past years, immunoprecipitation and mass spectrometry methods have been refined to confirm and also identify novel GATA1-interaction partners. In combination with chromatin immunoprecipitation (ChIP) experiments, this will allow for elucidation of the functional changes of distinct GATA1-protein complexes along with binding to target gene loci to regulate erythroid differentiation.

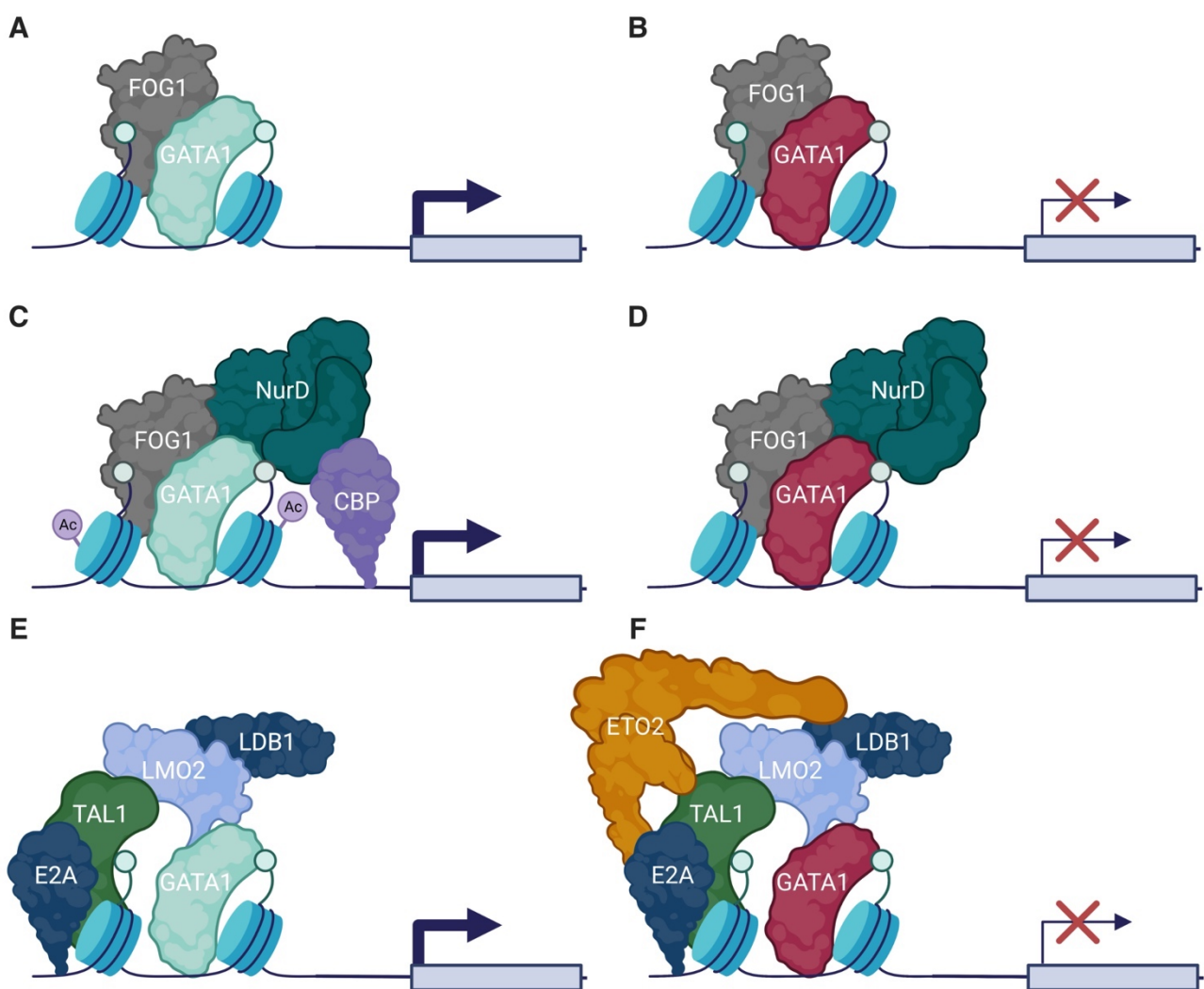


Figure 4. Model of GATA1-protein complexes to activate or repress erythroid differentiation. GATA1 forms activating protein complexes through interaction with either **A)** FOG1 alone, **C)** FOG1, the NurD complex and CBP/EP300; or **E)** as part of a pentameric GATA1/E2A/TAL1/LDB1/LMO2 complex. GATA1 forms repressive protein complexes via **B)** FOG1 alone or in combination with **D)** the NurD complex. **F)** Recruitment of the CBFA2T3 (ETO2) corepressor to the pentameric complex also results in the formation of a repressive complex. (Ac; acetylation, created with Biorender.com).

1.3. GATA1 regulation in erythropoiesis

GATA1 is probably the best-studied hematopoietic TF and is considered the master regulator of erythropoiesis as it regulates all aspects of maturation and functionality of RBCs. Maintaining GATA1 levels has been shown to be crucial during erythropoiesis and dysregulation results in ineffective erythropoiesis and/or leukemic transformation ⁽⁵⁰⁾.

In humans and mice, *GATA1* expression levels fluctuate. HSCs express low levels of *GATA1* mRNA, upon further maturation. *GATA1* expression slightly increases in the MEP and BFU-E. In CFU-Es *GATA1* levels peak, which activates the expression through the binding of GATA1 to the EPOR promoter ⁽⁸⁴⁾. EPOR-mediated signals induce the GATA1 transcriptional program, resulting in a series of changes in protein-protein interactions and post-translational modifications (PTM) to increase GATA1 binding to chromatin and enhance EPO-dependent erythroid maturation ^(85,86). After this, starting from the Pro-EB stage, *GATA1* levels gradually decline while cells fully mature into erythrocytes ^(50,87). Protein levels follow a similar dynamic, although their levels increase faster and decline steeper starting at the Pro-EB stage than mRNA levels (**Fig. 5**) ⁽⁶⁹⁾.

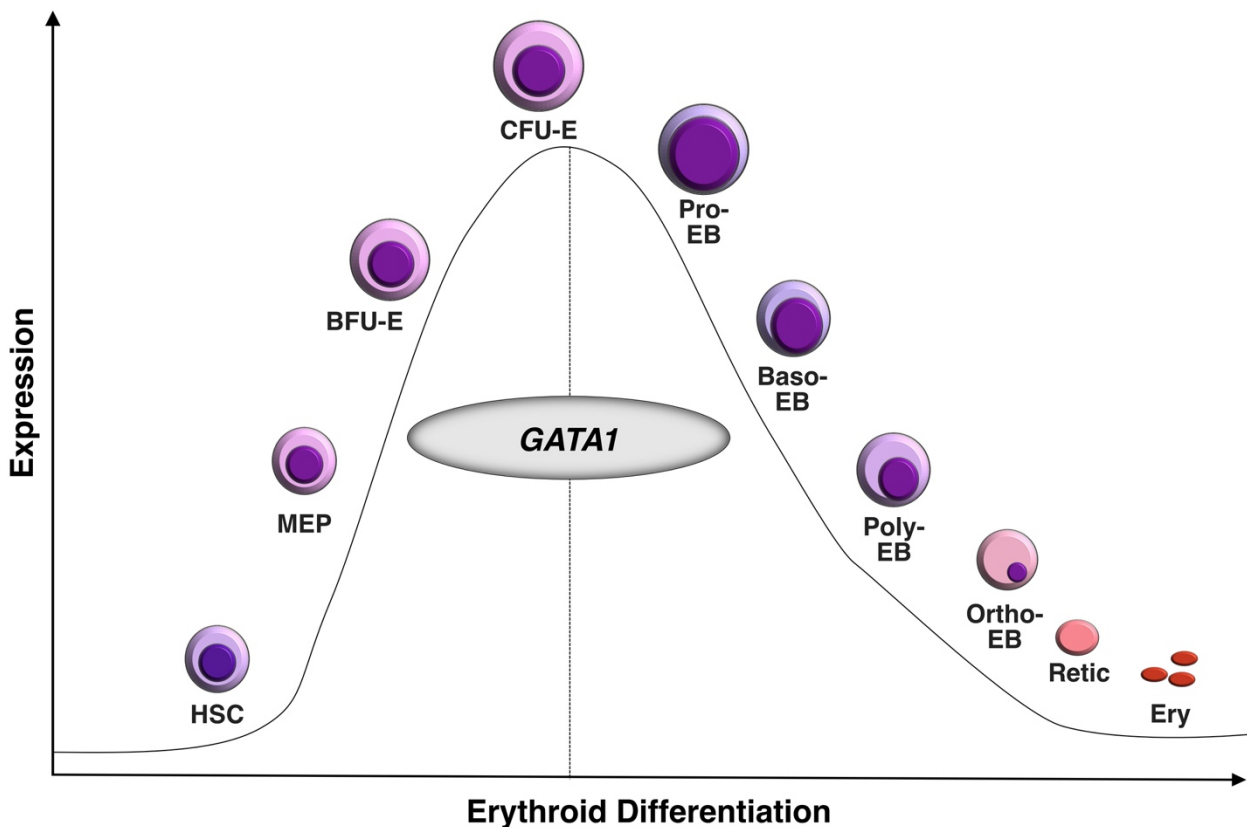


Figure 5. Dynamic expression of GATA1 during erythroid differentiation. During early stages, *GATA1* mRNA increases and reaches its peak levels at the CFU-E stage, whereas during terminal differentiation *GATA1* is downregulated. Protein expression follows a similar trend. Adapted from ⁽⁵³⁾.

1.3.1. Regulation of *GATA1* mRNA translation

GATA1 mRNA is translated into two isoforms: a full-length protein and a shorter version (so-called *GATA1s*), which lacks the N-terminal transactivation domain (N-TAD) resulting in functional aberrations ⁽⁵³⁾. Generally, *GATA1* mRNA has a less complex secondary structure than other hematopoietic TFs such as RUNX Family Transcription Factor 1 (RUNX1), which is associated with increased translational efficiencies and susceptibility to changes in ribosomal levels ^(50,88).

1.3.2. Regulation of *GATA1* protein levels through posttranslational modifications

Appropriate *GATA1* protein levels are controlled by a variety of mechanisms, in particular by multiple PTMs, including phosphorylation, acetylation, ubiquitination, and sumoylation ⁽⁵⁰⁾. The HAT CBP/EP300 acetylates *GATA1* which is required for *GATA1* binding to chromatin and activation of transcriptional targets that are necessary for terminal erythroid differentiation ^(89,90). Acetylation occurs on two lysine residues of the C-ZnF (245-252a.a. and 208-31a.a.) ⁽⁹¹⁾. Hereby, *GATA1* is stabilized to chromatin through its interaction with bromodomain protein 3 (BRD3) ^(92,93). Phosphorylation of *GATA1* occurs at seven serine residues, of which six seem to be constantly phosphorylated. Only the S310 residue which is located C-terminal of the C-ZnF is erythroid-specific ^(85,94). *GATA1* phosphorylation has been proposed to increase binding to chromatin and enhanced erythroid differentiation ⁽⁹⁵⁾. Chromatin-bound *GATA1* becomes phosphorylated, which results in targeting for degradation and ubiquitination of *GATA1* ⁽⁹⁶⁾. This degradation involves heat shock proteins such as the HSP27 protein chaperone, which thereby controls *GATA1* protein levels during erythroid differentiation ⁽⁹⁷⁾. Interestingly, sumoylation seems not to affect *GATA1* protein levels and the functional relevance remains unclear ^(98,99). Overall, the precise combinatorial network between *GATA1* PTMs in regulating protein levels is not entirely understood.

1.3.3. Regulation of *GATA1* protein levels through caspase-mediated cleavage

During erythropoiesis, *GATA1* protein is targeted by caspases ⁽¹⁰⁰⁾. This process occurs in an EPO-dependent manner: when EPO is present, *GATA1* is protected from cleavage by interaction with the heat shock 70kDa protein (HSP70) inside the nucleus. This results in the survival and proliferation of erythroblasts. Under EPO starvation, HSP70 will be exported out of the nucleus, and *GATA1* is no longer protected. This leads to cleavage of *GATA1* by caspase 3, differentiation arrest, and apoptosis ^(50,100,101). However, upon terminal erythroid differentiation, EPO is no longer needed and *GATA1* protein is cleaved to enable enucleation and maturation into fully functional erythrocytes. Therefore, caspase-mediated cleavage of *GATA1* is a double-edged sword and the right timing is essential to prevent cell death and allow normal erythroid differentiation.

Notably, in the context of globally impaired hematopoietic differentiation in diseases such as myelodysplastic syndromes (MDS) Frisan *et al.* found reduced GATA1 protein expression most likely resulting from defective nuclear localization of HSP70 and increased caspase 3-mediated cleavage. The expression of a mutant nucleus-targeted HSP70 was able to protect GATA1 and rescued erythroid differentiation ⁽¹⁰²⁾. Although studies have uncovered caspase-mediated mechanisms, the exact regulation of GATA1 degradation and the particular function of the chaperone system remains unclear. This opens interesting areas of research to understand protective GATA1 mechanisms that influence erythroid differentiation.

1.3.4. Dysregulated GATA1 protein levels can result in hematological malignancies

While normal erythroid differentiation depends on appropriate stage-specific GATA1 levels, aberrant expression can result in malignant transformation. The extreme effects are shown by mouse models with either complete loss or extremely high GATA1 levels which both result in embryonically lethal anemia ^(103,104). In particular, *Gata1*-null mice die *in utero* (E10.5-11.5) due to failure of erythroid precursors to mature beyond the proerythroblast stage resulting in severe anemia ⁽⁶¹⁾. Two *Gata1* mouse models have been created that result in the expression of 5% (*Gata1*^{0.05} ⁽¹⁰⁵⁾) or 20% (*Gata1*^{low} ⁽¹⁰⁶⁾) of normal *Gata1* levels. *Gata1*^{0.05} hemizygous males die from anemia due to ineffective primitive erythropoiesis between E11.5-E12.5. Heterozygous female mice were born at slightly lower mendelian frequencies as expected and often recovered from transient anemia due to X inactivation. However, females developed erythroleukemia between 3-6 months characterized by accumulation of immature erythroid progenitors or at advanced age B-Cell leukemia⁽¹⁰⁵⁾. *Gata1*^{low} mice, with physiologic GATA1 levels of 20% die of anemia between E13.5-14.5, with a few male mice born alive with severe anemia, which resolves around 4-5 weeks after birth. At 10 months, *Gata1*^{low} mice develop a myeloproliferative disease, characterized by severe thrombocytopenia and fibrosis of the BM, with defects in mast cell differentiation ⁽¹⁰⁶⁾. This suggests that reduced GATA1 levels are insufficient to promote primitive erythropoiesis but supports adult erythropoiesis and protects the cells from apoptosis although, erythroid differentiation is impaired. Overall, properly regulated GATA1 levels seem necessary to support normal blood development, which also becomes obvious in humans. Downregulation of GATA1 protein levels results in ineffective erythropoiesis characterized by expansion of progenitor cells and impaired production of erythrocytes and is often described in beta-thalassemia, MDS, Diamond-Blackfan anemia, 5q syndrome, and myelofibrosis ⁽¹⁰⁷⁾. In this disease context, downregulation of GATA1 protein levels can be the consequence of impaired translation of GATA1, impaired GATA1 protection by HSP70, or expression of a GATA1s ^(50,108).

Notably, not only reduced GATA1 levels affect erythropoiesis, also GATA1 overexpression results in dysfunctional erythropoiesis. A mouse model was generated with an X-linked *Gata1* transgene

resulting in its overexpression (GATA1-OX) ⁽¹⁰⁴⁾. In GATA1-OX male mice GATA1 overexpression in all erythroid cells beyond the Pro-EB stage impaired erythroid differentiation causing cell cycle arrest and apoptosis leading to anemia and death *in utero*. Strikingly, female mice developed normally, which can be explained by the phenomena of X-inactivation, only disrupting one *Gata1* allele. Thus, heterozygous female mice only had 50% of intrinsically defective erythroid progenitors which overexpressed GATA1, whereas the other 50% expressed normal levels. It was proposed that the defective maturation impaired erythroblasts responded to a yet unknown signal sent out by the wildtype (WT) cells, called “red cell differentiation signal” (REDS). However mixing WT fetal liver cells with GATA1 overexpressing maturation-impaired cells did not restore differentiation, indicating that REDS activity might also depend on cell-non-autonomous mechanisms. Notably, in the BM, erythropoiesis occurs in specialized units so-called erythroblastic islands ⁽¹⁰⁹⁾. These contain a central macrophage surrounded by erythroid cells at all stages of differentiation except mature erythrocytes. Immature cells are located close to the macrophage whereas differentiating cells can be found on the periphery. Macrophages phagocytize and degrade extruded nuclei from Ortho-EBs and reticulocytes can then enter the circulation. As a potential mechanism, it was suggested that differentiating Ortho-EBs send REDS signals towards adjacent overexpressing GATA1 maturation-impaired erythroblasts, which triggers them to undergo terminal differentiation (**Fig. 6**) ^(104,110). However, until now, the exact origin of REDS remains unknown, giving yet another example of many unclear mechanisms about GATA1 regulation throughout erythropoiesis.

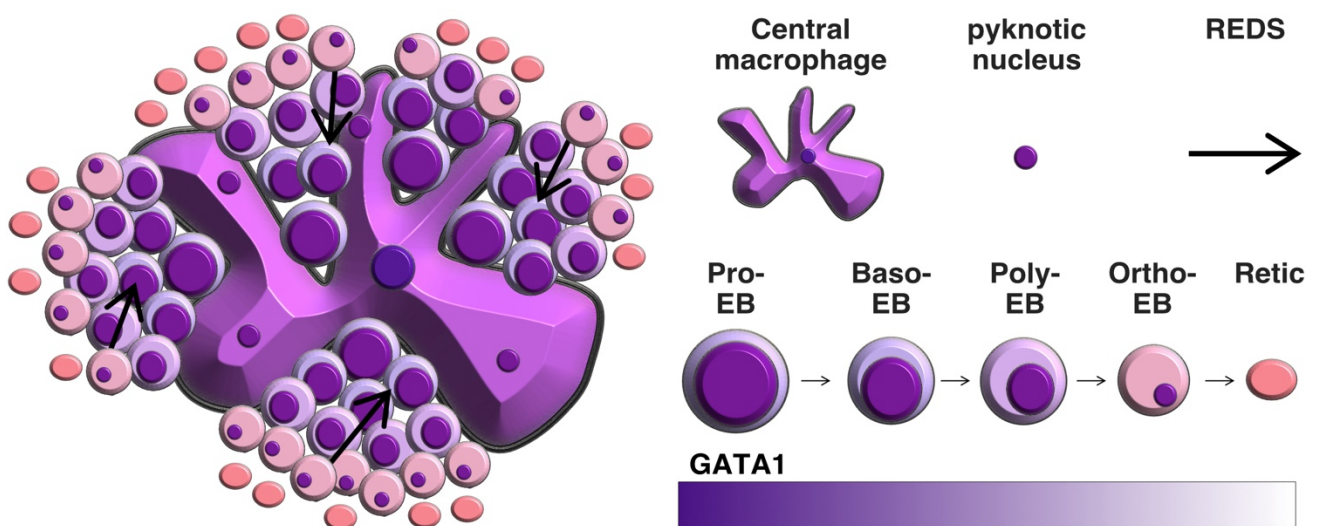


Figure 6. Terminal differentiation within an erythroblastic island. A central macrophage is surrounded by erythroid cells at various stages of differentiation. Arrows indicate the putative REDS signal, being produced by more differentiated Ortho-EBs (expressing lower GATA1) which signal to less differentiated erythroid cells (expressing higher GATA1). The colored bar indicates high expression (dark) towards low expression (light) of GATA1 during the differentiation process. Adapted from ⁽¹⁰⁴⁾.

1.4. Acute myeloid leukemia (AML)

Acute myeloid leukemia (AML) is an aggressive blood cancer characterized by the accumulation of poorly differentiated progenitor cells of the myeloid lineage. Expansion of malignant cells at the expense of normal hematopoiesis resulting in cytopenias that cause fatigue, anemia, bleeding, infections, and weight loss/anorexia ⁽¹¹¹⁾. There are two classification systems commonly used to diagnose AML. In 1976, the French-American-British (FAB) system was classifying AML based on the morphology of leukemic blasts and distinguished 8 subtypes (M0-M7) (**Table 1**) ⁽¹¹²⁾. M0-M4 are characterized as an accumulation of cells with maturation signs of the myeloid, M6 of the erythroid, and M7 of the megakaryocytic lineage.

Table 1. French-American-British (FAB) classification system of myeloid leukemias, including frequencies and outcomes of patients (adapted from: <https://www.cancer.org/cancer/acute-myeloid-leukemia>)

FAB subtype	Name	% of adult AML patients	Prognosis compared to average for AML
M0	Undifferentiated acute myeloblastic leukemia	5%	worse
M1	Acute myeloblastic leukemia with minimal maturation	15%	average
M2	Acute myeloblastic leukemia with maturation	25%	better
M3	Acute promyelocytic leukemia	10%	best
M4	Acute myelomonocytic leukemia	20%	average
M5	Acute monocytic leukemia	10%	average
M6	Acute erythroid leukemia	5%	worse
M7	Acute megakaryocytic leukemia	5%	worse

A more commonly used system from the World Health Organization (WHO) classifies AML not only based on morphologic appearance but also by taking into account genetic and molecular causes and how treatment affects disease progression and outcome. The classification was revised in 2017 but has been adapted recently. Based on the 2022 revision, the classification is structured systematically by first looking at the lineage, then the dominant clinical features, and lastly the dominant biological characteristics. Clinical features include descriptions such as acute, chronic, dysplastic and proliferative. Biological characteristics include gene fusions, rearrangements and mutations. This latest 5th edition of WHO classification separates AML in two major subtypes: “AML with defining genetic abnormalities” and “AML defined by differentiation” (**Table 2**). This eliminates the previously used term of AML-NOS (not otherwise specified), under which AMLs based on differentiation were listed. In earlier editions, AML was considered by the occurrence of more than 20% blasts. However, in the 5th edition, this blast cutoff value was removed instead emphasis was laid on assessing morphology and genetics more closely to determine disease origin. Therefore, the subclass “AML with defining genetic abnormalities” does not require 20% blasts with the exception of AML with

BCR::ABL1 fusion or AML with *CEBPA* mutation, to avoid overlap with chronic myeloid leukemia (CML). New and uncommon AML subtypes with rare fusions, will be defined as “AML with other defined genetic alterations” ⁽¹¹³⁾.

Table 2. 5th edition of WHO classification of AML based on ⁽¹¹³⁾.

Acute myeloid leukaemia with defining genetic abnormalities
Acute promyelocytic leukaemia with <i>PML::RARA</i> fusion
Acute myeloid leukaemia with <i>RUNX1::RUNX1T1</i> fusion
Acute myeloid leukaemia with <i>CBFB::MYH11</i> fusion
Acute myeloid leukaemia with <i>DEK::NUP214</i> fusion
Acute myeloid leukaemia with <i>RBM15::MRTFA</i> fusion
Acute myeloid leukaemia with <i>BCR::ABL1</i> fusion
Acute myeloid leukaemia with <i>KMT2A</i> rearrangement
Acute myeloid leukaemia with <i>MECOM</i> rearrangement
Acute myeloid leukaemia with <i>NUP98</i> rearrangement
Acute myeloid leukaemia with <i>NPM1</i> mutation
Acute myeloid leukaemia with <i>CEBPA</i> mutation
Acute myeloid leukaemia, myelodysplasia-related
Acute myeloid leukaemia with other defined genetic alterations
Acute myeloid leukaemia, defined by differentiation
Acute myeloid leukaemia with minimal differentiation
Acute myeloid leukaemia without maturation
Acute myeloid leukaemia with maturation
Acute basophilic leukaemia
Acute myelomonocytic leukaemia
Acute monocytic leukaemia
Acute erythroid leukaemia
Acute megakaryoblastic leukaemia

AML is a very heterogeneous disease that mostly arises *de novo*. AML can also develop as secondary neoplasm (sAML) from other hematologic malignancies, like MDS or myeloproliferative neoplasms (MPN). In addition, more and more frequently are therapy-related AML (t-AML) developing as a long-term consequence of genotoxic therapies (chemo-, and/or radiotherapy) ⁽¹¹⁴⁾. T-AML accounts for 7-8% of all AML cases and is associated with shorter survival than *de novo* AML ⁽¹¹⁵⁾.

Early cytogenetic and mutational analysis suggested that AML is the consequence of the cooperation of at least two functionally complementary genetic lesions leading to the formulation of a “two-hit model” by Gilliland and Griffin in 2002 ⁽¹¹⁶⁾. This model was based on 2 major groups of genetic alterations: class I mutations (e. g. *FLT3*, *KIT*, *RAS*), which usually activate signaling and provide a proliferative advantage, and class II mutations (e.g. *CEBPA*), which impair normal hematopoietic differentiation. The first hit usually occurs in genes belonging to class II, leading to impaired differentiation. Class I mutations typically occur as a later event leading to the leukemic transformation. This ultimately led to the concept of impaired differentiation and increased proliferation in AML. However, none of these mutations alone is sufficient to cause leukemia, so cooperation is

needed to initiate the disease. However, next-generation sequencing (NGS) approaches revealed more than 2 mutations per patient, which reclassified the “two-hit model” and led to the concept of a “multi-step model” that leads to malignant transformation ⁽¹¹¹⁾. In a collaborative effort, the Cancer Genome Atlas (TCGA) project took advantage of the most modern sequencing and genotyping tools to characterize 200 *de novo* AML patients ⁽¹¹⁷⁾. This redefined genomic lesions into 9 subgroups: class I – class IX (**Table 3**) ⁽¹¹⁸⁾.

Table 3. Identification of the (epi)genetic landscape of adult *de novo* AML based on ^(117,118).

Class	Type of genetic alteration	Affected genes
I	Transcription factor fusions	<i>RUNX1::RUNX1, CBFB::MYH6, PML::RARA, DEK::NUP214, MLL fusions</i>
II	NPM1 mutations	<i>NPM1</i>
III	Tumor suppressor gene mutations	<i>TP53, WT1, PHF6</i>
IV	DNA-methylation-related gene mutations	<i>TET2, IDH1, IDH2, DNMT3A</i>
V	Activated signaling gene mutations	<i>FLT3, KIT, RAS</i>
VI	Chromatin-modifying gene mutations	<i>ASXL1, EZH2</i>
VII	Myeloid transcription factor gene mutations	<i>CEBPA, RUNX1</i>
VIII	Cohesion-complex gene mutations	<i>STAG2, RAD21, SMC1, SMC2</i>
IX	Spliceosome-complex gene mutations	<i>SRSF2, U2AF35</i>

Despite efforts to characterize the epi(genetic) landscape of AML, the standard therapy has not drastically changed since the 1970s. As such, treatment consists of the “3+7” regime of chemotherapy. 3 days of daunorubicin and 7 days of cytarabine to debulk the initial leukemic burden followed by hematopoietic stem cell transplantation ^(119,120). However, this harsh treatment approach can be poorly tolerated especially in older patients. Since 2017, the US Food and Drug Administration (FDA) has approved new drugs, particularly targeting affected genes. FMS-like tyrosine kinase (FLT3) mutations are found in more than 30% of AML cases. FLT3 inhibitors such as midostaurin, gilteritinib, sorafenib, and quizartinib have been used in FLT3-mutated AML patients in combination with standard chemotherapy which resulted in improved survival rates than chemotherapy alone ^(119,120). Mutations in the isocitrate dehydrogenase genes IDH1&2 occur in 6-10% and 8-13% of AML cases, respectively ⁽¹²⁰⁾. Ivosidenib targets mutant IDH1, whereas enasidenib targets mutant IDH2. These have been used in combination with chemotherapy in young and fit patients but also in elderly patients (age>75) too unfit for intensive chemotherapy ⁽¹²¹⁾. An enhanced effect of IDH1/2-mutated AML was also seen when venetoclax was given. Venetoclax inhibits the B-cell lymphoma 2 (BCL2) and has been established as epigenetic therapy with hypomethylating agents such as azacytidine or decitabine but also improved patient outcomes when administered with low-dose cytarabine ⁽¹¹⁹⁾. This relatively mild treatment regimen is therefore used in older patients that are too unfit for intensive chemotherapy and has been implemented widely in the clinic ⁽¹²²⁾. Over the past years, AML-based immunotherapies are getting increasing attention. The success of such a therapy lies in identifying

the key target antigen predominantly expressed on the surface of leukemic cells. For AML there have been attempts to target CD33 (Siglec3) which led to the approval of gemtuzumab ozogamicin (GO) ⁽¹²³⁾. However, since potentially interesting antigens on leukemic cells are also expressed on normal hematopoietic cells, therapeutic opportunities are limited. It becomes obvious that understanding AML pathology is needed to develop specific targeted therapies.

1.5. Acute erythroleukemia (AEL)

1.5.1. History, diagnosis and characteristics of AEL

Acute erythroleukemia (AEL) is a rare (1-5%) but very aggressive type of AML generally characterized by the accumulation of immature erythroid precursors. The prognosis of AEL is reported as very poor with a median survival of four to six months ⁽¹²⁴⁾. AEL is mostly attributed to the elderly, whereby males are more commonly affected (65%), but can occasionally also be found in infants ^(125,126).

In 1917, Giovanni Di Guglielmo was the first one that characterized a patient having abnormal proliferation of immature erythroid progenitors. He called the condition “eritroleucemia” and hypothesized that it arose from a lesion in early hematopoietic cells. Over the years, he further defined the concept and separated the disease into “acute erythremic myelosis”, characterized by acute expansion of only erythroid cells, “chronic erythremic myelosis” and “erythroleukemia” which next to abnormal erythroid cells also presented with myeloid blasts in the blood ⁽¹²⁷⁾. Since erythroleukemia also occasionally showed different amounts of myeloblasts in the BM, around 1950, William Dameshek grouped all cases under the umbrella of myeloproliferative diseases as “Di Guglielmo`s Syndrome” ⁽¹²⁸⁾. He suggested that the syndrome progressed sequentially from a pure erythremic phase to an erythromyeloblastic phase into a myeloblastic phase.

In 1970, “Di Guglielmo`s Syndrome” was not considered a myeloproliferative disease anymore but instead became classified as acute leukemia in the FAB system ⁽¹²⁹⁾. Over the next 20 years, classification of the disease was difficult, since the FAB classification required a myeloid blast count of more than 30% in the BM to be considered leukemia, which did not take into account the predominant erythroid expansion. This resulted in the reclassification of AEL as a subtype of MDS, however, this grouping did not appear to be precise enough to characterize the subtype of erythroleukemia that is only composed of erythroid elements. This then led to the establishment of two subtypes: erythroid/myeloid leukemia (M6a) and pure erythroid leukemia (PEL/M6b) ⁽¹³⁰⁾. M6a patients were generally characterized as having more than 50% nucleated erythroid cells along with more than 20% of non-erythroid myeloblasts. PEL patients were defined as having more than 80 % of immature differentiation-impaired erythroblasts ⁽¹³¹⁾.

In the recently published 5th edition of the WHO classification, PEL (now referred to as AEL), remains a distinct AML subtype with an erythroid predominance of more than 80% and at least 30% of immature proerythroblasts ⁽¹¹³⁾. It becomes obvious that diagnosing AEL is challenging, as phenotypically and also genetically it closely overlaps with MDS. This is also attributed to the fact, that some AEL patients develop disease secondary to other hematological malignancies such as MDS, MPN, chronic myeloid leukemia (CML), or in response to chemotherapy as sAML ⁽¹³²⁾. Next to common symptoms, such as weight loss, fever, and night sweats, anemia is a common clinical feature, attributed to reduced hemoglobin levels. Furthermore, 20-30% of patients develop hepatosplenomegaly. In contrast to other AMLs, white blood counts remain usually in the normal range ⁽¹³²⁾. Since some cases of AEL are morphologically difficult to diagnose, flow cytometry and immunohistochemistry might help to dissolve the phenotype. Generally, Glycophorin A (CD235a), a sialoglycoprotein of the red blood cell membrane was shown to be expressed by tumor cells in 78% of AEL patients. Less specific but still useful are CD71 and CD36 surface markers for erythroblasts, which are also occasionally found on megakaryocytes and monocytes. Nevertheless, immunophenotypic characterization is valuable since it describes an undifferentiated erythroleukemia being Kit⁺, CD36⁺ and CD235a⁻ and a more mature type as Kit⁺, CD36⁺ and CD235a⁺ (**Fig. 2**) ⁽¹³³⁾. Notably, in addition to erythroid markers one group recently suggested that AEL cells can be distinguished from normal erythroblasts by assessing CD4, CD7, CD14, CD34, CD38, and HLA-DR as surface markers ⁽¹³⁴⁾.

Nevertheless, this characterization does not take into account all biological features and differential diagnosis still requires cytogenetics and molecular analysis to understand the molecular AEL pathogenesis.

1.5.2. Genetic alterations in AEL

Owing to its difficult diagnosis and rarity, the pathogenesis of AEL still remains poorly understood. Intrinsic cytogenetic and molecular alterations are suggested to underlie the current concept of AEL. Generally, erythroleukemic cells are characterized by impaired differentiation through mutations of transcriptional or chromatin regulators and by uncontrolled self-renewal and proliferation through mutations affecting cellular signal transduction pathways. Leukemic blasts from AEL patients often harbor complex karyotypes (50%) and chromosomal alterations were found in at least 75% of AEL patients ^(124,135). Although no unique specific chromosomal abnormalities were identified, loss of chromosomes 5 and 7 was the most prevalent. Two chromosomal translocations t(1;16)(p31;q24) and t(11;20)(p11;q11) have been found in very young children resulting in NFIA-CBFA2T3 and ZMYND8-RELA fusion genes, respectively ⁽¹³⁶⁻¹³⁸⁾. However, AEL is generally a disease affecting adults older than 60 years and frequently develops secondary to MDS therefore BCR-ABL fusions

are often associated with it (124,139). Furthermore, recurrent mutations were identified in NPM1, epigenetic modifiers (TET2, ASXL1, DNMT3A, and IDH2), signaling pathway proteins (JAK-STAT, RAS), and transcriptional regulators (RUNX1, FLT3). Strikingly, although relatively rarely mutated in other AML forms, *TP53* seems to play a prominent role in AEL pathogenesis as it was found to be mutated in more than 30% of AEL and almost all cases of PEL patients (140–142). Except for the high prevalence of *TP53* mutations, no exclusive AEL-associated driver mutation has been reported which shows the heterogeneity and simultaneously diagnostic dilemma of the disease (**Fig. 7**).

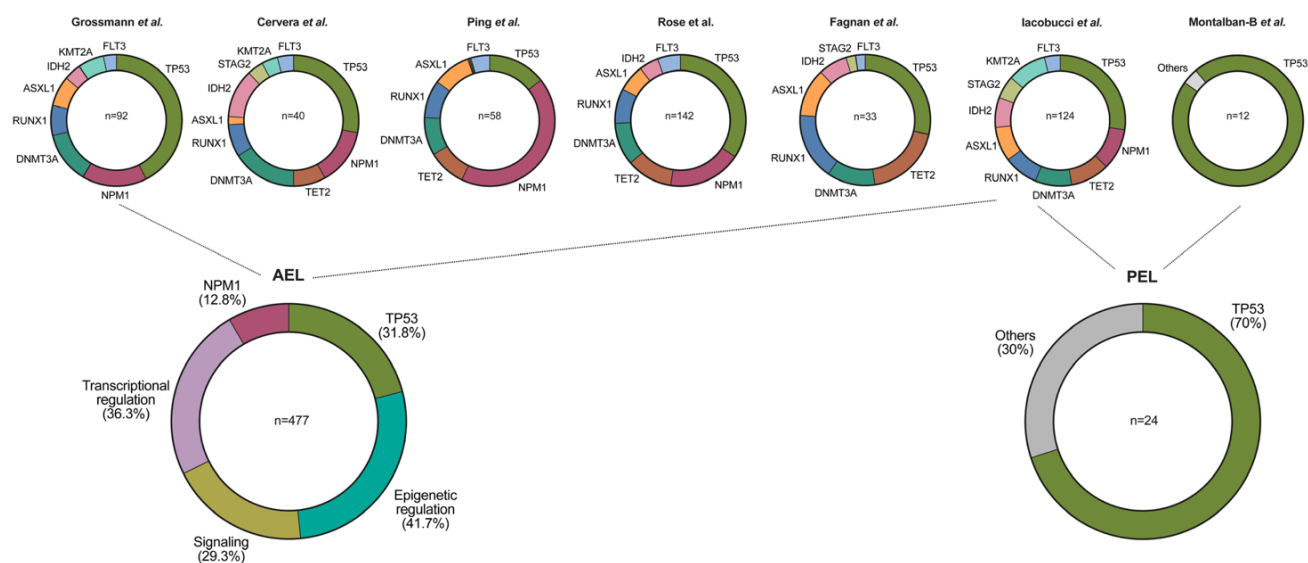


Figure 7. The genetic landscape of human AEL. The most prevalent genetic mutations reported in human AEL patients. *TP53* mutations were found in $\geq 30\%$ of AEL and in the majority of PEL patients. Note: only the studies by Iacobucci *et al.* and Montalban-Bravo *et al.* included patients with leukemia diagnosed as PEL (according to the WHO 2016 classification). Extracted from (124).

1.5.3. Models to study AEL

First insights into the biology of AEL emerged over half a century ago from avian and murine tumor viruses-induced erythroleukemia models. In addition, several transgenic mouse models started to unravel mechanisms of leukemic transformation through the malfunctioning of epigenetic modifiers, signaling effectors, or transcription factors. These models allow to investigate the molecular mechanisms that drive AEL. A more detailed description of models and mechanisms of erythroleukemia is provided in a recent review article, which I was also involved in (**Appendix 1**) (124). Here, I will only highlight a few approaches how one can model AEL.

The first models have mostly been discovered accidentally while studying oncogenic retroviruses in birds(143,144). The avian erythroblastosis retrovirus (AEV) induced erythroblastosis in young chickens. The genome of AEV consists of an oncogenic variant of the thyroid hormone receptor alpha ($TR\alpha$) called *v-erbA* and the mutated epithelial growth factor (EGF) receptor, *v-erbB*. Molecularly, *v-*

erbA interacts with SCF-activated tyrosine kinase Kit and cooperates with *v-erbB* to arrest erythroid differentiation and induce leukemia. Interestingly, although unable to get activated by hormone, *v-erbA* retains the ability to interact with co-repressors which suggests that recruitment of repressive complexes might actively repress erythroid differentiation⁽¹⁴⁵⁾. Most likely the best-characterized AEL model was established in 1957 by Charlotte Friend⁽¹⁴⁶⁾. She unexpectedly induced a transplantable AEL-like phenotype by injecting Ehrlich`s carcinoma cells into mice. These cells carried a mouse leukemia virus, called Friend virus (FV). Following inoculation, AEL evolves in two stages. In the first stage, erythroblasts start to proliferate due to constitutive activation of the EpoR. The second stage is characterized by loss of function of p53 accompanied by activation of the transcription factor PU.1 which results in erythroleukemic transformation⁽¹⁴⁷⁾. Permanent cell lines derived from these FV-induced mouse leukemias were established and are called murine erythroleukemia cells (MEL), which can be induced to differentiate into more mature hemoglobinized erythroid cells with various chemical agents such as dimethyl sulfoxide (DMSO) or hexamethylene bisacetamide (HMBA) and are thus widely used to study erythroid differentiation *in vitro*⁽¹⁴⁸⁾. Numerous other leukemia cell lines with “erythroid features” have been established over the past 30 years, which are arrested at different stages of differentiation⁽¹⁴⁹⁾. Most cell lines have been derived from secondary AELs. The best-studied AEL cell line is K562, derived from a patient with CML-blast crisis. The HEL cell line was established from a 30-year-old-man with AEL in relapse after treatment for Hodgkin lymphoma in 1980. In 1989, the F36P cell line was established from a 68-year-old-man with AEL secondary to MDS and subtype refractory anemia with excess of blasts (RAEB). KMOE2 was established from a 2-year-old girl with erythroblastosis and is therefore considered the only pure AEL cell line^(131,149) (**Fig. 8**). Since AEL is rare with limited access to primary patient-derived material these cell lines represent surrogate models to study molecular mechanisms of AEL. Furthermore, the cell lines also represent the heterogeneity of the human disease context.

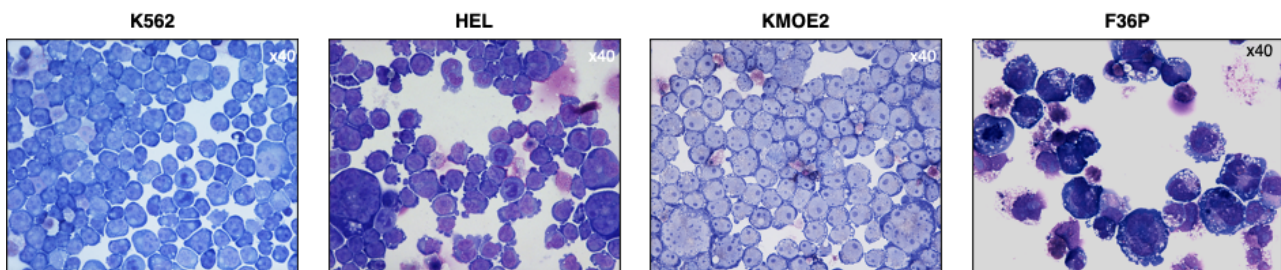


Figure 8. Selection of cell lines with erythroid features used in this project. Morphology was assessed on Wright-Giemsa-stained cytospin preparations.

1.5.4. AEL as an epigenetic disease

Three mouse models have been described, that connected epigenetic mechanisms to AEL development. First, loss of chromatin helicase remodeler (HELLS/SMARCA6) was associated with DNA hypomethylation, and upon transplantation of HELLS-deficient cells into irradiated mice, a subset developed an AEL-like disease. Furthermore, HELLS-deficient mice showed loss of DNMT3B binding to chromatin and elevated levels of PU.1 mRNA and protein⁽¹⁵⁰⁾. Although the disease phenotype was not in-depth characterized and was not fully penetrant, these observations suggest that aberrant chromatin organization may initiate malignant erythropoiesis. Second, more recently, Hochedlinger and colleagues described that hematopoietic expression of a doxycycline-inducible H3 lysine -to methionine point mutation (H3^{K36M}) led to a global reduction of H3K36me1/2/3 marks with a slight elevation of H3K27me3 levels, which rarely- co-occur on the same histone. H3K27me3 is mediated and maintained by the PRC2 supporting a repressive chromatin state⁽¹⁵¹⁾. Mice with the H3^{K36M} oncohistone transgene displayed severe anemia, thrombocytopenia, splenomegaly, and accumulation of erythroid progenitor cells⁽¹⁵²⁾.

One of the HKMT responsible for H3K36me1/2 is the nuclear receptor interacting SET domain 1 (NSD1). The NSD1 protein contains two nuclear receptor interacting domains (NIDs), two proline-tryptophan-tryptophan-proline (PWWP) domains, five plant homeodomain zinc fingers (PHD), a C-terminal (C5HCH) plant homeo-domain (PHD) finger and a catalytic (Su(var)3-9, Enhancer-of-zeste, Trithorax (SET) domain (**Fig. 9A**). The SET domain hereby sets the histone marks and methylates H3K36me1/2 predominantly at intergenic regions allowing recruitment of DNMT3A. This facilitates H3K36me3 by other HKMTs and binding of DNMT3B to active gene bodies and a generally active chromatin state (**Fig. 9B**). Ablation of NSD1 results in loss of H3K36me2 marks at intergenic regions, which allows spreading of polycomb repressive PRC2-mediated H3K27me3 marks and redistribution of DNMT3A-mediated DNA methylation to active gene bodies (**Fig. 9C**)⁽⁴³⁾. Third, to better understand its function, our lab inactivated NSD1 in mouse hematopoietic cells. Hereby, we modified a previously established conditional *Nsd1* allele (floxed exon 5, *Nsd1^{fl/fl}*) and crossed it with the *Vav1-iCre* transgenic line that cleaves the floxed allele in hematopoietic and some endothelial cells starting during around E12.5 of fetal hematopoiesis. At an average of 6-25 weeks (median 91 days; n=24), 100% of the mice developed a fully penetrant AEL-like disease. Phenotypically, the mice developed significant splenomegaly and showed extensive infiltration of BM, spleen, and other organs by erythroid progenitor cells. Analysis of peripheral blood revealed a reduction in mature erythrocytes, lower hemoglobin levels, reticulocytosis, thrombocytopenia, and the appearance of abnormal blast-like cells with a blue cytoplasm. Attempts to understand the molecular mechanisms we found that despite abundant GATA1 protein, expression of some of its activated primary erythroid targets like *Hba*, *Hbb*, or *Gypa* were significantly reduced. These findings indicate that ablation of *Nsd1* results in

a lethal disease in mice that phenocopies many aspects of human AEL⁽¹⁵³⁾. Collectively, these studies suggest a complex combinatorial interplay between chromatin architecture, along with changes in DNA methylation and aberrant expression of TFs might be critical for AEL development.

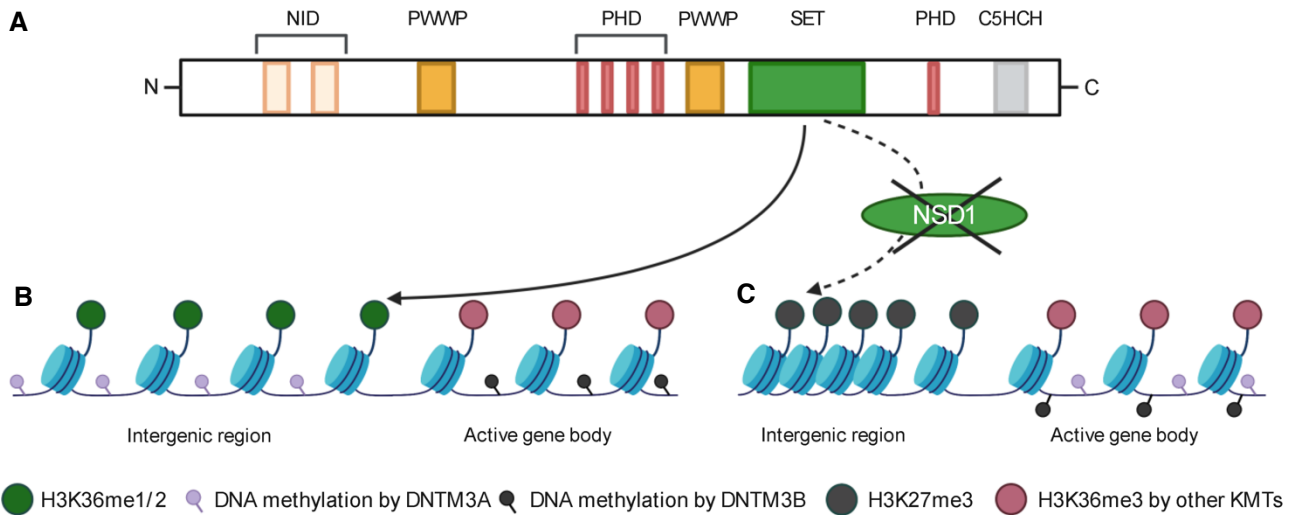


Figure 9. NSD1 protein structure and epigenetic regulation. A) NSD1 contains two nuclear receptor interacting domains (NID), two proline-tryptophan-tryptophan-proline (PWWP), five plant homeodomain zinc fingers (PHD), the catalytic Su(var)3-9, enhancer-of-zeste, Trithorax (SET) and the C-terminal C5HCH (Cys-His) domain. **B)** The NSD1 SET domain methylates H3K36me1/2 predominantly at intergenic regions allowing recruitment of DNMT3A and facilitates H3K36me3 by other HKMTs allowing recruitment of DNMT3B to active gene bodies. **C)** Ablation of NSD1 results in loss of H3K36me2 marks, which allows spreading of polycomb repressive PRC2-mediated H3K27me3 marks at intergenic DNA and redistribution of DNMT3A-mediated DNA methylation to active gene bodies. Extracted from⁽⁴³⁾.

Chapter 2

*Dissecting molecular mechanisms of
impaired differentiation in erythroleukemia*

2.1. Aims and Objectives

AEL is characterized by uncontrolled accumulation of poorly differentiated progenitor cells of the erythroid lineage. The overall aim of my thesis was to characterize molecular mechanisms that are responsible for impaired differentiation in AEL. Hereby, I was involved in two projects: first in the characterization of the NSD1 HKMT as an unexpected regulator of normal and malignant erythropoiesis (2.2.), and second, in the characterization of aberrantly formed GATA1 protein complexes that contribute to impaired maturation in AEL cells (2.3.).

2.2. NSD1 is a critical regulator of erythroid differentiation

Our laboratory studies molecular lesions initiating and/or maintaining acute myeloid leukemia (AML). While studying a chromosomal translocation t(5;11)(q35;p15) leading to the expression of a fusion between the nuclear pore protein 98 (NUP98) and the nuclear receptor interacting SET domain protein 1 H3K36 histone lysine methyltransferase (NSD1) associated with pediatric AML, we wondered about the function of NSD1 in normal hematopoiesis. To address this question, we first assessed the effects of knocking down NSD1 in human CD34⁺ cells and found increased self-renewing capacity and the formation of dense reddish colonies with a proerythroblast-like morphology. Unfortunately, deeper functional studies were hampered by the fact that cells could not be expanded in culture. Therefore, a mouse model was created in which we conditionally ablated the *Nsd1* gene in the hematopoietic system of the mouse. We found that inactivation of *Nsd1* during late fetal development unexpectedly induced a fully penetrant and transplantable disease closely mimicking human AEL. To identify the underlying molecular mechanisms, we aimed to rescue the phenotype by viral expression of either *Nsd1*-WT or a catalytically SET-domain mutant *Nsd1*. Only the *Nsd1*-WT ORF transduced *Nsd1*^{-/-} erythroblasts restored terminal differentiation associated with increased chromatin binding and activation of target genes of the erythroid master regulator GATA1. While the loss of the *Nsd1* catalytical domain severely impaired GATA1 transactivation activity and erythroid differentiation, overexpression of exogenous GATA1 was sufficient to partially overcome the differentiation block (**Appendix 2**, Leonards, Almosaileakh, Tauchmann *et al.*, 2020⁽⁴²⁾). Additional experiments of our work on *Nsd1* in hematopoiesis and erythroleukemia are provided in **Chapter 3**.

Based on the involvement of NSD1 in developmental disorders and the observed increased cell size of *Nsd1*^{-/-} erythroblasts, we generated and characterized a novel mouse model where *Nsd1*^{fl/fl} mice were crossed with an *UBC-Cre-ERT2* mouse line to dynamically follow the impact of NSD1 inactivation over time. Notably, *NSD1* monoallelic germline loss of function mutations have been described in a disease called SOTOS syndrome. SOTOS syndrome is generally characterized by distinct craniofacial appearance, childhood overgrowth, and learning disabilities with an increased risk to develop cancer. However, the mechanisms of how loss of function NSD1 mutations contribute to

overgrowth remain unclear. With this mouse model, we aimed to have a more controllable system to investigate the immediate changes occurring upon *Nsd1* ablation (**Chapter 4**).

2.3. Is GATA1 a key player in erythroleukemia?

To better understand the molecular mechanisms that control human AEL, we, in collaboration with the group of Thomas Mercher, have characterized the genetic and transcriptional landscape of 33 AEL patients ⁽¹⁴⁰⁾. Strikingly, >25% of AEL patients aberrantly expressed transcriptional co-regulators such as SKI, ERG, and CBFA2T3, which when ectopically expressed in WT murine erythroid progenitors blocked erythroid differentiation and functionally interfered with GATA1 activity resulting in decreased chromatin accessibility of GATA1-binding sites, suggesting that human AEL could be driven by GATA1 dysfunction.

Theoretically, impaired GATA1 function and thus impaired erythroid differentiation could be the consequence of (**Fig. 10**):

1. Repressive chromatin architecture, the inability of GATA1 complexes to bind
2. Alterations in GATA1 post-translational modifications
3. Aberrant GATA1-protein-protein interactions

Since multiple studies have shown that terminal erythroid maturation is controlled by GATA1 acting in transcriptionally active and repressive complexes, we primarily focused on investigating differences between normal and malignant erythroblasts, by looking into their proteome and GATA1 interactome. To identify novel regulators that control erythroid differentiation, we performed a targeted CRISPR/Cas9 screen of genes that encode for proteins overlapping between the proteome and GATA1-IP/MS analysis (**Chapter 5**).

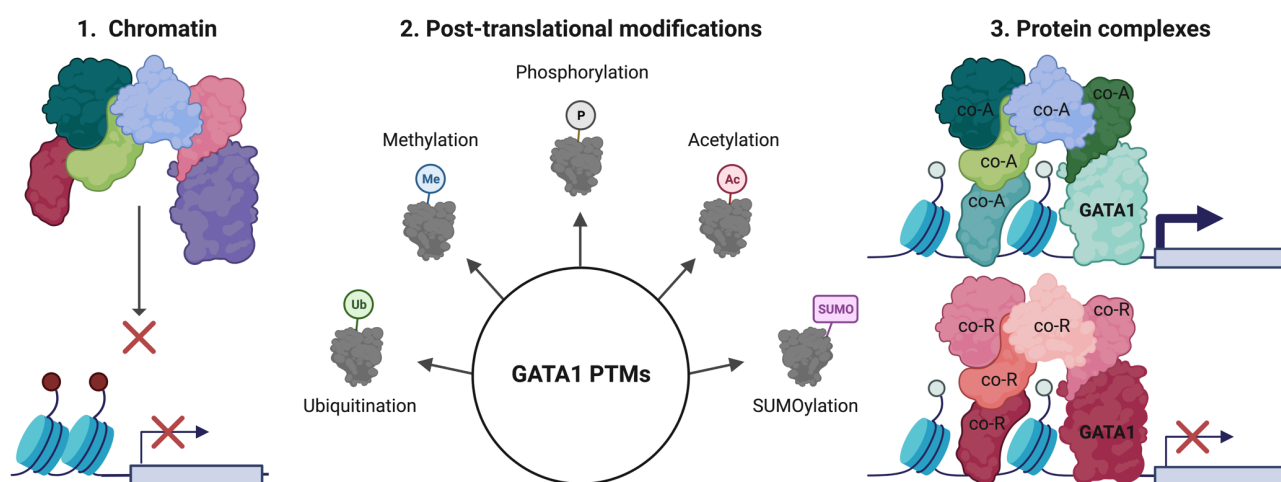


Figure 10. Impaired GATA1 activity as potential cause of impaired AEL differentiation. Alterations could affect 1) GATA1 binding to chromatin or result in 2) aberrant GATA1 posttranslational modifications (PTM). Another potential mechanism could be aberrant 3) GATA1 protein-protein interactions, such as trapping of GATA1 into repressive proteins to prevent activation of erythroid gene transcription (created with Biorender.com).

Chapter 3

Nuclear interacting SET domain protein 1 inactivation (NSD1) impairs GATA1-regulated erythroid differentiation and causes erythroleukemia

Leonards, K.* , Almosailleakh, M.* , Tauchmann, S* . et al. Nuclear interacting SET domain protein 1 inactivation impairs GATA1-regulated erythroid differentiation and causes erythroleukemia. *Nat Commun* **11**, 2807 (2020). <https://doi.org/10.1038/s41467-020-16179-8>

* equal contribution

3.1. Abstract

The nuclear receptor binding SET domain protein 1 (NSD1) is recurrently mutated in human cancers including acute leukemia. We found that NSD1 knockdown alters the erythroid clonogenic growth of human CD34⁺ hematopoietic cells. Ablation of *Nsd1* in the hematopoietic system of mice induced a transplantable erythroleukemia. *In vitro* differentiation of *Nsd1*^{-/-} erythroblasts was majorly impaired despite abundant expression of GATA1, the transcriptional master regulator of erythropoiesis, and associated with impaired activation of GATA1-induced targets. Retroviral expression of wildtype NSD1, but not a catalytically-inactive NSD1^{N1918Q} SET-domain mutant induced terminal maturation of *Nsd1*^{-/-} erythroblasts. Despite similar GATA1 protein levels, exogenous NSD1 but not NSD1^{N1918Q} significantly increased the occupancy of GATA1 at target genes and their expression. Notably, exogenous NSD1 reduced the association of GATA1 with the co-repressor SKI, and knockdown of SKI induced differentiation of *Nsd1*^{-/-} erythroblasts. Collectively, we identified the NSD1 methyltransferase as a novel regulator of GATA1-controlled erythroid differentiation and leukemogenesis (**Appendix 2**)⁽⁴²⁾.

3.2. Experimental contribution

Katharina Leonards initially characterized the erythroleukemic phenotype in *Nsd1*^{-/-} mice. I obtained more replicates (**Appendix 2, Fig. 2**)⁽⁴²⁾. To address the molecular mechanisms, Marwa Almosaillekh retrovirally expressed a WT NSD1 and a catalytically-inactive NSD1^{N1918Q} SET-domain mutant. With these cells, I performed functional assays, to investigate changes in gene and protein expression upon induced erythroid differentiation (**Appendix 2, Fig. 7**)⁽⁴²⁾. Since the expression of master erythroid regulator GATA1 was not significantly changed, I performed ChIP-sequencing and immunoprecipitations/mass spectrometry (IP/MS), which showed that NSD1 increased GATA1 chromatin binding and changes GATA1 protein interactions (**Appendix 2, Fig. 8**)⁽⁴²⁾.

3.3. Additional experiments; unpublished findings (*not included in the manuscript*)

3.3.1. GATA1 interactome in undifferentiated and 4 days differentiated MEL cells

NSD1 is a rather large protein/cDNA (>8Kb) limiting the efficacy of viral gene transfer and expression. Primary malignant cells can be obtained from diseased *Nsd1*^{-/-} mice and maintained in a so-called “extensively self-renewing erythroblast” (ESRE) – like culture system keeping the cells in a pro-erythroblastic state⁽¹⁵⁴⁾. In contrast, WT BM-derived erythroblasts only have restricted expansion potential (3-4 weeks maximum). In addition, when we grew WT-*Nsd1*^{fl/fl} BM-derived cells under ESRE conditions, we realized that the cultures always contained cells that also expressed myeloid markers, making a comparison with the pure erythroblast culture of diseased *Nsd1*^{-/-} mice problematic.

In order to establish a system that allows us to investigate potential GATA1 interaction partners in murine malignant erythroid cells, we, therefore, took advantage of MEL cells, which are similar to our primary *Nsd1*^{-/-} erythroblasts, as they constitutively express high levels of GATA1 protein. Furthermore, exposure to DMSO induces partial erythroid differentiation, making them an interesting dynamic model to assess GATA1 interaction changes upon induced differentiation⁽¹⁴⁷⁾.

To compare changes in potential GATA1 protein interactions, we exposed MEL cells to 2% DMSO for 4 days to allow erythroid differentiation and prepared nuclear lysates to perform IP/MS analysis with a highly specific anti-GATA1 antibody (**Fig. 11A**). Before digestion with trypsin, a fraction of the nuclear eluates was analyzed by Western Blotting to validate the pull-down efficiency. No GATA1 was immunoprecipitated with the control IgG antibody as the protein was left in the unbound fraction. In the case of the GATA1 antibody, a successful pull down of GATA1 was achieved with only 4% of the eluate (**Fig. 11B**). Two independent nuclear extracts incubated with GATA1 or control antibody were analyzed by shotgun liquid chromatography-MS. Protein abundances were analyzed by “MS1 intensity-based (Label Free) quantitation” and protein fold changes between conditions were statistically analyzed with SafeQuant (in collaboration with Thomas Bock, Proteomics Core Facility, Biozentrum, Basel). In total, 455 proteins were identified by MASCOT search against the UniProt Mus musculus database. Unsupervised clustering showed that duplicate samples, as well as control and GATA1 treated samples, clustered together (**Fig. 11C**). We determined cut-off values to reduce the number of potential candidates and normalized to the control antibody (n=2, p-value < 0.05, log₂ratio > 2). In total, 26 proteins were significantly enriched in undifferentiated MEL cells, whereas 17 proteins were identified in 4 days differentiated MEL cells (**Fig. 11D**).

After 4 days differentiation of MEL cells, the pattern of identified proteins seemed to reflect erythroid maturation with upregulation of hemoglobin subunit alpha (HBA) (log₂ratio of 5.2), beta-2 (HBB2) (log₂ratio of 3.95) and hemogenin (HEMGN) (log₂ratio of 4.93) (**Fig. 11D, E**). In undifferentiated MEL cells, the most significant putative GATA1-interacting proteins were previously characterized to have transcriptional repressive functions such as BAF chromatin remodeling complex subunit BCL11A, MAP3K12 binding inhibitory protein 1 (MBIP), nuclear transcription factor, X-box binding 1 (NFX1), nuclear receptor corepressor 2 (NCOR2) and CBFA2T3 (**Fig. 11D, F**). The top candidates that were significantly downregulated after 4 days (log₂ratio < -2.2) were NCOR2, which reached a 5-fold lower expression, and HDAC3, which was 4-fold down-regulated. Another transcriptional repressor that interacted with GATA1 was CBFA2T3, which was found 3.5-fold less abundant in differentiated cells (**Fig. 11F**). Collectively, our findings show that significant changes in putative GATA1 protein interactions occur upon DMSO-induced differentiation of MEL cells. Potential interactions with repressive proteins are significantly decreased during erythroid differentiation, suggesting that GATA1 might be trapped in co-repressors. Collectively, this data displays a dynamic change of GATA1 protein

and differentiated (right) MEL cells (p-value <0.05; green: log2ratio>2; orange: log2ratio>3; red: log2ratio>4) **E** Dynamic changes of significantly differentially enriched proteins in differentiating (purple) MEL cells (log2ratio>2; p-value<0.05; n=2) and comparing them to their enrichment in undifferentiated/maintenance (blue) MEL cells. **F** Dynamic changes of significantly differentially enriched proteins in undifferentiating (blue) MEL cells (log2ratio>2; p-value<0.05; n=2) and comparison to their enrichment in differentiated (purple) MEL cells.

3.3.2. GATA1 interactome in *Nsd1*^{-/-} erythroblasts retrovirally overexpressing *mGata1*

Our observation that overexpression of a murine *Gata1* ORF (*mGata1*) released the differentiation block of *Nsd1*^{-/-} erythroblasts suggests that inactivation of *Nsd1* may lead to the formation of aberrant GATA1 complexes that are structurally and/or functionally reconstituted by *mGata1* overexpression. To identify changes in protein interactions, we enriched for GATA1 in *Nsd1*^{-/-} cells by IP with and without *mGata1* overexpression 24 hours after induction of differentiation (**Fig. 12A**). This approach was chosen, as we were unable to enrich for sufficient GATA1 from control *Nsd1*^{fl/fl} mouse erythroblasts grown 12-24h in a differentiation-inducing medium (not shown). After 24h differentiation, upregulation of erythroid surface markers CD71 and TER119 was observed, accompanied by a more differentiated morphology (**Fig. 12B**). Furthermore after 24h, *mGata1*-transduced *Nsd1*^{-/-} erythroblasts displayed a red pellet whereas *mock*-transduced cells remained white strongly suggesting differentiation-associated hemoglobin production (**Fig. 12C**). Overexpression of HA-tagged exogenous *mGata1* was visualized by Western blotting. GATA1 could be successfully pulled-down in *mock* as well as *mGata1* transduced cells (**Fig. 12D**). In total, 365 proteins were identified by MASCOT search against the UniProt *Mus musculus* database. Unsupervised clustering showed that lysates immunoprecipitated in maintenance medium cluster separately from samples in differentiation (n=3; log2ratio > 2; p-value < 0.05) (**Fig. 12E**). In *mock*-transduced *Nsd1*^{-/-} erythroblasts, GATA2 and cysteine rich protein 1 (CRIP1) were the most significantly GATA1 interacting proteins. CRIP1 belongs to the LIM/double zinc finger protein family and is an oncogene promoting proliferation, migration, and invasion in a variety of cancers, such as breast, ovarian, colorectal, and thyroid carcinomas ⁽⁵⁾. GATA2 is a well-described transcription factor in normal and malignant hematopoiesis and is important for the proliferation and maintenance of progenitor cells. Ribosomal protein S28 (RPS28) was significantly pulled down with GATA1 in *mGata1*-transduced differentiating erythroblasts. Interestingly, RPS28 was found to be mutated in Diamond Blackfan anemia and results in reduced translation of *Gata1* mRNA and erythroid defects, which can be overcome by increasing GATA1 protein levels ⁽⁶⁾. Solute Carrier Family 4 Member 1 (SLC4A1) was another protein highly enriched in *mGata1*-expressing *Nsd1*^{-/-} erythroblasts. It is expressed in particular at the plasma membrane from basophilic erythroblasts and associated with erythroid differentiation (**Fig. 12F**). Collectively, this data suggests a shift of GATA1 protein interactions from a repressive towards a more activated cell state upon exogenous *mGata1*-induced *Nsd1*^{-/-} cell differentiation.

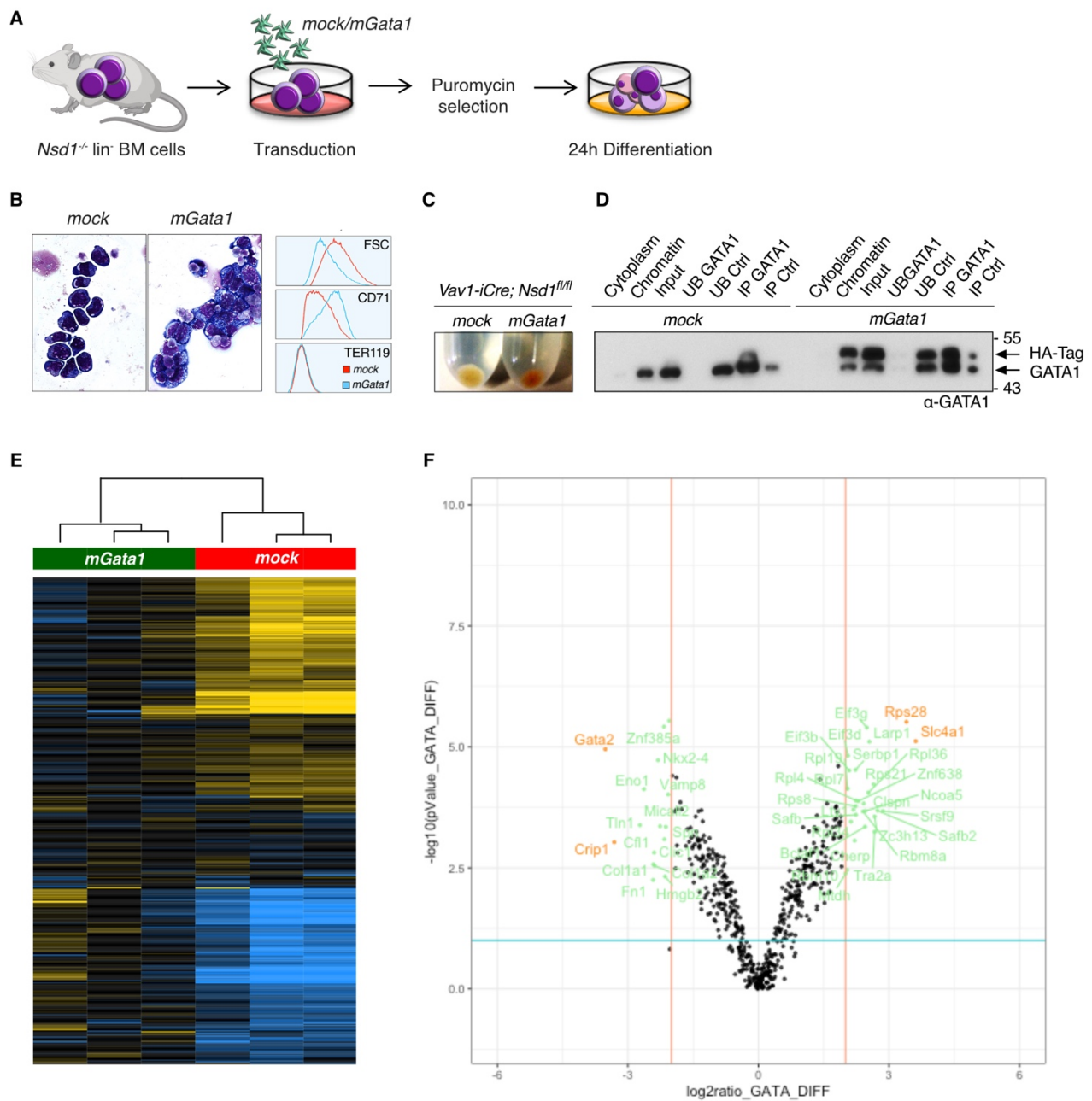


Figure 12. Dynamic changes of GATA1 protein interactions during GATA1-induced *Nsd1*^{-/-} cell differentiation. **A)** Experimental setup: lineage-depleted *Nsd1*^{-/-} erythroblasts are transduced in maintenance medium with *mock* or *mGata1* construct and selected 48h in puromycin before plating cells in differentiation medium. **B)** *mGata1*-transduced cells show more differentiated morphology accompanied by decreased cell size (FSC) and increased CD71 and TER119 erythroid surface marker expression. **C)** *mGata1*-transduced cells show hemoglobinization of the pellet. **D)** Confirmation of successful enrichment shown by Western Blotting in *mock*- and *mGata1*-transduced *Nsd1*^{-/-} erythroblasts. **E)** Unsupervised clustering shows clustering of *mock* and *mGata1* conditions (3 independent replicates per group). **F)** Volcano plot identifies differentially enriched proteins between undifferentiated/*mock* (left) and differentiated/*mGata1* (right) *Nsd1*^{-/-} erythroblasts (p-value <0.05; green: log₂ratio>2; orange: log₂ratio>3).

3.3.3. The EZH2 small molecule inhibitor GSK126 depletes H3K27me3, but does not induce differentiation of *Nsd1*^{-/-} erythroblasts

NSD1 is a histone methyltransferase that mono- and di-methylates H3K36me at intergenic regions allowing the recruitment of DNMT3A. This facilitates the addition of H3K36me3 by the SET domain containing 2, histone lysine methyltransferase (SETD2/ KMT3A) allowing recruitment of DNMT3B to active gene bodies⁽⁴³⁾. Therefore, ablation of NSD1 should result in changes in histone marks, which would allow the spreading of PRC2-mediated H3K27me3 marks at intergenic DNA and redistribution of DNMT3A-mediated DNA methylation to active gene bodies (**Fig. 9**). EZH2 is the enzymatic powerhouse of the PRC2 complex and essential for repressing transcription by setting H3K27me3 marks. To test whether the repressive chromatin state could be overcome and erythroid differentiation could be induced by removing repressive PRC2-set H3K27me3 marks, we treated *Nsd1*^{-/-} cells with GSK126, a previously characterized rather selective small molecule EZH2 inhibitor⁽¹⁵⁵⁾. First, we re-assessed global histone changes of *Nsd1*^{-/-} vs WT fetal liver cells in maintenance medium, which however did not show any changes in H3K27me3 and H3K36me3 but a clear reduction in H3K36me1, confirming that Nsd1 is essential to mono-methylate H3K36 whereas trimethylation can also be achieved through other HKMTs (**Fig. 13A**). *Nsd1*^{-/-} erythroblasts cultured in differentiation medium were treated with increasing concentrations of GSK126 for 48h and 72h. No changes in H3K36me2 protein expression, but a reduction of H3K27me3 protein expression was observed with increasing doses of GSK126, confirming the specificity of the inhibitor (**Fig. 13B**). Notably, reducing H3K27me3 did not reduce cell growth nor did it show any effects on erythroid differentiation after 48h of GSK126 treatment (**Fig. 13C, D**). This suggests that removing repressive chromatin marks is insufficient to restore erythroid differentiation when NSD1 is absent.

3.3.4. DZNep treatment reduces cell growth and induces erythroid differentiation in *Nsd1*^{-/-} erythroblasts

It has been shown that 3-Deazaneplanocin A (DZNep), an S-adenosylhomocysteine hydrolase inhibitor, targets EZH2 for degradation and induces apoptosis in various hematopoietic cell lines⁽¹⁵⁶⁾. Fujiwara *et al.* observed that treating the K562 AEL cell line and primary human CD34⁺ HSPCs with DZNep resulted in reduced cell growth and induced partial erythroid differentiation⁽¹⁵⁷⁾. Therefore, I also assessed whether DZNep may also induce terminal differentiation of *Nsd1*^{-/-} erythroblasts. Cells treated with increasing doses of DZNep for 48h, showed an exponential decrease in cell growth and reduced proliferation visible by a more reddish color of the medium (**Fig. 13E, F**). Furthermore, 0.1 and 0.2 μ M DZNep resulted in a shift towards a higher CD71⁺ cell population and increased TER119 expression compared to 0.2% DMSO treated cells (**Fig. 13G, H**). This suggests that DZNep is able to overcome the erythroid differentiation block of *Nsd1*^{-/-} erythroblasts. However, based on Fujiwara

et al. this activity may not solely depend on EZH2 inhibition but also on the reduction of the corepressor ETO2 (CBFA2T3) protein levels upon DZNep treatment. Therefore, the broader effects on other yet-to-be-determined PRC2-dependent and independent proteins remain to be elucidated.

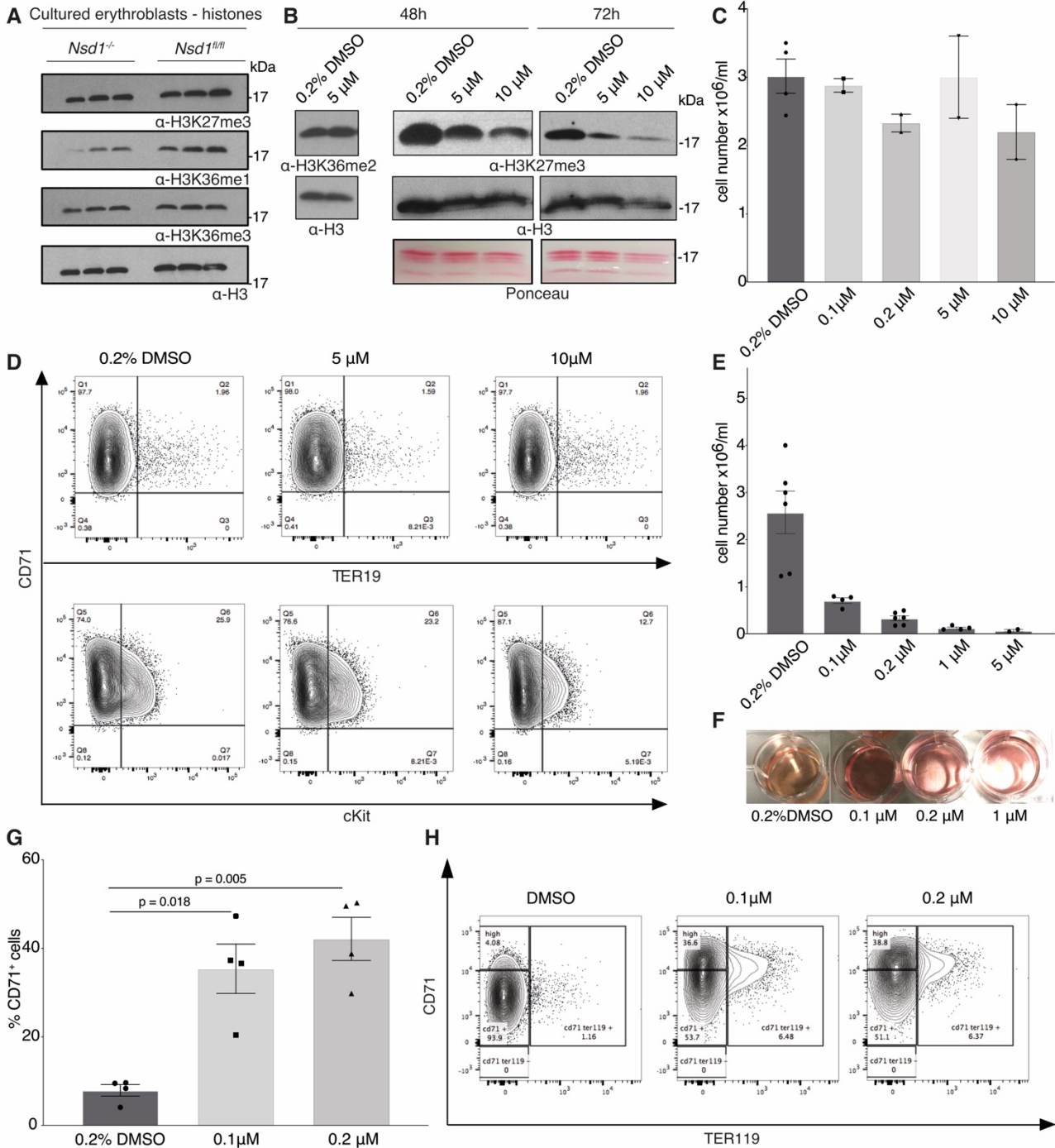


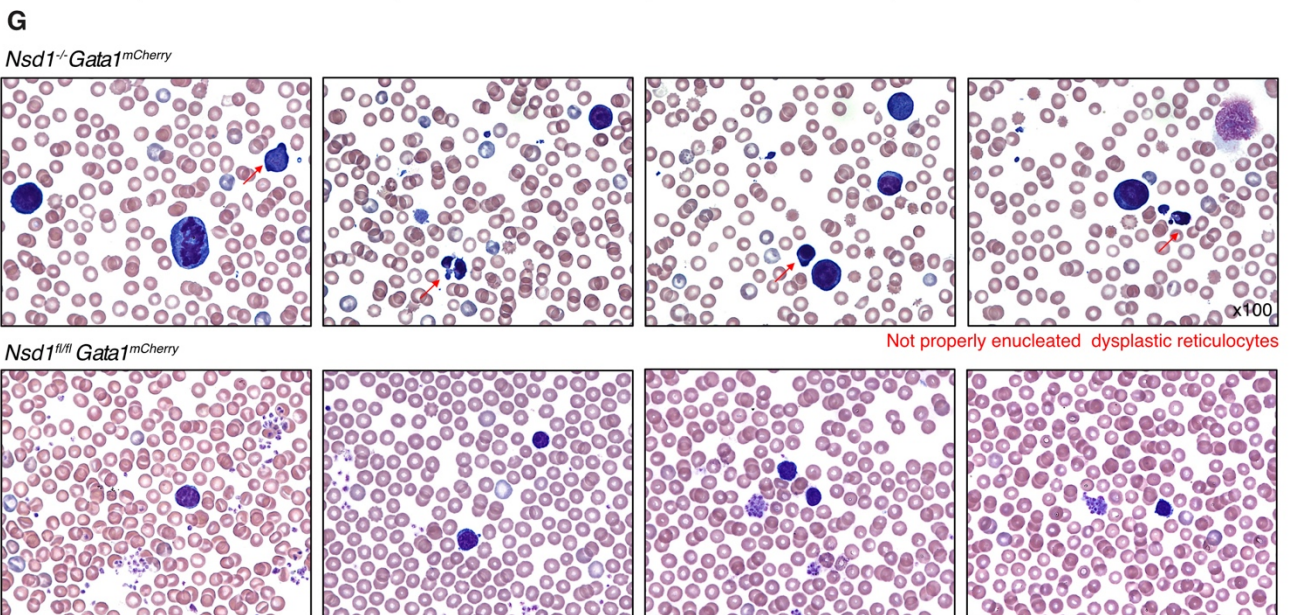
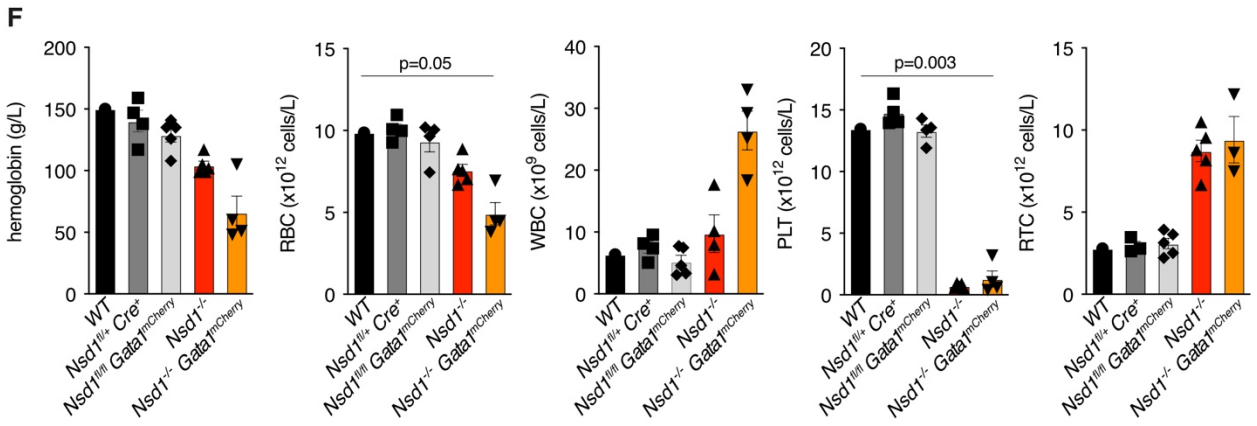
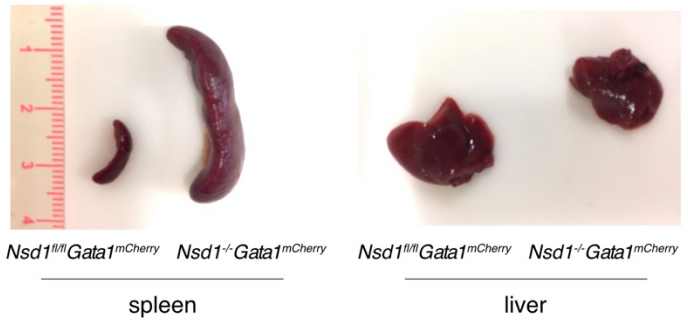
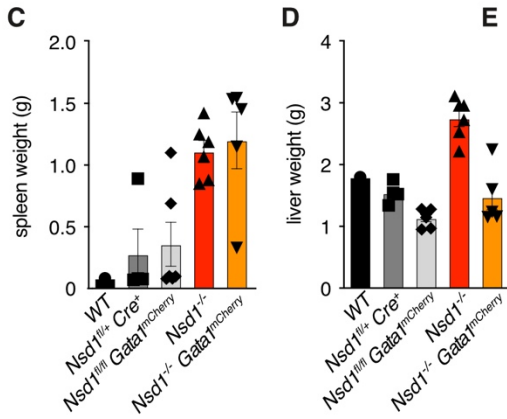
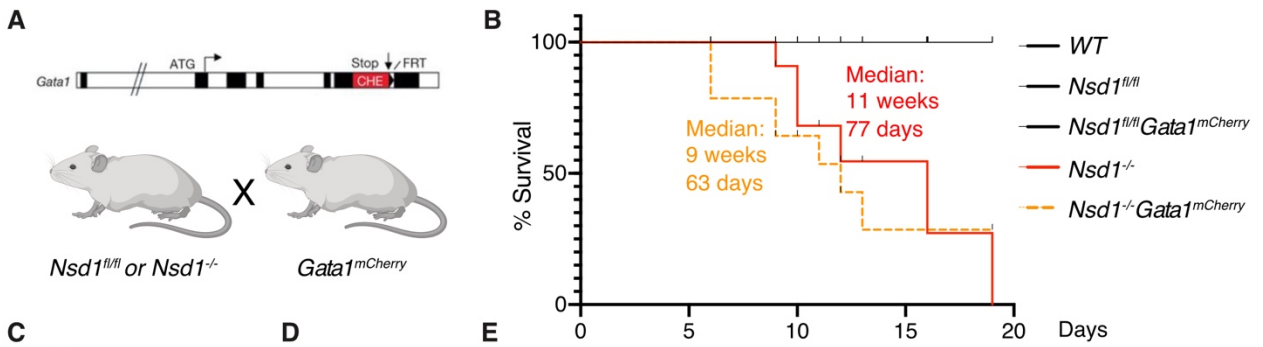
Figure 13. Effects of GSK126 and DZNep on differentiation in *Nsd1*^{-/-} erythroblasts. A) Western blot analysis of histone extracts obtained from fetal liver cultured erythroblasts in maintenance medium. Blots are probed with antibodies against H3K27me3, H3K36 mono(me1), and tri(me3) methylation. Histone H3 is used as a loading control. **B)** *Nsd1*^{-/-} erythroblasts cultured in differentiation medium are treated with 5 or 10 μM GSK126 for 48h and 72h. 0.2% DMSO is used as treatment control. Histone extracts are subjected to immunoblotting for analysis of H3K36me2 and H3K27me3. H3 and Ponceau staining are used as loading

controls. **C)** Cell number of *Nsd1*^{-/-} erythroblasts cultured in differentiation medium and treated with increasing concentrations of GSK126 for 48h. **D)** Representative flow cytometry panel of CD71, cKit, and TER119 stained populations obtained after 48h GSK126 treatment of *Nsd1*^{-/-} erythroblasts (n=2 per group). **E)** Cell number of *Nsd1*^{-/-} erythroblasts cultured in differentiation medium and treated with increasing concentrations of DZNep for 48h. **F)** Representative image of cell proliferation of cells treated for 48h with increasing concentrations of DZNep. **G)** Percentage (%) of CD71⁺ cells upon 48h DZNep treatment. **H)** Representative flow cytometry panel of CD71, and TER119 stained populations obtained after 48h DZNep treatment of *Nsd1*^{-/-} erythroblasts (n=4 per group).

3.3.5. Tracing GATA1 expression in *Nsd1*^{-/-} mice

Our work indicated that although GATA1 is abundantly expressed in *Nsd1*^{-/-} erythroblasts, its function is impaired. In order to have a model to sort out only GATA1 expressing erythroblasts and compare the differences of GATA1 between *Nsd1*^{fl/fl} and *Nsd1*^{-/-} mice, we crossed them with a transgenic reporter mouse strain in which *Gata1* was directly fused to a *mCherry* fluorochrome (kindly provided by Timm Schröder, ETH, Basel, Switzerland) ⁽¹⁵⁸⁾. In fact, the *mCherry* was fused to the C-terminus of GATA1, allowing live quantification of the intensity by flow cytometry and imaging of endogenous expression levels of GATA1 in hematopoietic cells (**Fig. 14A**).

Surprisingly, *Nsd1*^{-/-} *Gata1*^{mCherry} mice seem to develop the AEL-like disease faster than *Nsd1*^{-/-} mice reflected by a median survival of 9 weeks (63 days) compared to 11 weeks (77 days) (**Fig. 14B**). Two mice died after 6 weeks without any symptomatic signs of disease. The others developed severe splenomegaly but strikingly no hepatomegaly was ever observed in diseased *Nsd1*^{-/-} mice (**Fig. 14C, D, E**). Symptomatic *Nsd1*^{-/-} *Gata1*^{mCherry} mice had clearly reduced red blood cells (RBC), platelets (PLT), and hemoglobin and clearly increased “immature” reticulocytes (RTC) similar to *Nsd1*^{-/-} mice. However, *Nsd1*^{-/-} *Gata1*^{mCherry} mice also had increased white blood cells (WBC) that were never observed in *Nsd1*^{-/-} mice (**Fig. 14F**). On blood smears we not only find larger cells with a proerythroblast-like morphology but also the appearance of dysplastic “mega” reticulocytes, which seem to be unable to properly enucleate (**Fig. 14G**). Flow cytometric analysis of BM (**Fig. 14H, I**) and spleen (**Fig. 14J, K**) from *Nsd1*^{-/-} *Gata1*^{mCherry} mice revealed a significant increase in the number of cells expressing CD71⁺/cKit⁺ and CD71⁺/TER119⁺ erythroid markers, which was higher compared to *Nsd1*^{-/-} mice. Conclusively, *Nsd1*^{-/-} *Gata1*^{mCherry} mice display a more severe and different phenotype, visualized by increased WBC counts and no hepatomegaly which has not been observed in *Nsd1*^{-/-} mice. We hypothesize that the C-terminal fusion of *mCherry* somehow further impairs the function of GATA1, which becomes evident, particularly in absence of *Nsd1*. Although the erythroid lineage seems not affected in *Gata1*^{mCherry} mice under steady-state conditions, the situation seems to change upon stress, like ablation of *Nsd1* ⁽¹⁵⁸⁾. Therefore, these experiments were discontinued.



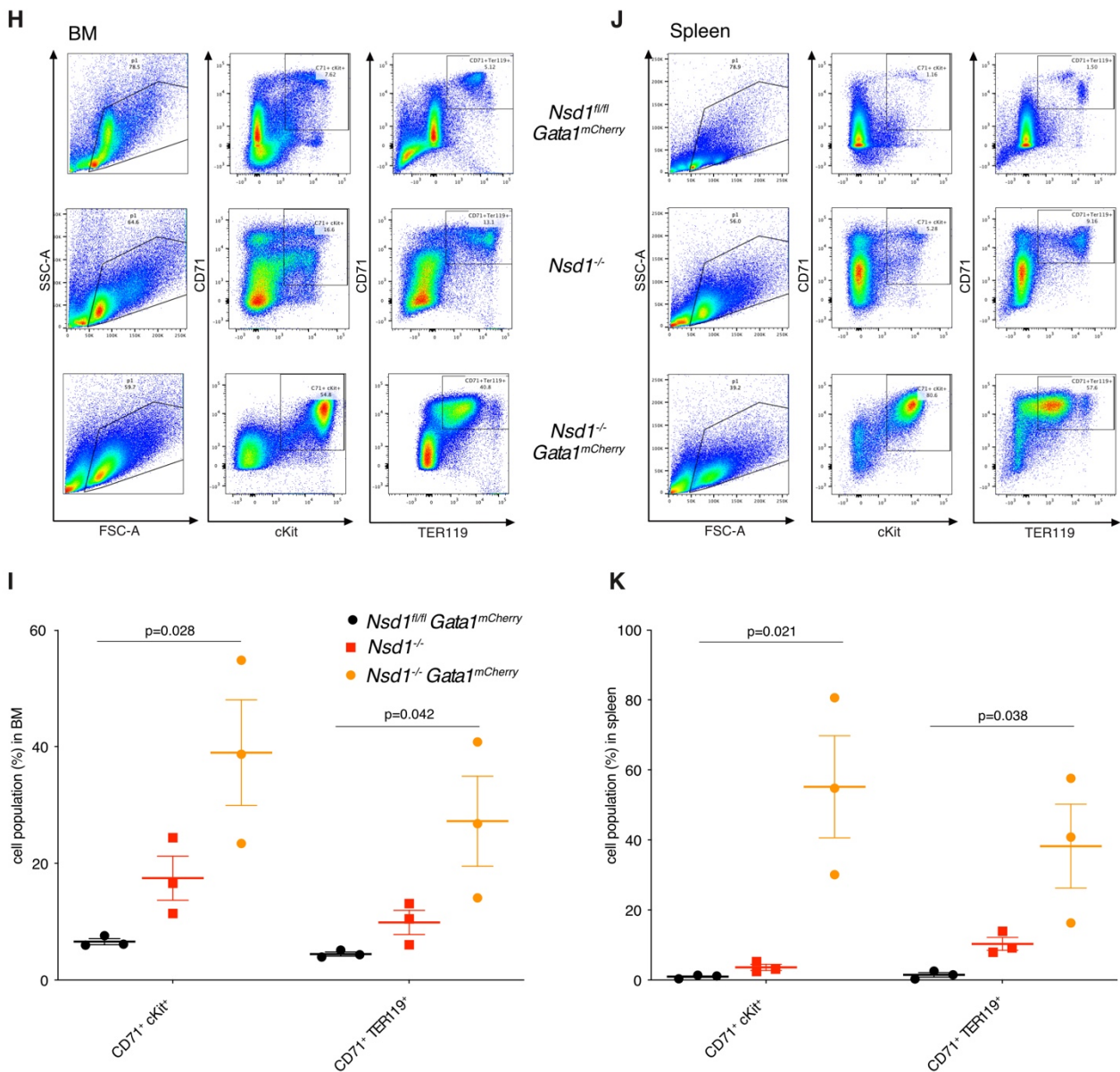


Figure 14. The presence of a *Gata1^{mCherry}* reporter alters the phenotype of *Nsd1^{-/-}* mice. **A) Endogenous gene loci after knock-in taken from ⁽¹¹⁾. Mice extracted from Biorender.com. Exons are displayed in black boxes. *Nsd1^{fl/fl}* or *Nsd1^{-/-}* mice are crossed with knock-in mice for *Gata1^{mCherry}*. **B)** Kaplan Meier plot of disease-free survival of *WT* (n=1, black line), *Nsd1^{fl/fl}* (n=4, black line) *Nsd1^{fl/fl} Gata1^{mCherry}* (n=6, black line), *Nsd1^{-/-}* (n=6, red line) and *Nsd1^{-/-} Gata1^{mCherry}* (n=5, orange line) mice. Median survival of *Nsd1^{-/-}* mice is 77 days and for *Nsd1^{-/-} Gata1^{mCherry}* 63 days. *WT*, *Nsd1^{fl/fl}*, and *Nsd1^{fl/fl} Gata1^{mCherry}* mice never develop any disease. **C)** Spleen and **D)** liver weight of *WT* (n=1, black), *Nsd1^{fl/fl}* (n=4, dark grey) *Nsd1^{fl/fl} Gata1^{mCherry}* (n=6, light grey), *Nsd1^{-/-}* (n=6, red) and *Nsd1^{-/-} Gata1^{mCherry}* (n=5, orange). **E)** Representative image of spleen and livers of symptomatic *Nsd1^{-/-} Gata1^{mCherry}* mice in comparison to *Nsd1^{fl/fl} Gata1^{mCherry}* mice. **F)** Peripheral blood counts measured at terminal workup. RBC: red blood cells; WBC: white blood cells; PL: platelets; RTC: reticulocytes. **G)** Representative images of Wright Giemsa-stained peripheral blood smears of symptomatic *Nsd1^{-/-} Gata1^{mCherry}* mice (upper panels) in comparison to *Nsd1^{fl/fl} Gata1^{mCherry}* mice (lower panels). Not properly enucleated dysplastic reticulocytes are shown with red arrows. **H)** and **J)** Representative flow cytometry panels of SSC-A/FSC-A population (left) CD71/cKit population (middle) and CD71/TER119 population (right) in BM (**H**) and spleen (**J**) of *Nsd1^{fl/fl} Gata1^{mCherry}*, *Nsd1^{-/-}* and *Nsd1^{-/-} Gata1^{mCherry}* mice. **I)** and **K)** Statistical evaluation of CD71⁺/cKit⁺ and CD71⁺/TER119⁺ population from BM (**I**) and spleen (**K**) of *Nsd1^{fl/fl} Gata1^{mCherry}*, *Nsd1^{-/-}* and *Nsd1^{-/-} Gata1^{mCherry}* mice (n=3 per group).**

Chapter 4

NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer

Tauchmann, S.; Schwaller, J. NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer. Life **2021**, *11*, 877. <https://doi.org/10.3390/life11090877>

4.1. Rationale

Increasing evidence indicates that NSD1 plays a central role in several human diseases. In particular, a variety of somatically acquired *gain-* and *loss-of-function* NSD1 genetic alterations are recurrently found in human cancers. Therefore, I wrote a review of the current knowledge about the biochemistry, cellular function, and role of NSD1 in human diseases (**Appendix 3**)⁽⁴³⁾.

Notably, one of these NSD1 *loss-of-function* mutation-induced diseases is the SOTOS hypergrowth syndrome, which is characterized by overgrowth, distinct craniofacial appearance, intellectual disabilities, and an increased risk to develop cancer^(159–161).

Interestingly, overexpression of *NSD* (fly homolog to NSD1) in *Drosophila melanogaster* resulted in a reduction in body size at the larval stage and atrophic wings⁽¹⁶²⁾. Deletion of *NSD* in *Drosophila* showed an increase in the body size of larvae and neurological disabilities, such as lower memory performance and motor defects, phenocopying many aspects of the overgrowth phenotype seen in SOTOS patients⁽¹⁶³⁾. Strikingly, during the characterization of the erythroleukemic phenotype of *Nsd1*^{-/-} mice we realized that the erythroblasts increased in cell size compared to control erythroblasts (**Appendix 2, Fig. 4B**)⁽⁴²⁾. So far, the mechanisms of how the loss or gain of *Nsd1* contributes to cell size changes remain unclear. Based on previous findings, we hypothesized that *Nsd1* ablation may play an important role in controlling cell size as well as cell cycle genes and thereby contribute to hypergrowth as well as tumorigenesis.

4.2. Experimental approach

To experimentally address this hypothesis, we created a novel mouse model by crossing *Nsd1^{fl/fl}* mice with an *UBC-Cre-ERT²* mouse line (obtained from Tobias Derfuss DBM, Basel). Through induction with 4-hydroxytamoxifen (TAM; H7904, Sigma) the Cre-ER^{T2} fusion protein traverses the nucleopore and cuts the *loxP* sites resulting in targeted deletion of the floxed *Nsd1* (**Fig. 15A**). This allows to dynamically follow the impact of *Nsd1* inactivation on cell size.

4.3. Results

We first isolated lineage marker-depleted BM cells from adult *Nsd1^{fl/fl};Cre-ERT²* or control *Nsd1^{fl/+};Cre-ERT²* mice and immediately induced with 1 μM and 5 μM TAM in maintenance medium, which resulted in complete cleavage of both *Nsd1* alleles in TAM-treated *Nsd1^{fl/fl};Cre-ERT²* and one allele of *Nsd1^{fl/+};Cre-ERT²* erythroblasts after 4 days (**Fig. 15B**). Prolonged culture in maintenance medium of uninduced *Nsd1^{fl/fl};Cre-ERT²* and *Nsd1^{fl/+};Cre-ERT²* for 11 days resulted in the expansion of proerythroblasts, characterized by CD71⁺/TER119⁻ surface marker expression. Furthermore, 5 μM TAM-treated *Nsd1^{fl/fl};Cre-ERT²* cells showed a decrease in the myeloid Gr1⁺/Mac1⁺ population. (**Fig. 15C**). 2 days of differentiation resulted in increased TER119 and reduced Kit expression of untreated erythroblasts. Cleavage of *Nsd1* impaired erythroid differentiation displayed by CD71⁺/Kit⁺/TER119⁻ surface marker expression after 2 days in differentiation (**Fig. 15D**). These findings confirm previous findings in the *Nsd1^{fl/fl};Vav-iCre (Nsd1^{-/-})* mouse model that loss of *Nsd1* favors outgrowth of immature erythroblasts and impairs erythroid differentiation. Moreover, 16-day old TAM treated cells showed a slight increase in cell size illustrated by a shift towards a higher FSC-A population (**Fig. 15E**). Along with this, *Nsd1* expression was associated with more resting cells in the G₀ phase, whereas loss of *Nsd1* resulted in a reduction of G₀ phase and accumulation of cells in the G₁ phase of the cell cycle (not statistically significant) (**Fig. 15F, G**). Collectively, this data suggests that loss of *Nsd1* correlates with increased cell growth and differences in cell cycle. However, the exact mechanisms remain to be elucidated. Generally, this model will serve as a platform for in-depth integrative molecular analysis addressing gene expression, DNA methylation, and chromatin accessibility upon induced *Nsd1* ablation.

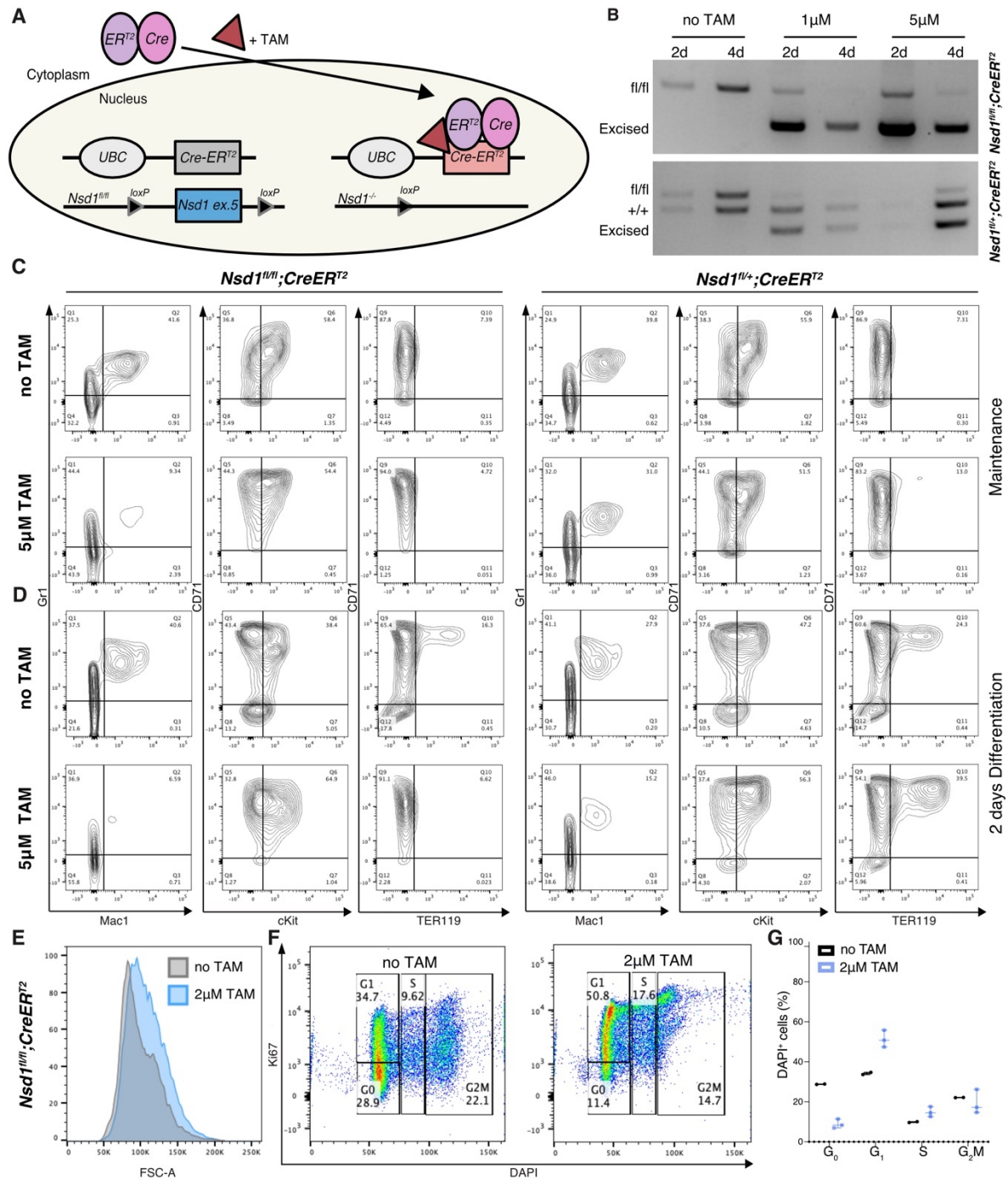


Figure 15. Ablation of *Nsd1* in vitro reduces myeloid cells, blocks erythroid differentiation, and decreases the number of quiescent cells. **A)** Schematic representation of the inducible *Nsd1*^{fl/fl};CreER^{T2} ablation system. 4-hydroxytamoxifen (TAM) results in a conformational change of the CreER^{T2} fusion protein that allows it to migrate into the nucleus and cut the loxP sites flanking exon 5 of *Nsd1*. **B)** Representative image of *Nsd1* cleavage of *Nsd1*^{fl/fl};CreER^{T2} (n=3) and *Nsd1*^{fl/+};CreER^{T2} (n=2) erythroblasts cultures exposed 2 and 4 days to 1 μM and 5 μM TAM. **C)** and **D)** Representative flow cytometry analysis of *Nsd1*^{fl/fl};CreER^{T2} (n=2) and *Nsd1*^{fl/+};CreER^{T2} (n=1) erythroblasts cultured **C)** 11 days in maintenance medium or **D)** 2 days in differentiation medium without (no TAM) or with (5 μM) TAM. **E)** FSC-A profile of *Nsd1*^{fl/fl};CreER^{T2} (n=2) erythroblasts cultured 16 days in maintenance medium without or with 2 μM TAM. **F)** Representative flow cytometry analysis of cell cycle populations (G₀, G₁, S, G₂M) of *Nsd1*^{fl/fl};CreER^{T2} erythroblast cultured 16 days in maintenance medium without (n=2) or with 2 μM (n=3) TAM. **G)** Quantification of the cell cycle populations of DAPI⁺ cells of *Nsd1*^{fl/fl};CreER^{T2} erythroblast cultured 16 days in maintenance medium without (n=2) or with 2 μM (n=3).

Chapter 5

Dissecting GATA1 protein interactions involved in impaired differentiation of malignant erythroblasts

Running title: Dissecting GATA1 complexes in normal and malignant human erythroblasts

Samantha Tauchmann^{1,2}, Frederik Otzen Bagger^{1,2,3}, Thomas Bock^{2,4}, Rathick Sivalingam^{1,2}, Thomas Eder⁵, Elizabeth Heyes⁵, Alexandre Fagnan⁶, Marieke van Lindern⁷, Thomas Mercher⁶, Florian Grebien⁵, Juerg Schwaller^{1,2}

University Children's Hospital Basel¹ & Department of Biomedicine, University of Basel, Basel Switzerland²; University of Copenhagen, Center for Genomic Medicine, Copenhagen University Hospital, Copenhagen, Denmark³; Proteomics Core Facility, Biozentrum, University of Basel, Basel, Switzerland⁴; University of Veterinary Medicine, Institute of Medical Biochemistry, Vienna, Austria⁵; INSERM U1170, Institut Gustave Roussy, Université Paris-Saclay, Université de Paris, Villejuif, France ⁶; Sanquin Research, Academic Medical Center, University of Amsterdam, the Netherlands⁷

Manuscript in preparation

5.1. Introduction

Erythropoiesis is a multi-step process starting from hematopoietic stem cells (HSC) differentiating into committed megakaryocytic-erythroid progenitors (MEP), burst-forming unit-erythroid (BFU-E) and colony-forming unit-erythroid (CFU-E) cells ⁽¹³⁾. The maturation of erythroid progenitors is ushered by the first morphologically identifiable proerythroblasts that terminally differentiate ultimately generating reticulocytes and mature erythrocytes ^(164,165). Erythropoiesis is controlled by several transcription factors (TFs) that drive distinct regulatory networks by recruiting coactivators or corepressors to modulate gene expression ^(26,27,53,70). So far, these networks were mostly studied by transcriptomics but more recent work revealed limited correlations between mRNA and the corresponding protein levels in erythroid cells ⁽¹⁶⁶⁾. This particularly accounts for erythroid transcriptional master regulators such as GATA1, KLF1, TAL1, LMO2, LDB1, and GFI-1B, which are known to be functionally regulated by multiple posttranscriptional mechanisms and protein-protein interactions ^(69,88). Notably, quantitative mass spectrometry studies found the lowest correlation between mRNA and protein levels in the early stages of erythropoiesis. Strikingly, these proteomics studies unraveled a marked disbalance between highly abundant transcriptional co-repressors facing much lower amounts of transcriptional co-activator proteins ^(69,167).

Malignancies of the erythroid lineage are relatively rare. First described in 1917 by Cesare Di Guglielmo, acute erythroleukemia (AEL) accounts for 1-5% of acute myeloid leukemia (AML) and is generally associated with a poor outcome ^(124,127). The cellular hallmark of AEL is impaired terminal differentiation and uncontrolled expansion of erythroid progenitor cells. The highly variable clinical presentation of AEL resulted in continuous discussions about its classification. Some AEL patients present with a heterogenous mixture of myeloid and erythroid features, while some less frequent cases present with >80% of erythroid progenitor cells considered as pure erythroid leukemia (PEL). AEL occurs either *de novo* or develops secondary from other hematological malignancies such as myelodysplastic syndromes (MDS), myeloproliferative neoplasms (MPN), or therapeutic exposure to genotoxic agents suggesting that the disease reflects a continuum of MDS and AML with erythroid hyperplasia ⁽¹⁶⁸⁾. Even though no single specific and highly prevalent driver mutation has been associated with AEL, deep sequencing studies allowed to define the genetic landscape of AEL. Recurrent mutations revealed at least three genetic subgroups including patients with *TP53* alterations, those with mutations in epigenetic regulators like *DNMT3A* and *TET2*, and cases with a generally very low mutational burden ^(140,141,169–172). Many AEL transcriptomes are characterized by aberrant expression of transcriptional (co)-regulators such as *SKI*, *ERG*, and *CBFA2T3*, that upon ectopic expression, impair differentiation of erythroid progenitor cells associated with functional interference with the master transcription factor GATA1 ⁽¹⁴⁰⁾. Altered chromatin modifications can also result in impaired erythroid differentiation as shown upon genetic inactivation of the nuclear receptor-

binding SET domain protein 1 (NSD1) histone methyltransferase leading to reduced H3K36 methylation, impaired binding of GATA1 to chromatin and an AEL-like phenotype in mice ⁽⁴²⁾. Similarly, transgenic overexpression of a histone 3 lysine 36 to methionine onco-histone mutation impaired erythroid differentiation and induced a very similar AEL-like phenotype in mice ⁽¹⁷³⁾.

5.2. Rationale

These observations suggest that altered protein and/or chromatin interactions may interfere with proper GATA1 function and consequently impair differentiation of erythroid progenitor cells ⁽¹⁷⁴⁾. Therefore, we hypothesize that impaired terminal erythroid differentiation in AEL might be the consequence of aberrant yet unknown GATA1 protein interactions. The aim of this project is to identify novel regulators that can be targeted to overcome the differentiation block in AEL.

5.3. Experimental approach

To explore this possibility, we compared the proteomes of three human AEL cell lines (F36P, K562, KMOE2) and primary AEL patient cells with HUDEP2 cells and primary human erythroblasts (hEBST) that both are capable of *in vitro* terminal erythroid differentiation, using a tandem mass tag (TMT)-based approach. In parallel, we investigated the GATA1 interactomes by immunoprecipitation (IP) followed by mass spectrometry (MS). Significantly associated and/or expressed proteins were functionally explored in a targeted CRISPR/Cas9 screen to assess genes responsible for impaired erythroid differentiation in K562 AEL cells.

5.4. Results

5.4.1. Proteome analysis separates human AEL cells from human erythroblasts with full differentiation potential

Recent studies indicated that quantitative protein changes may determine cell fate choices, and activate or repress transcriptional programs critical for erythroid differentiation^(175,176). As to date no study explicitly compared the proteomes of normal- and transformed erythroid cells, we first assessed quantitative proteomic differences in human erythroleukemic cells compared to human erythroblasts that terminally differentiate in permissive medium. For the latter, we established primary human erythroblasts (hEBST) from peripheral blood mononuclear cells from healthy donors⁽¹⁷⁷⁾. We also analyzed HUDEP2 cells that are derived from human CD34⁺ umbilical cord blood cells after immortalization with the human papillomavirus type 16-E6/E7 oncoprotein⁽¹⁷⁸⁾. We compared terminal erythroid differentiation of hEBST and HUDEP2 cells in permissive medium by assessing CD71, CD235a erythroid markers, and DRAQ5 (infrared fluorescent DNA dye) staining to determine their enucleation potential. Despite some kinetic differences, both cell types showed almost complete erythroid differentiation and we, therefore, designated them as normal erythroblasts (**Fig. 16**).

As models of fully transformed erythroblasts, we chose 3 established human cell lines: K562 (BCR-ABL⁺, erythroid blast crisis from chronic myeloid leukemia, CML), F-36P (AEL from MDS), KMOE-2 (pediatric PEL) and primary blasts from an AEL patient (LAM49) expanded in immunodeficient NSG mice^(140,149). Phenotypically, all cell lines expressed > 99% CD71⁺, whereas CD235a was the highest expressed in K562 and F36P and slightly less in HUDEP2 and hEBSTs, whereas in KMOE2 cells it was entirely absent (**Suppl. Fig. 1**). These AEL cells did not show any significant signs of terminal maturation when grown in permissive medium that induced terminal differentiation of hEBST or HUDEP2 cells (**data not shown**).

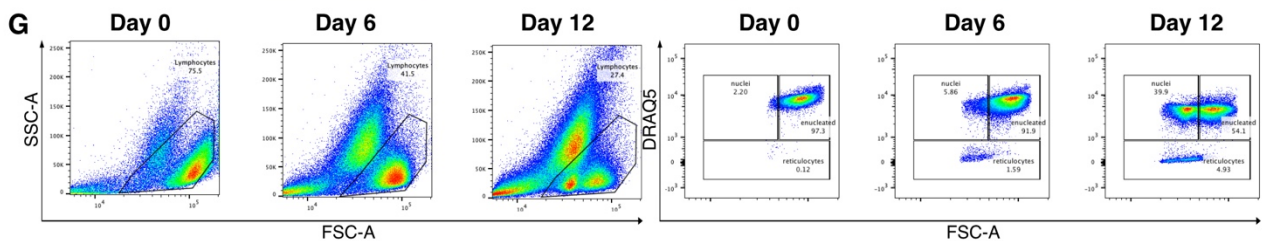
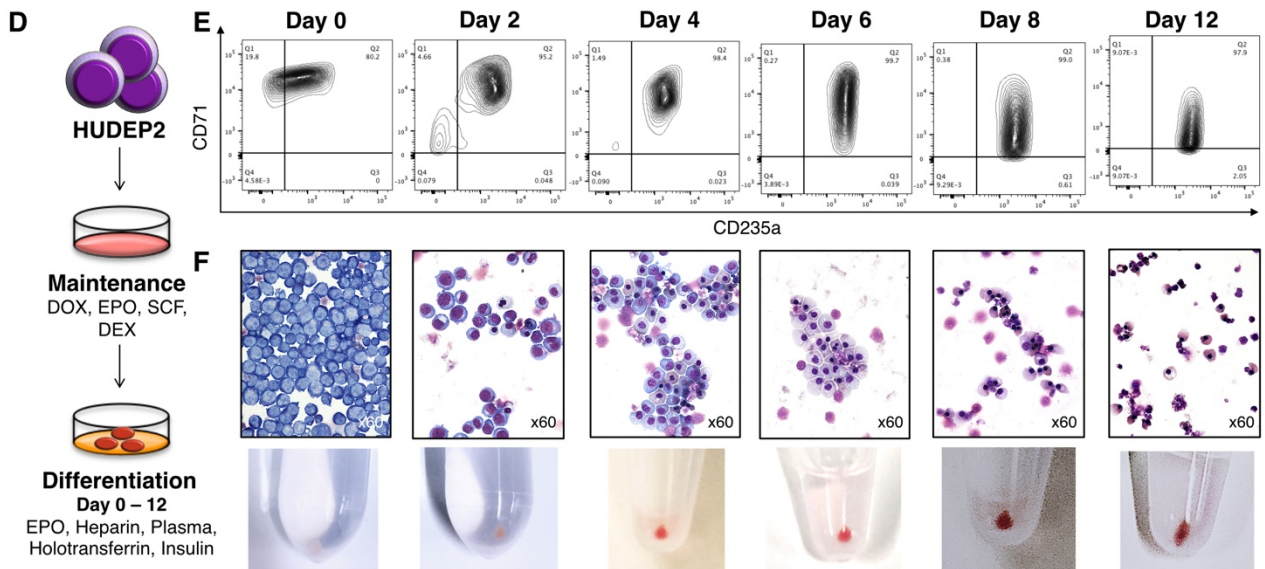
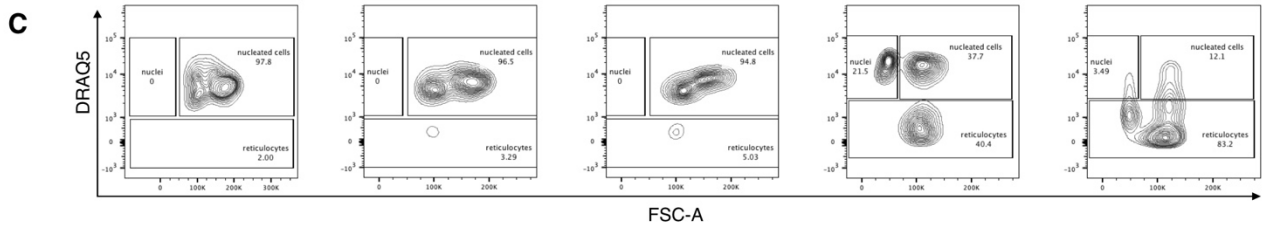
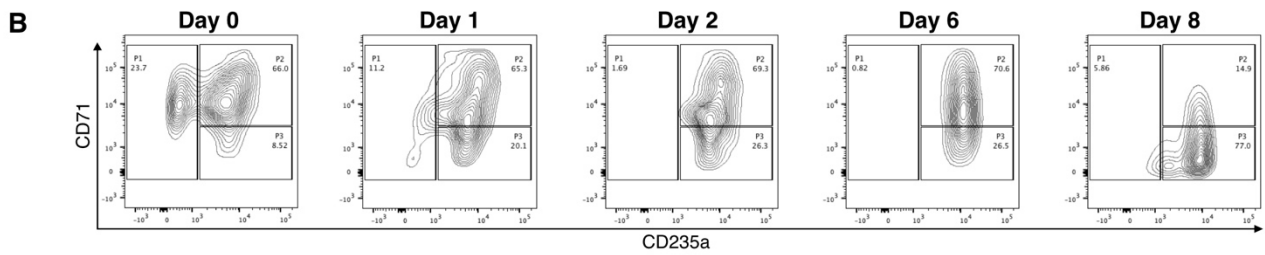
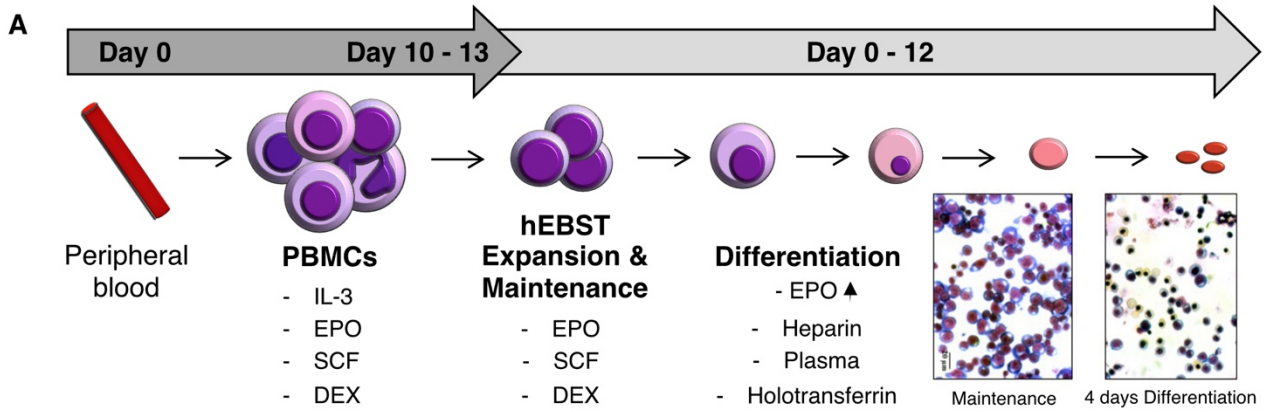


Figure 16. Expansion and differentiation of primary erythroblasts (hEBSTs) and HUDEP2 cells. A) 2-step culture system to establish primary human erythroblasts (hEBSTs). The erythroblast culture is established after 10-13 days by the expansion of peripheral blood mononuclear cells (PBMCs) from healthy donors in permissive medium. IL-3 is added to support stem cells and obtain a higher yield of hEBSTs. Maintenance and expansion depend on EPO, SCF, and dexamethasone (DEX), while differentiation relies on the removal of DEX and SCF and an increase of EPO, in addition to holotransferrin, plasma, and heparin. **B)** Representative images of flow cytometry analysis of CD71 and CD235a expression of hEBSTs after 0,1,2,6 and 8 days of differentiation. **C)** Representative images of flow cytometry analysis of CD71 and CD235a and of enucleation rate assessed by DRAQ5 staining in hEBSTs after 0,1,2,6 and 8 days of differentiation. **D)** 2-step culture system of the immortalized erythroid progenitor cell line HUDEP2. **E)** Representative images of flow cytometry analysis of CD71 and CD235a in HUDEP2 cells after 0,2,4,6,8 and 12 days of differentiation. **F)** Representative images of HE-stained cytosots and cell pellets obtained from HUDEP2 cells after 0, 2, 4, 6, 8, and 12 days of differentiation. **G)** Representative images of flow cytometry analysis of SSC-A/FSC-A (left) and DRAQ5/FSC-A (right) in HUDEP2 cells after 0, 6, and 12 days of differentiation.

To quantify steady-state protein expression, we isolated cellular nuclei followed by trypsin/Lys-C digestion, performed tandem mass tag (TMT) labeling, and analyzed the pooled peptide mixes by LC-MS/MS (**Fig. 17A**). Overall, we detected 6774 commonly expressed proteins, 370 proteins appeared more abundant in AEL cells, while 291 proteins were more abundant in the control erythroblasts ($\log_{2}FC > 1$; $q < 0.05$) (**Suppl. Table 1**). Unsupervised hierarchical cluster analysis displayed that the samples formed two branches separating the AEL cell lines from normal erythroblasts with the patient-derived AEL cells being mostly different from hEBSTs (**Fig. 17B**). Principle component analysis (PCA) showed that all AEL cells clustered together, whereas HUDEP2 and hEBSTs grown in maintenance medium as well as 2 and 5 days in differentiation-permissive medium clustered further apart (**Fig. 17C**). 38 proteins were significantly more abundant in leukemic cells ($\log_{2}FC > 2$; adj. p-value < 0.05) (**Suppl. Table 2**). Top hit proteins included QPRT, TBCA, ICAM1, and ENO1. (**Fig. 17D**). Quinolinate phosphoribosyltransferase (QPRT) is involved in NAD⁺ synthesis and has been found to have anti-apoptotic properties in leukemic cells ⁽¹⁷⁹⁾. Tubulin-specific chaperone A (TBCA) is involved in protein folding. Intercellular adhesion molecule 1 (ICAM1) was shown to be expressed on hematopoietic progenitors, but not on more mature cells of the erythroid or myeloid lineage ⁽¹⁸⁰⁾. Enolase 1 (ENO1) is a multifunctional oncoprotein overexpressed in a variety of cell types ⁽¹⁸¹⁾. 48 proteins were most significantly differentially abundant in normal erythroblasts ($\log_{2}FC > -2$; adj. p-value < 0.05) (**Suppl. Table 3**). Interestingly, the transferrin receptor (TFRC/CD71) appeared as one of the most abundantly expressed proteins in HUDEP2 and hEBST. TFRC/CD71 is known to be expressed in immature erythroid progenitors but also has been used as a marker of proliferating cells ⁽¹⁸²⁾. TF (transferrin), an iron-binding transport protein was significantly highly abundant in normal erythroblasts. Another iron-binding protein, Radical S-Adenosyl Methionine Domain Containing 2 (RSAD2), was also more abundant ⁽¹⁸³⁾. Furthermore, a variety of solute carrier organic anion transporter proteins that are

essential for transmembrane transport have been identified, such as SLC25A37, SLCO4C1, and SLC4A1⁽¹⁷⁵⁾. Other proteins found to be more abundant in normal erythroblasts included the prostaglandin E synthase (PTGES) and the lysine methyltransferase 2C (KMT2C aka MLL3) usually associated with chromatin remodeling and active gene transcription ⁽¹⁸⁴⁾. Among the erythroid transcriptional master regulators, only GATA1 and GFI-1B were significantly more abundant in normal erythroblasts (**Suppl. Table 4**). In leukemic erythroblasts, CBFA2T3 was more abundant but did not reach significance (logFC=0.9). Other transcription factors are known to play an essential role in erythropoiesis such as Ikaros family zinc finger 1 (IKZF1), KLF transcription factor 1 (KLF1), Lim domain-binding 1 (LDB1) or the TAL BHLH Transcription Factor 1 (TAL1) appeared not significantly differentially expressed (**Fig. 17E, Suppl. Table 4**). Applying Metascape analysis for TRRUST (www.grnpedia.org/trrust ⁽¹⁸⁵⁾), a database for transcriptional regulatory networks emphasized that the majority of proteins significantly (logFC >1; adj. p-value<0.05) found in leukemic erythroblasts are regulated by Specificity Protein 1 (SP1), a transcription factor involved in cell differentiation, cell growth, DNA damage, and chromatin remodeling and acting as an activator or repressor of transcription (**Fig. 17F**) ^(186,187). In normal erythroblasts the majority of significantly (logFC <-1; adj. p-value<0.05) identified proteins was predicted to be regulated by GATA1 (**Fig. 17G**).

Collectively, our data show differences in the global proteome between normal and leukemic erythroblasts and furthermore suggest that GATA1 is a major player that could explain impaired erythroid differentiation of AEL cells.

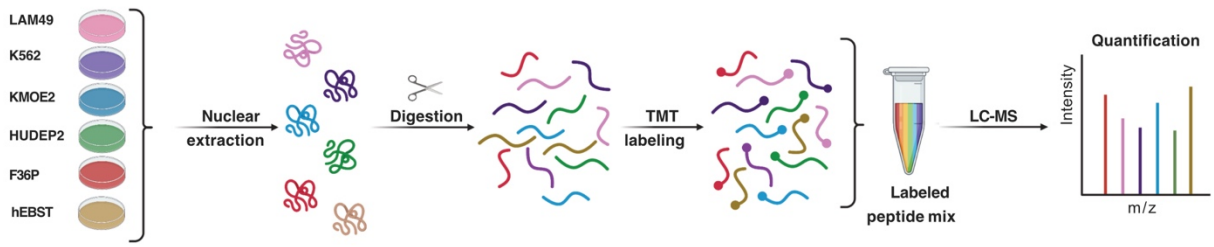
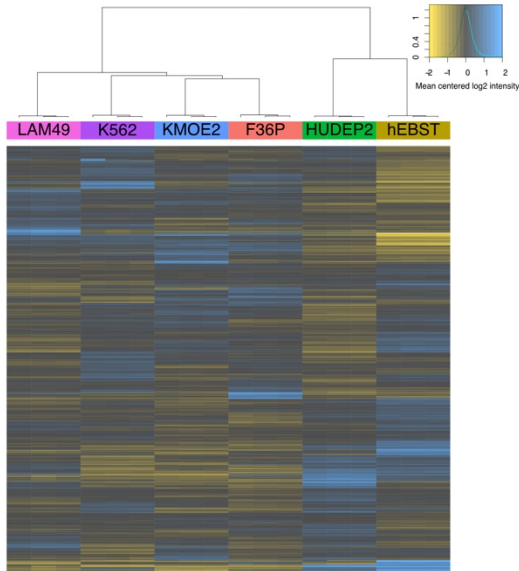
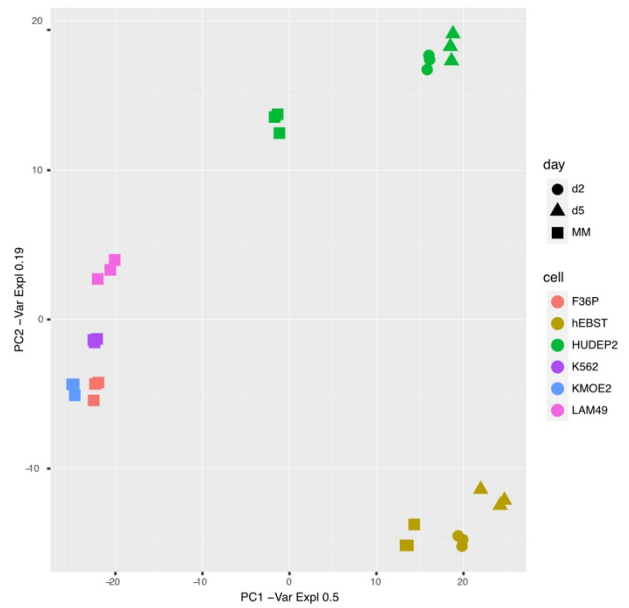
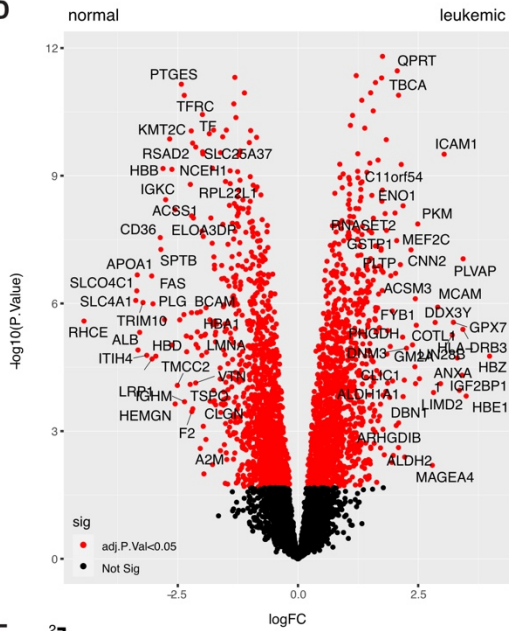
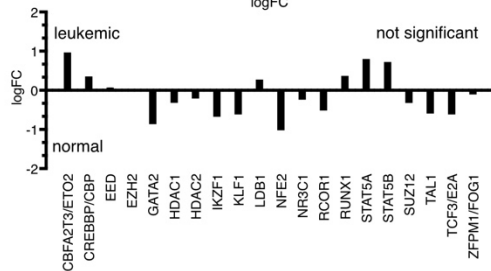
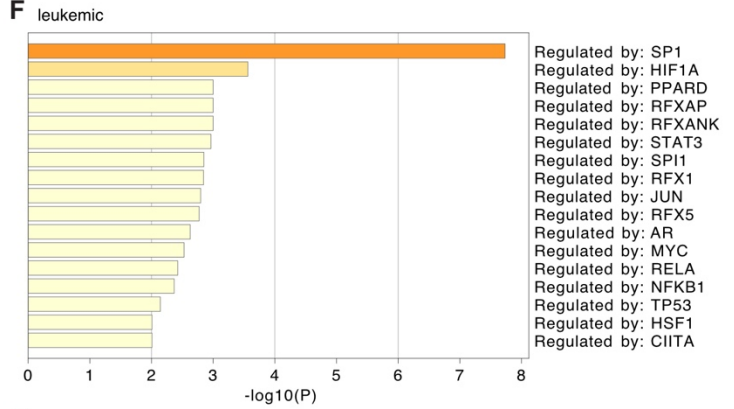
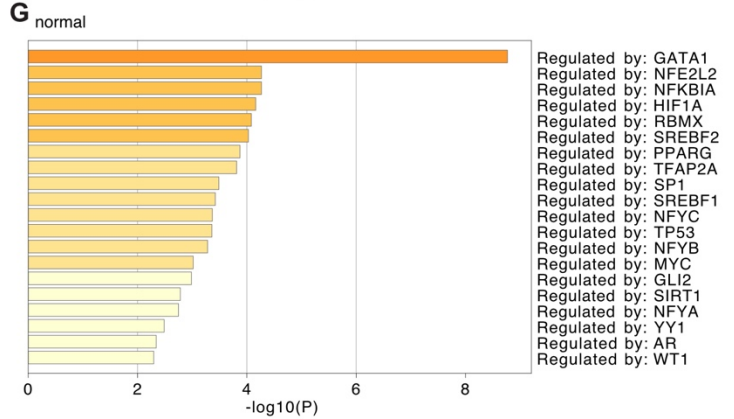
A**B****C****D****E****F****G**

Figure 17. Proteome analysis separates human AEL cells from human erythroblasts with full differentiation potential. **A)** Experimental setup: Extraction of cells followed by digestion of nuclear extracts into peptides and labeling with isobaric tags. All samples are pooled together, fractionated, and analyzed using an unbiased data-dependent MS approach (n=3 per cell type). **B)** Unsupervised hierarchical clustering of all cell lines based on log2ratio. **C)** Principle component analysis (PCA) of all cell lines including HUDEP2 and hEBST in 2 and 5 days in differentiation. **D)** Differential protein expression from normal (left) and leukemic (right) erythroblasts, (n=3, logFC>2, adj. p-value<0.05) **E)** LogFC of known erythroid transcription factors. Positive values show more expression in leukemic cells. Negative values show more expression in normal cells. All proteins are not significant. **F)** TTRUST (Transcriptional regulatory relationships unraveled by sentence-based text mining) analysis by Metascape in leukemic erythroblasts (based on proteins with logFC >1; adj. p-value<0.05). **G)** TTRUST (Transcriptional regulatory relationships unraveled by sentence-based text mining) analysis by Metascape in normal erythroblasts (based on proteins with logFC <-1; adj. p-value<0.05).

5.4.2. Exogenous GATA1 expression induces erythroid differentiation

Since we observed that GATA1 protein levels were more abundant in normal erythroblasts suggested functional importance. Therefore, we opted for a cellular model in which artificially increasing the amount of GATA1 protein may relieve at least partially the observed differentiation block of AEL cells (**Fig. 18A**). Hereby, we observed that lentiviral GATA1 overexpression induced significant partial erythroid differentiation of K562 cells as shown by reduced cell growth and increased hemoglobinization associated with a more differentiated morphology accompanied by increased *HBA* and *ALAS1* mRNA expression with a reduction in cell size (FSC-A) and an increase in CD71 and CD235a expression by flow cytometry (**Fig. 18B-G**). Similar to K562, GATA1 overexpression also induced signs of partial erythroid differentiation in HEL cells (**Suppl. Fig. 2A-D**).

In parallel, we overexpressed GATA1 in HUDEP2 cells. HUDEP2 maturation is normally achieved through a permissive medium that allows the cells to fully mature into erythrocytes. In contrast, upon GATA1 overexpression, cells already differentiated in maintenance medium (in which differentiation is blocked by dexamethasone), resulting in a gradual increase of CD71 and CD235a expression and differentiated morphology (**Suppl. Fig. 2E-J**).

These observations suggest, that AEL might be the consequence of alterations in transcription factors that affect erythropoiesis, potentially by affecting/impairing GATA1 activity.

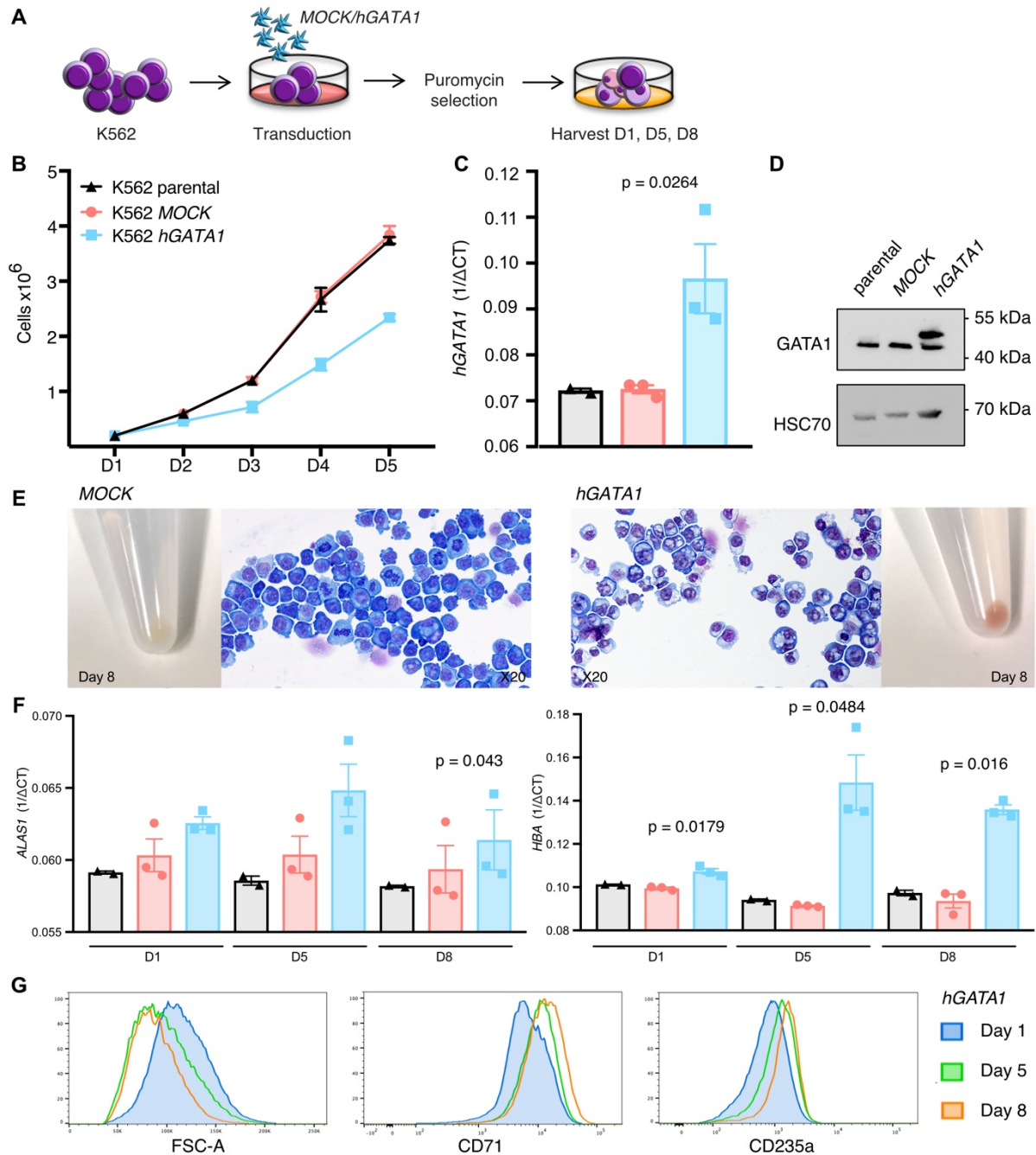


Figure 18. Effects of exogenous GATA1 expression in K562 cells. **A)** Experimental setup: K562 cells are transduced with *hGATA1* (*pSIN4-EF1a-GATA1-IRES-puro*) or *MOCK* (*pSIN4-EF1a-IRES-puro*) virus, selected with puromycin and harvested at 1, 5 and 8 days for further analysis. **B)** Cell growth of parental, *MOCK*, or *hGATA1*-transduced K562 cells 5 days after puromycin selection. **C)** Quantitative RT-PCR of *hGATA1* from parental, *MOCK*, and *hGATA1*-transduced K562 cells. CT values are normalized to *GAPDH* expression and shown as relative expression using $1/\Delta CT$ ($n=2$ parental; $n=3$ *MOCK* and *hGATA1*). **D)** Western Blot analysis of protein levels of GATA1 and HSC70 as loading control in parental, *MOCK*, and *hGATA1*-transduced K562 cells. **E)** Representative images of HE-stained cytoplots obtained from *MOCK* and *hGATA1*-transduced K562 cells and cell pellets 8 days after puromycin selection. **F)** Quantitative RT-PCR of *ALAS1* and *HBA* of parental, *MOCK*, and *hGATA1*-transduced K562 cells 1, 5 and 8 days after puromycin selection. CT values are normalized to *GAPDH* expression and shown as relative expression using $1/\Delta CT$ ($n=2$ parental; $n=3$ *MOCK* and *hGATA1*). **G)** Flow cytometry analysis of FSC-A, CD71, and CD235a expression in *hGATA1*-transduced K562 1, 5, and 8 days after puromycin selection.

5.4.3. Differential GATA1 interactome in normal and leukemic erythroblasts

To better understand how GATA1 exerts its function in synergy with other proteins, we isolated nuclear extracts, immunoprecipitated with GATA1 or IgG control antibodies followed by stringent washes, peptide digestion, and LC-MS/MS analysis (n=3/type) (**Fig. 19A**). PCA showed for all cell lines a similar clustering pattern independent whether they were immunoprecipitated with GATA1 or IgG antibody. Furthermore, primary erythroblasts were positioned close to F36P and KM0E2 whereas LAM49 and K562 appeared more separate (**Fig. 19B**). Analysis of the lysates from all six cell types identified 1616 proteins, of which 125 were differentially associated with GATA1 depending on the cell identity (**Suppl. Table 5**). 54 proteins were preferably pulled-down in the AEL group, whereas 71 proteins were more enriched in normal erythroblasts (adj. p-value <0.5) (**Fig. 19C, Suppl. Table 6**). One critical hematopoietic transcriptional regulator that we pulled-down with GATA1 in normal hEBST is IKZF1 which is required for normal development of blood cell lineages and can function as an activator or repressor by recruitment of chromatin remodelling complexes ⁽¹⁸⁸⁾. Another important TF that we found and is known to be associated with GATA1 is Runx family transcription factor 1 (RUNX1) which is essential for HSPC differentiation in a complex containing LDB1, TAL1, and CBFA2T3 in erythroid cells ⁽¹⁸⁹⁾. The identity of these GATA1-associated TFs in normal hEBSTs was validated by Western Blotting (**Fig. 19C, D**). Strikingly, we found the SKI proto-oncoprotein to be differentially enriched in all AEL cell samples and could also validate this by Western blotting (**Fig. 19C, E**). SKI has been previously characterized as a GATA1 interactor and negative regulator of erythroid differentiation cooperating with other oncogenes in cellular transformation ^(190,191). Among the other proteins, pulled-down with GATA1 exclusively in AEL cells (and never been linked to erythropoiesis) was Leucine Rich Pentatricopeptide Repeat Containing (LRPPRC) a multifunctional protein involved in metabolism and overexpressed in various human cancers ⁽¹⁹²⁾. We also identified chaperones containing TCP1 subunits 4&7 (CCT4&7) more enriched in AEL cells, which generally assist protein folding. Interestingly, TCP1 expression was previously reported to be significantly elevated in AML patients ⁽¹⁹³⁾. We also found the karyopherin's KPNB1 (importin- β) and KPNA2 (importin- α), which mainly regulate nuclear protein import, and have been found to be aberrantly expressed in many human cancers ⁽¹⁹⁴⁾.

Collectively, we identified several GATA1-associated proteins that significantly differ between normal and malignant erythroblasts, including known but also novel potential GATA1 interactors.

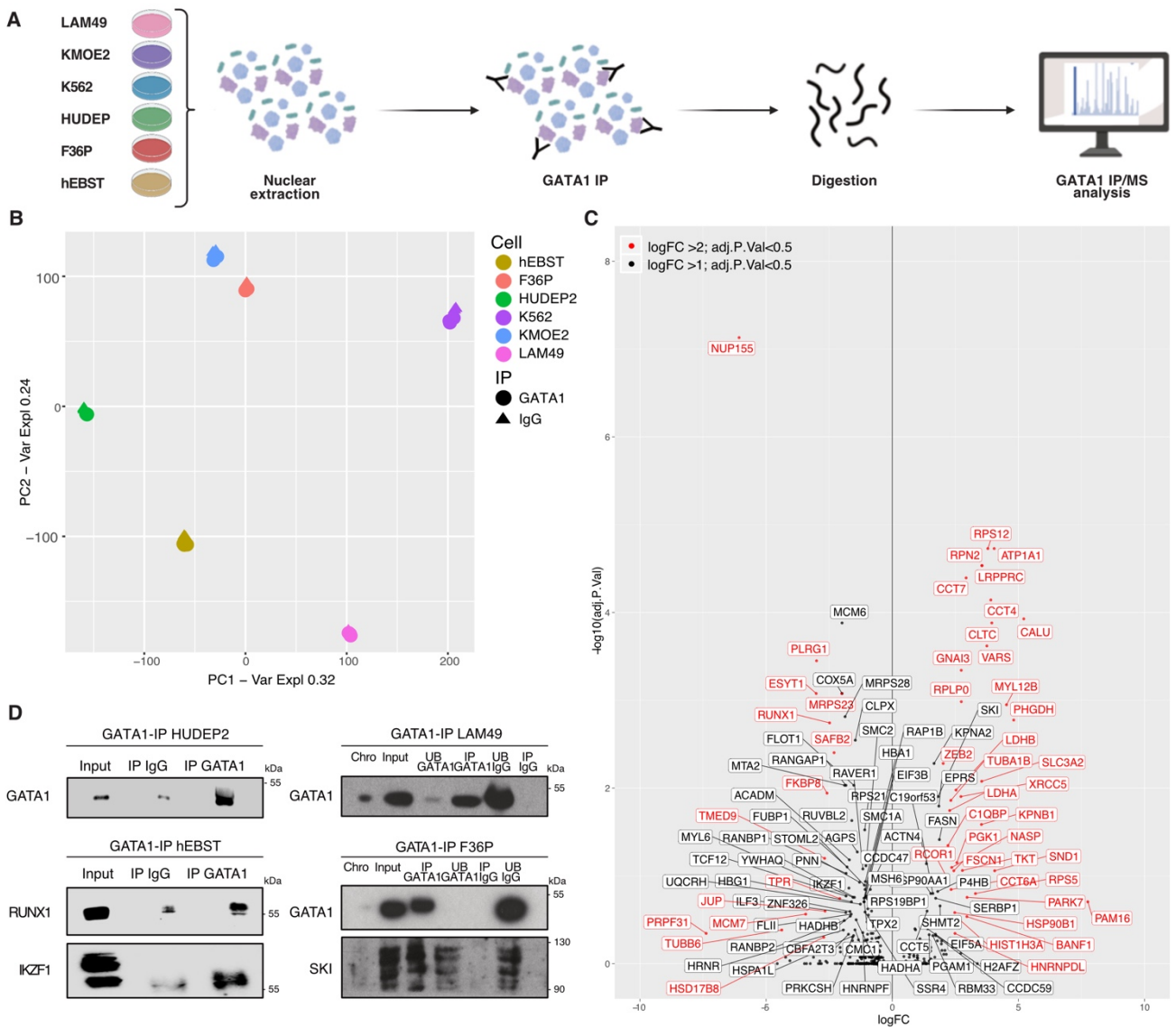


Figure 19. Comparative analysis of GATA1 protein interactions in normal (hEBST, HUDEP2) and leukemic (K562, KMOE2, F36P, LAM49) erythroblasts. A) Extraction of nuclear lysates, followed by immunoprecipitation (IP) with either GATA1 or unspecific (IgG) antibody. Tryptic digestion into peptides and analysis using an unbiased data-dependent MS approach (n=3 per cell type). **B)** PCA-analysis of all GATA1 and IgG IPs from all cell types (GATA1 n=3; IgG n=3). **C)** Differential protein expression from GATA1 IP/MS analysis in normal (left) and leukemic (right) erythroblasts, normalized to GATA1 and IgG control (n=3, $\log_{2}(\text{FC}) > 1$; $\text{adj. p-value} < 0.5$). **D)** GATA1 IP in HUDEP2 followed by Western blotting of GATA1 to confirm successful enrichment in normal erythroblasts (upper panel). GATA1 IP in primary hEBST followed by Western blotting of RUNX1 (middle panel) and IKZF1 (lower panel). **E)** GATA1 IP in primary AEL cells (LAM49) followed by Western blotting of GATA1 (upper panel) to confirm successful enrichment in leukemic cells. IP of GATA1 in F36P cells followed by Western blotting of GATA1 (middle panel) and SKI (lower panel).

5.4.4. A targeted CRISPR/Cas9 screen identifies GATA1-associated proteins/genes that maintain the differentiation block of K562 cells

Integrated analysis of the quantitative proteome- and the differential GATA1-pulldown experiments revealed 116 proteins that were either significantly differentially expressed and/or precipitated (**Fig. 20A**). To identify potential GATA1 protein interactions that contribute to blocked erythroid differentiation we decided to perform a targeted CRISPR screen in *SpCas9*-K562 cells. We first checked whether the *SpCas9*-K562 cells showed similar signs of erythroid differentiation upon exogenous GATA1 expression as the parental K562 cell line (**Suppl. Fig. 3A-E**). We then targeted all 116 candidate genes with two different sgRNA (co-expressing the iRFP670 fluorochrome) and assessed cell growth and erythroid maturation of the cells (in duplicates) over several days (4, 6, 8, 10 & 12 days) (**Fig. 20B**). While the fraction of iRFP670⁺ cells remained stable in cells expressing the sgRNA targeting the negative control non-essential adeno-associated virus integration site 1 gene (sgAAVS1), depletion was observed when the positive control sgRNA targeting the essential 60S ribosomal protein L17 gene (sgRPL-17) was transduced. Normalizing depletion score and MFI values to sgAAVS1 (as described in detail in the materials and methods section) revealed that sgRNA-mediated inactivation of the 116 candidate genes resulted in 4 distinct readouts (**Fig. 20C, Suppl. Table 7**).

Targeting “cluster-1” genes (n=62) did result in neither significant cell depletion nor differentiation. In this cluster were genes encoding for several transcription factors such as IKZF1, Zinc Finger E-Box Binding Homeobox 2 (ZEB2), Metastasis Associated 1 Family Member 2 (MTA2), and Far Upstream Element Binding Protein 1 (FUBP1). FUBP1 regulates MYC proto-oncogene expression by binding to the MYC promoter and can act as an activator or repressor of transcription⁽¹⁹⁵⁾. MTA2 is involved in repression and activation and is known to form a complex with GATA1⁽⁷⁰⁾. IKZF1 is well known for regulating hematopoietic cell differentiation and regulates transcription often through HDAC-dependent complexes and targets the two chromatin remodeling complexes: NurD and SWI/SNF^(196,197). ZEB2/SIP1 is a transcriptional repressor of E-cadherin⁽¹⁹⁸⁾.

Inactivation of “cluster-2” genes (n=21) significantly reduced cell survival without affecting erythroid differentiation. Genes belonging to this cluster encode for RAN-binding proteins 1&2, which regulate the transport of proteins across nuclear membranes and a variety of ribosomal proteins such as RPS5, RPS21, MRPS23, and RPL4^(199,200).

Targeting of “cluster-3” genes (n=11) not only induced erythroid differentiation as seen by increased CD71 and CD235a expression but also reduced cell viability. Among the genes of this cluster are RUNX1, Structural Maintenance of Chromosomes 1A (SMC1A), nucleoporin 93 (NUP93), and SERPINE1 mRNA binding protein (SERBP1). RUNX1 is essential for the emergence of hematopoietic stem cells (HSC)⁽¹⁸⁹⁾. SMC1A is involved in chromosome cohesion during cell cycle

and DNA repair ⁽²⁰¹⁾. NUP93 is a nuclear pore complex (NPC) protein important for its assembly ⁽²⁰²⁾. SERBP1 was previously proposed to regulate mRNA translation and has been found to be differentially expressed during erythropoiesis ^(203,204).

In contrast, the inactivation of “cluster-4” genes (n=22) displayed no changes in cell survival but resulted in significant signs of erythroid differentiation. Amongst these was CBFA2T3, which showed the strongest impact on erythroid differentiation. CBFA2T3 has been previously characterized as a repressive co-factor in GATA1- and/or NCOR-containing complexes controlling erythroid differentiation ^(81,205,206). In addition, aberrant CBFA2T3 expression was seen in some primary AEL samples and its overexpression blocked *in vitro* erythroid differentiation ⁽¹⁴⁰⁾. These observations indicate that aberrant CBFA2T3 levels may contribute to the differentiation block in some AEL patients by functionally impairing GATA1. Additionally to CBFA2T3, we identified 21 other genes that when targeted by sgRNA resulted in significant signs of erythroid differentiation. Among them, RCOR1, which is part of the CoREST transcriptional co-repressor complex formed by histone deacetylases 1 & 2 (HDAC1/2) and the lysine demethylase 1A (LSD1) and essential for normal myelo-erythroid lineage differentiation ⁽⁸³⁾. Notably, shRNA-mediated *Rcor1* knockdown increased hemoglobin, *e-band3*, and *Gypa* expression in murine erythroleukemia (MEL) cells ⁽⁸²⁾. Moreover, we identified Parkinson-associated deglycase (PARK7/DJ-1), which was found to be overexpressed in leukemia patient samples and inhibition in cell lines resulted in reduced proliferation and enhanced drug sensitivity ⁽²⁰⁷⁾. Interestingly, in “cluster-4”, we also observed that the TCP1 proteins CCT4 and CCT7 resulted in signs of erythroid differentiation. Another protein found to be significantly increasing CD235a expression upon CRISPR/Cas9-mediated inactivation was KPNA2. Strikingly, KPNA2 forms a complex with the KPNB1, which we also found pulled down with GATA1 in AEL cells. However, loss of KPNB1 resulted in significant cell death and is therefore found in “cluster-2” of the CRISPR/Cas9 screen.

Collectively, we found several GATA1-associated proteins that may contribute to impaired erythroid differentiation in AEL cells.

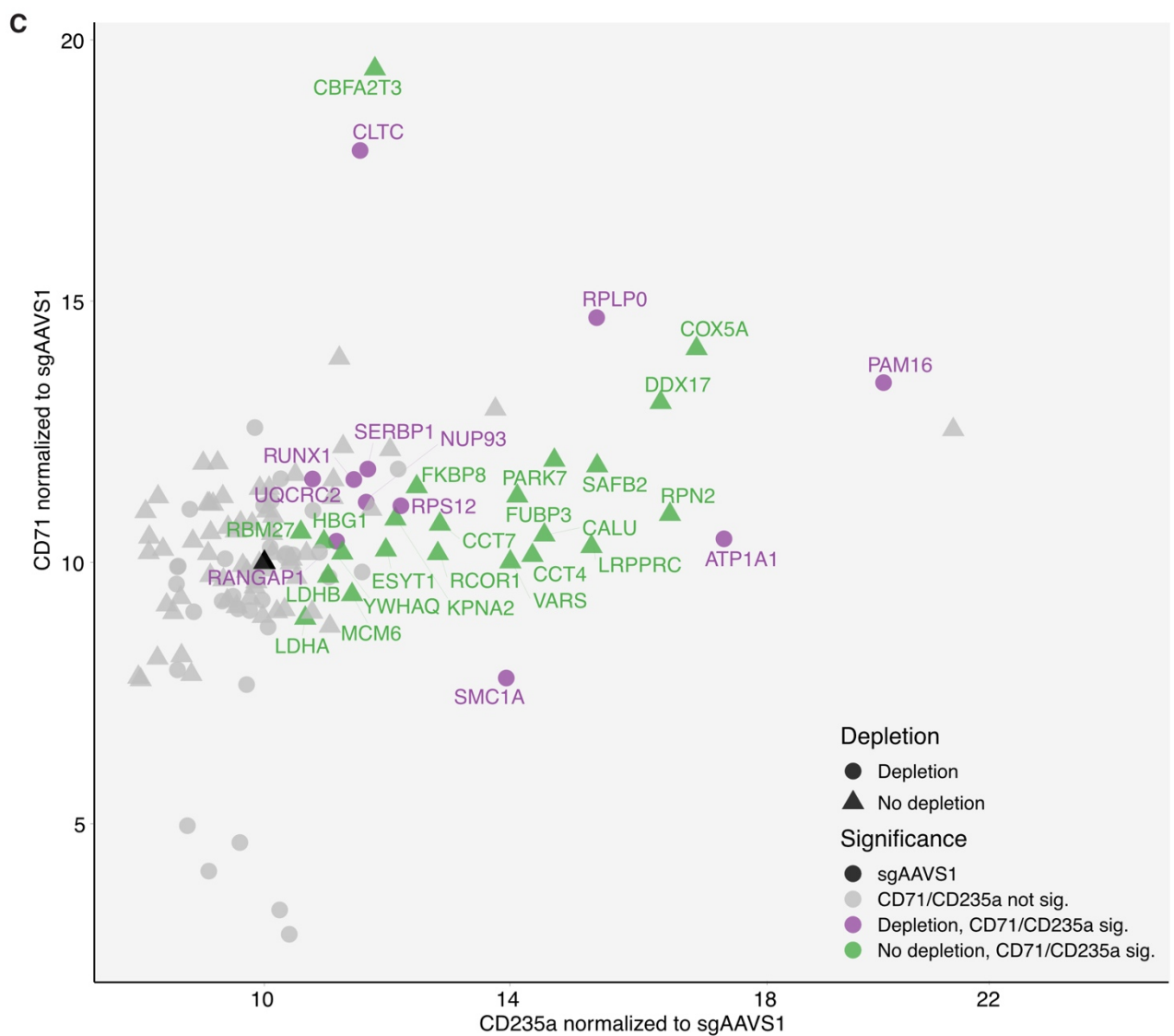
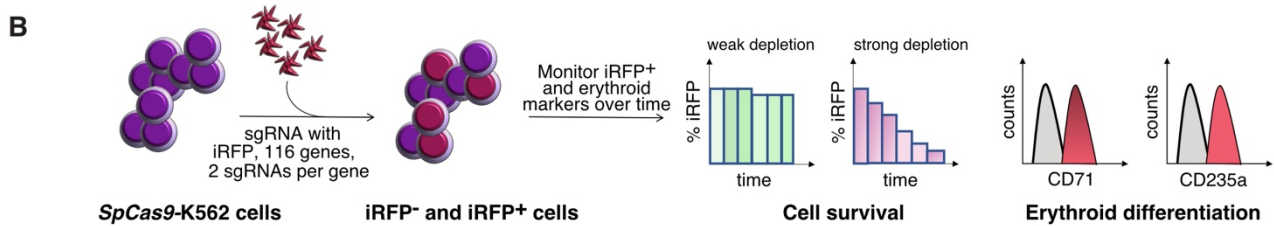
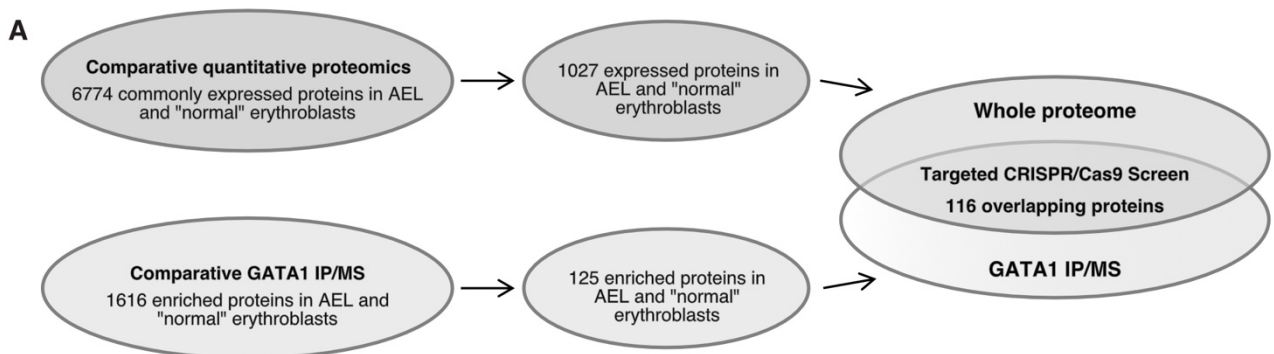


Figure 20. A targeted CRISPR/Cas9 screen identifies GATA1-interacting proteins/genes that can overcome the differentiation block in *SpCas9*-K562 cells. A) Results of whole proteome and GATA1-interactome identifies 116 overlapping proteins which are inactivated by a targeted CRISPR/Cas9 screen. **B)** Experimental setup of the CRISPR/Cas9 screen: each gene is targeted with 2 sgRNAs in *SpCas9*-K562 cells. sgRNA-expressing cells (iRFP670+) are monitored over time (days 4,6,8,10,12) and depletion is determined for each sgRNA based on the % of iRFP expression. Erythroid differentiation is assessed by CD71 and CD235a surface marker expression. **C)** Statistical evaluation of CRISPR/Cas9 screen. 4 clusters can be identified based on CD71/CD235a marker expression (MFI) and depletion (% iRFP670).

5.4.5. Characterization of KPNA2, RCOR1, and CCT4 as GATA1 interactors that contribute to the erythroid differentiation block

To further characterize selected candidates, we first assessed their nuclear expression in AEL vs normal erythroblasts by Western Blotting. RCOR1 protein expression appeared to be higher in K562, F36P, KMOE2 and hEBST, whereas LAM49 and HUDEP2 showed lower levels. In contrast, we found significantly higher protein levels of KPNA2 and CCT4 in all AEL cell types compared to HUDEP2 and hEBST, whereas PARK7 protein levels were not significantly different in AEL compared to normal erythroblasts (**Fig. 21A, Suppl. Fig. 4A**).

To confirm the capability to induce erythroid differentiation, we reduced their expression by targeting them with doxycycline (Dox)-regulated mir-shRNAs in K562 cells (**Fig. 21B, Suppl. Fig. 4B-D**). Expression of KPNA2, RCOR1, and CCT4 mir-shRNAs resulted in a significant reduction in cell growth accompanied by upregulation of *HBA* expression (**Fig. 21C, D**). Knockdown of KPNA2 showed the strongest effects on hemoglobinization and a shift of both erythroid markers CD71 and CD235a (**Fig. 21E, F**). Knockdown of RCOR1, PARK7, and CCT4 had less strong phenotypes (**Fig. 21F, Suppl. Fig. 4D-F**). In parallel, we virally overexpressed *KPNA2* and *RCOR1* in HUDEP2 cells. *RCOR1*- and *KPNA2*-transduced HUDEP2 cells showed increased cell growth and reduced expression of *HBA* compared to *MOCK*-transduced HUDEP2 cells (**Fig. 21G, H**).

In conclusion, these experiments suggest that by reducing the expression of potentially repressive proteins, erythroid differentiation can be initiated, whereas overexpression results in a more immature phenotype.

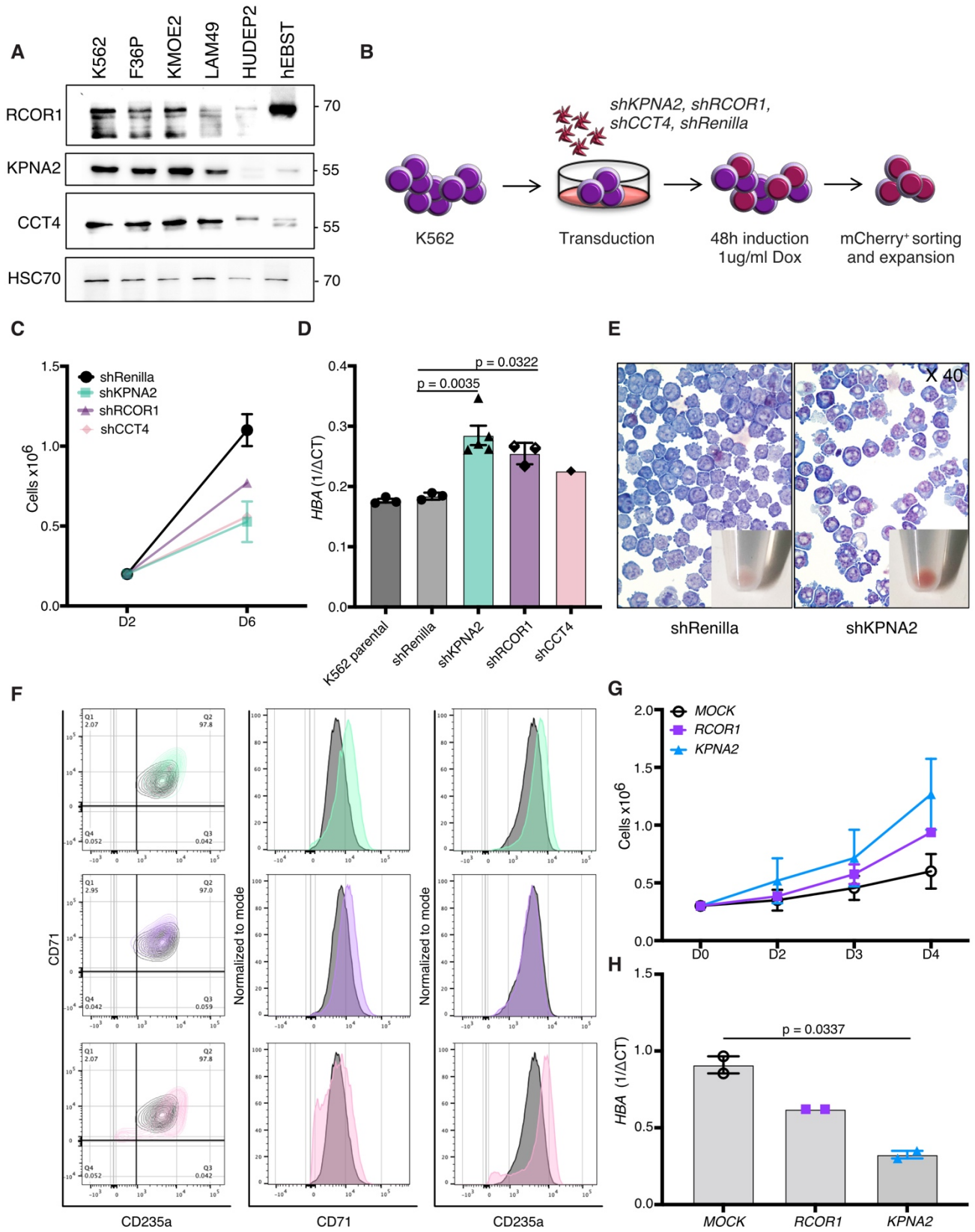


Figure 21. Characterization of KPNA2, RCOR1, and CCT4 as GATA1 interactors contributing to the erythroid differentiation block. **A)** Protein expression of RCOR1, KPNA2, and CCT4 in AEL (K562, F36P, KMOE2, LAM49) and normal (HUDEP2, hEBST) erythroblasts. HSC70 is used as a loading control. **B)** Experimental setup: K562 cells are transduced with mir-shRNAs (shKPNA2, shRCOR1, shCCT4) or control (shRenilla), induced with doxycycline (Dox), and sorted 48h after induction. Cells are expanded for at least 6 days before analysis. **C)** Growth curve of shKPNA2 (n=5), shRCOR1 (n=3), shCCT4 (n=1) and shRenilla (n=3)-transduced K562 cells. **D)** Quantitative RT-PCR of *HBA* of parental (n=3), shRenilla (n=3) shKPNA2 (n=5), shRCOR1 (n=3) and shCCT4 (n=1)-transduced K562 cells 6 days after transduction. CT values are normalized to *GAPDH* expression and shown as relative expression using $1/\Delta\text{CT}$. **E.** Representative images of HE-stained cytoplots obtained from shRenilla and shKPNA2 transduced K562 cells and cell pellets 8 days after transduction. **F)** Representative images of flow cytometry analysis of CD71 and CD235a expression in shKPNA2 (n=5), shRCOR1 (n=3), and shCCT4 (n=1)-transduced K562 cells 6 days after transduction. **G)** Proliferation of *KPNA2* (n=2), *RCOR1* (n=2), and *MOCK* (n=2) transduced HUDEP2 cells expanded in maintenance medium after sorting. **H)** Quantitative RT-PCR of *HBA* of *MOCK* (n=2), *KPNA2* (n=2), *RCOR1* (n=2) transduced and sorted HUDEP2 cells expanded in maintenance medium. CT values were normalized to *GAPDH* expression and shown as relative expression using $1/\Delta\text{CT}$.

5.4.6. coREST inhibition enhances erythroid differentiation

As outlined above, we found that inactivation of RCOR1 induced partial erythroid differentiation of K562 cells not only by CRISPR/Cas9 inactivation but also by inducible mir-shRNA-based knockdown (**Fig. 20 & 21**). To assess a therapeutic potential, we treated K562 cells with Corin, a small molecule that inhibits the coREST complex by interaction with LSD1 and HDAC1⁽²⁰⁸⁾. After 3 days we observed a dose-dependent reduction of cell growth, accompanied by increased HBA and CD71 expression, similar to previous experiments (**Fig. 22A-C**). Notably, 2 days of treatment of HUDEP2 cells reduced cell growth due to cell viability, but decreased CD235a expression, suggesting that mechanisms controlling erythroid differentiation in normal cells are different than in AEL cells (**Suppl. Fig.5A, B**).

This suggests that by targeted interfering with co-repressive complexes, erythroid differentiation can be initiated.

Collectively, these validation experiments confirm and strengthen the feasibility of our CRISPR/Cas9 screen to identify GATA1 interactors that contribute to the erythroid differentiation block. Our work so far suggests that impaired erythroid differentiation of AEL cells involves aberrant GATA1 protein complexes. Findings implied that interactions of GATA1 to the epigenetic regulator RCOR1 and to the KPNA2 importin contribute to the differentiation block. Future experiments aim to validate protein interactions by co-IPs and/or proximity ligation assays (PLA). In the next steps, we aim to assess the effects of gene inactivation of critical candidates on GATA1 binding to target gene loci to identify novel regulators of terminal differentiation in normal and malignant erythropoiesis.

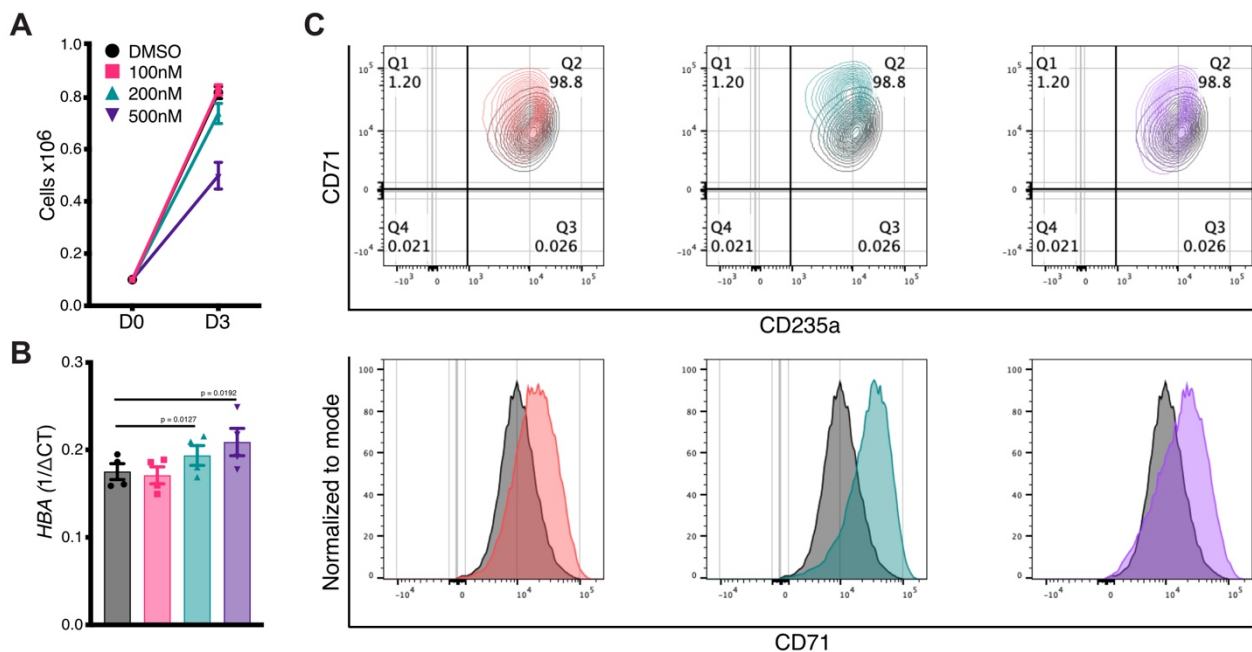


Figure 22. Exposure to the coREST complex inhibitor Corin results in signs of erythroid differentiation in K562 cells. **A)** Growth curve of Corin-treated K562 cells (DMSO, 100nM, 200nM and 500nM; n=4). **B)** Quantitative RT-PCR of *HBA* after 3 days: DMSO, 100nM, 200nM, and 500nM treated K562 cells. CT values are normalized to *GAPDH* expression and shown as relative expression using $1/\Delta CT$. **C)** Representative images of flow cytometry analysis of CD71 and CD235a expression after 3 days of treatment.

5.5. Additional experiments (will most likely not be included in the manuscript)

5.5.1. Characterization of the differentiating proteome of hEBST and HUDEP2 cells

To investigate the remodeling of the proteome during erythroid differentiation, we cultured primary erythroblasts (hEBSTs) and HUDEP2 cells in maintenance medium and under conditions that favor the maturation of cells. For this analysis, we took hEBSTs and HUDEP2 cells in maintenance medium and from day 2 and day 5 in differentiation-inducing medium (**Fig. 23A**).

Overall, we identified 6774 proteins in all conditions. Unsupervised hierarchical cluster analysis displayed that the samples formed two branches clearly separating hEBST and HUDEP2 cells. The biggest differences were obtained for cells kept in differentiation-induced conditions whereas both cell lines kept in maintenance medium clustered closer together (**Fig. 23B**). In pairwise comparisons the overall proteome significantly changed the most between D0 and D5 (**Fig. 23C**). Whereas less significant changes were observed comparing D0 with D2 and D2 with D5 (**Fig. 23D, E**). The most significantly changing proteins were erythroid-specific proteins, such as SPTA1, SPTB and ANK1, which are components of the erythrocyte plasma membrane. Furthermore, ALB (albumin) the most abundant protein in human blood was significantly more expressed in differentiating cells. Iron binding

and oxygen binding proteins, HBB and HBA1 as well as the SLC mediated-transmembrane transporter SLC4A1, were significantly identified and also found in another proteomic study ⁽¹⁷⁵⁾.

We next clustered differentially expressed proteins during all different stages into 4 major groups based on their trend:

Cluster	Day 0 to Day 2	Day 2 to Day 5
Down-down	down	down
Down-up	down	up
Up-down	Up	down
Up-up	Up	up

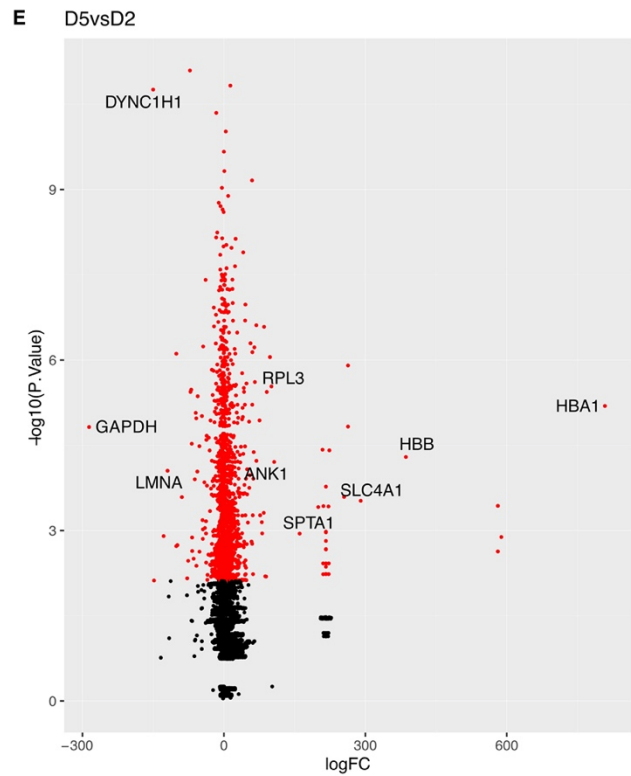
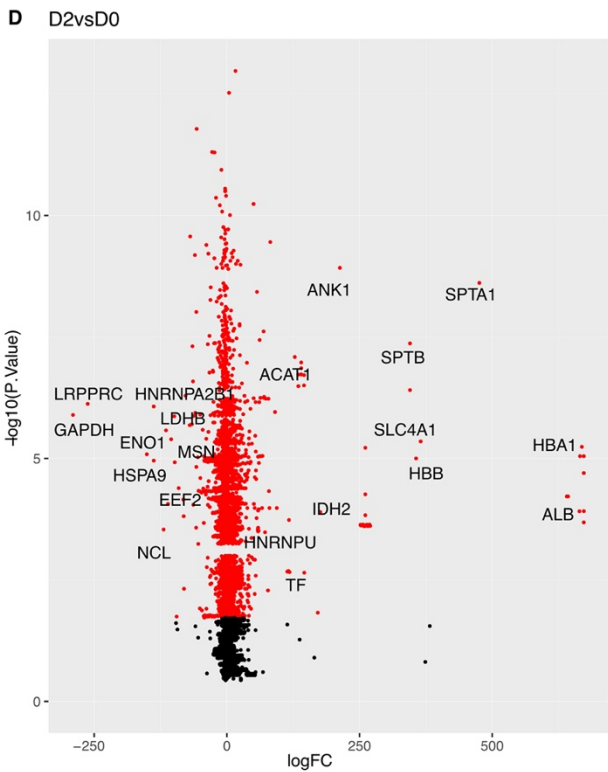
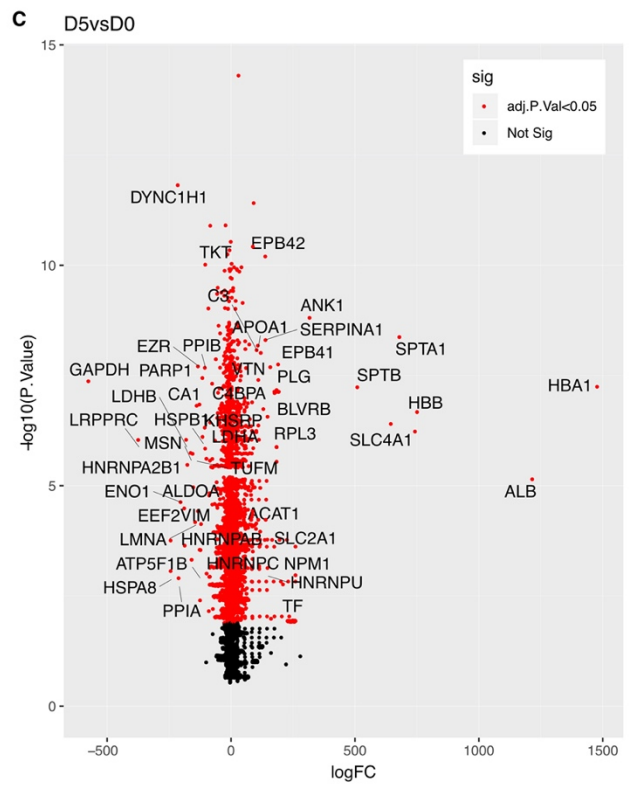
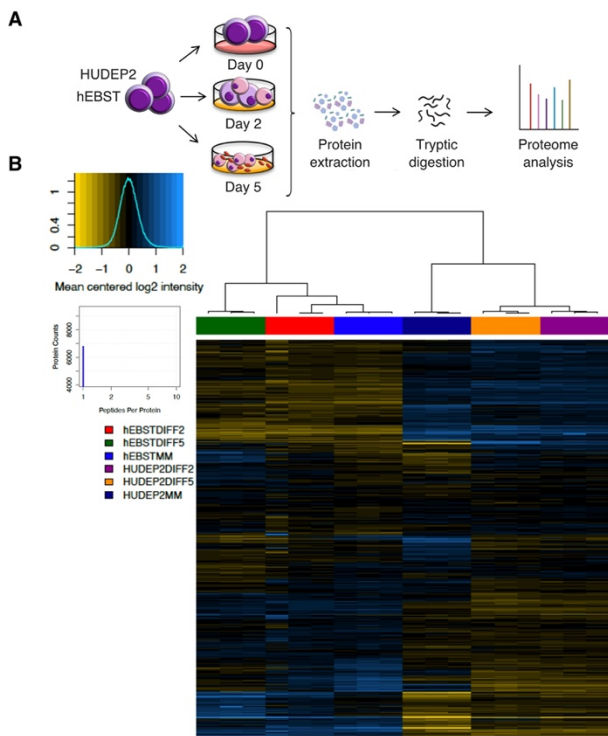
This pairwise comparison based on the DESeq2 package for specific time points was conducted by Liquing Zheng (University of Copenhagen, Master Program Bioinformatics) (p-value < 0.05). Significant proteins were further assessed with metascape.org for Pathway & Process Enrichment (standard settings). 433 proteins appeared in the “down-down” regulated cluster. Pathways and proteins were mainly related to mRNA metabolic process, ribonucleoprotein complex biogenesis and regulation of translation (**Fig. 23F**).

116 proteins belonged to the “down-up” cluster, with the top enriched pathway associated with the complement system. The complement system mainly helps to remove dead cells or debris that is created during erythroid maturation and is only needed at later stages of maturation explaining the inverse relationship of down-up regulation (**Fig. 23G**).

134 were significantly enriched in the “up-down” cluster. These proteins mainly have a chromatin remodeling and organization function and are involved in cell cycle processes which have to happen in early stages of cell maturation. The RB-TAL1-E2A-LMO2-LDB1 complex can also be significantly identified, which is a well described transcriptional activating complex during erythroid differentiation (**Fig. 23H**).

490 proteins of the “up-up” cluster were increased at Day 2 and Day 5. Erythroid progenitors consume a high amount of iron that must reach the mitochondrial matrix to support hemoglobinization. Iron trafficking occurs via transferrin receptor mediated endocytosis, which is also clathrin dependent. Autophagy is an essential process that takes place at final stages of differentiation to clear unnecessary organelles such as ribosomes and mitochondria from the blood and dysfunction hampers correct erythroid maturation (**Fig. 23I**) ⁽²⁰⁹⁾.

Taken together, we obtained a quantitative proteome-based view of differentiating erythroblasts. The identified proteins and cellular processes corroborate previously described findings on erythroid differentiation and therefore this dataset can be used for more detailed functional studies.



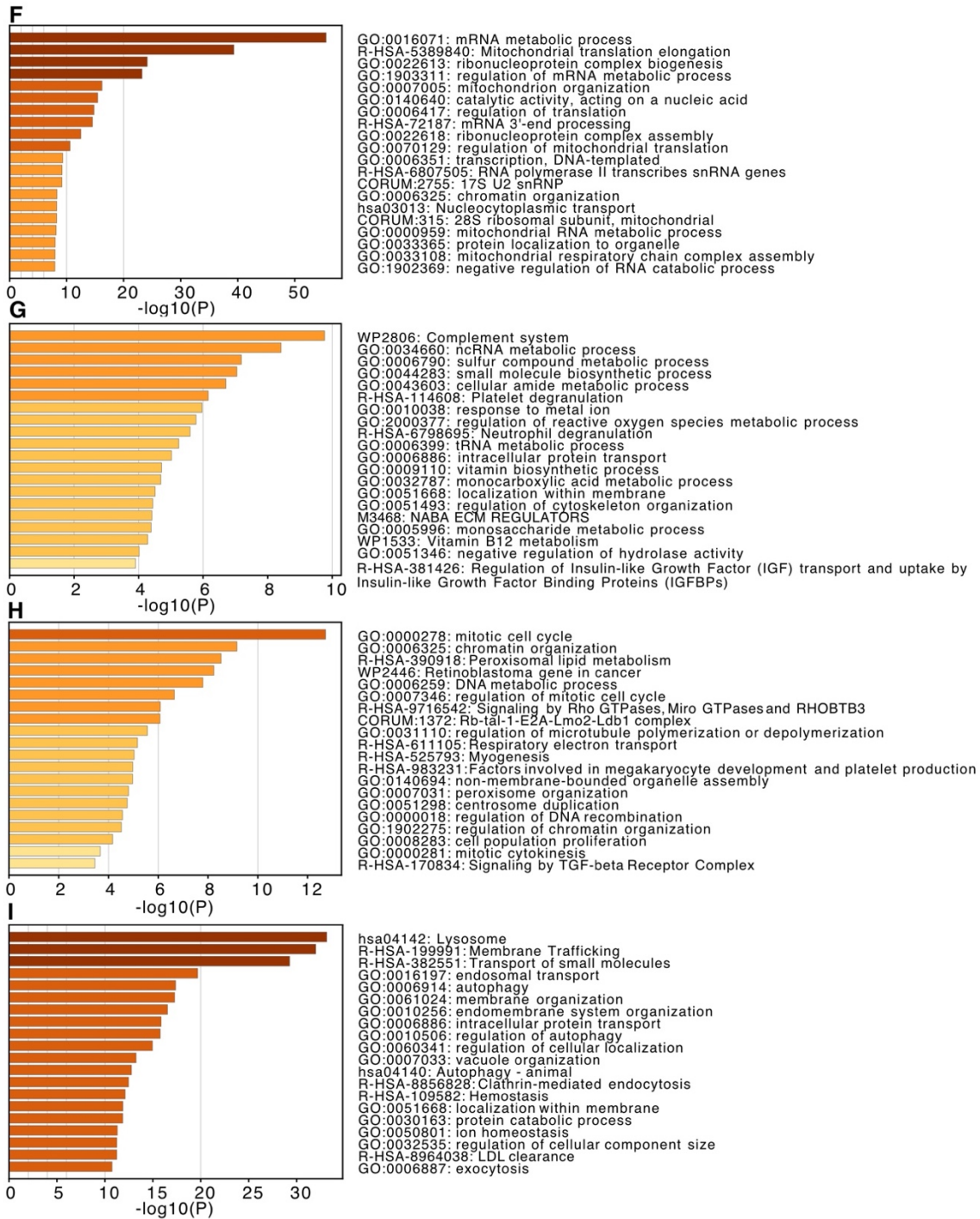


Figure 23. Dynamic proteome remodeling of normal erythroblasts (hEBST, HUDEP2). **A)** Experimental setup: primary erythroblasts (hEBST) or HUDEP2 cells are differentiated for 0, 2 or 5 days and protein is isolated followed by tryptic digestion of nuclear extracts into peptides and labeling with isobaric tags. All samples are pooled together, fractionated, and analyzed using an unbiased data-dependent MS approach (n=3). **B)** Unsupervised hierarchical clustering of all conditions. **C)** Differential protein expression from day 0 (left) and day 5 (right) normal erythroblasts, (adj. p-value<0.05). **D)** Differential protein expression from day 0 (left) and day 2 (right) normal erythroblasts, (adj. p-value<0.05). **E)** Differential protein expression from day 2 (left) and day 5 (right) normal erythroblasts, (adj. p-value<0.05). **F)** Metascape analysis of “down-down” proteins (down at day 2 and down at day 5). **G)** Metascape analysis of “down-up” proteins (down at day 2 and up at day 5). **H)** Metascape analysis of “up-down” proteins (up at day 2 and down at day 5). **I)** Metascape analysis of “up-up” proteins (up at day 2 and up at day 5).

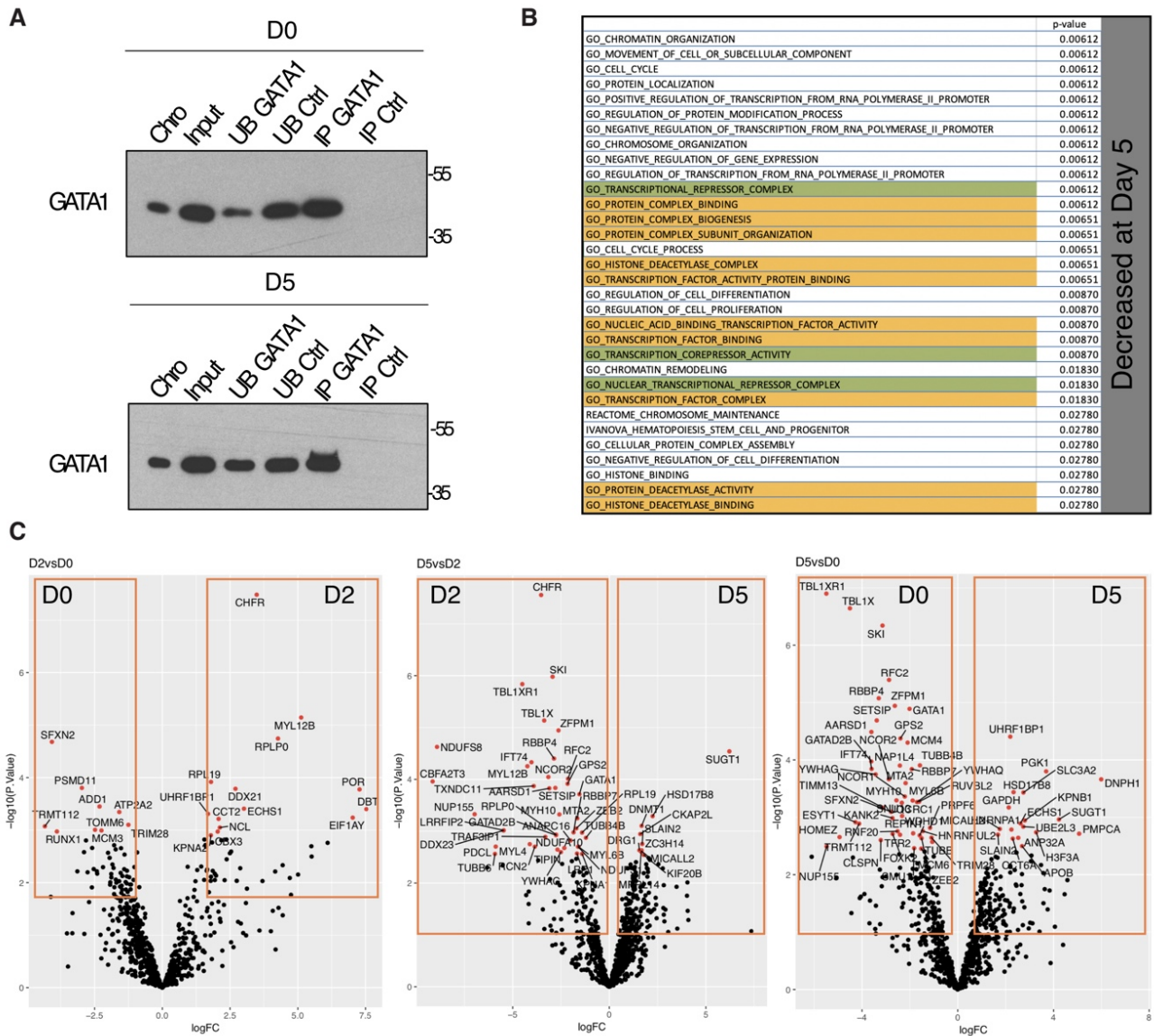
5.5.2. Characterization of the differentiating GATA1 interactome of hEBSTs

To study the dynamics of GATA1 protein complexes during differentiation of normal human erythroblasts we established erythroblast cultures from peripheral blood mononuclear cells (PBMCs) of healthy donors. For this analysis, we took hEBSTs expanded in maintenance medium (day 0), and cells at 2 and 5 days after induced differentiation.

We were able to immunoprecipitate sufficient GATA1 from cells grown in maintenance medium as well as 5 days after induced differentiation (**Fig. 24A**). Comparing GATA1-immunoprecipitated proteins between Day 0 and Day 5 by Gene Ontology (GO) analysis showed that pathways such as “transcriptional repressor complex/activity” as well as protein complex binding/biogenesis” were significantly decreased in differentiating hEBSTs (**Fig. 24B**). Pairwise comparisons of GATA1 interaction partners during erythroid differentiation revealed only small changes between day 0 and day 2. More significant changes in the GATA1 interactome were seen when comparing day 2 and day 5, whereas the most significant changes were observed when comparing day 0 and day 5 (**Fig. 24C**).

At day 0 a variety of proteins were pulled-down including the known transcriptional repressors such as SKI, NCOR1, NCOR2, TBL1X, TBL1XR1, and HDAC3. Furthermore, ZFPM1/FOG1, which can act as a transcriptional repressor and activator was found to be more enriched in maintenance. After 5 days of differentiation, the ubiquitin-like containing PHD and RING finger domains 1-binding protein (UHRF1BP1 aka ICBP90) which is a negative regulator of cell growth and known to bind to histone deacetylases was significantly enriched ⁽²¹⁰⁾.

Overall, this data suggests that GATA1 protein complexes undergo a dynamic switch from a repressive state to active confirmation potentially through the removal of these repressors which induces the maturation of erythroblasts. The functional importance in combination with interaction and binding on chromatin remains to be elucidated.



Decreased at Day 5

Figure 24. GATA1 IP/MS analysis during *in vitro* maturation of hEBSTs. A) GATA1 is successfully immunoprecipitated in erythroblasts maintained (day 0) or differentiated for 5 days. **B)** Gene Ontology (GO) analysis shows significant GO-terms that decreased on day 5. **C)** GATA1 IP/MS analysis in human erythroblasts kept 13 days in maintenance (day 0), 2 days (day 2), or 5 days in differentiation (day 5). (n=3, adj. p-value<0.05, logFC>2).

5.5.3. mRNA and protein levels of erythroid master transcription factors do not correlate

To better characterize the role of GATA1 in the transformed state characterized by blocked erythroid maturation, we first compared its expression in human AEL and non-AEL cell lines. In contrast to non-AEL cells, *GATA1* mRNA and protein appeared abundantly highly expressed in AEL cell lines (**Fig. 24A, F**). In contrast, we did not observe any significant correlation of mRNA and protein levels of other erythroid differentiation-related transcriptional co-regulators such as SKI, CBFA2T3 or NCOR1&2, which could separate AEL from non-AEL cells (**Fig. 24B-E, F**). However, protein levels of CBFA2T3 were significantly higher in all AEL cell lines and NCOR1&2 as well as SKI also showed a trend towards higher protein expression in AEL cells (**Fig. 24F**).

This shows, that TFs seem to play a role in AEL and furthermore displays the limitation of models solely based on mRNA measurements without integration of protein data. The exclusively high mRNA and protein expression of GATA1 in AEL suggests a driving mechanism in erythroleukemia.

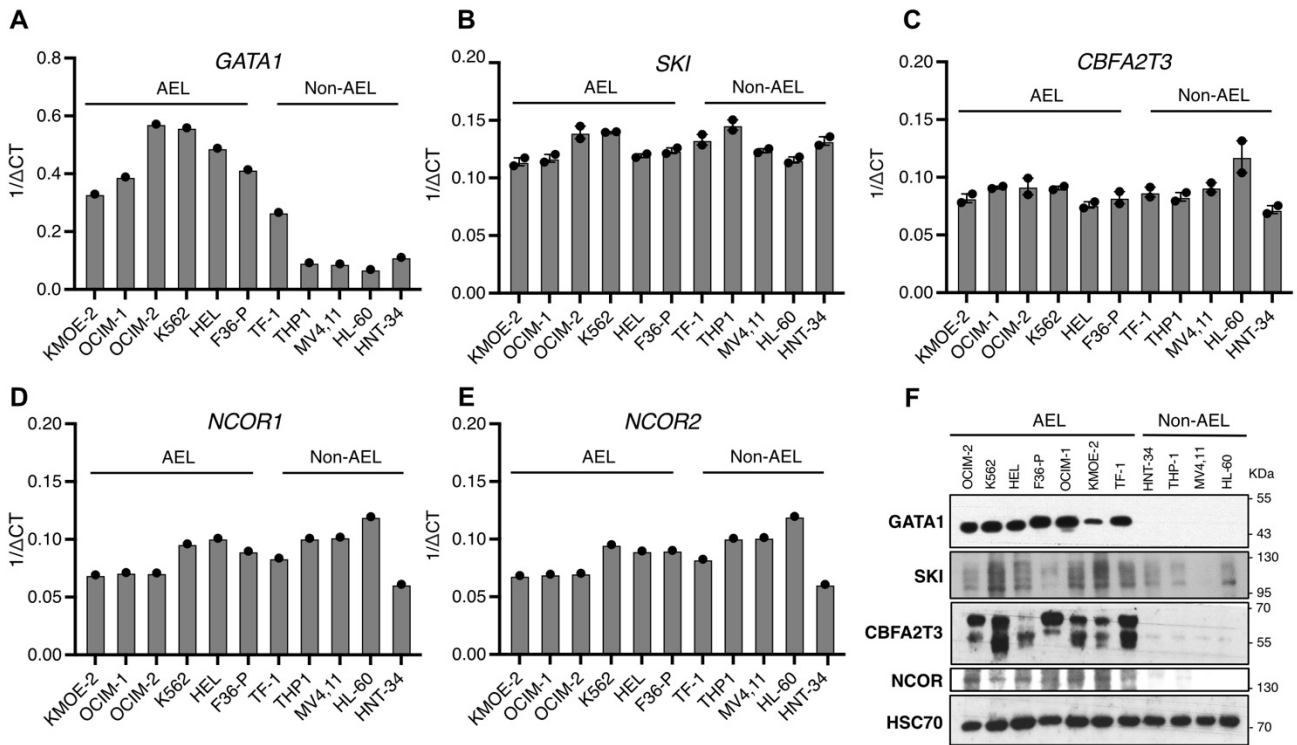


Figure 25. Protein and mRNA levels of transcriptional co-regulators do not always correlate in human AML cell lines. Quantitative RT-PCR of **A) *GATA1*** (n=1), **B) *SKI*** (n=2), **C) *CBFA2T3*** (n=3), **D) *NCOR1*** (n=1) and **E) *NCOR2*** (n=1) from AEL and non-AEL cell lines. CT values are normalized to *GAPDH* expression and shown as relative expression using $1/\Delta CT$. **F)** Protein expression of GATA1, SKI, CBFA2T3, and NCOR in AEL and non-AEL cell lines. HSC70 is used as a loading control.

5.5.4. Effects of exogenous NUP155-FL and deletion mutant expression in hEBST and HUDEP2 cells

NUP155 was found to be the most highly enriched protein in the GATA1 IP/MS analysis in normal erythroblasts (**Fig. 19**). NUP155 is a nuclear pore complex protein regulating movements across the nuclear envelope. NUP155 was initially found to be mutated in atrial fibrillation leading to changes in nuclear localization. The resulting reduction of the protein inhibited the export of Hsp70 mRNA and the import of HSP70 protein in cardiomyocytes. NUP155 knockout mice die before E8.5⁽²¹¹⁾. So far, it is not clear if structural NUPs, such as NUP155, participate in gene regulation in hematopoietic cells. Notably, our interrogation of public databases revealed that the highest *NUP155* mRNA levels were found in erythroid progenitor cells (www.genevestigator.com). Likewise, among cells from the hematopoietic system, erythroid progenitors seem to express the highest *NUP155* mRNA levels (<https://servers.binf.ku.dk/bloodspot/>). We, therefore, hypothesized that NUP155 is somehow involved in normal erythropoiesis also because similar to GATA1, it seemed to be downregulated upon induced erythroid differentiation of primary erythroblasts (**Fig. 26A**).

To obtain a first functional insight into the role of NUP155 in erythroid cells, we tried to functionally interfere with potentially dominant-negative acting mutants as reported by Busayavalasa *et al.* in cardiomyocytes ⁽²¹²⁾. Hereby, we virally expressed full-length NUP155, Δ N- or Δ C-mutants in HUDEP2 cells (**Fig. 26B, C**). Overexpression of NUP155 seemed to be most efficient when cells were transduced with the Δ C mutant (**Fig. 26D, E**). Investigating the erythroid surface markers CD71 and CD235a there was a significant reduction of the double positive population in NUP155FL, Δ N or Δ C mutant in HUDEP2 cells, with the lowest expression in the Δ C mutant cells. All of the transduced cells seemed to be slightly bigger, with a shift towards lower CD235a expression and a slight increase in CD71 compared to the control (**Fig.26F-H**).

This suggests, that cells are more immature, however, when transferring them in differentiation medium, although delayed, cells were still able to mature. Similar experiments in hEBSTs followed the same trend- Whereas control-transduced hEBSTs showed to be very differentiated after 19 days in culture, NUP155-FL-transduced cells seemed to have a delay and appeared to be slightly more immature (**Fig. 26I, J**).

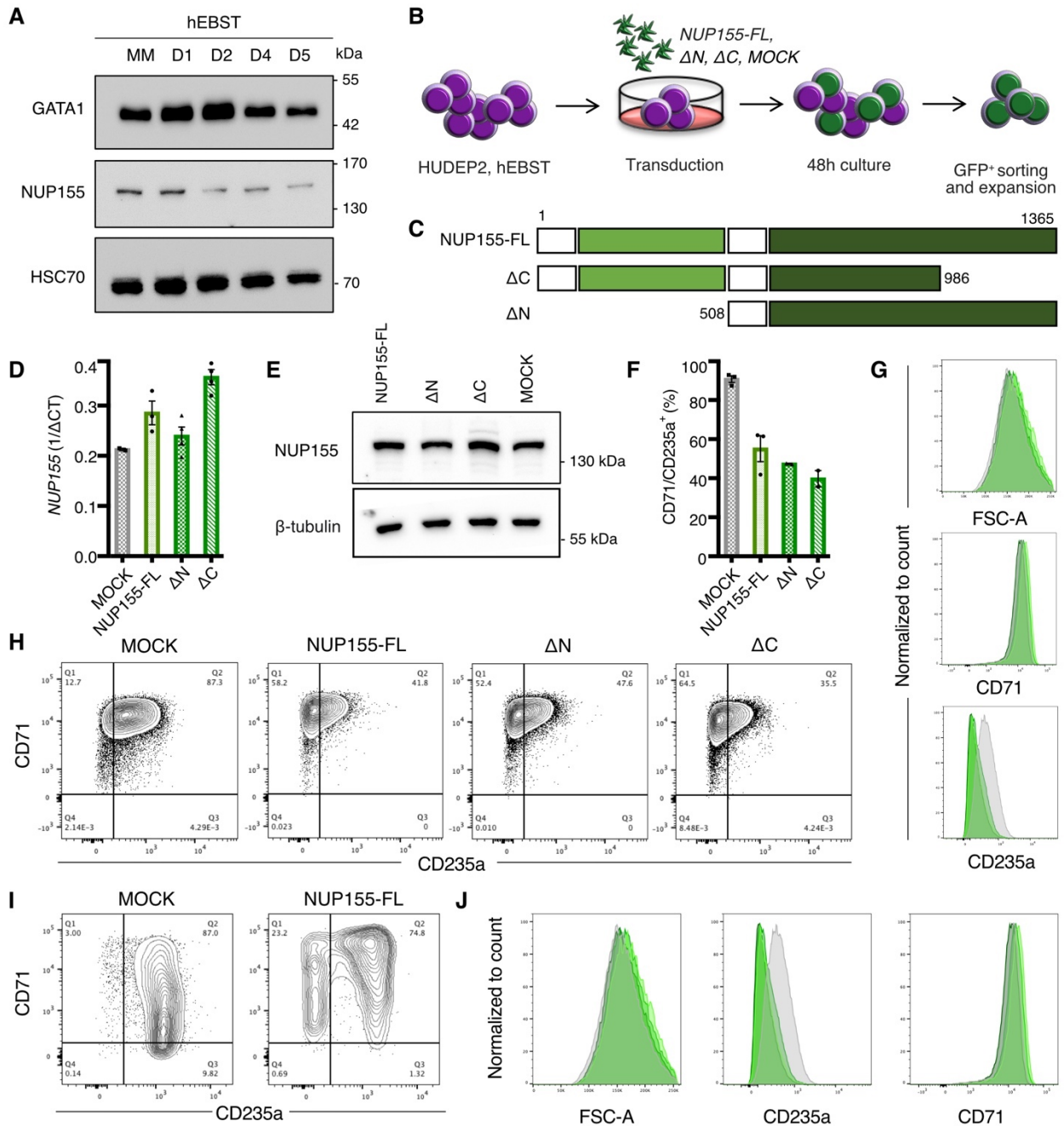


Figure 26. Effects of exogenous NUP155-FL and deletion mutant expression in hEBST and HUDEP2 cells. **A**) Western Blot analysis of GATA1, NUP155 and HSC70 as loading control in primary erythroblasts (hEBST) in maintenance and after 1, 2, 4 and 5 days of differentiation. **B**) Experimental setup: hEBST or HUDEP2 cells are transduced with NUP155-FL (*pMSCV-NUP155-FL-IRES-GFP-puro*); ΔN (*pMSCV-ΔN-IRES-GFP-puro*), ΔC (*pMSCV-ΔC-IRES-GFP-puro*) and MOCK (*pMSCV-IRES-GFP-puro*) for 48h following by GFP+ sorting, recovering and downstream analysis. **C**). Design of NUP155 deletion mutants according to ⁽²¹²⁾. **D**) Quantitative RT-PCR of *NUP155* in HUDEP2 cells. CT values were normalized to *GAPDH* expression and shown as relative expression using 1/ΔCT (n=3 MOCK, NUP155-FL; n=4 ΔN, ΔC). **E**) Western Blot analysis of protein levels of NUP155 and β-tubulin as loading control in HUDEP2 cells. **F**) Flow cytometry analysis of CD71+/CD235a+ population in HUDEP2 cells. **G**) and **J**) Histograms of flow cytometry analysis measuring FSC-A, CD71, and CD235a expression 8 days after sorting in **(G)** HUDEP2 cells and **(J)** hEBSTs. **H**) and **I**) Representative images of flow cytometry analysis of CD71 and CD235a expression in **(H)** HUDEP2 cells and **(I)** 19-day-old hEBSTs (n=2).

5.5.5. NUP155 knockdown in K562 induces erythroid differentiation and enhances erythropoiesis in CD34⁺ CB cells

To modulate NUP155 expression levels in erythroid cells, we explored 4 Dox-inducible NUP155-targeting mir-shRNA knockdown vectors (“936”, “2496”, “331” and “3733”, obtained from the Mercher laboratory) in K562 cells. Cells were transduced, treated for 2 days with 1 ug/ml doxycycline (Dox), and sorted for mCherry⁺ cells followed by immediate RNA and protein extraction (**Fig. 27A**). Sorted cells were maintained in culture on Dox for the following 5 days. Hereby, we observed that shNUP155-transduced cells slowed down in growth dramatically compared to MOCK-transduced cells, except for the “3311” mir-shRNA, which also showed a slightly reduced knockdown efficiency on mRNA and protein level (**Fig. 27B-D**). 2 days after sorting all shNUP155-transduced cells showed an increase in CD71, whereas increased expression of CD235a was only observed starting day 5 accompanied by hemoglobinization of the pellet compared to control transduced K562 cells (**Fig. 27E-I**). Overall, 3 (“936”, “2496”, “3733”) out of 4 mir-shRNAs seem to be efficiently targeting NUP155 expression and showed signs of erythroid differentiation in K562 cells.

Next, we wanted to assess the effects in hEBSTs and therefore isolated CD34⁺ cells from cord blood (CB) and transduced them after 1 day in culture with the construct expressing the “3311” mir-shRNA or MOCK vector control. We selected “3311” as this construct showed the least impairment on cell growth of K562 cells but still upregulation of HBA. We induced the knockdown on day 2 in culture and performed flow cytometry on day 4. We realized that comparison between different CBs was difficult as maturation status was already different in the MOCK-transduced cells. Therefore, we decided to compare mCherry⁺ and mCherry⁻ shNUP155 CD34⁺ CB cells, which showed significant differences. A clear increase in CD34⁺/CD36⁺ population was visible in the shNUP155 mCherry⁺ cells, which was further accompanied by an increase in a CD34⁺ high, CD71⁺ high, and CD235a⁺ high population with the most striking effect on almost 100% expression of CD36⁺ cells (**Fig. 27J, K**).

Conclusively, the knockdown of NUP155 induced erythroid differentiation in K562 cells and also resulted in enhanced erythroid maturation of human CD34⁺ CB cells *in vitro*. The overexpression and knockdown studies of NUP155 are promising for further studies. The interplay between NUP155/HSP70 and GATA1 during erythropoiesis still remains to be explored.

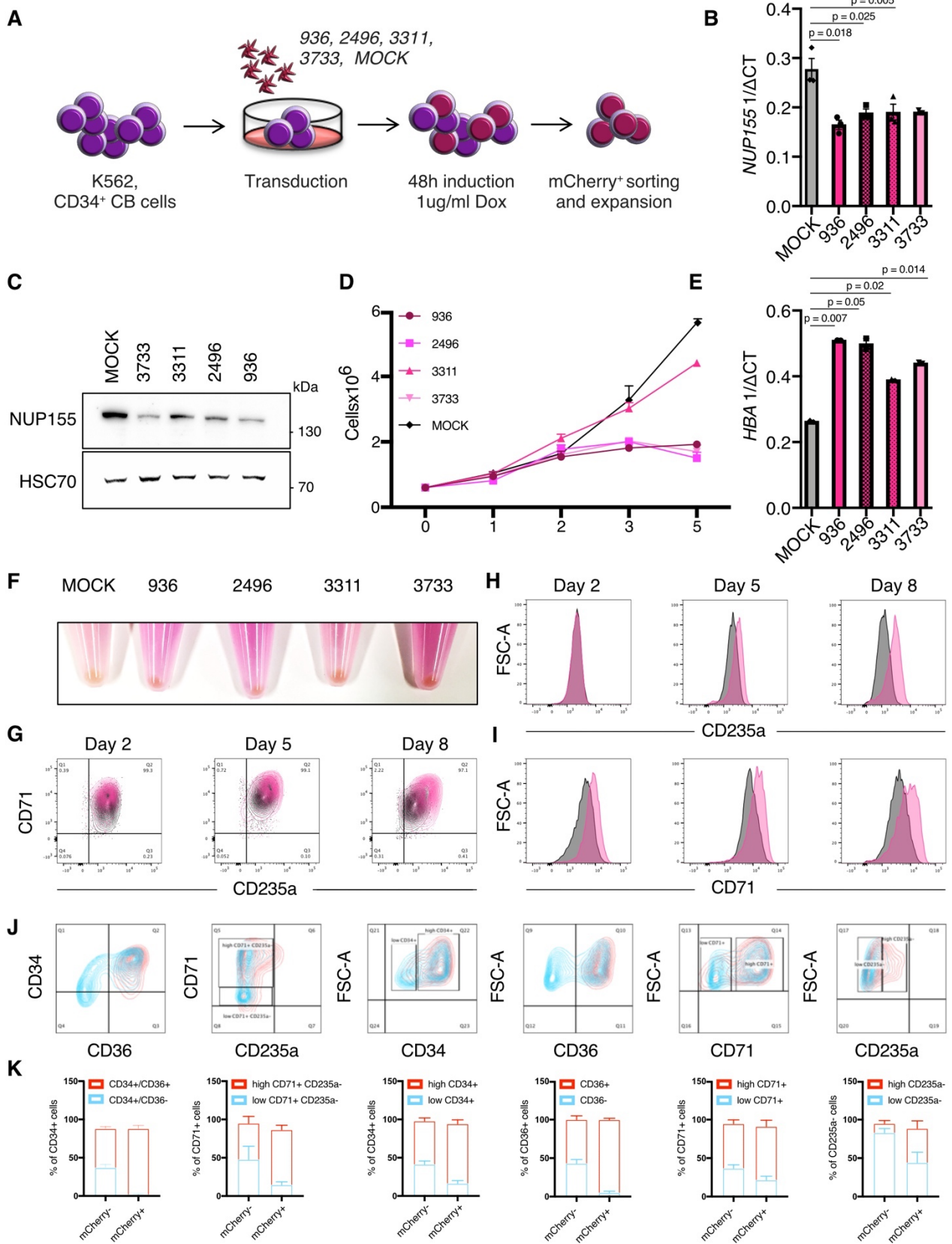


Figure 27. NUP155 knockdown in K562 induces erythroid differentiation and enhances erythropoiesis in CD34⁺ CB cells. **A)** Experimental setup: K562 cells are transduced with DOX-inducible NUP155 shRNA constructs (*pLT3-cherry-shNUP155-936/2496/3311/3733*) for 48h followed by mCherry⁺ sorting, recovering, and downstream analysis. **B)** Quantitative RT-PCR of *NUP155*. CT values are normalized to *GAPH* expression and shown as relative expression using $1/\Delta\text{CT}$ (n=3). **C)** Western Blot analysis of protein levels of NUP155 and HSC70 as loading control. **D)** Cell growth of transduced *pLT3cherry-shNUP155-936/2496/3311/3733* K562 cells followed for 5 days (n=2). **E)** Quantitative RT-PCR of *HBA*. CT values were normalized to *GADPH* expression and shown as relative expression using $1/\Delta\text{CT}$ (n=2). **F)** Cell pellets of transduced *LT3cherry-shNUP155-936/2496/3311/3733* K562 cells 5 days after sorting. **G)** Flow cytometry analysis showing **H)** CD235a and **I)** CD71 expression 2, 5, and 8 days after sorting, representative image for all shRNA clones is shown by *936*-transduced K562 cells. **J)** Representative images of flow cytometry analysis showing CD34, CD36, CD71, and CD235a expression of *3311*-transduced CD34⁺ cord blood (CB) cells 3 days after transduction. **K)** Quantification of the percentages of flow cytometry analysis of transduced *shNUP155-3311*-transduced mCherry⁻ and mCherry⁺ populations (CD34/CD36 n=3; CD71/CD235a n=2; CD34 = n=3; CD36 n=4; CD71 n=3; CD235a n=4).

Chapter 6

General Discussion

The research of my thesis aimed to study the mechanisms behind AEL, which is a rare but very aggressive blood cancer characterized by the uncontrolled proliferation of poorly differentiated erythroid cells. In particular, I tried to characterize molecular mechanisms that are responsible for the differentiation block in AEL.

First, I was involved in the functional characterization of the unexpected AEL phenotype we observed upon targeted inactivation of the *Nsd1* methyltransferase in the murine hematopoietic system. Here we were able to show that in absence of *Nsd1*, transactivation of the erythroid master transcription factor GATA1 was impaired ⁽⁴²⁾. Unexpectedly, we also observed that ablation of the *Nsd1* gene resulted in seemingly enlarged erythroblasts. Therefore, we created another mouse line that allowed us to conditionally ablate *Nsd1* in all cells to track size changes upon ablation in embryonic fibroblasts or erythroblasts.

Second, I was also involved in a collaborative effort that defined the epigenomic landscape of human AEL. Here, we found that in a significant fraction the tumor cells carry lesions which impair GATA1 function. As it is well-established that GATA1 controls terminal erythroid maturation as part of transcriptionally activating and repressing complexes, I aimed to compare potential GATA1 protein interactions in normal and malignant erythroblasts. To functionally characterize putative differential GATA1 interactors, I performed a targeted CRISPR/Cas9 screen in an AEL cell line. Ongoing experiments are validating the novel molecular players we identified.

6.1. NSD1, a lysine methyltransferase in leukemia and overgrowth syndromes

Does NSD1 control cell size?

In addition to its role in leukemia and multiple types of other cancers, *loss-of-function* mutations of *NSD1* have been proposed to be a molecular correlate of the SOTOS congenital overgrowth syndrome ⁽²¹³⁾. Strikingly, we found that loss of the *Nsd1* gene in fetal liver hematopoietic cells in *VaviCre;Nsd1^{fl/fl}* mice resulted in seemingly enlarged erythroblasts (**Appendix 2, Fig. 4B**) ⁽⁴²⁾. To further explore this observation in different cell types, we generated a tamoxifen-inducible mouse line by crossing *Nsd1^{fl/fl}* mice with an *UBC-Cre-ERT²* mouse line to dynamically follow size changes over time. Isolation of adult erythroblasts showed that the cultures from non-induced *Nsd1^{fl/fl};Cre-ERT²* and *Nsd1^{fl/+};Cre-ERT²* control mice always contained cells expressing erythroid surface markers (CD71⁺/Ter119^{+/-}) and additionally also cells expressing myeloid surface markers (Gr1^{+/-}/Mac1^{+/-}). Induction of *Nsd1* cleavage, however, resulted in the disappearance of myeloid cells over time and the proliferation of a pure erythroblastic population, which we also observed in our *Vav-iCre;Nsd1^{fl/fl}* mouse model (**Fig. 15C, D**). We did not explore whether induced ablation of *Nsd1* in *Nsd1^{fl/fl};Cre-ERT²* mice *in vivo* will also induce the AEL-like phenotype.

NSD1 catalyzes mono- and di-methylation of H3K36 which is essential for several fundamental cellular functions including DNA repair, genomic stability, transcription regulation, or cell cycle progression ⁽²¹⁴⁾. The connection between overgrowth and NSD1 deficiency was so far only studied in flies by CRISPR/Cas9-mediated deletion of the NSD1 homolog (*Drosophila dNSD*) ⁽¹⁶³⁾. Strikingly, *dNSD* inactivation resulted in increased body size of *Drosophila* larvae. Upon TAM-induced ablation in *Nsd1^{fl/fl};Cre-ERT²* erythroblasts, we observed a reduction of the G₀ phase, and accumulation of cells in the G₁ phase of the cell cycle potentially increasing growth of cells (**Fig. 15F, G**). Strikingly, a very recently published paper (October 2022), claimed that another H3K36me HKMT so-called SETD2 regulates cell size ⁽²¹⁵⁾. They performed Dox-inducible SETD2 knockdown in human retinal pigment epithelial cells transduced with the telomerase reversed transcriptase gene (*RPE1-hTERT*) and found an increased cellular protein content which correlated with increased cell volume and size. Furthermore, SETD2 depletion in mouse embryonic stem (ES) cells by endogenously tagging with a dTAG (degradation tag) corroborated their findings suggesting that size changes are independent of the cell type. Although cell cycle disruption was only seen in one of their models by a slight increase in cells in G₁ and a decrease in G₂M upon SETD2 loss, this strongly supports our finding in *Nsd1^{fl/fl};Cre-ERT²* erythroblasts that showed slight alterations in cell cycle. Investigation of cyclin-dependent kinase and genes involved in cell cycle regulation by mRNA might help to elucidate the contribution of cell cycle in our model.

How cell size is regulated by NSD1 or SETD2 is still unclear. It has been proposed that both enzymes are also able to methylate non-histone substrates which opens the scope for multiple, yet to-be-defined involved pathways ^(215,216). *In vitro* studies mostly in 293T cells suggested that NSD1 can methylate linker histones (e.g. H1.5-K168) and 50 non-histone proteins that are being recognized by the SET domain of NSD1 ⁽²¹⁶⁾.

Gene expression profiles of skin fibroblasts from 15 SOTOS patients with *NSD1* mutations displayed reduced expression of genes involved in cell cycle G₂M checkpoints, which usually prevents cells from entering mitosis and stops cell proliferation. On the other hand, genes involved in the regulation of nuclear division, meiotic cell cycle, and kinetochore organization were significantly upregulated suggesting that *NSD1* is important for proper cell cycle regulation, particularly at the G₂M checkpoint ⁽²¹⁷⁾.

Interestingly, germline 5q35 micro-duplications leading to *NSD1* overexpression were associated with undergrowth phenotypes ⁽²¹⁸⁾. Modeling this phenotype in *Drosophila* led the authors to the conclusion that NSD1 overexpression was associated with the downregulation of mTOR signaling suggesting that NSD1 gene dosage might be critical to control cell growth. As branched-chain amino acids, such as leucine, have been found to directly activate mTOR, the *Drosophila* larvae were supplied with leucine which triggered mTOR reactivation and rescue of the phenotype ⁽²¹⁹⁾. This work

suggests that loss of *NSD1* results in the upregulation of mTOR-related pathways, which could be therapeutically exploited in SOTOS patients by reducing leucine in their diet or targeting mTOR by small molecules.

So far, dysregulated erythropoiesis has not been reported in SOTOS patients. It is however not clear whether one can expect a significant erythroid phenotype, as SOTOS individuals carry heterozygous *NSD1* gene mutations. Notably, our *VaviCre;Nsd1^{fl/+}* heterozygous mice did not exhibit a hematopoietic phenotype and never developed any disease ⁽⁴²⁾. Whether SOTOS mutations somehow affect the expression/function of the remaining WT allele remains to be explored. Alternatively, cells from different organs/lineage might be differentially affected by reduced or increased *NSD1* gene dosage.

Since material from SOTOS patients is very limited, we wondered whether we can model the increased cell size change upon *Nsd1* inactivation in non-hematopoietic cells. We isolated mouse embryonic fibroblasts (MEFs) from *Nsd1^{fl/fl};Cre-ERT2* embryos and conditionally ablated *Nsd1 in vitro* by providing Tamoxifen (TAM) to the culture. This project was continued by a master's student. His results suggested immediate loss of *Nsd1* (2 or 4 days after TAM treatment), significantly increased cell growth, and eventually also increased size of MEFs. However, it appeared that measuring cell size changes needs to be tightly controlled as MEFs also enlarge with every passage, and reproduction of the master student's observations was not possible by two independent researchers. Investigating the immediate cellular changes in erythroblasts after 2 or 4 days of induced *Nsd1*-ablation only showed a slight but not significant increase in cell size (**data not shown**). This suggests that the significantly increased cell size observed upon fetal inactivation of *Nsd1* may not happen immediately after ablation. While the connection between *Nsd1* gene dosage and increased cell size is clearly given, the exact contribution and time frame need further investigation.

NSD1 at the interface of histone modification, DNA methylation, and the PRC2 complex

In our hematopoietic *Nsd1^{-/-}* mouse model, we observed a decrease of H3K36me1 levels (**Fig. 13A**). As we did not clearly observe globally reduced H3K36me3 protein levels, partial compensation by other H3K36 mono- and/or di-methyltransferases such as NSD2, NSD3, ASH1L, and SMYD2 or the tri-methylase SETD2 can be assumed (**Fig. 13A**) ^(163,214). Another study found that NSD1-mediated H3K36me2 marks are essential for the recruitment of the DNA methyltransferase DNMT3A and subsequent DNA methylation at intergenic regions. They showed that loss of *Nsd1* in mouse ES cells not only reduced H3K36me2 at intergenic regions but also resulted in re-localization of the DNA methyltransferase DNMT3A to H3K36me2-modified gene bodies. These regions with reduced H3K36me2 lacking DNMT3A were also characterized by DNA hypomethylation ⁽²²⁰⁾. Notably, impaired intergenic DNMT3A localization and hypomethylated DNA are also commonly found in SOTOS

patients⁽²¹³⁾. This is particularly interesting as NSD1-mediated H3K36me2 marks were shown to co-localize with H3K27me2, however upon NSD1 loss spreading of H3K27me3 marks can occur^(151,221).

H3K27me2/3 marks are set by the Polycomb repressive complex 2 (PRC2), which consists of the core SET domain proteins EZH1/2, EED, and SUZ12⁽²²²⁾. Dysregulation of PRC2 genes, such as SUZ12 and EED was also seen in patients with SOTOS syndromes⁽²²³⁾. These authors also computationally explored the binding of EZH2 to chromatin and showed that not only did it bind to hypomethylated regions but in particular to genes that are downregulated in SOTOS⁽²²³⁾. The higher occupancy of these genes by EZH2 strongly suggests an essential role of PRC2-mediated silencing in the SOTOS phenotype. Abnormal PRC2 activity has been described in additional overgrowth syndromes illustrated by EZH2 mutations frequently found in WEAVER syndrome⁽²²⁴⁾. Furthermore, EED mutations frequently occur in COHEN-GIBSON syndrome and SUZ12 mutations have been observed in IMAGAWA-MATSUMOTO syndrome, which strongly suggests that disruption of the PRC2 is a general mechanism underlying overgrowth syndromes^(225,226). However, so far, it is not clear how NSD1, EZH2, EED, and SUZ12 mutations result in similar phenotypic characteristics. Additional investigation of these phenotypes at the epigenomic level is necessary.

To address the involvement of the PRC2 complex in the context of *Nsd1* ablation, we tested two EZH2 small molecule inhibitors GSK126 and DZNep in *Nsd1*^{-/-} erythroblasts (**Fig. 13**)^(155,227). DZNep is a nonspecific PRC2 inhibitor, reducing several components of the PRC2 complex such as EZH2, EED and SUZ12, and has been shown to exhibit anti-tumor effects in various cancers⁽²²⁸⁾. GSK126 is a highly selective S-adenosyl-methionine-competitive small molecule inhibitor of the EZH2 methyltransferase activity and has shown promising effects in xenograft cancer mouse models⁽²²⁹⁾. However, the exact mechanisms of their activity in erythroid cells are still poorly understood. We chose to target in particular EZH2, as, as outlined above, loss of NSD1 results in PRC2-mediated spreading of H3K27me3. In addition, PRC2 dysfunction in multiple cancers (e.g. breast cancer, glioblastoma, melanoma) is largely based on EZH2 overexpression rather than high levels of EED or SUZ12^(222,230). Accordingly, equivalent to EZH2 overexpression, *gain-of-function* mutations have been described in B-cell lymphomas and generally result in increased H3K27me3 levels⁽²³¹⁾. In AML, EZH2 is commonly overexpressed in high-risk MDS and AML⁽²³²⁾. However, in myeloid malignancies, most frequently occurring in MDS/MPN (10-13%) or myelofibrosis (13%) patients, also *loss-of-function* mutations of *EZH2* have been observed⁽²³³⁾. Therefore, it seems that EZH2 has a bivalent role and is not only considered a tumor suppressor gene for leukemia initiation, but also an oncogene in leukemia maintenance⁽²³⁴⁾. We observed that treatment with the selective EZH2 inhibitor GSK126 reduced H3K27me3 as expected, but did not overcome the differentiation block of *Nsd1*^{-/-} erythroblasts (**Fig. 13B-D**). This suggests that removing repressive marks is not sufficient to achieve an active chromatin state and that other EZH2-independent mechanisms might be involved. Because

PRC2 seems to silence GATA1 target genes during erythroid differentiation, Fujiwara and colleagues postulated that DZNep treatment may also influence erythropoiesis ⁽²³⁵⁾. Interestingly, they found upregulation of erythroid genes upon DZNep exposure of K562 AEL cells or normal human CD34⁺ cells. We could corroborate these findings, as DZNep treatment of *Nsd1*^{-/-} erythroblasts reduced cell growth and induced partial erythroid differentiation (**Fig. 13E-H**). To identify the underlying mechanisms, they performed siRNA-mediated EZH2 knockdown which surprisingly did not upregulate erythroid genes in K562 cells, suggesting that solely inhibition of EZH2 is insufficient to restore an active chromatin state and could also explain the inability of GSK126 to induce erythroid differentiation. Thus, one can speculate that other proteins might be responsible for the induction of erythroid target genes. Indeed, Fujiwara *et al.* found reduced levels of the corepressor CBFA2T3 protein but not on mRNA level, which emphasizes the need to look into the proteome rather than transcriptome to understand the mechanisms of erythroid differentiation. In line with these findings, a more recent publication demonstrated that combining an HDAC1 inhibitor (Entinostat) with an EZH1/2 inhibitor (UNC1999) restored erythroid differentiation and alleviated transcriptional repression in SPI1 overexpressing erythroleukemic mouse cells ⁽²³⁶⁾. One of the proposed SPI1-activated genes was CBFA2T3. They also linked this study to GATA1, since SPI1 required GATA1 to bind to chromatin and exert its repressive function. Therefore, one can postulate that protein interactions and/or chromatin binding determine cell fate decisions, and it would be interesting to look into the proteome, GATA1 interactome, and chromatin occupancy upon DZNep and/or GSK126 treatment with and without an HDAC inhibitor in our *Nsd1*^{-/-} erythroblasts.

6.2. Profiling the proteomic landscape in normal and malignant erythropoiesis

The need to combine transcriptomic and proteomic data is crucial when studying erythropoiesis, since it is a gradual process driven by a variety of transcription factors (TFs), resulting in the activation of transcriptional programs and dramatic changes in chromatin architecture and abundance of proteins.

Gene regulatory networks (GRNs) have been computationally created that integrate multiple transcriptomic datasets and thereby predict the cellular response to signals from the environment leading to the production of RBCs. While it is possible to identify autoregulation, feedback loops, and regulation of other target genes by particular TFs, this approach looks only at gene expression programs and does not take into account that proteins are the actual drivers of transcriptional regulation ⁽¹⁶⁶⁾. Therefore, several labs (Brand, Frayne, and Sankaran) have investigated normal red blood cell formation on the proteome level. To identify the temporal proteomic changes during erythropoiesis in single cells, Brand and colleagues measured 16 endogenous TFs and 11 cell surface markers by a multidimensional single-cell proteomic approach using CyTOF at 13 sequential time

points ⁽¹⁶⁷⁾. Quantification of proteins by mass spectrometry became very powerful in the last years and various techniques have been established to identify proteomic dynamics. Relative quantification allows to follow changes in the abundance of proteins across different samples or different time points and can be achieved by isotopically labeling peptides or label-free approaches. The Brand lab has used iTRAQ (isobaric tag for relative and absolute quantification) to label peptides to study erythropoiesis ⁽⁶⁹⁾. The Frayne lab has used a slightly different labeling tag called TMT (tandem mass tag) to compare the proteome of adult and cord erythroid cells ⁽²³⁷⁾. The Sankaran lab also applied TMT to measure ribosomal protein abundance in Diamond Blackfan anemia patients ⁽²³⁸⁾.

To address the changes in the proteome of hEBSTs and HUDEP2 cells upon induced erythroid differentiation I applied TMT labeling (**Fig. 23**). As expected, the most significantly identified proteins becoming more abundant upon differentiation were erythroid-specific, such as Spectrin-alpha; erythrocytic 1 (SPTA1), Spectrin-beta; erythrocytic (SPTB) and Ankyrin 1 (ANK1), which are components of the erythrocyte plasma membrane. Furthermore, Albumin (ALB) the most abundant protein in human blood was significantly more expressed in differentiating cells. Iron binding and oxygen binding proteins, Hemoglobin subunit beta (HBB) and Hemoglobin subunit alpha1 (HBA1) as well as the Solute Carrier Family 4 Member 1 (SLC4A1) transmembrane transporter were also significantly captured by assessing the proteomic landscape during erythropoiesis of CD34⁺ HSPCs by the lab of Mathias Mann ⁽¹⁷⁵⁾. Our data is coherent with other studies that address the proteomic changes during erythropoiesis and therefore shows the feasibility of the TMT approach in our hands ^(69,175,237). Ideally, one would combine all available datasets to identify protein regulatory networks and obtain a more comprehensive picture of the proteome during erythropoiesis.

I also used TMT to assess quantitative proteomic differences in human erythroleukemic cells compared to normal human erythroblasts (hEBST, HUDEP2) as to date no such data appeared available (**Fig. 17**). Two of the most significantly abundant proteins in leukemic erythroblasts were enolase 1 (ENO1) and the intercellular adhesion molecule 1 (ICAM1). ENO1 is a transcriptional co-repressor generally associated with promoting glycolysis and energy metabolism in cancer cells and is overexpressed in more than 70% of cancers ^(181,239). Strikingly, ENO1 has been found as one of the most overexpressed genes in AML patients ⁽²⁴⁰⁾. ICAM1 is expressed on hematopoietic progenitors including BFU-E, CFU-E, CFU-GM, and pro-erythroblasts, whereas its expression decreases upon erythroid differentiation, overall supporting the presence of impaired differentiation of transformed erythroblasts ⁽¹⁸⁰⁾. Metascope analysis for TTRUST (www.grnpedia.org/trrust⁽¹⁸⁵⁾), which is a database that computationally calculates transcriptional regulatory networks suggested that the 6.8% of significantly identified proteins from leukemic erythroblasts were regulated by Specificity Protein 1 (SP1). SP1 belongs to the Sp/KLF family of TFs and is known to regulate cell proliferation, differentiation, and growth, therefore, playing a key role in cancers ⁽²⁴¹⁾. Strikingly, *Sp1* knockdown

induced erythroid differentiation in K562 cells ⁽²⁴²⁾. To emphasize, we did not find SP1 protein itself in our TMT analysis, but SP1 came up as the most significant transcriptional regulatory network suggesting that proteins in leukemic erythroblasts regulated by SP1, eventually impair erythroid differentiation.

On the counterpart, in normal erythroblasts, we identified proteins that were mainly regulated by GATA1. As the master regulator of erythroid differentiation, a critical level of GATA1 is needed to induce transcription of erythroid genes thereby allowing cells to differentiate. Several mouse models have shown that reduced GATA1 levels led to severe anemias and impaired erythropoiesis. Furthermore, *loss-of-activity* GATA1 mutations resulting in the expression of a short GATA1 isoform (GATA1s) have been described in Diamond Blackfan anemia and acute megakaryoblastic leukemia (AMKL) in children with Down's syndrome with erythroid defects ^(50,243). Therefore, it can be assumed that blocked differentiation of leukemic erythroblasts involves reduced GATA1 levels and/or function, respectively. Indeed, comparing the proteome of normal and leukemic erythroblasts, we identified more GATA1 protein in the first group. However, we found that AEL cells generally express more GATA1 compared to other non-AEL cell lines (**Fig. 25**), suggesting that it is not a simple question of GATA1 abundance but rather linked to its activity that seems to be disturbed.

6.3. Dissecting GATA1-interacting proteins in normal and malignant erythroid cells

GATA1 overexpression partially induces erythroid differentiation of human AEL cells

To prove, that GATA1 activity is impaired we overexpressed GATA1 in two AEL cell lines, K562 and HEL. In both cell lines, similar to GATA1 overexpression in *Nsd1*^{-/-} erythroblasts we observed at least partial erythroid differentiation of AEL cells (**Fig. 18 & Suppl. Fig. 2A-D**). However, it remains unclear how erythroid differentiation was induced. One can speculate that only the additional exogenous GATA1 is able to activate erythroid genes by finding its way to chromatin. Therefore, investigating GATA1 binding to chromatin could be performed, since endogenously and exogenously expressed GATA1 could be discriminated based on the HA-tag that was cloned in the cDNA construct. As GATA1 is mainly found in the form of protein complexes and depending on its partners activates or represses transcription of erythroid genes, another possible scenario could be that exogenously expressed GATA1 titrates away GATA1-bound co-repressors which could prevent it from exerting its activating functions. Therefore, we investigated GATA1 interaction partners in a variety of cell models.

GATA1 interacting proteins upon differentiation of mouse and human erythroid cells

We established a variety of dynamic murine and human models to study GATA1 interactions in steady state and differentiating conditions. In contrast to previous studies, we characterized GATA1

complexes in erythroid cells by pulling down endogenous GATA1 thereby preventing interferences of epitope tags with cellular and molecular functions. We are however fully aware that we have most likely not captured all GATA1 interacting proteins by our IP/MS analysis since low abundant, unstable, and/or poorly ionized peptides will most likely be lost during the experimental preparations. Nevertheless, we managed to enrich GATA1-associated proteins in MEL cells which can be induced to differentiate by supplying the cultures with DMSO (**Fig. 11**). Since we found that GATA1 overexpression is able to induce partial erythroid differentiation in *Nsd1*^{-/-} erythroblasts, we compared immunoprecipitated GATA1 interacting proteins in *Nsd1*^{-/-} erythroblasts overexpressing GATA1 versus mock control (**Fig. 12**). Moreover, we performed GATA1 IP/MS analysis in *Nsd1*^{-/-} erythroblasts expressing a SET-mutant *Nsd1* (differentiation impaired) or a WT-*Nsd1* (differentiation induced) construct (**Appendix 2, Fig. 8**)⁽⁴²⁾. We furthermore explored GATA1 interactions in differentiating primary human erythroid cells (hEBSTs), which were established from PBMC from healthy donors (**Fig. 24**).

Whereas slight differences in pulled-down proteins exist between all these models, independent of normal or malignant erythroblasts, in steady-state (maintenance) conditions, we found GATA1 associated with transcriptional co-repressors including CBFA2T3, NCOR1, NCOR2, HDAC3, and SKI. In differentiating conditions, these associations seem to be lost, suggesting that in steady-state conditions, GATA1 is situated in a mostly repressive environment. Importantly, the inactivation of some of these co-repressors such as CBFA2T3 and SKI has been shown to induce erythroid differentiation: shRNA-mediated knockdown of CBFA2T3 in K562 and CD34⁺ cells resulted in upregulation of hemoglobin *HBB*, *HBA*, and *ALAS2* expression, whereas CBFA2T3 overexpression reduced CD34⁺ erythroid differentiation⁽²⁰⁵⁾. It is unclear whether CBFA2T3 is able to entirely block erythroid differentiation of human CD34⁺ cells or if differentiation is only delayed. Furthermore, we found that reduced SKI levels in *Nsd1*^{-/-} erythroblasts were able to induce erythroid differentiation⁽⁴²⁾. Interestingly, this has previously been shown by Ueki *et al.* in K562 cells, where SKI was able to interact with GATA1 thereby facilitating repression of erythroid differentiation⁽¹⁹⁰⁾.

These data support our hypothesis, that erythroid differentiation could be a “stoichiometrically-controlled” event mediated by removing repressive proteins from GATA1 to activate transcription. Of note, we always identified more GATA1 interaction partners in erythroblasts kept in maintenance compared to differentiation-induction conditions. This could be explained by the fact that co-repressors seem to be more abundant than co-activators in the nucleus of erythroid cells⁽⁶⁹⁾. It is therefore likely that we were unable to identify interactions of GATA1 with co-activators simply due to their rarity in the nucleus.

Identification of novel GATA1-associated proteins that maintain the differentiation block of malignant erythroblasts

By comparing GATA1 protein complexes we identified several proteins that significantly differed between normal and malignant erythroblasts, including known but also novel potential GATA1 interactors (**Fig. 19**). To obtain an unbiased view on their functional role in erythroid differentiation, we decided to target these proteins in a CRISPR/Cas9-based screen. Based on efficient transduction as well as the observation that GATA1 overexpression resulted in partial erythroid differentiation, we used K562 cells for the screen. We established *SpCas9*-K562 cells (stably expressing Cas9) and tested whether these cells still retain the permissiveness to differentiate when overexpressing GATA1 (**Suppl. Fig. 3**). We targeted 116 candidate genes that we found in the proteome and GATA1-interactome studies which led to the identification of 4 clusters based on cell survival and erythroid differentiation (**Fig. 20**). Although we mainly focused our attention on genes that only showed effects on erythroid differentiation (“cluster-4”), it would also be worth to dig deeper into genes/proteins that when inactivated result in differentiation as well as reduced cell survival (“cluster-3”). Amongst “cluster-4” genes, one of the genes that had the strongest impact on erythroid differentiation was CBFA2T3, confirming previous findings and also the validity of our approach. However, CBFA2T3 activity does not seem to be specific for malignant cells, as K562 as well as human CD34⁺ cells react to its reduced expression ⁽²⁰⁵⁾. Notably, we found that the association of CBFA2T3 to GATA1 was more abundant in normal than transformed erythroblasts, suggesting that although it is functionally important for normal erythroid differentiation, increased expression is most likely not sufficient to transform cells. Therefore, we focused our attention on proteins that were significantly more enriched in GATA1 pull-downs of AEL cells and showed effects on erythroid differentiation in our CRISPR/Cas9 screen. Among them were proteins such RCOR1, KPNA2, CCT4, CCT7, and PARK7, which will be discussed in detail in the following paragraphs.

Notably, although we found the SKI co-repressor more enriched in the GATA1 pulldowns from AEL cells and our previous experiments in *Nsd1*^{-/-} cells showed functional relevance, inactivation of SKI did not affect the differentiation or survival of *SpCas9*-K562 cells. This could be due to the selection of sgRNA sequences leading to potentially insufficient KO efficiency. As SKI is very well established as a repressive GATA1-interacting protein we did not test additional sgRNAs. However, this observation indicates that we cannot exclude that some genes of “cluster-1” are potentially exerting effects on the cell phenotype.

RCOR1 and the coREST transcriptional repressor complex

RCOR1 is part of a transcriptional coREST repressor complex that is formed together with HDAC1/2 and LSD1/KDM1A⁽²⁴⁴⁾. LSD1 is frequently highly expressed in many cancers, including AML, and has therefore been extensively studied⁽²⁴⁵⁾. *Lsd1* gene inactivation in mice resulted in embryonic lethality before E7.5, whereas conditional deletion in mice led to failure of multilineage hematopoiesis, affecting granulopoiesis, erythropoiesis, and thrombopoiesis^(82,246). As previously mentioned LSD1 is a histone H3K4 demethylase and is generally recruited to chromatin to silence transcription by forming complexes with other proteins⁽⁴⁶⁾.

Flag-TAL1 IP/MS analysis in K562 cells revealed that LSD1 interacts with TAL1, RCOR1, and HDAC1/2⁽²⁴⁷⁾. Notably, TAL1 is part of the activating pentameric GATA1 complex together with LDB1, E2A, and LMO2 (**Fig. 28A**). They also found that TAL1 recruits LSD1 to repress target genes, whereas knockout of LSD1 in MEL cells resulted in de-repression of these TAL targets. As we never found GATA1 interacting with LSD1 or HDAC1/2 we could envision another hypothetical model in which repression of the pentameric complex is achieved through binding of LSD1 to TAL1 and RCOR1 serving as a bridge to connect to GATA1 and simultaneously binding HDAC1/2 (**Fig. 28B**). The exact interactions, however, would need to be validated by additional co-IPs.

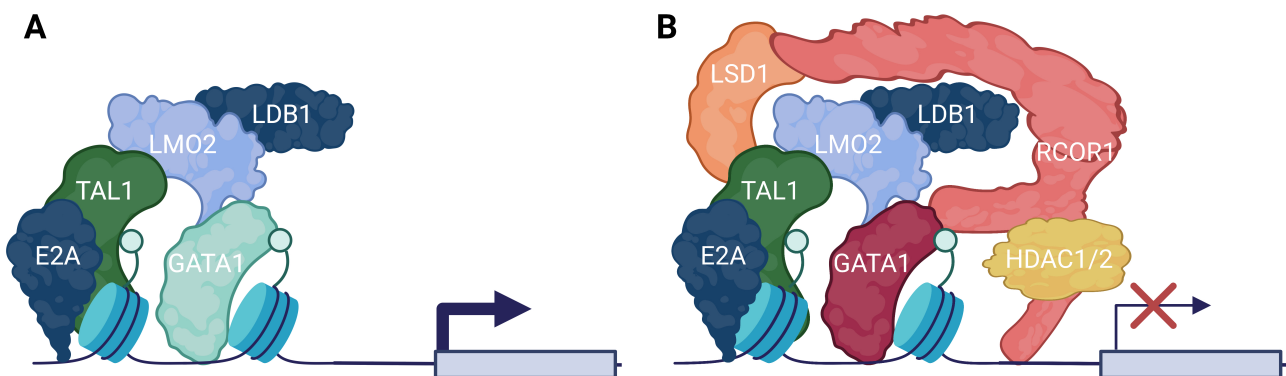


Figure 28. Hypothetical model of GATA1 protein interactions involving the coREST transcriptional corepressor complex. A) The pentameric GATA1/E2A/TAL1/LDB1/LMO2 complex activates transcription. **B)** Transcriptional silencing is achieved by binding of LSD1 to TAL1, RCOR1 bridging to GATA1, and binding of HDAC1/2 to RCOR1 (created with Biorender.com).

Earlier data from the Orkin lab indicate that gene repression is exerted by recruitment and association of LSD1 and RCOR1 to GFI-1B⁽⁸²⁾. GFI-1B controls gene expression during the development and maturation of erythrocytes and megakaryocytes⁽²⁴⁵⁾. Two other groups found that GATA1 interacts with GFI-1B. Rodriguez *et al.* found that GFI-1B was pulled down with GATA1 in MEL cells and they proposed that GATA1 forms a complex with GFI-1B to repress proliferation-associated genes such

as *Myc* and *Myb* ⁽⁷⁰⁾. Huang *et al.* co-immunoprecipitated flagged GFI-1B upon co-expression of Myc-GATA1 in 293T cells ⁽²⁴⁸⁾. Of note, we did not detect GFI-1B in our GATA1 IP/MS analyses and therefore primarily focused our work on RCOR1. Orkin and co-workers showed that siRNA-mediated knockdown of LSD1 impaired induced maturation of MEL cells, whereas knockdown of RCOR1 did not ⁽⁸²⁾. However, in the megakaryoblastic cell line L8057, both LSD1 and RCOR1 inhibition impaired differentiation. These findings suggest that RCOR1 has different functions depending on the lineage context. Therefore, I recently explored the effects of RCOR1 inhibition and overexpression in human erythroid cell lines. I reduced RCOR1 expression by doxycycline-regulated mir-shRNAs in K562 cells, which confirmed and extended our findings of the CRISPR/Cas9 screen to induce erythroid differentiation based on an increased erythroid surface marker and *HBA* expression (**Fig. 21A-F**). On the contrary, overexpression of RCOR1 in HUDEP2 cells showed a shift towards a more immature phenotype characterized by reduced *HBA* expression (**Fig. 21G, H**). We also explored the possibility of targeting the coREST complex with an inhibitor that targets HDAC1 and LSD1 simultaneously, called Corin. Corin is a synthetic hybrid derivative from the class I HDAC inhibitor entinostat and an LSD1 inhibitor (tranylcypromine analog)⁽²⁰⁸⁾. It was suggested that binding of the Corin warhead to one target, enhances the accumulation of the warhead to the other target. Furthermore, Corin showed superior anti-proliferative effects in comparison to monofunctional inhibitors and less toxicity in melanoma, squamous cell carcinoma cell lines, and patient-derived diffuse intrinsic pontine glioma (DIPG) cell lines^(208,249). In this context, gene ontology analysis revealed that Corin induced a transcriptional signature that was seen in the normal brain compared to DIPG, associated with brain development and neurogenesis and upregulation of neuronal differentiation genes, suggesting that Corin can overcome the differentiation block seen in DIPG ⁽²⁴⁹⁾.

Therefore, we hypothesized that Corin would release the coREST-mediated repression and induce differentiation also in erythroid cells. In line with this, I found that Corin treatment of K562 cells favored erythroid differentiation, seen by an increase of CD71 and upregulation of *HBA* mRNA expression (**Fig. 22**). Interestingly, Anastas *et al.* previously reported that Corin increased HDAC-targeted H3K27ac and LSD1-targeted H3K4me1 levels in particular at genomic regions which had binding sites for RCOR1 ⁽²⁴⁹⁾. It would therefore also be interesting to look into RCOR1 binding and H3K27ac and H3K4me1 marks by ChIP-Seq in erythroleukemic cells treated with Corin.

Notably, Ishikawa and colleagues recently reported that a LSD1 inhibitor (T-3775440), which disrupts LSD1-GFI-1B interaction, led to trans-differentiation of AEL and AMKL cells towards granulomonocytic-like lineage cells measured by an increase of the granulocytic marker CD86 and a decrease in erythroid surface marker CD235a ⁽²⁵⁰⁾. So far, we have not explored the possibility of trans-differentiation upon Corin treatment in AEL cells, yet. However, it is interesting that we mostly see effects on CD71, whereas CD235a remained unaffected. One can therefore speculate that

disruption of the LSD1-GFI-1B axis may push the cells towards a monocytic phenotype, whereas interference with the coREST complex by either combinatorial inhibition of LSD1 and HDAC or knockout of RCOR1 itself seems to favor differentiation into erythroid cells. It remains to be elucidated how GATA1 comes into this game and if/how GATA1 and RCOR1 binding to chromatin is altered upon Corin treatment and/or RCOR1 inhibition.

Regulation of GATA1 by the karyopherin KPNA2?

The transport of proteins in and out of the nucleus is a tightly controlled process, which is exerted by karyopherins. Karyopherins are nuclear transport proteins that serve as cargo to import or export proteins via the nuclear pore complex (NPC) ⁽²⁵¹⁾. This is particularly important for proteins >40-60 kDa since they cannot passively pass the nuclear membrane but need to be bound to cargo proteins ⁽²⁵²⁾. Overexpression of importins and exportins has been reported in hematologic malignancies. Especially exportin 1 (XPO1) has gained increasing interest in the past years and selective inhibitors have been developed. The FDA has approved Selinexor for multiple myeloma and diffuse large B cell lymphoma ⁽²⁵³⁾. However, how Selenexor exerts its functions is still largely undefined. Supposably, in AML, XPO1 results in cytoplasmic localization of tumor suppressors, cell cycle regulators, and DNA damage response proteins ⁽²⁵⁴⁾. Thus, XPO1 inhibition favors the return of these proteins into the nucleus to exert their anticancer functions and promote differentiation and growth arrest of cells ⁽²⁵⁵⁾. In the context of beta-thalassemia, a major hemoglobinopathy resulting in reduced production of β -globin and ineffective erythropoiesis, it has been shown that XPO1 transports HSP70 out of the nucleus ^(256,257). Treatment with an XPO1 inhibitor (KPT-251), however, increased the amount of nuclear HSP70, which protected GATA1 from caspase-3 mediated cleavage and thereby favored erythroid differentiation of thalassemic erythroblasts ⁽²⁵⁷⁾.

In our CRISPR/Cas9 screen, we found that the inactivation of karyopherin KPNA2 (aka Importin- α , IPOA1) significantly increased erythroid differentiation of *SpCas9*-K562 cells. KPNA2 generally associates with KPNB1 (Importin- β , IPOB) to mediate nuclear import (**Fig. 29A**). Interestingly, we also identified KPNB1 in our GATA1 IP/MS analysis, however, CRISPR-mediated inactivation resulted in significant cell death without signs of differentiation. KPNB1 has been reported to be highly expressed in several cancers (e.g. cervical, breast cancer, multiple myeloma), highlighting a potential role in disease maintenance ⁽²⁵⁸⁾. Although aberrant KPNA2 expression was associated with solid tumors such as lung cancer, breast cancer, or colorectal cancers, its role in leukemia remains to be elucidated ⁽¹⁹⁴⁾. Interestingly, upon ATRA treatment that led to granulocytic differentiation of HL-60 cells, expression of KPNA2 was reduced ⁽²⁵⁹⁾. However, it is so far unclear if differentiation reduces KPNA2 levels or vice versa. Furthermore, these studies do not correlate protein levels to expression on mRNA level. Notably, we found KPNA2 protein highly abundant in all of our AEL cell lines compared to normal

control erythroblasts (**Fig. 21A**), suggesting that KPNA2 is indeed dysregulated in leukemia. Whether there is any kind of lineage association needs to be elucidated.

Another importin, importin 11 (IPO11, aka RanBP11) was recently identified in a CRISPR screen as a potentially novel therapeutic target in AML by Nachmias *et al.* ⁽²⁶⁰⁾. Of note, they also found KPNB1 as an essential target in their CRISPR screen, but with a lesser enrichment in AML stem cells. Moreover, they showed that IPO11 knockdown reduced cell growth and viability and increased differentiation measured by non-specific esterase (NSE) staining. Since importins have assigned similar functions, their results strengthen our findings.

Generally, KPNA2 was shown to recognize specific signals so-called nuclear localization signals (NLSs) on the target protein that needs to be imported. Interestingly, the Sankaran lab found that in congenital anemia GATA1 is mutated and that these mutations affect the GATA1 NLS sequence RKASGKGKKKR at aa307 ⁽²⁶¹⁾. Mutant GATA1 was 40% less present in the nucleus and increased in the cytoplasm, suggesting that nuclear localization was impaired. Since we did not observe any reduction of GATA1 protein levels in nuclear extracts of AEL cell lines compared to normal erythroblasts suggests that nuclear import is unaffected. Accordingly, treatment with ivermectin which particularly interrupts the dimerization of KPNA2 and KPNB1 and thereby inhibits the KPNA2/KPNB1-dependent transport (**Fig. 29B**) did not result in any signs of erythroid differentiation in leukemic K562 and KMOE2 cells (**data not shown**) ⁽²⁶²⁾. Nevertheless, future experiments will need to investigate how the nuclear localization of GATA1 is affected by reducing KPNA2 levels with microscopy and/or Western blotting of different cellular fractions.

Importantly, GATA1 is not the only TF and importins can regulate nuclear localization and subsequent binding of other TFs to chromatin. In undifferentiated mouse ES cells, it has been shown that TFs controlling neural differentiation are retained in the cytoplasm, thereby preventing TF activity in the nucleus ⁽²⁶³⁾. When KPNA2 is downregulated, cell differentiation can proceed, by which mechanisms this downregulation occurs is not clear yet. These data suggest that importins could modulate transcription, by regulating the import of specific TFs, which generally act in concert with other co-regulators.

To assess the impact of the GATA1 mutation on protein interactions Sankaran and co-workers performed IP/MS analysis from GATA1 WT and GATA1 mutant cells and found that well-known binding partners such as FOG1 and proteins of the NurD complex seem to be conserved. We also found FOG1/ZFPM1 and proteins of the NurD complex interacting with GATA1, however, they were not significantly differentially enriched in AEL versus normal erythroblasts (**Suppl. Table 5**). Next, they investigated if the NLS is important for GATA1 chromatin occupancy since the mutants show altered chromatin accessibility. Interestingly, mutant GATA1 showed less enrichment at genes that usually get upregulated upon erythroid differentiation such as *SLC4A1* or *HBB*.

Translated into AEL, this finding would support the hypothesis that transcriptional regulation (involving KPNA2) rather than nuclear import might primarily be affected. One working hypothesis could be that erythroid differentiation is impaired due to the inability of the complex to dissociate, which prevents the binding of “free” GATA1 to loci that are critical for terminal maturation (**Fig. 29C**). We found that loss of KPNA2 results in erythroid differentiation, most probably through dissociation of the complex and GATA1 becoming “free” to bind. Such hypothesis will need to be explored by comparing GATA1 binding to chromatin upon KPNA2 knockdown (**Fig. 29D**). However, so far no one has explored mechanistically, how GATA1 is transported in and outside the nucleus in the first place.

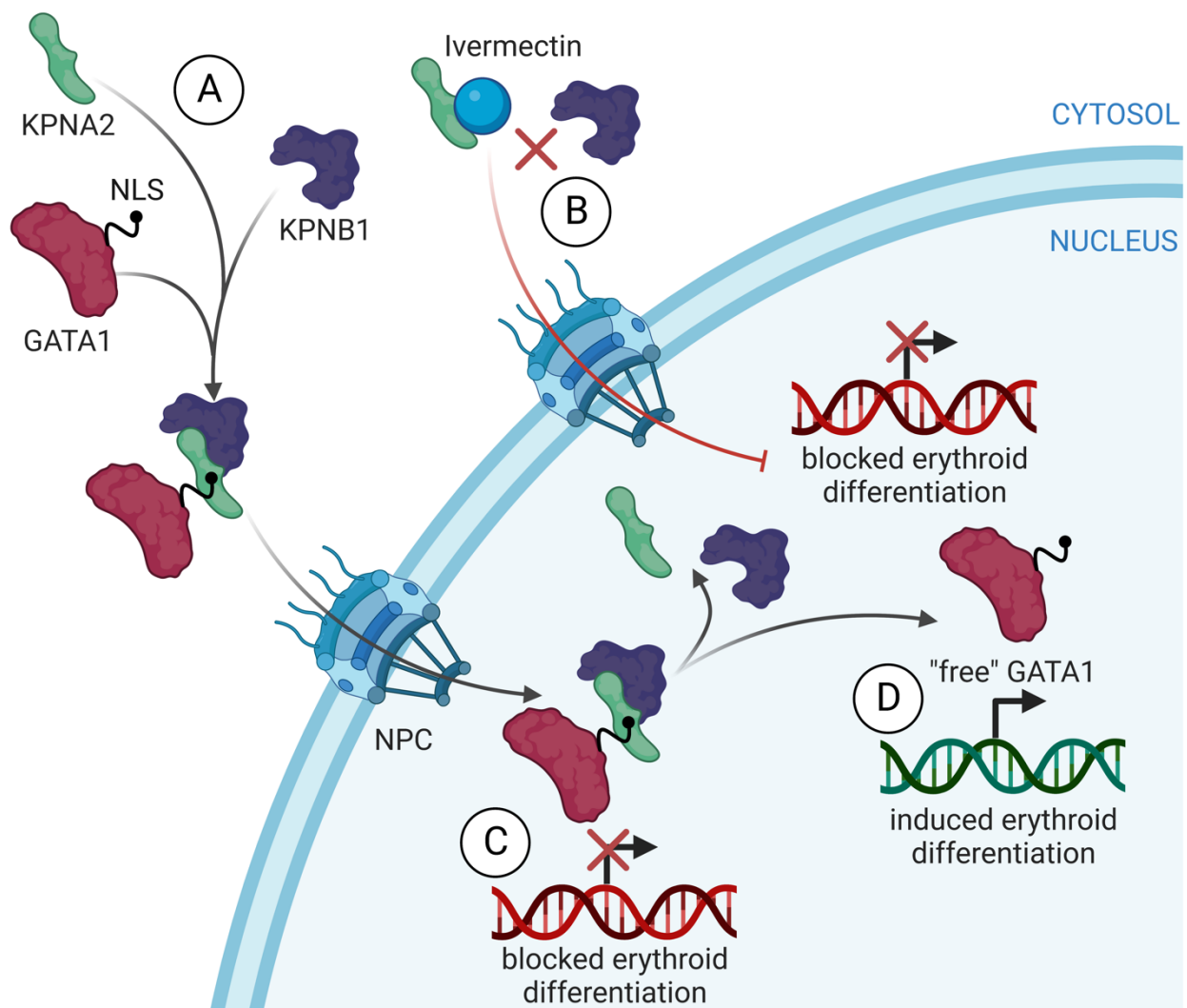


Figure 29. Hypothetical model GATA1 regulation by Karyopherins. A) KPNA2 associates with KPNB1. KPNA2 recognizes the nuclear localization signal (NLS) on the client protein, in this case, GATA1, and forms a complex that is being transported through the nuclear pore complex (NPC) inside the nucleus. **B)** Pharmacological targeting: Ivermectin binds to KPNA2 to prevent dimerization of KPNA2 with KPNB1 thereby preventing nuclear import. Treatment with ivermectin shows no effects on erythroid differentiation of K562 and KMOE2 cells. **C)** Hypothetically erythroid differentiation could be impaired due to the inability of the complex to dissociate, which prevents the binding of “free” GATA1 to chromatin and the activation of gene transcription. **D)** Loss of KPNA2 induces erythroid differentiation, hypothetically through dissociation of the complex and GATA1 becoming “free” to bind (created with Biorender.com).

Regulation of GATA1 function by chaperones?

Chaperone protein complexes are of particular interest in erythroid development as they were shown to regulate protein synthesis of immature erythroblasts, hemoglobin synthesis, and degradation of unnecessary proteins at late stages of erythropoiesis ⁽²⁶⁴⁾. Erythroid maturation is generally considered a cellular stress situation and therefore associated with increased release of reactive oxygen species (ROS) ⁽²⁶⁵⁾. ROS increases in response to EPO which together with iron has to be imported into the cell to ensure proper hemoglobinization. Thus, during erythroid differentiation efficient cytoprotective systems are crucial to shield ROS-related toxic effects and reduce cellular stress. In particular, heat shock proteins serve as stress reducers during erythroid differentiation. As outlined before (**Chapter 1.3.3.**) erythroblast survival is regulated by the interplay of GATA1, EPO, and HSP70. EPO stimulates the translocation of HSP70 into the nucleus, where it can bind directly to GATA1 and prevents it from being cleaved by caspase-3 (**Fig. 30A**). It is not clear if GATA1 forms a complex with HSP70 already in the cytosol, or if they are separately transported through the NPC. If cells are deprived of EPO or if insufficient HSP70 is translocated inside the nucleus the unprotected GATA1 gets degraded (**Fig. 30B**) ⁽¹⁰¹⁾. The exact mechanisms of how HSP70 protects GATA1 from being cleaved have not been explored. Interestingly we found HSP70/HSPA1L more enriched in the GATA1-pulldowns of normal erythroblasts that maintain their differentiating potential, confirming its functional relevance. Notably, we identified several chaperone proteins including the Parkinson-associated deglycase PARK7 (aka DJ-1), Chaperonin Containing TCP1 Subunit 4 and 7 (CCT4 and CCT7) in our GATA1 pulldown assay, that upon CRISPR-mediated inactivation showed significant induction of erythroid surface markers in K652 cells (**Fig. 20**).

Liu *et al.* observed that PARK7 expression was elevated in MDS and AML patients, suggesting that PARK7 might be relevant in disease maintenance or progression ⁽²⁰⁷⁾. siRNA-mediated PARK7 knockdowns in K562 and HL60 cell lines did not induce significant effects of differentiation measured by expression of myeloid surface markers CD11b, CD15, and CD15. The effect on erythroid differentiation remained unexplored. In contrast, another group claimed that treating HL60 cells with diallyl disulfide (DADS) resulted not only in the downregulation of PARK7 but also myeloid differentiation. However, it remains unclear whether downregulation of PARK7 was the cause of differentiation or vice versa ^(266,267). Due to these obscure findings, we knocked-down PARK7 with our doxycycline-inducible mir-shRNA viral construct, which confirmed our CRISPR/Cas9 screen results of induced partial erythroid differentiation in K562 cells (**Suppl. 4F, G**). However, we did not observe any significant differences in protein expression of PARK7 between normal versus leukemic erythroblasts (**Suppl. Fig. 4A**).

We found two additional chaperones, CCT4 and CCT7, enriched in our GATA1 IP/MS analysis in AEL cells. They are part of the so-called chaperonin containing TCP1 ring complex (TRiC) which consists totally of 8 CCTs. Previous studies showed that CCT1 and CCT2 promoted the proliferation of breast cancer cells and others linked CCT1 to AML drug resistance ^(193,268). It has notably been shown that CCTs cooperate with other chaperones such as HSP70 to enhance client protein binding/folding but also stabilization of transcription factors to modulate gene transcription ⁽²⁶⁹⁾. The myeloid leukemia factor 1 (MLF1) has been identified to act in chaperone complexes and loss of chaperones altered transcription factor stability and gene expression ⁽²⁷⁰⁾. Overexpression of MLF1 inhibited EPO-induced differentiation of the J2E erythroleukemia cell line associated with an erythroid to myeloid lineage switch ⁽²⁷¹⁾. Although MLF1 did not drop out in our screen, we saw that targeting CCT4 resulted in significant cell death induced by lentiviral knockdown, accompanied by increased erythroid surface markers and *HBA* expression. Interrogation of the Depmap indicated that all components of the TRiC seem essential and therefore most likely not suitable for selective targeting. So far, it is not clear whether loss of CCT4 or other TRiC proteins may result in the activation of caspase-mediated GATA1 cleavage which has been observed in the context of HSP70. Since we found GATA1 associated with a variety of chaperones, one could speculate that in AEL cells, chaperone complex proteins are dysfunctional and GATA1 might be trapped in chaperones which might perturb its function and erythroid differentiation. Whether this is due to protein misfolding, alterations in cytoplasmic/nuclear localization or modulation of gene transcription remains to be explored (**Fig. 30C**).

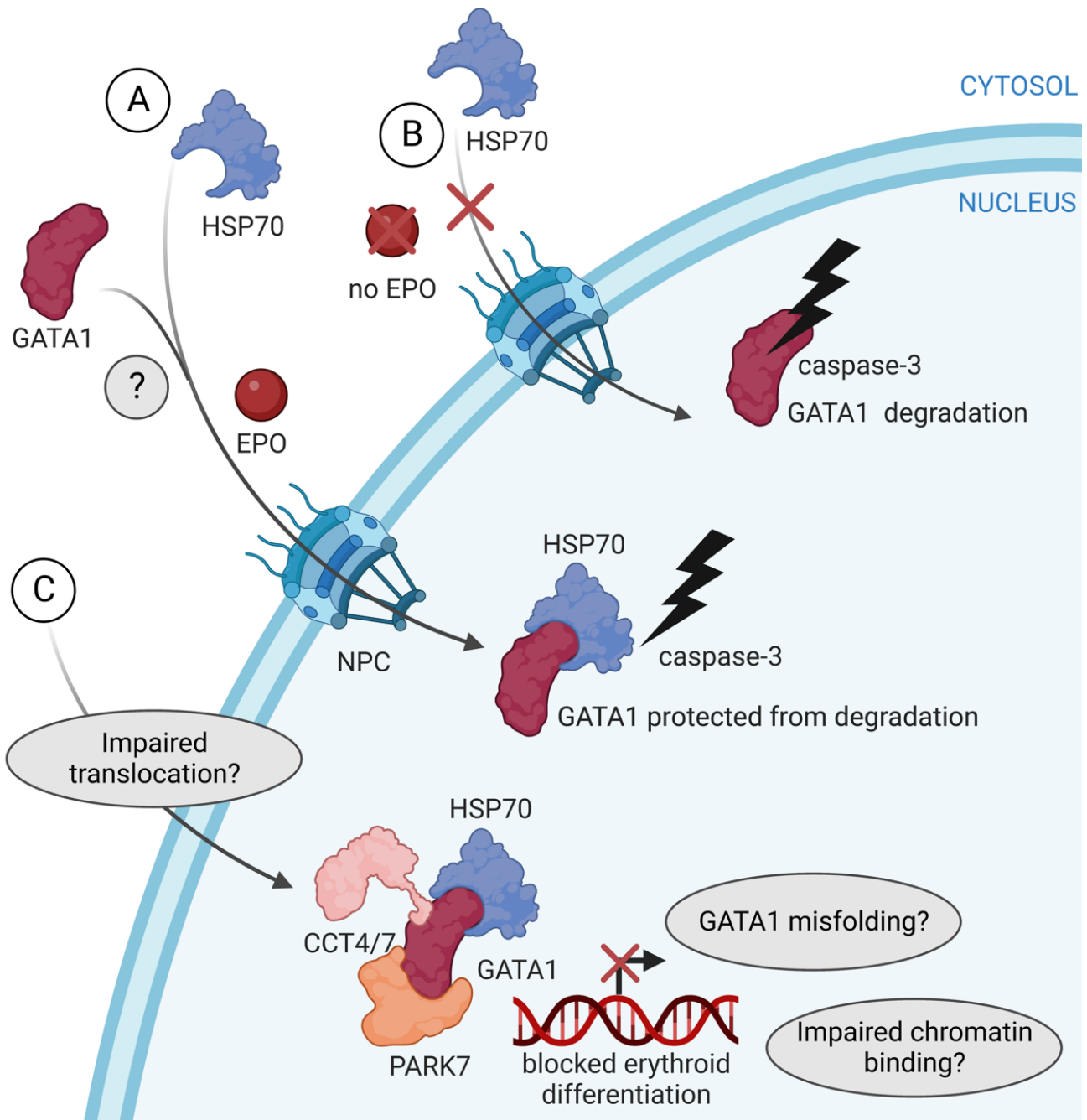


Figure 30. Model of GATA1 protein interactions involving chaperones. A) EPO stimulates the translocation of HSP70 into the nucleus, where it can bind directly to GATA1 and prevents it from caspase-3-mediated cleavage. It is not clear how GATA1 translocates into the nucleus. **B)** EPO deprivation or insufficient HSP70 translocation leads to cleavage of GATA1 by caspase-3. **C)** Hypothetically, GATA1 might be associated with a variety of chaperones which inhibits erythroid differentiation. The exact mechanisms remain to be explored. Possible mechanisms could involve alterations in cytoplasmic/nuclear localization, misfolding or modulation of gene transcription by impaired binding (created with Biorender.com).

6.4. CRISPR/Cas9 screening in K562 AEL cells: limitations

CRISPR screens have become a powerful tool to explore effects on cell identity. Genome-wide CRISPR “dropout” screens take advantage of a predefined pooled sgRNA library and measure the depletion of sgRNAs to investigate potential therapeutic targets for cancer treatments in large-scale ⁽²⁷²⁾. CRISPR screens are very popular tools as they allow us to easily assess the consequences of gene inactivation rather than repressing genes at the posttranscriptional level (shRNA). A systematic comparison of CRISPR/Cas9 screens to shRNA-based strategies to identify essential genes showed high precision, but nevertheless often identified different genes suggesting that each method has its advantages and disadvantages to identify cellular phenotypes ⁽²⁷³⁾. This makes it often difficult to draw general conclusions based on publicly available datasets and emphasizes the need to execute in-house screens.

In erythroid cells so far, screens have mostly been performed in HUDEP2 and K562 cells ^(175,274–276). However, most of these screens only look into dependencies (“dropout”), whereas we were mainly interested in erythroid differentiation capabilities. Therefore, we designed a targeted CRISPR/Cas9 screen specifically to understand which GATA1 interacting proteins may contribute to the differentiation block in AEL cells. Overall, we selected 116 potentially interesting interactors, that we targeted with 2 sgRNAs per gene. We took advantage of the “Vienna Bioactivity CRISPR score” (www.vbc-score.org) to predict and select sgRNAs for efficient production of *loss-of-function* mutations. Since we did not perform any sequencing or confirmed loss of expression upon inactivation, we cannot exclude random off-target effects. However, this database seemed to outperform previous prediction tools ⁽²⁷⁷⁾. We initially also considered using HUDEP2 cells for our screen, however stable expression of the *SpCas9* construct is demanding since cells are prone to differentiate easily upon any manipulation, even in maintenance conditions (personal communication with Mitchell J. Weiss, MD, PhD, Department of Hematology, St. Jude Children`s Research Hospital, Memphis, USA). For this reason, clonally selected stable *SpCas9*-K562 cells seemed to be a superior model. We were however aware that a flow cytometry read-out for our screen would be difficult to assess since K562 cells already express the erythroid CD71 and CD235a surface markers, making it challenging to observe any significant shifts towards a more mature phenotype and the necessity for further validation.

To get an unbiased overview, we designed the statistical analysis of our screen by calculating the mean fluorescent intensity (MFI) from both surface markers and performed t-tests normalized to the control sgAAVS1 construct (not affecting differentiation). It is possible, that through these calculations, we may have missed some targets that could induce erythroid differentiation of the cells. Targets were also excluded/not significant if only one construct was reactive and the other one not. Despite these possible limitations, we found attractive genes that we further validated by the expression of inducible

mir-shRNAs. Targeting of RCOR1, CCT4, KPNA2, and PARK7 resulted in consistency of the flow cytometry read-outs upon CRISPR-mediated inactivation or reduced expression. We were also able to follow our selected targets in a more coherent manner by assessing cell growth and *HBA* expression as well as by morphological inspection of the “reddish” cell pellet. This shows, that considering all the limitations, our results seem to be consistent and reproducible independent of *loss-of-function* or knockdown of target genes. Although we focused our analysis in particular on genes/proteins that were more enriched in AEL cell lines, it would be interesting to also explore the genes/proteins that had a phenotypical role in normal erythroblasts to get a better fundamental understanding of processes controlling normal erythropoiesis.

6.5. Concluding remarks

AEL is a rare but highly aggressive form of AML associated with poor prognosis driven by complex yet poorly understood molecular mechanisms. Multiple efforts have been undertaken to characterize the molecular landscape of AEL patients ^(140,141,172). Although genetic lesions have been identified, it still remains unclear how they relate to the development of AEL. Alterations in repressive transcription factors such as ERG, GFI1, CBFA2T3, and SPI1 have been found to be aberrantly expressed in some AEL patients ⁽¹⁴⁰⁾. Interference with these TFs revealed alterations in GATA1 binding and dysregulated erythroid differentiation, suggesting that GATA1 function is central to AEL leukemogenesis ⁽¹⁷⁴⁾.

These findings provided a rationale to functionally characterize molecular mechanisms that could be responsible for impaired AEL differentiation in several models. Analysis of the AEL phenotype in *Nsd1*^{-/-} mice suggests that histone methylation, in particular, H3K36me is essential for normal red blood cell development. Reduced H3K36me seemed to alter GATA1 protein interactions and chromatin interaction essential for erythroid differentiation ⁽⁴²⁾.

Next, we brought these findings into the perspective of human AEL by characterization of GATA1 complexes as potential key players of blocked differentiation. Our work so far suggests that impaired erythroid differentiation of AEL cells involves aberrant GATA1 protein complexes. We identified interaction of GATA1 to the epigenetic regulator RCOR1 and to the importin KPNA2 that contribute to impaired AEL differentiation, however, the detailed molecular mechanisms remain to be clarified.

Taken together, the findings allow us to speculate that blocked erythroid differentiation of AEL cells is the result of aberrant histone modifications (and DNA methylation), protein interactions, and function of erythroid master TFs like GATA1. Restoring impaired GATA1 function may allow to release the differentiation block of the cells and reduce the disease burden. Future AEL therapies will most likely combine the restoration of erythroid differentiation together with the inhibition of growth and survival pathways linked to EPO/JAK2/STAT5 signaling. There is hope that we can improve the therapy and outcome for this rare but very aggressive AML form in the future.

Chapter 7

Additional Information

7.1. Material and methods

Statistical analysis

Prism 9 software (Graphpad, La Jolla, CA, USA) was used for all statistical analyses, and error bars are generally shown as mean \pm SEM (standard error of the mean). Experiments were performed mostly in triplicates. The individual sample size is reported in the figure legends. Significance was calculated with unpaired (different cell of origin) or paired (same parental cell) two-tailed t-test, assuming equal variance, unless otherwise noted.

Genetically modified mice

All mice were maintained under specific pathogen-free conditions housed in individually ventilated cages at the animal facility of the Department of Biomedicine (Basel, Switzerland) with free access to food and water in accordance with Swiss Federal Regulations.

To inactivate the *Nsd1* gene in the hematopoietic system, the floxed *pgk-neomycin* selection cassette was deleted from mice with an *NSD1^{+L3}* allele, leaving two loxP sites flanking the largest coding exon 5, (here referred to as *Nsd1^{fl/fl}*). *Nsd1^{fl/fl}* mice were intercrossed with a *Vav1-iCre^{tg/+}* transgenic strain leading to constitutive ablation of the NSD1 gene in fetal and adult hematopoiesis (*Vav1-iCre;Nsd1^{fl/fl}* or also called *Nsd1^{-/-}*). Symptomatic *Vav1-iCre;Nsd1^{fl/fl}* mice were sacrificed by CO₂ asphyxia and cervical dislocation 7-17 weeks after birth once they reached an “animal welfare score” \geq 3, based on appearance, natural and provoked behavior, and body weight.

The knock-in of mCherry as a fusion into the endogenous GATA1 locus mouse (*C57BL/6;JM8N4-Gata1^{mCherry}*) was a kind gift from Timm Schröder. We crossed this strain with the *Vav1-iCre;Nsd1^{fl/fl}* strain to obtain *Nsd1^{-/-}GATA1^{mCherry}* transgenic mice.

The *Nsd1^{fl/fl};CreERT2⁺* transgenic mouse line was created by crossing *Nsd1^{fl/fl}* mice with a *UBC-CreERT2* mouse line obtained from Tobias Derfuss (DBM, Basel). Through induction with tamoxifen (TAM) the Cre-ER fusion protein cuts the *loxP* sites resulting in targeted deletion of the floxed target sequences in the cells.

Pups were genotyped with toe-derived DNA using the KAPA2G Fast HotStart Genotyping Mix kit (Cat:KK5621, KapaBiosystems, Basel, Switzerland) (**Suppl. Table 8**).

Isolation and lineage marker depletion of BM-derived hematopoietic cells

Muscle, tendons, and fat bodies were removed and total bone marrow was harvested by crushing femur, tibia, and ilium in RPMI containing 10% FCS and 1% penicillin/streptomycin (P/S) and then filtered through a 40 μ m cell strainer (Cat. 352340, BD, New Jersey, US). Red blood cells (RBC) were lysed with ammonium-chloride potassium lysis buffer (150mM NH₄Cl, 10 mM KHCO₃, and 0.1 mM

EDTA, pH 8.0) for 10 minutes on ice. After 5 minutes at 1500rpm centrifugation, cells were resuspended in PBS plus 10% FCS and counted.

Direct Lineage Cell Depletion Kit (mouse) (Cat. 130-090-858, Miltenyi Biotec, Bergisch Gladbach, Germany) was used to deplete the cells of mature hematopoietic cells resulting in an enriched population of stem and progenitor cells ("Lin⁻"). Cells were incubated briefly with microbeads conjugated to monoclonal antibodies against CD5, CD11b, CD45R/B220, Anti-Gr-1, 7-4, and Ter119 for 10 minutes at 4°C. The labeled cells were then separated by placing LS columns (Cat. 130-042-401, Miltenyi Biotec, Bergisch Gladbach, Germany) in a magnetic field of a MACS® separator. Lineage-negative unlabeled cells were washed and cells were plated in "maintenance medium".

Isolation of fetal liver hematopoietic cells

Mice were mated for two nights. The pregnant mice were sacrificed at day 13.5/14.5 post-coitus by CO₂ suffocation and cervical dislocation. The uterine horns were dissected and each embryo was subsequently separated from its placenta and embryonic sac. The embryos were kept on ice to anesthetize them for further processing. The head of the embryo was dissected and used for the genotyping. The fetal liver was dissected to establish erythroblastic cultures. After isolation, the fetal livers were resuspended in PBS and filtered through a regular flow cytometry tube (35µM). After centrifugation of the cells for 5 minutes at 1500rpm, the red blood cells (RBC) were lysed with lysis buffer (150mM NH₄Cl, 10 mM KHCO₃, and 0.1mM EDTA, pH 8.0) for 15 minutes on the ice. The cells were then centrifugated again for 5 minutes at 1500rpm and were resuspended and cultured in "maintenance medium".

Culture of primary murine erythroblasts

Culture conditions for primary murine erythroblasts were based on ⁽¹⁵⁴⁾. Lin⁻ BM cells or fetal liver-derived hematopoietic cells were expanded for 7 days in "maintenance medium" containing StemSpan SFEM (Cat. 9650, StemCell Technologies, Vancouver, Canada), 1% P/S, 10⁻⁶ M dexamethasone (Cat. 265005, Calbiochem, Sigma Aldrich, Buchs, Switzerland) to block differentiation, 40 ng/ml human insulin-like growth factor 1 (IGF1) (Cat. 1001-11, Peprotech, London, UK), 0.4% cholesterol (Cat. 12531-018 Gibco, Thermo Fisher Scientific Reinach, Switzerland), 100 ng/ml murine SCF (Cat. 250-03, Peprotech, London, UK) and 2U/ml human EPO (Eprex 4000, 9096976, Pharmacy of University Hospital Basel) to establish murine primary erythroblasts (**Suppl. Table 9**). Cells were passaged every second day to 0.3-0.5 x10⁶ cells/ml; keeping the cells below 1.5 x10⁶ cells/ml. Terminal maturation of erythroblasts was initiated by resuspension of the cells in "differentiation medium" containing IMDM (Cat. 31980022 Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 1% P/S, 10% FCS, 10% PFHMII (Cat. 12040077, Gibco, Thermo Fisher Scientific,

Reinach, Switzerland) 5% human PDS (0.45 μ M filtered, Blood donation Centre, University Hospital Basel), 1:100 monothioglycerol (Cat. M6145, Sigma Aldrich, Buchs, Switzerland), 100ng/ml mSCF and 2U/ml hEPO (**Suppl. Table 10**).

Cultures of established cell lines

The Friend Leukemia virus strain 745 PC-4 was used to establish mouse erythroleukemia (MEL) cells (provided by I. Vizirinakis, Thessaloniki, Greece). MEL cells were maintained in suspension in DMEM (Cat. 61965059, Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 10% FCS (Cat. 2-01F10-I, BioConcept, Allschwil, Switzerland), and 1% Penicillin Streptomycin (P/S) (Cat. 14140, Gibco, Thermo Fisher Scientific, Reinach, Switzerland).

The human suspension cell line K562 (ACC-10, obtained from the Leibniz Institute DSMZ-German Collection of Microorganisms and cell cultures) was established from a patient diagnosed with chronic myeloid leukemia (CML) in acute blastic crisis ⁽²⁷⁸⁾. The human erythroleukemia (HEL) cell line (ACC-11; DSMZ-German Collection of Microorganisms and cell cultures) was established from a patient who developed erythroleukemia after treatment for Hodgkin lymphoma ⁽²⁷⁹⁾. K562 and HEL cells were cultured in RPMI 1640 (Cat. 61870, Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 10% FCS, and 1% P/S.

The AML-M6 cell line F36P (ACC-543, purchased from the Leibniz Institute DSMZ-German Collection of Microorganisms and cell cultures) was cultured in RPMI 1640 with 20% FCS, 1% P/S, and 10 ng/ml hGM-CSF (Cat. 300-03, Peprotech, London, UK). This cell line was obtained from a patient who had developed erythroleukemia secondary to MDS ⁽²⁸⁰⁾.

The human umbilical cord blood-derived erythroid progenitor-2 (HUDEP2) cell line was established from umbilical cord blood CD34⁺ hematopoietic cells and immortalized by an HPV16-E6/E7 expression system ⁽¹⁷⁸⁾. HUDEP2 cells were maintained in “HUDEP2 maintenance medium” (StemSpan SFEM StemCell Technologies, Cat. 9650, Vancouver, Canada), with 50ng/ml of human SCF (stock concentration: 10 μ g/ml, Peprotech, LOT: 04153440716, Lubio), 2% P/S, 3 U/ml of human EPO (EPREX 4000, 9096976, Pharmacy of University Hospital Basel), 10⁻⁶M of dexamethasone (Calbiochem, Sigma Aldrich, Cat. 265005, Buchs, Switzerland) and 1 μ g/ml of doxycycline (Takara Clontech, Cat. 631311) (combined protocol established from ^(178,281)) and maintained at cell densities between 0.1 and 0.8x10⁶ cells/ml, passaged every second day. Doxycycline was always supplied freshly to the medium for immediate use (**Suppl. Table 11**). To differentiate HUDEP2 cells, human plasma was first heat inactivated for 30min at 56°C, differentiation of HUDEP2 cells was initiated by resuspension of the cells in “HUDEP2 differentiation medium” containing IMDM (Cat. 31980022 Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 2% P/S, 5% human inactivated human plasma AB (0.45 μ M filtered, Blood donation Centre, University Hospital Basel), 10mg/ml human

insulin (Sigma, I9278), 3U/ml heparin (Biochrom, L6510), 500µg/ml holo-transferrin (Sigma, T4132) and 3U/ml EPO (**Suppl. Table 12**). Cell Banker-1 cryopreservative medium (Zenoaq, LOT: 210414, Japan) was used for freezing HUDEP2 cells.

Isolation of PBMCs to establish primary adult human erythroblasts (hEBST)

“Cellquin V4.5” medium with ingredient mix was obtained by Marieke von Lindern`s Lab, Sanquin Research Institute, Amsterdam. The medium consisted of IMDM, w/o: L-glutamine, w:3.024 g/L NaHCO₃ PANBiotech, P04-20250, 1:100 L-glutamine, 1:100 P/S. 1:1000 Na-pyruvate, 1:500 insulin, 1:1000 sigma lipid mixture, 1:167 transferrin and 1:200 HSA (produced at Sanquin Research Institute, Amsterdam) (**Suppl. Table 13**).

PBMCs were isolated to obtain primary human erythroblasts (hEBSTs) based on ⁽¹⁷⁷⁾. Blood was transferred into 50ml falcons and centrifuged at 700rpm at RT for 10 minutes with the brakes off. Plasma (upper layer) was aspirated, and the RBC and buffy coat were resuspended with 1:1 ratio PBS by inverting the tube. 15ml Ficoll was added to a fresh tube and overlaid with diluted blood. Tubes were centrifuged at 600g for 30 minutes without brake. This results in layer separation and the buffy coat can be carefully collected by aspirating the upper layer plasma up to 1ml before. The PBMCs were freed of remaining platelets by mixing with 20ml PBS and centrifuging at 1200 rpm for 10 minutes with the brake on. The supernatant was aspirated and pellet/PBMCs were plated in “Cellquin V4.5 complete” medium, which was always prepared freshly to be used on the same day, by adding 100ng/ml hSCF (ITK Diagnostics BV, Uithoorn, The Netherlands), 1 U/ml hEPO (Eprex 4000, 9096976, Pharmacy of University Hospital Basel and 1nM dexamethasone (Sigma, St. Louis, MO). Dexamethasone (D4902-25mg) was dissolved before in 31.85 ml EtOH (stock 2.10⁻³ M). The stock was diluted 1:1 to aliquot for further use. 1 ng/ml IL-3 (Miltenyi Biotec, Bergisch Gladbach, Germany) was added in the first culture round to support the outgrowth of myeloid cells. PBMC culture was started with 10 x 10⁶ in 10 cm² dish with 10ml total volume. The medium was carefully changed partially without disturbing the cells until Day 7 and erythroblasts were visible under the microscope. Cells were counted on Day 7 and passaged accordingly with a density of 0.5-0.7 x 10⁶ cells/ml. The concentration of cells should never exceed 2.5 x10⁶ cells/ml. Cells were frozen between days 7-10 for later use in 50% “Cellquin V4.5 complete”, 10% DMSO, and 40% FCS.

Between days 9-12, cells were ready to be differentiated. Cells were spun down at 1800 rpm and the supernatant was removed entirely. Cells were resuspended to a concentration of 2 x 10⁶ cells/ml into “Cellquin V4.5 complete” media with 10 IU/ml EPO, 5 IU/ml heparin (LEO Pharma BV, Breda, The Netherlands), 5% plasma (0.45 µM filtered, Blood donation Centre, University Hospital Basel) and 700µg/ml holotransferrin (Sanquin Research Institute, Amsterdam, The Netherlands). On days 2, 5, and 7/8 half of the medium was refreshed. Cells were fully differentiated on days 10-12.

Isolation of CD34⁺ cells to establish primary human fetal erythroblasts

Sep-Mate 50 tubes (Cat.-Nr: 15460; Stemcell Technologies) were filled with 15ml of Ficoll solution at the bottom. Cord blood was mixed with PBS in a 1:1 ratio by inverting and carefully pipetted in the Sep-Mate tubes. After centrifugation for 1200g for 10min with brakes off, plasma was aspirated carefully. The layer containing PBMCs was carefully poured into a tube containing RBC lysis buffer and mixed by inverting. Cells were centrifuged at 300g for 10 min. In total, this step was repeated 3 times until the pellet appeared to be white. Pellet was washed with PBS before continuing with isolation of CD34⁺ cells. Isolation was performed according to the manufacturer's protocol (Indirect CD34 Microbead Kit, human; (Cat.-Nr: 130-046-701; MACS)). Isolated cells were plated in "Cellquin V4.5 complete" medium along with 1:100 hSCF (100 ng/ml), 1:1000 hEPO (1U/ml) 10 ng/ml IL-3, 10ng/ml IL-6 and 10 ng/ml TPO (300-18, Peprotech). TPO was removed after the first plating. This medium will allow outgrowth and maturation into mature red blood cells within 14 days. It is possible to use differentiation medium described in the section before to increase the enucleation rate.

Cell morphology analysis

0.1-0.4 x 10⁶ cells were resuspended in PBS and spun down using a Shandon Cytospin 3 centrifuge (Cat. 5991040, Thermo Fisher Scientific, Reinach, Switzerland) onto a coated glass slide (Cat. 5991056, Thermo Scientific, Rheinach, Switzerland) for 3 min at 250rpm. Cytospots were stained with Wright-Giemsa at the diagnostic hematology laboratory of the University Hospital Basel and subsequently microscopically assessed.

Flow cytometry

0.2-0.5 x 10⁶ cultured cells were washed, resuspended in FACS buffer (0.5% BSA, 1mM EDTA in PBS), and incubated with fluorochrome-conjugated antibodies (**Suppl. Tables 14 & 15**) for 20-30 min at 4°C. Following this, cells were washed with PBS to remove nonspecific binding of antibodies. Stained cells were resuspended in FACS buffer containing 1µg/ml DAPI (Cat. D1306, Life Technologies, Paisley, UK) to exclude dead cells from the analysis. Data collected from each cell was assessed in two-dimension plots on LSR Fortessa (BD, New Jersey, USA). Further analysis was performed with FLOWJO software (Tree Star). Population gating was always performed according to an unstained sample.

Cell cycle analysis with flow cytometry

For cell cycle analysis 0.3 x 10⁶ cells were washed with cold PBS and fixed with 300 ul BD Cytifix/Cytoperm (Cat. 51-2090KZ, BDBiosciences, San Diego, CA) for 15min at 4°C. After incubation, cells were washed twice with 1x BD Perm/Wash (Cat. 51-2091KZ, BDBiosciences, San

Diego, CA). Cells were stained for 1.5h with Ki67 (1:100 PerCP.Cy5.5) and DAPI (1mM: 4ul per ml) at 4°C. Cells were washed with 1x BD Perm/Wash and resuspended in FACS buffer. Flow speed was set to low before acquisition of the sample. Further analysis was performed with FLOWJo software (Tree Star).

Plasmid preparation

For transformation, high-efficiency chemically competent MAX Efficiency® Stbl2™ cells (Cat. 10268019, Thermo Fisher Scientific, Reinach, Switzerland) were incubated with 50ng of plasmid DNA. To take up the plasmids, competent cells were heat-shocked for two minutes at 42°C followed by immediate transfer on ice for another two minutes. Luria-Bertani (LB) medium (Cat. L3022, Sigma Aldrich, Buchs, Switzerland) without antibiotics was added and cells were incubated for one hour at 37°C with constant shaking at 250rpm. For the selection of transformed cells, cell suspensions were spread on LB agar Petri plates (Cat. 214010, BD Bioscience, Le Pnd de Daix, France) containing 100µg/ml ampicillin (Cat. A9518-25G, Sigma Aldrich, Buchs, Switzerland) and allowed to grow ON at 37°C. Five colonies were picked and cultured in LB medium supplemented with ampicillin, and grown for 24h in a shaking incubator at 37°C at a revolution speed of 120rpm. Plasmid DNA was extracted and analyzed by restriction enzyme digestion, agarose gel electrophoresis, and sequencing. Positively identified clones were extracted and purified using an alkaline lysis plasmid midiprep (Cat. 740984.10, NucleoSpin® RNA Plus, Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. Briefly, cells were lysed, neutralized, washed, and eluted under high alkaline conditions. Extracted plasmids were sterilized by washing with 70% ethanol and stored in nuclease-free water at 4°C before transfection.

For knockdown experiments (shNUP155, shKPNA2, shRCOR1, shCCT4, shPARK7, shRenilla), conditional inducible mir-shRNAs (LT3-mCherry-mirsh-PGK-Puro-IRES-rtTA3 obtained from Thomas Mercher Lab, Paris, France) were used and induced with 1µg/ml doxycycline. For gain of function experiments, cDNAs/ORFs were cloned into retroviral/lentiviral vectors (**Suppl. Table 16**).

Production of viral particles in HEK293T-LX cells

HEK293T-LX cells (Takara Bio Europe, Saint-Germain-en-Laye, France) expressing the SV40 T-antigen for high transfectability were seeded at an optimal density of 4×10^6 cells on a 10cm tissue culture plate in DMEM supplemented with 10%FCS and 1%P/S to obtain logarithmic cell growth at the time of transfection with approximately 80% confluence. Medium was replaced with DMEM with only 10%FCS to prevent interference of the antibiotics with the virus transfection. For ecotropic retroviruses, HEK293T-LX cells were transiently transfected with packaging vector (*pIPAK6*) and respective targeting plasmids. In case of human cell transductions, MSCV-based retroviruses were

feline glycol-coated with *pCMV-Gag-Pol* vector (RV-111) and *pEnv/pG* (RD114) envelope vector for packaging. For amphotropic lentiviral production, HEK293T-LX cells were transiently transfected with *pMD2.G* envelope vector, *pMDLg/pPRE* packaging vector, and *pRSV-Rev* expression plasmid plus respective targeting plasmids.

Per transfection reaction 40µl of JetPRIME reagent (Polyplus transfection®, VWR, Cat. 114-15) and 500µl of JetPRIME buffer (Polyplus transfection®, VWR, Cat. 114-15) were added to the plasmid mixture before adding to the cells to facilitate delivery into cells. Transfected HEK 293T-LX cells were incubated at 37°C for 4h. Afterwards, the medium was changed with fresh DMEM with 10% FCS and 1%P/S, and cells were incubated at 37°C overnight (ON). Viral supernatant was harvested 24, 48, and 72h after transfection, passed through 0.2µm filters, and concentrated in Vivaspin-20 centrifugal concentrators (Sartorius, Cat. VS2042, LOT: 211800217) at 3000rpm at 4°C. Viral supernatant was shock-frozen in liquid N₂ and stored at -80°C or used freshly.

Viral transduction of murine erythroblasts

0.5 x 10⁶ murine erythroblasts were spin-infected in “maintenance medium” with ten times concentrated retrovirus in the presence of 5µg/ml polybrene (Cat. 10.7689, Sigma Aldrich, Buchs, Switzerland) for 90 minutes, 2500 rpm at 30°C. 4h after transduction, cells were washed with PBS and resuspended in pre-stimulated “maintenance medium”. Two days after transduction, cells were selected with 2µg/ml puromycin for 2 days (Cat. A11138-03, Gibco, Thermo Fisher Scientific, Reinach, Switzerland).

Viral transduction of human cells

Recombinant fibronectin coating was performed 24h before transduction (Retronectin, TaKaRa Bio. Inc., Cat. T100A/B v1201, Japan). 500µl of retronectin (1µg/µl, in PBS) was used to coat 24-wells ON at 4°C. The next day, retronectin was removed and coated wells were blocked with 500µl of PBS containing 2% of Bovine serum albumin (BSA) (Sigma Aldrich, Cat. A2153-100G, Buchs, Switzerland) for 30min at RT. Wells were washed once with PBS and 500µl of 60x concentrated virus was added. The plates were centrifuged for 2-3h at 3000rpm at 32°C. After centrifugation, the non-attached viral supernatant was removed. 0.2 x 10⁶ K562 cells, 0.2 x 10⁶ HUDEP2 cells, and 0.3-0.5 x 10⁶ hEBST or cord blood-derived-CD34⁺ cells were used per well in 500 µl medium, according to cell line. Cells were added and the plate was centrifuged for 5min at 1000rpm to allow attachment to the coating. Cells were incubated at 37°C ON for 24h before detachment, by stringently washing the wells with medium to obtain a 1:3 cell dilution which was transferred into a fresh 12-well plate. The cells were either selected with puromycin (1µg/ml, Gibco Thermo Fisher Scientific, Cat. A11138-03, Reinach, Switzerland) or induced with 1 µg/ml doxycycline for 48h.

RNA isolation, cDNA synthesis, and quantitative RT-PCR analysis

RNA was isolated from dry frozen cell pellets with the NucleoSpin® RNA plus Kit (Cat. 740955.50, Macherey-Nagel, Düren, Germany) according to manufacturer`s instructions. Eluted RNA was diluted to 1 µg for cDNA synthesis. cDNA was prepared by using a cDNA reverse transcription kit (Cat. 4368814, Applied Biosystems, Foster City, USA). The program was run for 10 min at 25°C, 2 hours at 37°C, 5sec at 85°C, and infinite 4°C. The resulting cDNA was diluted 1:10 and quantitative RT-PCR was performed with selected primers (**Suppl. Table 17**) using FastStart Universal SYBR Green Master (Rox) in an ABI prism 7500 Software detection system. Ct values were normalized to *Gapdh* expression and relative expression was quantified using the $1/\Delta Ct$ method.

Protein isolation

All buffers were supplied with Complete Mini proteinase inhibitor (Cat. 11836153001, Roche) to prevent degradation of the target protein. To lyse the cell membrane, cell pellets (5-15 x10⁶ cells) were incubated in hypotonic lysis buffer (10 mM HEPES; pH 7.9, 10 mM KCl; 1.5 mM MgCl₂) for 15 minutes on ice. After short centrifugation at 14000 rpm at 4°C, the cytoplasmic supernatant was kept for further analysis, and the nuclear lysate was dissolved in non-ionic nuclear lysis buffer (10 mM Tris pH 7.5, 100 mM NaCl, 2.5 mM MgCl₂, 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 (30sec sonication, 30sec pause) on a Bioruptor pico sonicator (Cat. B01060001, Diagenode, Seraing, Belgium). After 2h additional incubation on ice, the sample was centrifuged 15min at 4000 rpm at 4°C, resulting in nuclear supernatant and a pelleted chromatin fraction which were kept for further analysis.

The protein concentration was determined with a colorimetric protein assay, based on the Bradford method. The absorption was assessed with an ELISA reader at 595 nm wavelength and was proportional to the amount of bound dye and thus the concentration of the protein. A calibration curve was established with BSA dilutions of known concentration. Using the standard curve, the concentration of the samples was measured. An aliquot of the nuclear extract was saved as a positive control.

Western Blotting

Input/nuclear and/or cytoplasmic fractions were boiled in 4x Laemmli buffer (Cat. 161-0747, Biorad, München, Germany), and chromatin and/or immunoprecipitated proteins were dissolved in 1x Laemmli for 10 minutes at 70°C. All fractions were supplied with 0.5% DL-Dithiothreitol (Cat. D0632-1G, Sigma Aldrich, Buchs, Switzerland) and loaded on self-casted SDS-PAGE gels. Next, SDS-PAGE proteins were transferred to nitrocellulose membranes (Cat. 170-4158, Biorad, München, Germany)

with a semi-dry blotting system. Efficient transfer of proteins was evaluated with Ponceau-S and Coomassie brilliant blue staining. Membranes were blocked with 10% non-fatty milk in 1x PBS containing 0.1% Tween for two hours at room temperature. Membranes were exposed to primary antibodies in 2.5%NFM/PBS-1%Tween ON at 4°C (**Suppl. Table 18**). After 3 x 15 min washes in 1x PBS-T, membranes were incubated with appropriate secondary antibodies coupled to horse-radish peroxidase (HRP) for two hours at room temperature (**Suppl. Table 19**). After three additional washing steps, signals were visualized using Supersignal West Femto Max Substrate (Cat. 11859290, Thermo Scientific, Rheinach, Switzerland). Carestream Biomax Kodak films (Cat. Z373508-50EA, Sigma Aldrich, Buchs, Switzerland) were used for development.

Purification of proteins for TMT quantitative multiplexed proteomics

For comparative/quantitative multiplexed proteomics 50 µg nuclear lysate was supplied with 5% SDS, boiled for 10min at 95°C to dissolve poorly-soluble molecules. Lysates were further incubated with 10mM TCEP to break disulfide bridges and carbamidomethylation of cysteines was performed using 15mM chloroacetamide at 37°C for 30min at 500 rpm. Protein denaturation was achieved by acidification with phosphoric acid 1.2% and digestion was continued with the S-TRAP™ sample processing kit (ProtiFi, Huntington, NY, US). Peptides were dried and 10ug was taken for labeling with Tandem Mass Tag (TMT) pro 16plex label reagent set (ThermoScientific, A44520) according to the manufacturer's protocol. After labeling, samples were combined, fractionated, and cleaned up using C18 reversed-phase spin columns (Macrospin, Harvard Apparatus, MA, US) according to the manufacturer's protocol.

Immunoprecipitation

After cytoplasmic/nuclear/chromatin protein isolation, the supernatant containing the nuclear fraction was pre-cleared for 1 h at 4°C to remove non-specific contaminants bound to the Protein G dynabeads (Cat. 1004D, Thermo Scientific, Rheinach, Switzerland). The cell pellet containing chromatin fraction was kept for further applications. According to the calculation, the antibodies against GATA1 N6 (Cat. Sc-265, Santa Cruz; 2µg for 200µg protein) or IgG2a (Cat. Sc-3883, Santa Cruz; 8µg for 200µg protein) were pre-bound to the beads rotating for 15 minutes at room temperature to minimize co-elution of the antibodies. Immunoprecipitations were performed ON rotating at 4°C.

Purification of immunoprecipitated proteins for MS analysis

After immunoprecipitation, the bead-antibody-antigen complexes were washed 3 times in nuclear lysis buffer to remove nonspecific binding. Samples were subjected to trypsin-based (5 µg/ml, Promega) on-bead digestion in 1.6M urea/100 mM ammonium bicarbonate buffer at 27°C for 30

minutes. Supernatant eluates containing active trypsin were further incubated with 1mM TCEP at room temperature ON. Followed by this, carbamidomethylation of cysteines was performed next using 15mM chloroacetamide in the dark at 27°C for 30 minutes. The tryptic digest was acidified (pH<2) using 5% TFA and desalted using C18 reversed-phase spin columns (Macrospin, Harvard Apparatus, MA, US) according to the manufacturer's protocol. Dried peptides were dissolved in 0.1% formic acid before injection into the mass spectrometer.

MS analysis and label-free quantification

Samples were analyzed in duplicates or triplicates using a dual pressure LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) connected to an electrospray ion source (Thermo Fisher Scientific) as recently described (PMID: 23017020). Peptide separation was performed with an EASY nLC-1000 system (Thermo Fisher Scientific) equipped with an RP-HPLC column (75 μ m \times 30 cm) packed in-house with C18 resin (ReproSil-Pur C18–AQ, 1.9 μ m resin, Dr. Maisch GmbH). A linear gradient from 95% solvent A (0.15% formic acid, 2% acetonitrile) and 5% solvent B (98% acetonitrile, 0.15% formic acid) to 28% solvent B over 60 min at a flow rate of 0.2 μ l/min was set. 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 widths at half-maximum (at m/z 400) followed by 20 MS/MS scans in the linear ion trap of the most intense ions (TOP20) using rapid scan speed. Unassigned and singly charged ions were excluded from analysis by setting the dynamic exclusion duration to 30 seconds. MS1-based label-free quantification was performed using Progenesis Q1 software (Nonlinear Dynamics (Waters), version 2.0). MS raw files were imported into the software and analyzed with default parameter settings. MS/MS-data were exported from Progenesis Q1 in “mgf” format and searched with a target/decoy strategy against the UniProt database containing forward and reverse sequences of the proteome from *Mus musculus* (33,984 entries) or *Homo sapiens* (41,592 entries) using MASCOT (version 2.4.1). Search criteria required full tryptic specificity (cleavage after lysine or arginine residues) allowing for three missed cleavages. Carbamidomethyl of cysteine was specified as a fixed modification. Oxidation of methionine and acetyl of the N-terminus were specified as variable modifications. The mass tolerance was set to 10 ppm for precursor (intact peptide) ions and 0.6 isolation width (Da) for fragment ions to filter the set of candidate peptides. Results from UniProt were imported into Progenesis Q1 and the resulting peptide measurement list containing peak area values of identified peptides was exported. Processing of peptide and protein quantities between samples was performed using SafeQuant in-house R script (PMID: 23017020).

Targeted CRISPR/Cas9 screen

A) Preparation of the vector carrying the sgRNA

As a gRNA-expressing vector, we used *pLentiCRISPRv2* (494_hU6-IT-sgRNA-PGK-IRFP670, Grebien lab, Vienna). 5µg of *pLentiCRISPRv2* was digested with 3µl *BsmBI* (stock: 10'000 U/ml, NEB, Cat. R0580, U.S) in 5µl of NEB buffer 3.1 (stock:10X, NEB, Cat. B7203S, U.S) and amount of water to bring reaction mix to 50µl by incubating for 2h at 55°C. Dephosphorylation was achieved by adding 1µl of arctic phosphatase (AP) (stock: 5'000 U/ml, NEB, Cat. M0289S, U.S) in 6µl of AP buffer (stock: 10X, NEB, Cat. B0289S, U.S) before incubating for 1h at 37°C. Gel extraction with 0.7% agarose gel (0.7-1 mg of Agarose BioReagent (99%, Sigma-Aldrich, Cat. A9539, U.S) in 100ml of 1x TAE buffer (stock: 50X, Life Tech, Cat. 24710030, Austria) and 10µl of GelGreen Nucleic Acid Gel Stain (stock: 10'000X, Biotium, Cat. 41004, U.S) was performed to purify the gRNA vector. The sample was loaded with 6x loading dye and gel was run for 30min at 120V. Gel extraction was performed with MiniPex 3 in 1 Plasmid prep kit (IMP, Austria) by cutting bands from the gel under UV light and transferring them into Eppendorf tubes (max. 400µg per Eppendorf tube). For 100mg of gel, 300µl of buffer G from the MiniPex 3 in 1 Plasmid prep kit was added. The mixture was incubated and vortexed in between for 20min at 50°C until the gel was entirely dissolved. The homogenous mixture was loaded onto columns and centrifuged for 1min at 13000rpm, flow through was discarded and the process was repeated until the total amount of DNA sample was distributed equally in 4 columns. The columns were washed by adding the washing buffer (buffer W), incubated for 3min, and centrifuged for 1min at 13000 rpm. Flow through was discarded and the gRNA vector was eluted in an Eppendorf tube with 25µl of elution buffer (buffer EB) per column to get a total volume of 100µl. Concentration was brought to 50µg/µl by adding the respective amount of EB buffer.

B) Cloning of sgRNAs

Constructs were selected based on the highest Vienna Biocenter (VBC)-scores (<https://www.vbc-score.org>). Oligonucleotides were ordered from IDT (pre-mixed F&R "wet", 40µM concentration (each oligo) (**Suppl. Table 20**) in 100µl IDTE pH8 PCR plate, stored at -20°C) and thawed on ice.

Oligonucleotides were diluted 1:2 with ddH₂O. To phosphorylate and anneal each pair of oligonucleotides, oligonucleotides were mixed with 10x T4 ligation buffer supplemented with 10mM ATP (New England Biolabs, Massachusetts USA) and T4 Polynucleotide Kinase (PNK) (New England Biolabs, Massachusetts USA). The reaction was incubated in a thermocycler for 30 minutes at 37°C following 2 min at 95°C and ramped down at 25°C. Annealed oligonucleotides were diluted 1:1000 in ddH₂O. For the ligation, 50ng/µl of gRNA vector was mixed with the diluted oligonucleotides, 10 x T4 Ligation Buffer, T4 DNA ligase (both New England Biolabs, Massachusetts USA), and water. The reaction was incubated for 20 min at room temperature. The ligation reaction was diluted by

adding 5 μ l CCMB80 buffer. As competent cells, 10 μ l of One Shot™ Stbl3™ Chemically Competent *E. coli* bacteria (Invitrogen, Thermo Fisher Scientific, Cat. C737303) were used and ligation reaction was added and incubated for 5 min on ice followed by heat shocking on a thermocycler (42°C for 40 sec) and rapidly cooling down for 1 min on ice. SOC Medium (Super optimal broth with Catabolite repression, a nutritionally rich bacterial culture medium, Carl Roth, Cat. AE27.1) was then added to the transformed cells and incubated for 30min at RT for recovery of the cells. Recovered cells were streaked on 12cm square Petri dishes with LB agar (98%, Carl Roth, Cat. X964.4, Germany) containing 100 μ g/ml of carbenicillin (100mg/ml, Carl Roth, Cat. 6344.1) and incubated at 37°C ON. On the next day, single colonies were picked and incubated in 2 ml LB medium supplemented with 100mg/ml carbenicillin ON at 37°C. The following day, colonies were picked and cultivated in a deep well plate (2.2ml, square-wells, notches, Biozym, Cat. 710850) with 1.2ml of Terrific Broth (a nutritionally rich medium for the growth of bacteria, Carl Roth, Cat. X972.1) including 100 μ g/ml carbenicillin. Two wells per colony were used, deep well plates were sealed with a gas permeable seal and bacterial cultures were incubated ON at 37°C on a shaker. Deep-well plates were centrifuged at 1000xg for 10 min before removing the broth. The plasmid DNA was isolated using the MiniPex 3 in 1 Kit (Institute of Molecular Pathology, Vienna Austria) according to manufactory's instructions. Subsequently, plasmid DNA was analyzed by Sanger sequencing (Microsynth, Vienna Austria) using *PGK_R P246* primer (5'-GCTACTTCCATTTGTCACGTC-3'). The sequence of the isolated plasmid containing the gene product of interest was then analyzed with the Geneious Prime program (www.geneious.com).

C) Production and transduction of sgRNA-expressing virus

For the CRISPR/Cas9 screen, HEK 293T-LX cells were seeded at 0.4 x10⁶ cells/ml in 24-well plate wells with 1ml DMEM complete. Transfection master mix was prepared in 1.5 ml Eppendorf tubes. First JetPRIME buffer was added, then packaging plasmid and vortex for 10sec. Transfection mix was supplied with JetPRIME reagent and transferred into PCR strips. Targeting plasmid (construct) was added to the corresponding PCR strip tube and the mix was vortex 10sec and incubated 10min at RT. The master mix was added to the HEK 293T-LX cells (**Suppl. Table 21**). Transfected HEK 293T-LX cells were incubated for 4h before changing medium to DMEM with 10% FCS and 1% P/S. After 24h, transfected cells were centrifuged for 5min at 1500 rpm and lentiviral supernatant was harvested by avoiding touching the bottom of the wells. Unfiltered viral supernatant was directly used for transduction. Transductions were done in duplicates. 0.3x10⁶ cell/ml *SpCas9-K562* clone-1 cells were resuspended in 500 μ l RPMI complete per well. 10 ng/ml Polybrene (Sigma-Aldrich, Cat. 10.7689, Buchs, Switzerland) was added to enhance transfection and viral supernatant was added in a 1:1

ratio. Cells were spin infected by centrifugation of the plate for 90min at 1000g and incubated at 37°C ON. Cells were washed and resuspended in fresh RPMI complete medium 2 days after transduction.

D) Competition-based proliferation and differentiation assay

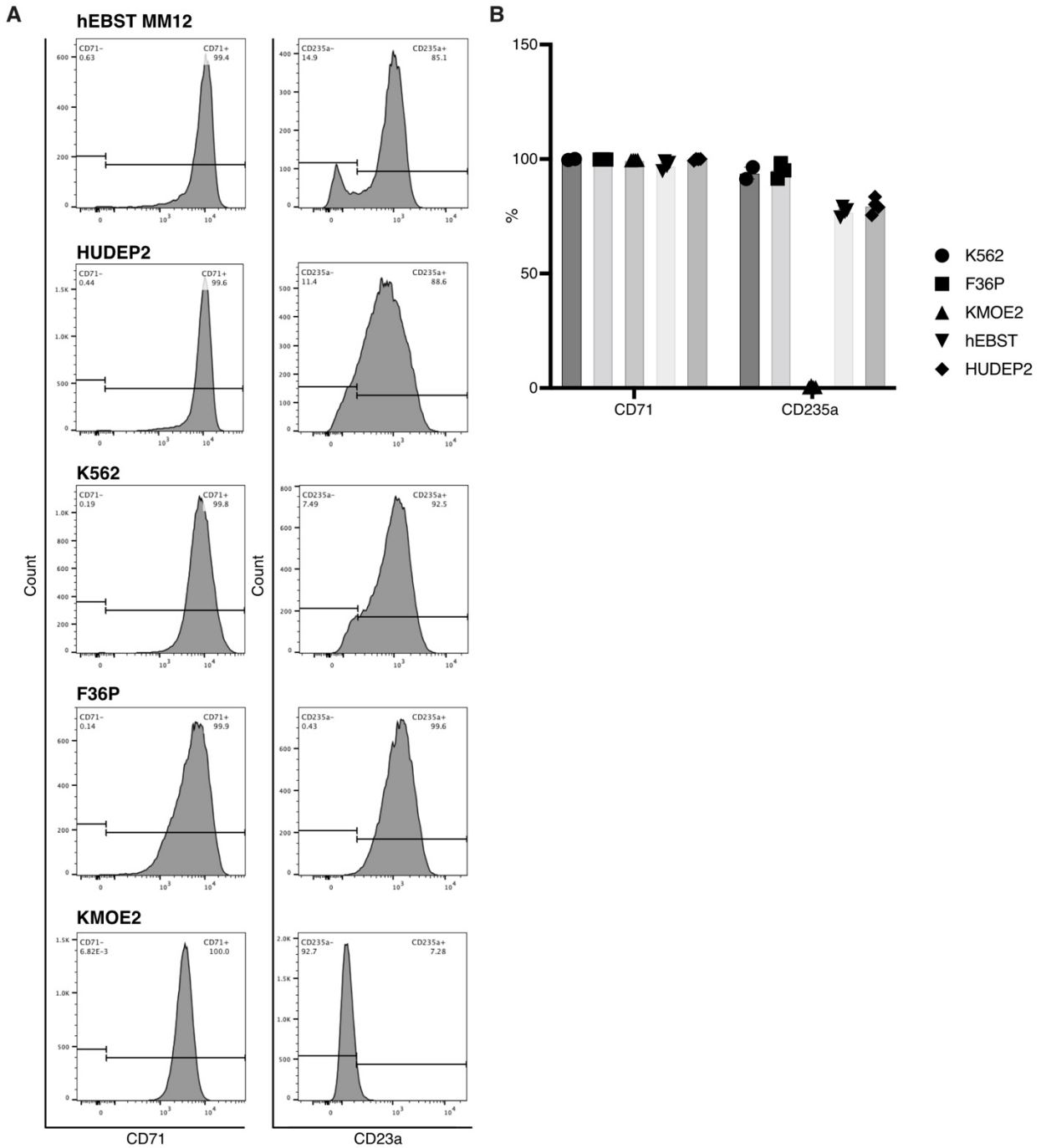
To assess the effect of gene knockout on cell proliferation lentiviral titers with an infection rate of 20-60% were aimed for to initiate a competition-based proliferation assay from a pool of transduced and non-transduced cells. The expression of sgRNAs was measured by the presence of the iRFP670 signal and the growth of transduced cells expressing iRFP670 was monitored over time. The level of iRFP670⁺ cells was measured at regular intervals by flow cytometry using the CytoFLEX (Beckman Coulter Life Sciences). 200µl of transduced cells was added to a round bottom 96-well plate and the plate was centrifuged for 5min at 300g. Cells were washed twice with PBS. Cells were stained for 20min at 4°C with FITC anti-human CD71 (BioLegend, Cat. 334103) or PE anti-human CD235a (BioLegend, Cat. 349106) antibodies at a 1:200 dilution in FACS buffer. Stained cells were washed and resuspended in 100µl of FACS buffer with DAPI (1:20000). Further analysis was performed with FlowJo software (Tree Star).

E) Statistical evaluation of cell depletion and erythroid differentiation

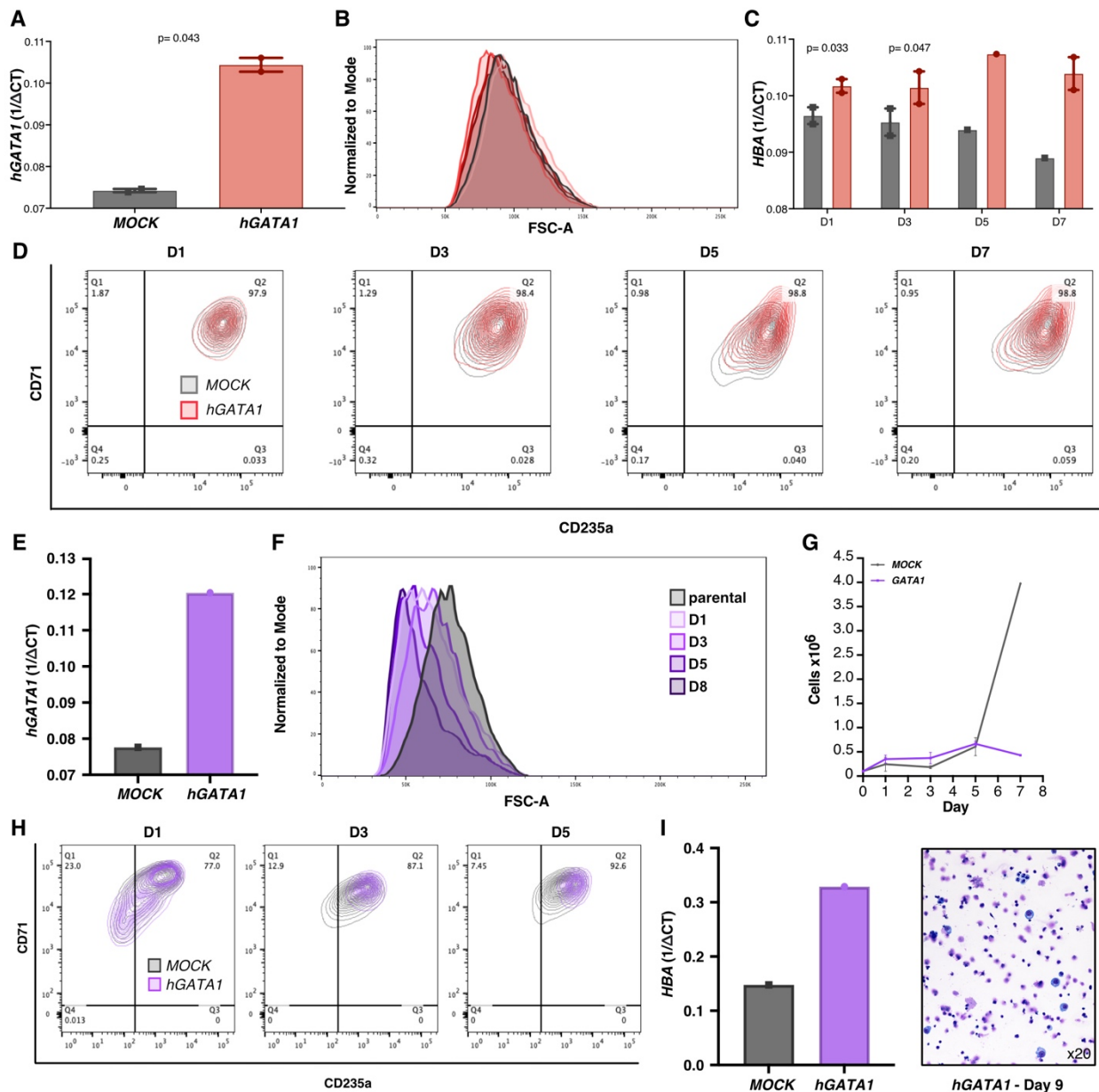
Depletion was measured by averaging the iRFP670⁺ percentages from both replicates (transductions #1 & #2) of each sgRNA construct (constructs #1 & #2). Values were normalized to the iRFP670 intensity of negative control sgRNA targeting adeno-associated virus integrate site 1 (sgAAVS1) on D4 (post-transduction) and averaged over the replicates. As positive control, the sgRNA targeting the essential 60s ribosomal protein L17 gene (sgRPL-17) was used. The iRFP670 values of the last day of measurement were averaged and normalized to *sgAAVS-1* on Day-4. Targeting of genes that result in $\geq 20\%$ reduction of the iRFP670⁺ signal was considered as being significantly depleted ($fc < 0.8$).

Erythroid differentiation was analyzed by calculating the mean fluorescent intensities (MFI) of CD71 and CD235a for iRFP670⁺ cells by FlowJo for each of the two replicates ((transductions #1 & #2). Then both constructs (constructs #1 & #2) were averaged for each time point. The shift of the MFIs was calculated by: $CD71 + CD235a - 200$. *sgAAVS1* values were all therefore all set to 0. (CD71: 100; CD235a: 100; - 100 = 0). Significance was assessed by performing t-tests followed by multiple testing corrections via adj. p-value. A resulting adj. p-value < 0.05 was considered significant.

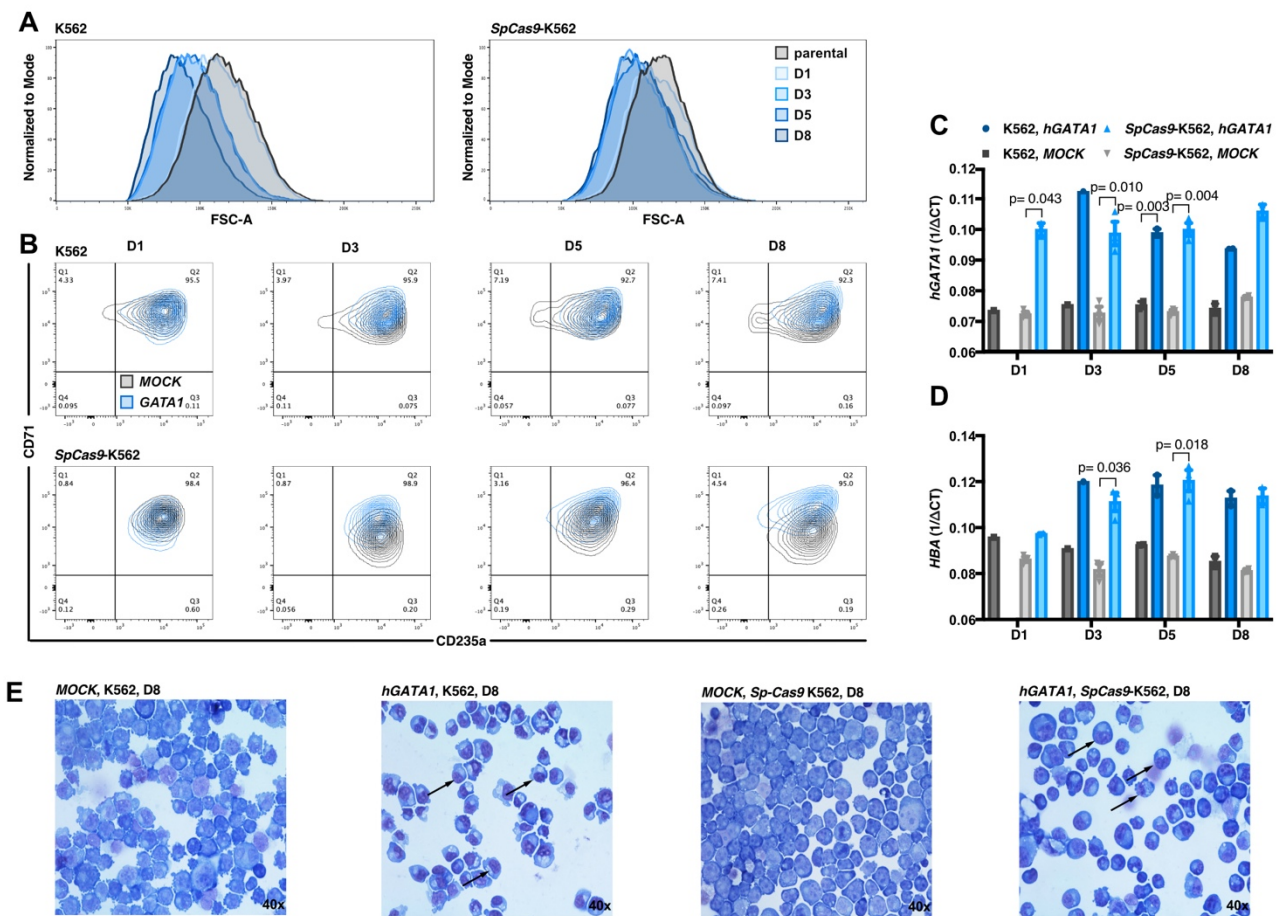
7.2. Supplementary figures



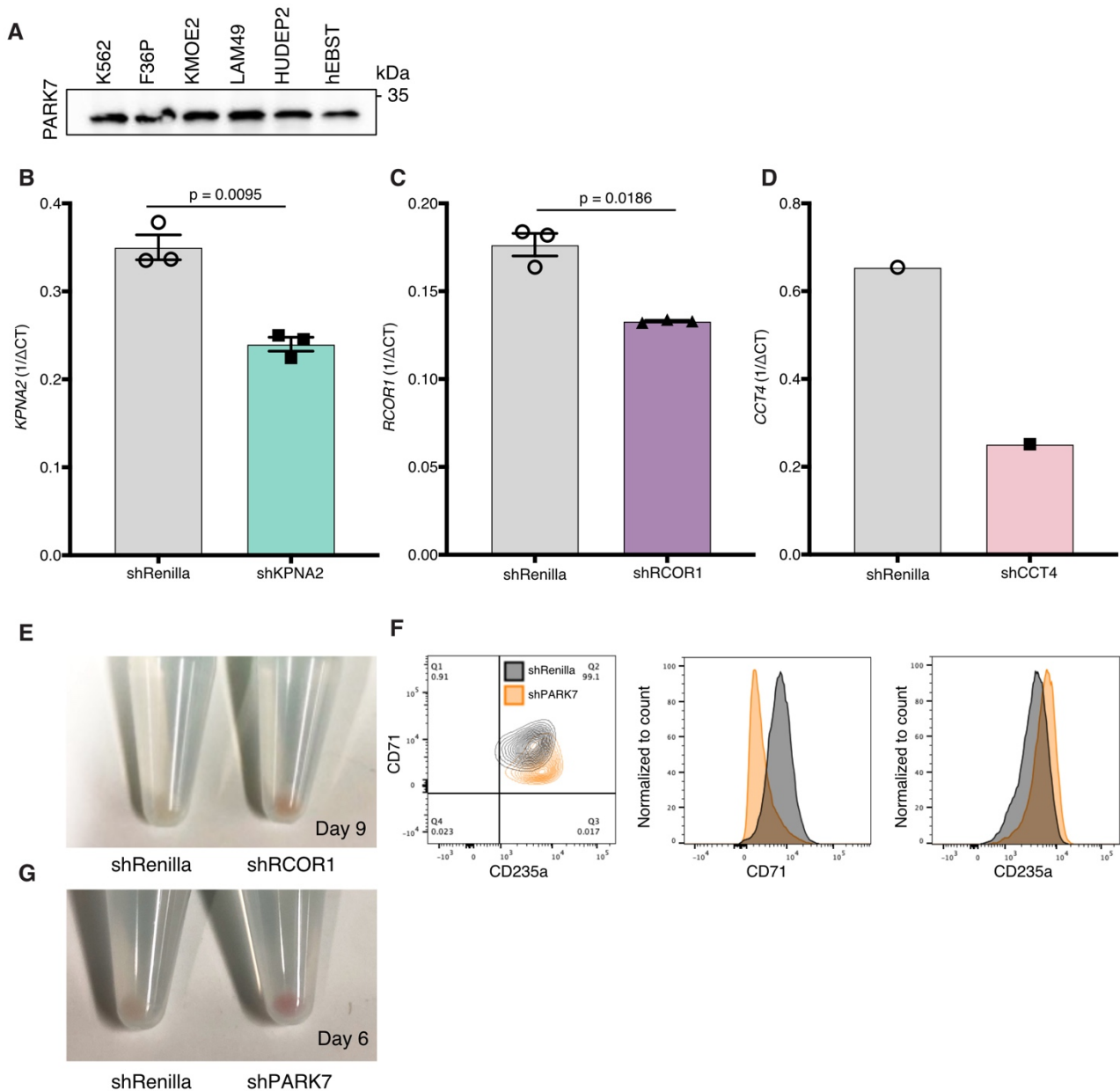
Supplementary Figure 1. Flow cytometric characterization of normal (hEBST, HUDEP2) and leukemic (K562, F36P, KMOE2) cell lines **A)** Representative images of flow cytometry analysis of CD71 and CD235a expression. **B)** Statistical evaluation of flow cytometry marker expression of CD71 and CD235a (K562: n=2, F36P: n= 3, KMOE2 n= 4, hEBST: n=4, HUDEP2: n=4).



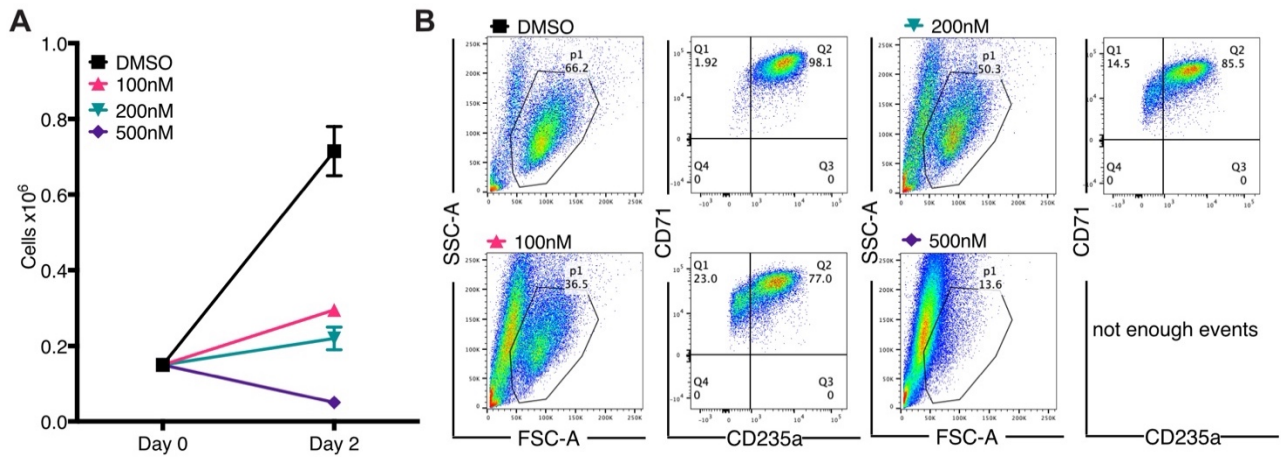
Supplementary Figure 2. Effects of exogenous GATA1 expression in HEL and HUDEP2 cells. A) Quantitative RT-PCR of *hGATA1* of *MOCK* and *hGATA1*-transduced HEL cells 1 day after puroposelection. CT values were normalized to *GAPDH* expression and shown as relative expression using $1/\Delta CT$ ($n=2$). **B)** Representative flow cytometry analysis of forward scatter (FSC-A) of *hGATA1*-transduced HEL cells D1, D3, D5 and D7 after puroposelection. **C)** Quantitative RT-PCR of *HBA* of *MOCK* and *hGATA1*-transduced HEL cells 1, 3, 5 and 7 days after puroposelection. CT values were normalized to *GAPDH* expression and shown as relative expression using $1/\Delta CT$ ($n=2$). **D)** Representative flow cytometry analysis of CD71 and CD235a expression in transduced HEL cells 1, 3, 5 and 7 days after puroposelection. **E)** Quantitative RT-PCR of *hGATA1* of *MOCK* and *hGATA1*-transduced HUDEP2 cells 1 day after puroposelection. CT values were normalized to *GAPDH* expression and shown as relative expression using $1/\Delta CT$ ($n=1$). **F)** Representative flow cytometry analysis of FSC-A of *hGATA1*-transduced HUDEP2 cells D1, D3, D5 and D8 after puroposelection ($n=2$). **G)** Growth curve of *MOCK* and *hGATA1*-transduced HUDEP2 cells 1-8 days after puroposelection. **H)** Representative flow cytometry analysis of CD71 and CD235a expression in transduced HUDEP2 cells 1, 3 and 5 days under puroposelection ($n=2$). **I)** Quantitative RT-PCR of *hGATA1* of *MOCK* and *hGATA1*-transduced HUDEP2 cells 1 day after puroposelection. CT values were normalized to *GAPDH* expression and shown as relative expression using $1/\Delta CT$ ($n=1$). **J)** Representative image of HE-stained cytospot obtained from *hGATA1*-transduced HUDEP2 cells 9 days after puroposelection. (Rathick Sivalingam performed experiment. Suppl. Fig 2. A-D, E, I).



Supplementary Figure 3. Effects of exogenous GATA1 expression in *SpCas9-K562* in comparison to K562 cells. A) Representative flow cytometry analysis of FSC-A) of *hGATA1*-transduced K562 and *spCas9-K562* cells D1, D3, D5 and D8 after puromycin selection. **B)** Representative flow cytometry analysis of CD71 and CD235a expression in transduced *hGATA1*-transduced K562 and *SpCas9-K562* cells 1, 3, 5 and 8 days after puromycin selection. **C)** Quantitative RT-PCR of *hGATA1* and **D)** *HBA* of *MOCK* and *hGATA1*-transduced K562 and *SpCas9-K562* cells 1, 3, 5 and 8 days after puromycin selection. CT values were normalized to *GAPDH* expression and shown as relative expression using $1/\Delta$ CT **E)** Representative images of HE-stained cytoplots obtained from *MOCK* and *hGATA1*-transduced K562 and *SpCas9-K562* cells 8 days after puromycin selection. (This experiment was performed by Rathick Sivalingam).



Supplementary Figure 4. Characterization of KPNA2, RCOR1, and CCT4 as GATA1 interactors that contribute to the erythroid differentiation block. **A)** Protein expression of PARK7 in AEL (K562, F36P, KMOE2, LAM49) and normal (HUDEP2, hEBST) erythroblasts. **B)** Quantitative RT-PCR of *KPNA2* from shRenilla and shKPNA-transduced K562 cells (n=3). CT values were normalized to *GAPDH* expression and shown as relative expression using 1/ Δ CT. **C)** Quantitative RT-PCR of *RCOR1* from shRenilla and shRCOR1-transduced K562 cells (n=3). CT values were normalized to *GAPDH* expression and shown as relative expression using 1/ Δ CT. **D)** Quantitative RT-PCR of *CCT4* from shRenilla and shCCT4-transduced K562 cells (n=1). CT values were normalized to *GAPDH* expression and shown as relative expression using 1/ Δ CT. **E)** Representative images of cell pellets obtained from shRenilla and shRCOR1-transduced K562 cells 9 days after transduction. **F)** Representative images of flow cytometry analysis of CD71 and CD235a expression in shPARK7-transduced K562 cells 6 days after transduction (n=2). **G)** Representative images of cell pellets obtained from shRenilla and shPARK7-transduced K562 cells 6 days after transduction.



Supplementary Figure 5. Exposure to the coREST complex inhibitor Corin results in reduced cell growth but does not differentiate HUDEP2 cells. A) Growth curve of Corin-treated HUDEP2 cells (DMSO, 100nM, 200nM and 500nM; n=2). **B)** Representative images of flow cytometry analysis of CD71 and CD235a expression after 2 days of treatment.

7.3. Supplementary tables

Suppl. Table 1. 6774 commonly expressed proteins → will be provided online upon publication

Suppl. Table 2. Most abundant proteins in AEL cells (logFC>2; adj.p-value<0.05)

Protein	logFC	P.Value	adj.P.Val
HBZ	3.98346541	1.7349E-05	0.0001729
HBE1	3.49779541	0.00015008	0.00098975
PLVAP	3.43452726	8.9492E-08	3.2508E-06
HLA-DRB3	3.42910683	4.0747E-06	5.7225E-05
IGF2BP1	3.41082604	4.8547E-05	0.00039759
LIMD2	3.35690339	0.00011022	0.00077238
ANXA1	3.31351564	1.9101E-05	0.00018641
GPX7	3.23123388	2.7821E-06	4.3719E-05
LIN28B	3.2087099	5.553E-06	7.1913E-05
ICAM1	3.03986632	3.1322E-10	5.7337E-08
GTSF1	2.96455037	7.681E-05	0.00057421
MCAM	2.9065255	1.2201E-06	2.3344E-05
DDX3Y	2.85243739	2.8017E-06	4.3826E-05
CKB	2.81849615	0.00012383	0.00084544
MAGEA4	2.79372381	0.00635202	0.01930979
ANXA6	2.52499615	5.8652E-05	0.00046338
PKM	2.49047554	1.3617E-08	8.0195E-07
FYB1	2.4598166	3.2585E-06	4.8292E-05
LPCAT1	2.44459076	7.6575E-05	0.00057372
ACSM3	2.44041942	7.7618E-07	1.6742E-05
GM2A	2.43128959	3.0756E-05	0.00027554
COTL1	2.38123967	9.4042E-06	0.00010869
MEF2C	2.35018789	5.5204E-08	2.2389E-06
DNM3	2.25696564	1.1307E-05	0.00012534
CLIC1	2.21128735	6.337E-05	0.00049285
ALDH1A1	2.18569401	0.00015449	0.00101117
ENO1	2.18500806	5.0891E-09	4.0551E-07
PHGDH	2.17569339	6.2072E-06	7.8436E-05
C11orf54	2.14807029	5.4404E-10	8.7733E-08
CNN2	2.12844564	1.2263E-07	4.2161E-06
DBN1	2.09270755	0.00064087	0.00312806
TBCA	2.0914755	1.2843E-11	7.9647E-09
ALDH2	2.0899643	0.00243554	0.00898471
QPRT	2.06403402	3.4415E-12	6.8323E-09
GSTP1	2.05204301	3.3481E-08	1.5118E-06
ARHGDIB	2.04285617	0.00074079	0.00352293
PLTP	2.02962994	1.9664E-07	5.9192E-06
RNASET2	2.0178017	7.5285E-09	5.3115E-07

Suppl. Table 3. Most abundant proteins in normal cells (logFC>-2; adj.p-value<0.05)

Protein	logFC	P.Value	adj.P.Val
LMNA	-2.0001507	1.6654E-05	0.0001681
PALM2AKAP2	-2.0164232	5.9336E-06	7.6115E-05
BCAM	-2.0278731	1.5714E-06	2.8535E-05
A2M	-2.032597	0.002553	0.00929504
ITGB4	-2.0804708	1.1041E-05	0.00012389
SQOR	-2.0932366	4.665E-06	6.3831E-05
HBA1	-2.0974757	1.6539E-06	2.9556E-05
CD82	-2.1212465	9.2014E-06	0.0001069
SLC25A37	-2.1265718	2.1223E-10	4.6369E-08
VTN	-2.1276441	7.3827E-05	0.00055931
ELOA3DP	-2.176002	9.8279E-09	6.4089E-07
CLGN	-2.1883133	0.00030015	0.00170119
RSAD2	-2.1912916	1.7127E-10	3.8667E-08
F2	-2.2083146	0.00034789	0.00190635
ACSS1	-2.2127982	8.7599E-09	5.9331E-07
AQP1	-2.2151333	1.6514E-06	2.9556E-05
TF	-2.2204693	8.8556E-11	2.6078E-08
C1QC	-2.2264561	6.2911E-06	7.92E-05
RPL22L1	-2.2390124	1.5991E-09	1.7469E-07
TSPO	-2.242439	7.9915E-05	0.00059194
SLC2A1	-2.3140371	1.3703E-05	0.00014604
SPTA1	-2.3224352	5.8471E-06	7.5147E-05
IGHM	-2.3601663	0.00020422	0.00125515
TFRC	-2.3652566	1.2935E-11	7.9647E-09
IGHG1	-2.3727942	1.8336E-06	3.2007E-05
PTGES	-2.426088	7.0808E-12	6.8512E-09
CFB	-2.459482	2.438E-06	4.008E-05
TMCC2	-2.5085083	8.4422E-05	0.00061815
HEMGN	-2.5541433	0.00022988	0.00137784
NCEH1	-2.6221203	7.0853E-10	1.0664E-07
KMT2C	-2.6691083	1.3689E-10	3.4338E-08
IGKC	-2.7527766	3.6574E-09	3.2467E-07
EPB42	-2.7749028	2.4098E-06	3.9949E-05
HBB	-2.8078394	6.7765E-10	1.0431E-07
SPTB	-2.8524329	5.3941E-08	2.2008E-06
CD36	-2.8651763	2.8439E-08	1.3376E-06
HBD	-2.9564703	1.7359E-05	0.0001729
PLG	-3.0164754	1.0223E-06	2.0485E-05
LRP1	-3.0340532	2.0171E-05	0.00019406
FAS	-3.0415907	2.264E-07	6.3893E-06
ITIH4	-3.1447835	1.6551E-05	0.00016765
TRIM10	-3.2236621	9.5423E-07	1.9467E-05
ALB	-3.2925778	4.7847E-06	6.4702E-05
GNPMB	-3.3273885	1.2437E-05	0.00013587
APOA1	-3.3429464	2.1398E-07	6.167E-06

SLCO4C1	-3.3534523	5.0688E-07	1.1797E-05
SLC4A1	-3.3739535	8.5187E-07	1.7974E-05
RHCE	-4.4510475	2.5963E-06	4.1719E-05

Suppl. Table 4. Erythroid transcription factors identified in whole proteome (only GATA1 and GFI-1B are significantly differentially expressed $\log_{2}FC > 1$ or $\log_{2}FC < -1$; $\text{adj. p-value} < 0.05$).

Protein	logFC	P.Value	adj.P.Val
CBFA2T3/ETO2	0.94221887	4.5112E-07	1.0873E-05
CREBBP/CBP	0.33247203	0.02526534	0.05750072
EED	0.04606689	0.7349809	0.80238969
EZH2	-0.0200647	0.89622471	0.92405693
GATA1	-1.0512091	0.00066525	0.00322991
GATA2	-0.8426445	4.0732E-06	5.7225E-05
GFI1B	-1.8333947	2.4316E-06	4.0072E-05
HDAC1	-0.298	0.14967235	0.23813268
HDAC2	-0.1886537	0.03094274	0.06780174
IKZF1	-0.65432	0.00225522	0.00846707
KLF1	-0.5946298	0.01315855	0.03433827
LDB1	0.25078885	0.10970209	0.18687708
NFE2	-0.99856	9.0553E-07	1.8756E-05
NR3C1	-0.2194504	0.19832685	0.29491029
RCOR1	-0.4935951	0.01115536	0.02999413
RUNX1	0.34665512	0.13024396	0.21458096
STAT5A	0.77657964	0.00934442	0.02591719
STAT5B	0.69968188	0.00048706	0.00250864
SUZ12	-0.3002967	0.07788083	0.14210315
TAL1	-0.5710173	0.00013986	0.00093512
TCF3/E2A	-0.595825	0.0019836	0.00761619
ZFPM1/FOG1	-0.0853314	0.55704818	0.65569818

Suppl. Table 5. 1616 identified proteins in GATA1-IP/MS analysis → will be provided online upon publication

Suppl. Table 6. 125 Differentially associated proteins with GATA1 sorted based on adj. p-value (positive logFC more enriched in leukemic erythroblasts, negative logFC more enriched in normal erythroblasts)

Protein	logFC	P.Value	adj.P.Val
NUP155	-6.0695	0.0000000	0.0000001
RPS12	3.7909	0.0000001	0.0000187
ATP1A1	4.0412	0.0000001	0.0000187
RPN2	3.5509	0.0000002	0.0000293
LRPPRC	3.5627	0.0000002	0.0000293
CCT7	2.9289	0.0000003	0.0000404
CCT4	3.9086	0.0000007	0.0000721
CALU	5.2150	0.0000013	0.0001182
MCM6	-1.9886	0.0000018	0.0001316
CLTC	3.9467	0.0000019	0.0001316
VARS	3.7515	0.0000038	0.0002407
PLRG1	-3.0002	0.0000061	0.0003552
GNAI3	2.7370	0.0000084	0.0004555
COX5A	-1.9868	0.0000176	0.0008376
MRPS23	-2.0027	0.0000185	0.0008376
ESYT1	-3.0137	0.0000191	0.0008376
RPLP0	2.7389	0.0000251	0.0010381
MYL12B	4.5299	0.0000288	0.0011255
MRPS28	-1.8695	0.0000416	0.0015380
RUNX1	-2.4827	0.0000541	0.0018096
PHGDH	4.8163	0.0000478	0.0016802
CLPX	-1.4638	0.0000892	0.0028519
SAFB2	-2.3052	0.0001293	0.0039521
adj.p-value<0.005			
KPNA2	1.65712792	0.00018249	0.00524315
ZEB2	2.02087564	0.00018646	0.00524315
FLOT1	-1.5224733	0.00037597	0.00930709
SLC3A2	3.54536488	0.00030999	0.00838163
RANGAP1	-1.8348691	0.0003669	0.00930709
FKBP8	-2.5845735	0.00050432	0.01143657
MTA2	-1.8881242	0.00038393	0.00930709
LDHB	2.52743057	0.00044926	0.01052757
XRCC5	2.72284348	0.00058688	0.01250227
SKI	1.81851247	0.00058663	0.01250227
EPRS	1.84868316	0.00079934	0.01605534
TUBA1B	2.32312034	0.00066901	0.01383274
LDHA	2.27679751	0.00092347	0.0180334
RUVBL2	-1.6018207	0.00123408	0.02344758
RPS21	-1.0896045	0.00166216	0.02996154
KPNB1	3.53656812	0.00140472	0.02598724
FASN	1.86963336	0.00219883	0.0386444
C1QBP	2.2096228	0.00263238	0.04513575
adj.p-value<0.05			
TMED9	-2.6800077	0.00386544	0.06319552

AGPS	-1.3836609	0.00319437	0.05346759
SND1	4.05979188	0.00650519	0.0874512
ACADM	-1.6995158	0.00411398	0.06573023
SMC2	-1.0425879	0.00513231	0.07676624
PGK1	2.56303785	0.00452322	0.07066268
NASP	2.35750125	0.00559517	0.08027355
C19orf53	1.35564384	0.00475973	0.07274116
FUBP1	-1.788642	0.00541195	0.07926246
RAVER1	-1.164332	0.00593552	0.08345343
FSCN1	2.44096254	0.00659305	0.0874512
RANBP1	-1.8274036	0.00719576	0.09367814
TKT	2.78140962	0.00630133	0.08685954
adj.p-value<0.1			
RCOR1	2.21484226	0.00867685	0.11090588
PNN	-1.2948989	0.0095853	0.11821873
STOML2	-1.1050384	0.01052224	0.12753678
HBA1	-1.0653606	0.00924802	0.11609565
CCT6A	2.33364155	0.0124635	0.14288994
ACTN4	1.61067535	0.0144522	0.15874843
HSP90AA1	1.52951149	0.01490332	0.15914265
P4HB	1.80303043	0.01348381	0.15046215
SMC1A	-1.0777097	0.01260196	0.14288994
PAM16	7.75307935	0.02287933	0.19724219
IKZF1	-1.7163343	0.0115603	0.13774392
CCDC47	-1.0259481	0.01653653	0.1706583
DDX17	-0.8672582	0.01255225	0.14288994
TCF12	-1.832182	0.01617619	0.16972926
PARK7	2.96534731	0.01788554	0.17571201
RPS5	3.29747543	0.01494085	0.15914265
EIF3B	-1.0638168	0.01818821	0.17571201
FUBP3	-0.9362811	0.02074244	0.18694792
RPS19BP1	1.24715018	0.01675025	0.1706583
MYL6	-1.5583539	0.01824606	0.17571201
RAP1B	-1.1183822	0.01911691	0.17980143
NUP93	-0.9904101	0.01968272	0.18142386
MSH6	-1.1196652	0.02217127	0.19724219
TPR	-2.0771185	0.01987146	0.18142386
SERBP1	1.72162675	0.01918223	0.17980143
YWHAQ	-1.4976776	0.01814894	0.17571201
HBG1	-1.6158877	0.02266076	0.19724219
SFXN1	-0.8132197	0.02498571	0.21004043
UQCRH	-1.262667	0.02509729	0.21004043
TPX2	-1.1137705	0.02300691	0.19724219
HIST1H3A	2.70522477	0.04634625	0.32258827
SHMT2	1.89465778	0.0265851	0.21987445
MPP1	-0.8185733	0.03189218	0.25191241
SSR4	-1.0781114	0.03341239	0.25811987

FLII	-1.6265262	0.0369014	0.27307038
CCT5	1.3818206	0.03142617	0.25191241
ILF3	-1.8392866	0.03456067	0.26208807
RPN1	-0.9688628	0.02903298	0.23732776
ZNF326	-1.6749374	0.03265241	0.25505157
BANF1	2.95753899	0.04056931	0.29102269
HSP90B1	2.47624486	0.03467168	0.26208807
JUP	-2.6620618	0.03175626	0.25191241
MCM7	-3.4288095	0.03649172	0.27291151
HADHA	-1.4295375	0.04452609	0.31618025
HADHB	-1.6636407	0.03969909	0.28771608
ZC3H13	-0.9967134	0.04899119	0.33557237
RANBP2	-1.9006749	0.04616956	0.32258827
PCBP1	-0.821326	0.04916636	0.33557237
EIF5A	1.39903544	0.06039442	0.40054032
ILF2	-0.7137939	0.05990187	0.40054032
RPL4	0.89987854	0.05944529	0.40054032
CMC1	-1.4778341	0.06303575	0.41209805
RBM27	0.92357116	0.06330952	0.41209805
TUBB6	-4.3778787	0.06533507	0.41378876
SDHA	-0.6695378	0.06504823	0.41378876
HRNR	-1.9360122	0.06476223	0.41378876
HNRNPF	-1.1668812	0.07057148	0.44296206
PRPF31	-7.3765129	0.073911	0.45223841
DAZAP1	-0.7678254	0.07462255	0.45223841
IMMT	-0.836631	0.07429402	0.45223841
HNRNPDL	2.48667495	0.07347725	0.45223841
CBFA2T3	-1.7332433	0.07787442	0.46237673
UQCRC2	-0.9770895	0.07816545	0.46237673
H2AFZ	1.71581414	0.07826861	0.46237673
PGAM1	1.4626955	0.07986766	0.46789137
HSPA1L	-1.532997	0.08340507	0.4787774
RBM33	1.61174112	0.082731	0.4787774
CCDC59	1.719617	0.08376902	0.4787774
HSD17B8	-2.7140962	0.08960545	0.49994154
MCM2	-0.5176671	0.08898705	0.49994154
PRKCSH	-1.5723007	0.0884503	0.49994154
adj.p-value<0.5			

Suppl. Table 7. Inactivation of genes by targeted CRISPR/Cas9 screen in *SpCas9*-K562 on Day 4. Genes were normalized to sgAAVS1 and assessed for their depletion by calculating the fold change (fc) and erythroid differentiation by calculating the MFI.

gene	Diff. Expressed to AAVs-1	fc	Differentiation	depletion
Cluster 1 (62 genes, excluding negative control sgAAVS1)				
AAVS1	ns	1	CD71/CD235a low or not sig. different to AAVs1	no
GNAI3	ns	0.95	CD71/CD235a low or not sig. different to AAVs1	no
IKZF1	ns	1.035	CD71/CD235a low or not sig. different to AAVs1	no
IMMT	ns	0.955	CD71/CD235a low or not sig. different to AAVs1	no
MSH6	ns	0.96	CD71/CD235a low or not sig. different to AAVs1	no
MTA2	ns	0.92	CD71/CD235a low or not sig. different to AAVs1	no
MYL6	ns	0.83	CD71/CD235a low or not sig. different to AAVs1	no
PHGDH	ns	0.975	CD71/CD235a low or not sig. different to AAVs1	no
SKI	ns	0.955	CD71/CD235a low or not sig. different to AAVs1	no
TCF12	ns	0.91	CD71/CD235a low or not sig. different to AAVs1	no
TUBB6	ns	0.845	CD71/CD235a low or not sig. different to AAVs1	no
ZEB2	ns	1	CD71/CD235a low or not sig. different to AAVs1	no
AAVS1	ns	1	CD71/CD235a low or not sig. different to AAVs1	no
AGPS	CD235a	0.925	CD71/CD235a low or not sig. different to AAVs1	no
CLPX	ns	1.19	CD71/CD235a low or not sig. different to AAVs1	no
TMED9	ns	1.115	CD71/CD235a low or not sig. different to AAVs1	no
XRCC5	both	1.035	CD71/CD235a low or not sig. different to AAVs1	no
AAVS1	ns	1	CD71/CD235a low or not sig. different to AAVs1	no
ACADM	ns	1.205	CD71/CD235a low or not sig. different to AAVs1	no
ACTN4	ns	1.02	CD71/CD235a low or not sig. different to AAVs1	no
C1QBP	ns	1.39	CD71/CD235a low or not sig. different to AAVs1	no
FASN	ns	1.285	CD71/CD235a low or not sig. different to AAVs1	no
FLOT1	ns	1.195	CD71/CD235a low or not sig. different to AAVs1	no
FSCN1	ns	1.15	CD71/CD235a low or not sig. different to AAVs1	no
FUBP1	ns	0.975	CD71/CD235a low or not sig. different to AAVs1	no
HBA1	ns	1.19	CD71/CD235a low or not sig. different to AAVs1	no
MRPS28	ns	1.085	CD71/CD235a low or not sig. different to AAVs1	no
NASP	ns	1.01	CD71/CD235a low or not sig. different to AAVs1	no
P4HB	ns	1.155	CD71/CD235a low or not sig. different to AAVs1	no
PGK1	ns	1.14	CD71/CD235a low or not sig. different to AAVs1	no
SLC3A2	ns	1.1	CD71/CD235a low or not sig. different to AAVs1	no
SND1	ns	1.14	CD71/CD235a low or not sig. different to AAVs1	no
STOML2	ns	0.95	CD71/CD235a low or not sig. different to AAVs1	no
AAVS1	ns	1	CD71/CD235a low or not sig. different to AAVs1	no
CMC1	CD235a	1.01	CD71/CD235a low or not sig. different to AAVs1	no
DAZAP1	ns	1.105	CD71/CD235a low or not sig. different to AAVs1	no
HNRNPDL	ns	1.255	CD71/CD235a low or not sig. different to AAVs1	no
HNRNPF	ns	1.155	CD71/CD235a low or not sig. different to AAVs1	no
MPP1	ns	1.24	CD71/CD235a low or not sig. different to AAVs1	no
PCBP1	ns	0.825	CD71/CD235a low or not sig. different to AAVs1	no
TPR	ns	0.87	CD71/CD235a low or not sig. different to AAVs1	no

AAVS1	ns	1	CD71/CD235a low or not sig. different to AAVs1	no
CCDC47	ns	1.34	CD71/CD235a low or not sig. different to AAVs1	no
CCDC59	ns	1.215	CD71/CD235a low or not sig. different to AAVs1	no
HADHA	ns	1.18	CD71/CD235a low or not sig. different to AAVs1	no
HADHB	ns	1.34	CD71/CD235a low or not sig. different to AAVs1	no
ILF3	ns	1.005	CD71/CD235a low or not sig. different to AAVs1	no
JUP	ns	1.21	CD71/CD235a low or not sig. different to AAVs1	no
MCM2	ns	0.825	CD71/CD235a low or not sig. different to AAVs1	no
MCM7	ns	0.805	CD71/CD235a low or not sig. different to AAVs1	no
PRKCSH	ns	0.925	CD71/CD235a low or not sig. different to AAVs1	no
RAVER1	ns	1.19	CD71/CD235a low or not sig. different to AAVs1	no
SDHA	ns	1.045	CD71/CD235a low or not sig. different to AAVs1	no
SFXN1	ns	1.225	CD71/CD235a low or not sig. different to AAVs1	no
SHMT2	ns	0.97	CD71/CD235a low or not sig. different to AAVs1	no
SSR4	ns	1.285	CD71/CD235a low or not sig. different to AAVs1	no
TKT	ns	1.115	CD71/CD235a low or not sig. different to AAVs1	no
ZC3H13	ns	0.915	CD71/CD235a low or not sig. different to AAVs1	no
AAVS1	ns	1	CD71/CD235a low or not sig. different to AAVs1	no
C19orf53	ns	0.95	CD71/CD235a low or not sig. different to AAVs1	no
EIF5A	both	1.065	CD71/CD235a low or not sig. different to AAVs1	no
HRNR	both	0.97	CD71/CD235a low or not sig. different to AAVs1	no
HSD17B8	both	0.915	CD71/CD235a low or not sig. different to AAVs1	no
HSP90AA1	both	1	CD71/CD235a low or not sig. different to AAVs1	no
HSP90B1	both	0.925	CD71/CD235a low or not sig. different to AAVs1	no
RBM33	CD235a	0.9	CD71/CD235a low or not sig. different to AAVs1	no
RPN1	CD235a	0.855	CD71/CD235a low or not sig. different to AAVs1	no
ZNF326	ns	0.9	CD71/CD235a low or not sig. different to AAVs1	no

Cluster 2 (21 genes, excluding positive control RPL-17 and CD71)

FLII	ns	0.685	CD71/CD235a low or not sig. different to AAVs1	yes
NUP155	ns	0.525	CD71/CD235a low or not sig. different to AAVs1	yes
RANBP2	ns	0.795	CD71/CD235a low or not sig. different to AAVs1	yes
RPL-17	ns	0.4	CD71/CD235a low or not sig. different to AAVs1	yes
SMC2	ns	0.645	CD71/CD235a low or not sig. different to AAVs1	yes
CD71	CD71	0.306930693	CD71/CD235a low or not sig. different to AAVs1	yes
PLRG1	ns	0.695	CD71/CD235a low or not sig. different to AAVs1	yes
RPL-17	ns	0.4	CD71/CD235a low or not sig. different to AAVs1	yes
RPS21	CD71	0.755	CD71/CD235a low or not sig. different to AAVs1	yes
RUVBL2	CD71	0.75	CD71/CD235a low or not sig. different to AAVs1	yes
CCT6A	ns	0.675	CD71/CD235a low or not sig. different to AAVs1	yes
CD71	ns	0.306930693	CD71/CD235a low or not sig. different to AAVs1	yes
KPNB1	ns	0.305	CD71/CD235a low or not sig. different to AAVs1	yes
PNN	ns	0.44	CD71/CD235a low or not sig. different to AAVs1	yes
RPL-17	ns	0.4	CD71/CD235a low or not sig. different to AAVs1	yes
CD71	ns	0.306930693	CD71/CD235a low or not sig. different to AAVs1	yes
ILF2	ns	0.58	CD71/CD235a low or not sig. different to AAVs1	yes
PGAM1	ns	0.67	CD71/CD235a low or not sig. different to AAVs1	yes
PRPF31	ns	0.65	CD71/CD235a low or not sig. different to AAVs1	yes

RPL-17	ns	0.4	CD71/CD235a low or not sig. different to AAVs1	yes
CD71	ns	0.306930693	CD71/CD235a low or not sig. different to AAVs1	yes
EPRS	ns	0.785	CD71/CD235a low or not sig. different to AAVs1	yes
RPL-17	ns	0.4	CD71/CD235a low or not sig. different to AAVs1	yes
RPL4	ns	0.215	CD71/CD235a low or not sig. different to AAVs1	yes
RPS5	ns	0.27	CD71/CD235a low or not sig. different to AAVs1	yes
TPX2	ns	0.57	CD71/CD235a low or not sig. different to AAVs1	yes
CCT5	CD235a	0.515	CD71/CD235a low or not sig. different to AAVs1	yes
EIF3B	both	0.595	CD71/CD235a low or not sig. different to AAVs1	yes
MRPS23	ns	0.77	CD71/CD235a low or not sig. different to AAVs1	yes
RANBP1	ns	0.665	CD71/CD235a low or not sig. different to AAVs1	yes
RPL-17	ns	0.4	CD71/CD235a low or not sig. different to AAVs1	yes

Cluster 3 (11 genes)

RUNX1	CD235a	0.795	CD71/CD235a sig. different to AAVs1	yes
SMC1A	CD235a	0.45	CD71/CD235a sig. different to AAVs1	yes
ATP1A1	CD235a	0.74	CD71/CD235a sig. different to AAVs1	yes
CLTC	both	0.635	CD71/CD235a sig. different to AAVs1	yes
RANGAP1	CD235a	0.785	CD71/CD235a sig. different to AAVs1	yes
RPLP0	both	0.32	CD71/CD235a sig. different to AAVs1	yes
NUP93	CD235a	0.705	CD71/CD235a sig. different to AAVs1	yes
PAM16	both	0.74	CD71/CD235a sig. different to AAVs1	yes
SERBP1	CD235a	0.695	CD71/CD235a sig. different to AAVs1	yes
RPS12	CD235a	0.725	CD71/CD235a sig. different to AAVs1	yes
UQCRC2	both	0.735	CD71/CD235a sig. different to AAVs1	yes

Cluster 4 (22 genes)

CBFA2T3	both	0.88	No depletion, CD71/CD235a sig. different to AAVs1	no
FKBP8	CD235a	0.96	No depletion, CD71/CD235a sig. different to AAVs1	no
RCOR1	CD235a	0.92	No depletion, CD71/CD235a sig. different to AAVs1	no
CALU	CD235a	1.09	No depletion, CD71/CD235a sig. different to AAVs1	no
CCT4	CD235a	0.925	No depletion, CD71/CD235a sig. different to AAVs1	no
CCT7	CD235a	0.82	No depletion, CD71/CD235a sig. different to AAVs1	no
ESYT1	CD235a	1.18	No depletion, CD71/CD235a sig. different to AAVs1	no
KPNA2	CD235a	1.1	No depletion, CD71/CD235a sig. different to AAVs1	no
LDHA	CD235a	1.115	No depletion, CD71/CD235a sig. different to AAVs1	no
LDHB	CD235a	1.095	No depletion, CD71/CD235a sig. different to AAVs1	no
LRPPRC	CD235a	0.885	No depletion, CD71/CD235a sig. different to AAVs1	no
MCM6	CD235a	0.97	No depletion, CD71/CD235a sig. different to AAVs1	no
RPN2	CD235a	0.96	No depletion, CD71/CD235a sig. different to AAVs1	no
SAFB2	both	1.145	No depletion, CD71/CD235a sig. different to AAVs1	no
VARS	CD235a	0.805	No depletion, CD71/CD235a sig. different to AAVs1	no
COX5A	both	0.975	No depletion, CD71/CD235a sig. different to AAVs1	no
DDX17	both	1.1	No depletion, CD71/CD235a sig. different to AAVs1	no
FUBP3	both	1.055	No depletion, CD71/CD235a sig. different to AAVs1	no
HBG1	CD235a	0.915	No depletion, CD71/CD235a sig. different to AAVs1	no
PARK7	CD235a	1.1	No depletion, CD71/CD235a sig. different to AAVs1	no
RBM27	CD235a	1.135	No depletion, CD71/CD235a sig. different to AAVs1	no
YWHAQ	CD235a	1.145	No depletion, CD71/CD235a sig. different to AAVs1	no

Suppl. Table 8. Primer sequence used for genotyping PCR

Target	Sequence forward (5' - 3')	Sequence reverse (5' - 3')
<i>Vav-iCre</i>	ctc tga cag atg cca gga ca	tga ttg cag gga tgg aca ca
<i>Nsd1 ex.4</i>	cta cct aca taa tga gtt cca gg	gcc aga gac ttc aag cag cca tt
<i>UBC Cre ERT2</i>	gac gtc acc cgt tct gtt g	agg caa att ttg gtg tac gg
<i>mGata1</i>	gca tgg acg agc tgt aca ag	agg tac tgc cca cct cta tc

Suppl. Table 9. Maintenance medium for murine erythroblasts

ingredient	Stock conc.	Final conc.	Volume for 50 ml
StemSpan SFEM			49 ml
Mouse SCF	100 µg	100 ng/ml	500 ul
P/S	100%	1%	500 ul
Cholesterol	100%	0.4%	200 ul
Human EPO	4000 U/ml	2 U/ml	25 ul
Human IGF1	0.1 mg/ml	40 ng/ml	20 ul
Dexamethasone	0.1 Molar	10 ⁻⁶ Molar	0.5 ul

Suppl. Table 10. Differentiation medium for murine erythroblasts

ingredient	Stock conc.	Final conc.	Volume for 50 ml
IMDM + glutamine (1x)			36.5 ml
PDS filtered	100%	5%	2.5 ml
PFHMII	100%	10%	5 ml
MTG	1.25 g/ml	1:100	63.5 ul
Mouse SCF	100 µg	100 ng/ml	500 ul
Human EPO	4000 U/ml	2 U/ml	25 ul
P/S	100%	1%	500 ul
FCS	100%	10%	5 ml

Suppl. Table 11. HUDEP2 maintenance medium

ingredient	Stock conc.	Final conc.	Volume for 50 ml
StemSpan SFEM			49 ml
Human SCF	100 µg	50 ng/ml	250 ul
P/S	100%	2%	1 ml
Human EPO	4000 U/ml	3 U/ml	37.5 ul
Dexamethasone	0.1 Molar	10 ⁻⁶ Molar	0.5 ul
Doxycycline	10 mg/ml	1 µg/ml	5 ul

Suppl. Table 12. HUDEP2 differentiation medium

ingredient	Stock conc.	Final conc.	Volume for 50 ml
IMDM+ glutamine (1x)			41 ml
P/S	100%	2%	1 ml
Human EPO	4000 U/ml	3 U/ml	37.5 ul
Holo-transferrin	5 mg/ml	500 µg/ml	5 ml
Heparin	5000 U/ml	3 U/ml	30 ul
Human insulin	10 mg/ml	20 µg/ml	50 ul
Inactivated human plasma AB	100%	5%	2.5 ml

Suppl. Table 13. "Cellquin V4.5" medium preparation

Ingredient	Stock conc.	Final conc.	Dilution factor	Vol for 1 bottle	
HSA (production)	20%	0.1%	1:200	2.59	mL
Transferrin	50 mg/mL	0.3 mg/mL	1:167	3.11	mL
Sigma lipid mixture	1000X	1X	1:1000	0.518	mL
Insulin	5 mg/mL	0.01 mg/mL	1:500	1.036	mL
Na.pyruvate	100 mM	0.1 mM	1:1000	0.518	mL
Pen/Strep	16.8 mM/6.9 mM (100X)	1X	1:100	5.00	mL
L-glutamine	200 mM	2 mM	1:100	5.00	mL
PAN IMDM				500	mL

Suppl. Table 14. Flow cytometric surface markers for murine cells

Marker	Clone	Fluorochrome	Conc.	Manufacturer
Anti-Mouse CD71	RI7217	Pe-Cy7	1:100	BioLegend
Anti-Mouse TER119	TER-119	APC	1:100	BioLegend
Anti-Mouse c-Kit	2B8	BV711	1:200	BD Horizon
Anti-Mouse Mac1	M1/70	BV510	1:200	BD Horizon
Anti-Mouse Gr-1	RB6-8C5	APC-Cy7	1:100	BD Pharmingen

Suppl. Table 15. Flow cytometric surface markers for human cells

Marker	Clone	Fluorochrome	Conc.	Manufacturer
Anti-human CD71	CY1G4	APC	1:200	BioLegend
Anti-human CD235a	GA-R2	BV421	1:200	BD Horizon
Anti-humanCD36	5-271	PerCP/Cy5.5	1:200	BioLegend
Anti-humanCD34	581	APC-Cy7	1:200	BioLegend

Suppl. Table 16. Plasmids

Short name	Full name	Internal plasmid bank number
Mouse MOCK	pMSCV-PGK-puro	15
mGata1	pMSCV-mGata1-PGK-puro	568
hGATA1	pSIN4-EF1a-GATA1-IRES-Puro	676
Human MOCK	pSIN4-EF1a-Empty vector	687
KPNA2	pMSCV-KPNA-IRES-GFP	717
RCOR1	pMSCV-RCOR1-IRES-GFP	Not in bank
NUP155 ΔC	pMSCV-NUP155-deltaC1-IRES-GFP	702
NUP155 ΔN	pMSCV-NUP155-deltaN1-HA-IRES-GFP	701
NU155 FL	pMSCV-NUP155-HA-IRES-GFP	700
MOCK	pMSCV-IRES-GFP (MIG)	14
LentiCas9-Blast	Plasmid #52962, Addgene	Gift from Grebien lab

Suppl. Table 17. Primer sequences used for real-time (RT- PCR analysis

Gene Target	Sequence forward (5'-3')	Sequence reverse (5'-3')
<i>mGapdh</i>	ATGACATCAAGAAGGTGGTG	CATACCAGGAAATGAGCTTG
<i>mGata1</i>	GTGTCCTCACCATCAGATTCCAC	TGCATTTGGGGAAGTGAAGA
<i>HA-mGata1</i>	ATGACGTGCCTGACTATGCC	TGCATTTGGGGAAGTGAAGA
<i>hGATA1</i>	AAACGGGCAGGTACTIONCAGTG	CGGTTACCTGGTGTAGCTT
<i>hHBA</i>	GACCTGCACGCGCACAAAGCTT	GCTCACAGAAGCCAGGAACCTG
<i>hALAS1</i>	GATGTCAGCCACCTCAGAGAAC	CATCCACGAAGGTGATTGCTCC
<i>hKPNA2</i>	TTTGAACGCAGTCGCCCTAC	GGGCAGCTGGTGTATTAGCA
<i>hRCOR1</i>	CCAGTAACCAGAAGCCTGTGAAG	AAGCCACCAGTTTCTCAGGAGG
<i>hCCT4</i>	CATGTCTCGACCTGTGGAAGCTG	CACTCTTGGAGAAAGCAGACTTG
<i>hGAPDH</i>	CCTGACCTGCCGTCTAGAAA	CTCCGACGCCTGCTTCAC

Suppl. Table 18. Primary antibodies used for Western Blotting

Antibody	Conc.	Clone	Manufacturer	Host species
H3K27me3	1:2000	07-449	Millipore	rabbit
H3K36me1	1:2000	5928	Cell Signaling	rabbit
H3K36me2	1:1000	C75H12-2901S	Cell signaling	rabbit
H3K36me3	1:2000	Ab9050	Abcam	rabbit
H3	1:5000	D1H2	Cell Signaling	rabbit
GATA1	1:1000	D52H6 XP	Cell Signaling	rabbit
CBFA2T3	1:500	ab33072	Abcam	rabbit
NUP155	1:500	GTX120945	GeneTex	rabbit
SKI	1:1000	G-8	Santa Cruz	mouse
IKZF1	1:500	E-2	Santa Cruz	mouse
RUNX1	1:1000	GTX129924	GeneTex	rabbit
NCOR	1:500	F-1	Santa Cruz	mouse
RCOR1/coREST	1:1000	H-8	Santa Cruz	mouse
KPNA2	1:1000	B-9	Santa Cruz	mouse
CCT4/TCP-1δ	1:1000	H-1	Santa Cruz	mouse
HSC70	1:1000	K-19	Santa Cruz	goat
HSC70	1:2000	GTX134584	GeneTex	rabbit

Suppl. Table 19. Secondary antibodies used for Western Blotting

Antibody	Conc.	Clone	Manufacturer	Host species
VeriBlot HRP	1:2000	131366	Abcam	goat
Mouse True Blot® Ultra Anti-Mouse IgG HRP	1:2000	18-8817-31	Rockland	rat
Goat TrueBlot® Ultra Anti-Goat IgG HRP	1:2000	18-8814-31	Rockland	mouse
Goat anti-Mouse IgG HRP	1:2000	31432	Pierce	mouse
Donkey anti-rabbit IgG, HRP	1:2000	AP182P	Millipore	rabbit

Suppl. Table 20. Sequences of sgRNA oligos used in CRISPR/Cas9 screen

Forward sgRNA	Sequence	Reverse sgRNA	Sequence
CBFA2T3_fwd_1	caccgGACGTGCAGCGGCTGCAACG	CBFA2T3_rev_1	aaacCGTTGCAGCCGCTGCACGTCc
CBFA2T3_fwd_2	caccgGAACTCGACATTGACGATCG	CBFA2T3_rev_2	aaacCGATCGTCAATGTTCGAGTTCc
FKBP8_fwd_1	caccgGAAGTGTCTGAACAACCTGG	FKBP8_rev_1	aaacCCAGGTTGTTTCAGACACTTCc
FKBP8_fwd_2	caccgGAAGCTGGAACCTTCCAACA	FKBP8_rev_2	aaacTGTGGAAAGGTTCCAGCTTCc
FLII_fwd_1	caccgGCCGAGCCAGATGAACACGT	FLII_rev_1	aaacACGTGTTTCATCTGGCTCGGCc
FLII_fwd_2	caccgGCCCCACAGGTCTACATGT	FLII_rev_2	aaacACATGTAGACCTGTGGGGGCc
GNAI3_fwd_1	caccgGATGATCGACCGCAACTTAC	GNAI3_rev_1	aaacGTAAGTTGCGGTCGATCATCc
GNAI3_fwd_2	caccgGATCGACCGCAACTTACGGG	GNAI3_rev_2	aaacCCCGTAAGTTGCGGTCGATCc
IKZF1_fwd_1	caccgAGGTGGTGGAGAGGTCTCTCG	IKZF1_rev_1	aaacCGAGGACCTCTCCACCACCTc
IKZF1_fwd_2	caccgTGAGCCCATGCCGATCCCCG	IKZF1_rev_2	aaacCGGGGATCGGCATGGGCTCAc
IMMT_fwd_1	caccgGACGATGTGCATGAGCAATG	IMMT_rev_1	aaacCATTGCTCATGCACATCGTCc
IMMT_fwd_2	caccgGAAGGGGAATCCAGACGAG	IMMT_rev_2	aaacCTCGTCTGGATTCCCCCTTCc
MSH6_fwd_1	caccgGTTACTGGACCAAATATGGG	MSH6_rev_1	aaacCCCATATTTGGTCCAGTAACc
MSH6_fwd_2	caccgAGGGACGTAACAACCCATCT	MSH6_rev_2	aaacAGATGGGTTGTTACGTCCCTc
MTA2_fwd_1	caccgAGATCCAACAGGAAGCACAG	MTA2_rev_1	aaacCTGTGCTTCTGTTGGATCTc
MTA2_fwd_2	caccgGATCAGGGCGAGATTAGAGT	MTA2_rev_2	aaacACTCTAATCTCGCCCTGATCc
MYL12B_fwd_1	caccgGCATCAAGGTATGCATCAGT	MYL12B_rev_1	aaacACTGATGCATACCTTGATGCc
MYL12B_fwd_2	caccgCCATGATGAATGAGGCCCCA	MYL12B_rev_2	aaacTGGGGCCTCATTTCATCATGGc
MYL6_fwd_1	caccgCCAGTGTGACAAGAACATGC	MYL6_rev_1	aaacGCATGTTCTTGTCCACTGGc
MYL6_fwd_2	caccgTGTTCCGGTCAAACAGCTGGA	MYL6_rev_2	aaacTCCAGCTGTTTGACCGAACAc
NUP155_fwd_1	caccgGAAGATGCTGAGAATCCAGG	NUP155_rev_1	aaacCCTGGATTCTCAGCATCTTCc
NUP155_fwd_2	caccgTGAATTAGAAGAAAAATGG	NUP155_rev_2	aaacCCATTTTTTCTTCTAATTCAc
RANBP2_fwd_1	caccgGATCCTAATGAGAAGAGACC	RANBP2_rev_1	aaacGGTCTCTTCTCATTAGGATCc
RANBP2_fwd_2	caccgAAACTCACCAGCATAATCTG	RANBP2_rev_2	aaacCAGATTATGCTGGTGAGTTTc
RCOR1_fwd_1	caccgCGTGATTGGGAACAAATCAG	RCOR1_rev_1	aaacCTGATTGTTCCCAATCAGGc
RCOR1_fwd_2	caccgGTCTAGTTGTCTCAGCACCG	RCOR1_rev_2	aaacCGGTGCTGAGACAACTAGACc
RPS19BP1_fwd_1	caccgGACCAGGACGAGAAGCACCG	RPS19BP1_rev_1	aaacCGGTGCTTCTCGTCCTGGTCc
RPS19BP1_fwd_2	caccgAAGTTCAGCAGGAATACTT	RPS19BP1_rev_2	aaacAAGTATTCCTGCTGGAACTTc
RUNX1_fwd_1	caccgGCTCCGTGCTGCCTACGCAC	RUNX1_rev_1	aaacGTGCGTAGGCAGCACGGAGCc

RUNX1_fwd_2	caccgGCCATCTGGAACATCCCCTA	RUNX1_rev_2	aaacTAGGGGATGTTCCAGATGGCc
SKI_fwd_1	caccgTGTGCGGCGGGTACATGAGG	SKI_rev_1	aaacCCTCATGTACCCGCCGCACAc
SKI_fwd_2	caccgTGCAACGCGCTGCTCTACGG	SKI_rev_2	aaacCCGTAGAGCAGCGCGTTGCACc
SMC1A_fwd_1	caccgGCTGGCATAGGTCAATGAGG	SMC1A_rev_1	aaacCCTCATTGACCTATGCCAGCc
SMC1A_fwd_2	caccgGTACAAGATCAACAACAAAG	SMC1A_rev_2	aaacCTTTGTTGTTGATCTTGACc
SMC2_fwd_1	caccgGAAAACTATAGAAAAAAGG	SMC2_rev_1	aaacCCTTTTTTCTATAGTTTTCc
SMC2_fwd_2	caccgTGAAATCACAGTAACAAGGC	SMC2_rev_2	aaacGCCTTGTTACTGTGATTTCAc
TCF12_fwd_1	caccgGAAAGTGCTAGCGAACATAc	TCF12_rev_1	aaacGTATGTTCCGCTAGCACTTTCc
TCF12_fwd_2	caccgAGTCTAGAACAGCAAGTCAG	TCF12_rev_2	aaacCTGACTTGCTGTTCTAGACTc
TUBB6_fwd_1	caccgGACCCGCGCCGAGGCTACGT	TUBB6_rev_1	aaacACGTAGCCTCCGGCCGGGTCc
TUBB6_fwd_2	caccgTGGCTCTAAGTCCACCAGGG	TUBB6_rev_2	aaacCCCTGGTGGACTTAGAGCCAc
ZEB2_fwd_1	caccgGTAGGGGCAGGTCAGCAGTT	ZEB2_rev_1	aaacAACTGCTGACCTGCCCTACc
ZEB2_fwd_2	caccgGCAAACTGTAGCTACAGAG	ZEB2_rev_2	aaacCTCTGTAGCTACACGTTTGc
ACADM_fwd_1	caccgGATGCTGTGCAGATACTTGG	ACADM_rev_1	aaacCCAAGTATCTGCACAGCATCc
ACADM_fwd_2	caccgAAGATGTGGATAACCAACGG	ACADM_rev_2	aaacCCGTTGGTTATCCACATCTTc
ACTN4_fwd_1	caccgGCAGTATGAACGCAGCATCG	ACTN4_rev_1	aaacCGATGCTGCGTTCATACTGCc
ACTN4_fwd_2	caccgGATCCAGGGCCCCACAACAT	ACTN4_rev_2	aaacATGTTGTGGGGCCCTGGATCc
AGPS_fwd_1	caccgGAATATCTATGGCAATATCG	AGPS_rev_1	aaacCGATATTGCCATAGATATTCc
AGPS_fwd_2	caccgGGAGAAGATAATGGACAGAG	AGPS_rev_2	aaacCTCTGTCCATTATCTTCTCCc
ATP1A1_fwd_1	caccgGCAACCAGTTATGATTACAA	ATP1A1_rev_1	aaacTTGTAATCATAACTGGTTGCc
ATP1A1_fwd_2	caccgGGTCTTACAGATGACCAAGT	ATP1A1_rev_2	aaacACTTGGTCATCTGTAAGACCc
BANF1_fwd_1	caccgGAATGGCTGAAAGACACTTG	BANF1_rev_1	aaacCAAGTGTCTTTCAGCCATTc
BANF1_fwd_2	caccgAAGGCATCCGAAGCAGTCCC	BANF1_rev_2	aaacGGGACTGCTTCGGATGCCTTc
C19orf53_fwd_1	caccgGAAGAAGATCGAACATGACG	C19orf53_rev_1	aaacCGTCATGTTTCGATCTTCTTCc
C19orf53_fwd_2	caccgGCAGCCTCTGAAAAGAATCG	C19orf53_rev_2	aaacCGATTCTTTTCAGAGGCTGCc
C1QBP_fwd_1	caccgATAATGACAGTCCAACACAA	C1QBP_rev_1	aaacTTGTGTTGGACTGTCATTATc
C1QBP_fwd_2	caccgGTTGGACTGTCATTATCCAG	C1QBP_rev_2	aaacCTGGATAATGACAGTCCAACc
CALU_fwd_1	caccgTGATGGTTAGAGATGAGCGG	CALU_rev_1	aaacCCGCTCATCTCTAACCATCAc
CALU_fwd_2	caccgGGATGGCAAGCTTACCAAGG	CALU_rev_2	aaacCCTTGGTAAGCTTGCCATCCc
CCDC47_fwd_1	caccgTGTGCCAACAGCAAATACGT	CCDC47_rev_1	aaacACGTATTTGCTGTTGGCACAc
CCDC47_fwd_2	caccgGAAGGAGCGAATCATGAATG	CCDC47_rev_2	aaacCATTGATGATTGCTCCTTTCc
CCDC59_fwd_1	caccgTAGCTGAAGAGGAAAGACAT	CCDC59_rev_1	aaacATGTCTTTCCTCTTCAGCTAc
CCDC59_fwd_2	caccgTTCGTGGGGAGCGTTCGCGA	CCDC59_rev_2	aaacTCGCGAACGCTCCCCACGAAc
CCT4_fwd_1	caccgGACAATTGATGACTGTGAGT	CCT4_rev_1	aaacACTCACAGTCATCAATTGTc
CCT4_fwd_2	caccgGCCAGGGAGAAAAAAGTGC	CCT4_rev_2	aaacGCAGTTTTTCTCCCTGGGCc
CCT5_fwd_1	caccgTGATTGCCATCGCAACAGGA	CCT5_rev_1	aaacTCCTGTTGCGATGGCAATCAc
CCT5_fwd_2	caccgGATAATCGTGTGGTGTATGG	CCT5_rev_2	aaacCCATACACCACACGATTATCc
CCT6A_fwd_1	caccgGTTAAACAGCTTAATCAGA	CCT6A_rev_1	aaacTCTGATTAAGCTGTTAAACc
CCT6A_fwd_2	caccgGACGTGCTAAGGACCAACCT	CCT6A_rev_2	aaacAGGTTGGTCTTAGCACGTCc
CCT7_fwd_1	caccgGGATCTGAAGAGGACAATGA	CCT7_rev_1	aaacTCATTGTCCTCTTCAGATCCc
CCT7_fwd_2	caccgGGCGCCGCCACGGAGAATGA	CCT7_rev_2	aaacTCATTCTCCGTGGCGGGCGCCc
CLPX_fwd_1	caccgTGAGGAGGTGATGAGCCGGA	CLPX_rev_1	aaacTCGGGCTCATCACCTCCTCAc
CLPX_fwd_2	caccgGAATTAGGGACTTCAAACAT	CLPX_rev_2	aaacATGTTTGAAGTCCCTAATTc
CLTC_fwd_1	caccgGTACTCCAGACACTATCCGT	CLTC_rev_1	aaacACGGATAGTGTCTGGAGTACc
CLTC_fwd_2	caccgTCGTCCTAGAACTGCATGG	CLTC_rev_2	aaacCCATGCAGTTTCTAGGACGAc

CMC1_fwd_1	caccgTGAGAGAAAAGGCCAAAGAG	CMC1_rev_1	aaacCTCTTTGGCCTTTTCTCTCAc
CMC1_fwd_2	caccgTTCTTATGGTAGTAAAATGC	CMC1_rev_2	aaacGCATTTTACTACCATAAGAAc
COX5A_fwd_1	caccgGATGCCTGGGAATTGCGTAA	COX5A_rev_1	aaacTTACGCAATCCCAGGCATCc
COX5A_fwd_2	caccgGGGCTCTGGAACCATATCAT	COX5A_rev_2	aaacATGATATGGTTCCAGAGCCCc
DAZAP1_fwd_1	caccgAATCGCAGGAAGCTCTTCGT	DAZAP1_rev_1	aaacACGAAGAGCTTCTGCGATTc
DAZAP1_fwd_2	caccgAGGGGTGGACACGATGTAGG	DAZAP1_rev_2	aaacCCTACATCGTGTCCACCCCTc
DDX17_fwd_1	caccgGAAAGAAGGAGATTACAGTG	DDX17_rev_1	aaacCACTGTAATCTCCTTCTTTc
DDX17_fwd_2	caccgGATTACACCCAGATCAACGT	DDX17_rev_2	aaacACGTTGATCTGGGTGTAATCc
EIF3B_fwd_1	caccgATGGCTCCACCAGGACACAG	EIF3B_rev_1	aaacCTGTGTCCTGGTGGAGCCATc
EIF3B_fwd_2	caccgGAACATCGCAGAGCACTACA	EIF3B_rev_2	aaacTGTAGTGCTCTGCGATGTTc
EIF5A_fwd_1	caccgTGCATACATACAGGTCCATC	EIF5A_rev_1	aaacGATGGACCTGTATGTATGCAc
EIF5A_fwd_2	caccgTCTCGACGATCTTACATGGC	EIF5A_rev_2	aaacGCCATGTAAGATCGTCGAGAc
EPRS_fwd_1	caccgGCTCAATCAGTGGTGCAATG	EPRS_rev_1	aaacCATTGCACCACTGATTGAGCc
EPRS_fwd_2	caccgAAGATCTGGCAAACCGAGC	EPRS_rev_2	aaacGCTCGGTTTTGCCAGATCTTc
ESYT1_fwd_1	caccgGAAGAACTCAACCCACAGTG	ESYT1_rev_1	aaacCACTGTGGGTGAGTTCTTc
ESYT1_fwd_2	caccgGAACAAGGAACCCAACCCTA	ESYT1_rev_2	aaacTAGGGTTGGGTTCTTGTTCc
FASN_fwd_1	caccgGTAGCCACGGAAGGGCATGG	FASN_rev_1	aaacCCATGCCCTTCCGTGGCTACc
FASN_fwd_2	caccgCGAGACCCCGAGACACTCGT	FASN_rev_2	aaacACGAGTGTCTCGGGTCTCGc
FLOT1_fwd_1	caccgGACGTTAGAGGGCCACCAGA	FLOT1_rev_1	aaacTCTGGTGGCCCTTAACGTc
FLOT1_fwd_2	caccgAGACGTTAGAGGGCCACCAG	FLOT1_rev_2	aaacCTGGTGGCCCTTAACGTCTc
FSCN1_fwd_1	caccgGTTCTGCGACTATAACAAGG	FSCN1_rev_1	aaacCCTTGTATAGTCGCAGAAc
FSCN1_fwd_2	caccgGAAGCAGGACAGGCGGTCCT	FSCN1_rev_2	aaacAGGACCGCCTGTCTGCTTc
FUBP1_fwd_1	caccgGGGGAATGCATATCCACAC	FUBP1_rev_1	aaacGTGTGGATATGCATTTCCCCc
FUBP1_fwd_2	caccgATGATGGGACAACACCCGAA	FUBP1_rev_2	aaacTTCGGGTGTTGTCCCATCATc
FUBP3_fwd_1	caccgAGAGTTCTGGGATTCCAGAG	FUBP3_rev_1	aaacCTCTGGAATCCCAGAACTCTc
FUBP3_fwd_2	caccgGATTGTAATAGGAAGAAACG	FUBP3_rev_2	aaacCGTTTCTTCTTATTACAATCc
H2AFZ_fwd_1	caccgACAACACTGACCTGCAAGC	H2AFZ_rev_1	aaacGCTTGCAGGTCAGTAGTTGTc
H2AFZ_fwd_2	caccgACGACCAGTCATGGACGTGT	H2AFZ_rev_2	aaacACACGTCCATGACTGGTGTc
HADHA_fwd_1	caccgGGCCCCAGCAAACCTTCAG	HADHA_rev_1	aaacCTGAAGTTTTGCTGGGGGCCc
HADHA_fwd_2	caccgTGTGGATGTAGCGAAACATG	HADHA_rev_2	aaacCATGTTTCGTACATCCACAc
HADHB_fwd_1	caccgGGAAGTGAAAACAAGCAATG	HADHB_rev_1	aaacCATTGCTTGTTCCTACTTCCc
HADHB_fwd_2	caccgTGAGTGACGAATAGGGACAT	HADHB_rev_2	aaacATGTCCCTATTCTGCTACTCAc
HBA1_fwd_1	caccgGTCGGCAGGAGACAGACCA	HBA1_rev_1	aaacTGGTGCTGTCTCCTGCCGACc
HBA1_fwd_2	caccgGGTTAAGGGCCACGGCAAGA	HBA1_rev_2	aaacTCTTGCCGTGGCCCTTAACc
HBG1_fwd_1	caccgGATGCTGGAGGAGAAACCCT	HBG1_rev_1	aaacAGGGTTTCTCCTCCAGCATCc
HBG1_fwd_2	caccgAAAGCTGTCAAAGAACCTCT	HBG1_rev_2	aaacAGAGGTTCTTTGACAGCTTc
HIST1H3A_fwd_1	caccgTTTACGAATAAGCAGTTCAG	HIST1H3A_rev_1	aaacCTGAACTGCTTATTCGTAAAc
HIST1H3A_fwd_2	caccgGTAATCACGCCCTCTCTCCG	HIST1H3A_rev_2	aaacCGGAGAGAGGGCGTGATTACc
HNRNPDL_fwd_1	caccgGAAAGCAGATACCATCAAAT	HNRNPDL_rev_1	aaacATTTGATGGTATCTGCTTTc
HNRNPDL_fwd_2	caccgTGGAGCTGGATTTAAATGG	HNRNPDL_rev_2	aaacCCATTTTAAATCCAGCTCCAc
HNRNPF_fwd_1	caccgGATAGGGCACAGGTACATTG	HNRNPF_rev_1	aaacCAATGTACCTGTGCCCTATCc
HNRNPF_fwd_2	caccgGATTGGGTGTTGAAGCACAG	HNRNPF_rev_2	aaacCTGTGCTTCAACACCCAATCc
HRNR_fwd_1	caccgGGAGTATGATACGTTGAACA	HRNR_rev_1	aaacTGTTCAACGTATCATACTCCc
HRNR_fwd_2	caccgTGACAGTGATGACGCCTTGT	HRNR_rev_2	aaacACAAGGCGTCATCACTGTCAc

HSD17B8_fwd_1	caccgAGGACACAACGACAGATGGT	HSD17B8_rev_1	aaacACCATCTGTGCTTGTGTCTc
HSD17B8_fwd_2	caccgGTCCTGTGCGGGCATCACCC	HSD17B8_rev_2	aaacGGGTGATGCCCGCACAGGACc
HSP90AA1_fwd_1	caccgGCTTTGTTGCAACATCTCAC	HSP90AA1_rev_1	aaacGTGAGATGTTGCAACAAAGCc
HSP90AA1_fwd_2	caccgGATCTGTCAAGCTTTCATAC	HSP90AA1_rev_2	aaacGTATGAAAGCTTGACAGATCc
HSP90B1_fwd_1	caccgGAAGCTTGTTCGTAACACGC	HSP90B1_rev_1	aaacGCGTTTTACGAACAAGCTTCc
HSP90B1_fwd_2	caccgTGTTGCCAGACCATCCGTAC	HSP90B1_rev_2	aaacGTACGGATGGTCTGGCAACAc
HSPA1L_fwd_1	caccgACTGCTGGGGACACTCACCT	HSPA1L_rev_1	aaacAGGTGAGTGCCCCAGCAGTc
HSPA1L_fwd_2	caccgTGGCACGGTAATCACTGCAT	HSPA1L_rev_2	aaacATGCAGTGATTACCGTGCCAc
ILF2_fwd_1	caccgTGCCCTGGGGAAACAAAGTCG	ILF2_rev_1	aaacCGACTTTGTTCCCCAGGGCAc
ILF2_fwd_2	caccgTGGCCCTAAACGTTGCATAC	ILF2_rev_2	aaacGTATGCAACGTTTAGGGCCAc
ILF3_fwd_1	caccgGAAATGGAGAAAGTATTAGC	ILF3_rev_1	aaacGCTAATACTTTCTCCATTTCc
ILF3_fwd_2	caccgAGGGCCTCGCCAGCACCCAT	ILF3_rev_2	aaacATGGGTGCTGGCGAGGCCCTc
JUP_fwd_1	caccgGTAGTTGATGAGATGCACAA	JUP_rev_1	aaacTTGTGCATCTCATCAACTAc
JUP_fwd_2	caccgGAACAGGACCGACTCCACAG	JUP_rev_2	aaacCTGTGGAGTCGGTCTGTTCc
KPNA2_fwd_1	caccgAGTATTAGCAGATACCTGCT	KPNA2_rev_1	aaacAGCAGGTATCTGCTAATACTc
KPNA2_fwd_2	caccgGCAGGAAAACCGCAACAACC	KPNA2_rev_2	aaacGGTTGTTGCGGTTTTCTGTCc
KPNB1_fwd_1	caccgAGCTCCTAGAGACTACAGAC	KPNB1_rev_1	aaacGTCTGTAGTCTCTAGGAGCTc
KPNB1_fwd_2	caccgTATCTCGAACAACTACTACTG	KPNB1_rev_2	aaacCAGTGTAGTTGTTTCGAGATAc
LDHA_fwd_1	caccgGGGGAACATGGAGATTCCAG	LDHA_rev_1	aaacCTGGAATCTCCATGTTCCCCc
LDHA_fwd_2	caccgTGCCATCAGTATCTTAATGA	LDHA_rev_2	aaacTCATTAAGATACTGATGGCAc
LDHB_fwd_1	caccgGCTAAAAGGATATACCAACT	LDHB_rev_1	aaacAGTTGGTATATCCTTTTAGCc
LDHB_fwd_2	caccgGAAGGAAGTGCATAAGATGG	LDHB_rev_2	aaacCCATCTTATGCACTTCCTTCc
LRPPRC_fwd_1	caccgGAATCCGACATGGTTACTGG	LRPPRC_rev_1	aaacCCAGTAACCATGTCCGATTTCc
LRPPRC_fwd_2	caccgAGCGCGGAGACCCACATCAA	LRPPRC_rev_2	aaacTTGATGTGGGTCTCCGCGCTc
MCM2_fwd_1	caccgGGAGAGTCCAGGCAAAGTGG	MCM2_rev_1	aaacCCACTTTGCTGGACTCTCCc
MCM2_fwd_2	caccgGCTGATCCGCACCAGTGGGG	MCM2_rev_2	aaacCCCCACTGGTGCGGATCAGCc
MCM6_fwd_1	caccgAGAGTCAGAGGACTTCATTG	MCM6_rev_1	aaacCAATGAAGTCTCTGACTCTc
MCM6_fwd_2	caccgGTTGATGGATATGAGACAGA	MCM6_rev_2	aaacTCTGTCTCATATCCATCAACc
MCM7_fwd_1	caccgGAGGAGCTGAGGCAAATTGC	MCM7_rev_1	aaacGCAATTTGCTCAGCTCCTCc
MCM7_fwd_2	caccgGATACAGCCGCCCTCCTGAG	MCM7_rev_2	aaacCTCAGGAGGGCGGCTGTATCc
MPP1_fwd_1	caccgGAAGATACCGATCAGACCA	MPP1_rev_1	aaacTGGTGCTGATCGGTATCTTCc
MPP1_fwd_2	caccgTGAAAAACAGTCTGTACGG	MPP1_rev_2	aaacCCGTACAGGACTGTTTTTCAc
MRPS23_fwd_1	caccgGCAGCCGGCTGGAAACCGTA	MRPS23_rev_1	aaacTACGGTTTCCAGCCGGCTGCc
MRPS23_fwd_2	caccgTGCTCCCTACGGTTTCCAGC	MRPS23_rev_2	aaacGCTGGAACCGTAGGGAGCAc
MRPS28_fwd_1	caccgGCTGTGTCGGACCCGTGCTG	MRPS28_rev_1	aaacCAGCACGGGTCCGACACAGCc
MRPS28_fwd_2	caccgGGAATACCAGAAAGGAACC	MRPS28_rev_2	aaacGGTTCCTTTCTGGTATTTCc
NASP_fwd_1	caccgGTGTTGGGAAACGCCTTGGA	NASP_rev_1	aaacTCCAAGGCGTTTCCAACACc
NASP_fwd_2	caccgAGAATGGAGAATGGTGTGTT	NASP_rev_2	aaacAACACACCATTCTCCATTCTc
NUP93_fwd_1	caccgTGTGAACAAGATGTTTCATGG	NUP93_rev_1	aaacCCATGAACATCTTGTTCACAc
NUP93_fwd_2	caccgAGCAAAGGTAATCAGAGTGC	NUP93_rev_2	aaacGCACTCTGATTACCTTTGCTc
P4HB_fwd_1	caccgGCTGCGGAAAAGCAACTTCG	P4HB_rev_1	aaacCGAAGTTGCTTTTCCGAGCc
P4HB_fwd_2	caccgAGAAGGTTCCGAGATCAGGT	P4HB_rev_2	aaacACCTGATCTCGGAACCTTCTc
PAM16_fwd_1	caccgAAGGTGAATGATAAATCCGT	PAM16_rev_1	aaacACGGATTTATCATTACCTTCc
PAM16_fwd_2	caccgGGATGAGGAACTCAAATCC	PAM16_rev_2	aaacGGATTTGAGTTCCTCATCCc
PARK7_fwd_1	caccgGATGTGGTGGTTCTACCAGG	PARK7_rev_1	aaacCCTGGTAGAACACCACATCc

PARK7_fwd_2	caccgTGCAAGCGCAAACCTCGAAGC	PARK7_rev_2	aaacGCTTCGAGTTTGCCTTGCAc
PCBP1_fwd_1	caccgAGAGATCCGCGAGAGTACGG	PCBP1_rev_1	aaacCCGTACTCTCGCGGATCTCTc
PCBP1_fwd_2	caccgAATCAGGGAGCCGCACTGGG	PCBP1_rev_2	aaacCCCAGTGC GGCTCCCTGATTc
PGAM1_fwd_1	caccgTCTTTATTGCTTAGGATCGC	PGAM1_rev_1	aaacGCGATCCTAAGCAATAAAGAc
PGAM1_fwd_2	caccgGGGCATTGTCAAGCATCTGG	PGAM1_rev_2	aaacCCAGATGCTTGACAATGCCCc
PGK1_fwd_1	caccgTAAGGTGCTCAACAACATGG	PGK1_rev_1	aaacCCATGTTGTTGAGCACCTTAc
PGK1_fwd_2	caccgCTGTGCCAAATGGAACACGG	PGK1_rev_2	aaacCCGTGTTCCATTTGGCACAGc
PHGDH_fwd_1	caccgGAAGGGGAAATCTCTCACGG	PHGDH_rev_1	aaacCCGTGAGAGATTTCCCCTTCc
PHGDH_fwd_2	caccgAGTGCCGCGAAGCACTACTTG	PHGDH_rev_2	aaacCAAGTGAGTTCTGCGGCACTc
PLRG1_fwd_1	caccgGCACCCGACAATCGATGTGT	PLRG1_rev_1	aaacACACATCGATTGTCGGGTGCc
PLRG1_fwd_2	caccgCTTCAACGACCTGAACACAA	PLRG1_rev_2	aaacTTGTGTTTCAGGTCGTTGAAGc
PNN_fwd_1	caccgGACCAGAAGAGAATCACGCC	PNN_rev_1	aaacGGCGTGATTCTCTTCTGGTCc
PNN_fwd_2	caccgTCAACAAGCCAAATATTCGC	PNN_rev_2	aaacGCGAATATTTGGCTTGTGAc
PRKCSH_fwd_1	caccgCGGTTGGAGGGGATATACAG	PRKCSH_rev_1	aaacCTGTATATCCCCTCCAACCGc
PRKCSH_fwd_2	caccgGGAACAGACGAGTACAACAG	PRKCSH_rev_2	aaacCTGTTGTA CTCTGTCTCCc
PRPF31_fwd_1	caccgGAAGCGGACATACCCCTGGG	PRPF31_rev_1	aaacCCCAGGGGTATGTCGGCTTCc
PRPF31_fwd_2	caccgGACGCTGACGACCATGATGG	PRPF31_rev_2	aaacCCATCATGGTCGTCAGCGTCc
RANBP1_fwd_1	caccgATGGAGCTGAAGCCCAACGC	RANBP1_rev_1	aaacGCGTTGGGCTTCAGCTCCATc
RANBP1_fwd_2	caccgTTCTGGGAGATCGTTCTCAG	RANBP1_rev_2	aaacCTGAGAACGATCTCCCAGAAc
RANGAP1_fwd_1	caccgTGGAGAGGGAAAAGCCAGGA	RANGAP1_rev_1	aaacTCTGGCTTTTCCCTCTCCAc
RANGAP1_fwd_2	caccgGCGTCTGGAAGGCAACACAG	RANGAP1_rev_2	aaacCTGTGTTGCCCTTCCAGACGCc
RAP1B_fwd_1	caccgAGGCGTTGGAAGTCTGCTT	RAP1B_rev_1	aaacAAGCAGACTTTCCAACGCCTc
RAP1B_fwd_2	caccgTAGGTTCCAATGATTCTTGT	RAP1B_rev_2	aaacACAAGAATCATTGGAACCTAc
RAVER1_fwd_1	caccgAGAGCCTCGGAGCCCCACAG	RAVER1_rev_1	aaacCTGTGGGGCTCCGAGGCTCTc
RAVER1_fwd_2	caccgGTGGCAGGCGGTCCACACAG	RAVER1_rev_2	aaacCTGTGTGGACCGCTGCCACc
RBM27_fwd_1	caccgGCTCCTTCAACAGTGCACGG	RBM27_rev_1	aaacCCGTGCACTGTTGAAGGAGCc
RBM27_fwd_2	caccgGACCAGTGCTACAACATAGT	RBM27_rev_2	aaacACTATGTTGTAGCACTGGTCc
RBM33_fwd_1	caccgGATCAGGGCCACATTTATGT	RBM33_rev_1	aaacACATAAATGTGGCCCTGATCc
RBM33_fwd_2	caccgGATGTCAGTGGGACCCATTC	RBM33_rev_2	aaacGAATGGGTCCCACTGACATCc
RPL4_fwd_1	caccgCGGAGTACACCGATATCAGT	RPL4_rev_1	aaacACTGATATCGGTGTA CTCCGc
RPL4_fwd_2	caccgGAAAAATAGATGTGTCGTGG	RPL4_rev_2	aaacCCACGACACATCTATTTTTTc
RPLP0_fwd_1	caccgAGACTGCTGCCTCATATCCG	RPLP0_rev_1	aaacCGGATATGAGGCAGCAGTCTc
RPLP0_fwd_2	caccgTGATGCGCAAGGCCATCCGA	RPLP0_rev_2	aaacTCGGATGGCCTTGGC CATCAc
RPN1_fwd_1	caccgTG TAGGCAACAATCACAGGG	RPN1_rev_1	aaacCCCTGTGATTGTTGCCTACAc
RPN1_fwd_2	caccgGTCATTGAAGTCTCTCACTG	RPN1_rev_2	aaacCAGTGAGAGACTTCAATGACc
RPN2_fwd_1	caccgGTATATTGCAAATACCGTAG	RPN2_rev_1	aaacCTACGGTATTTGCAATATACc
RPN2_fwd_2	caccgTGCTTGGTGAGGTAGTGAGT	RPN2_rev_2	aaacACTCACTACCTACCAAGCAc
RPS12_fwd_1	caccgCCTTTG TAAAATTGACAGAG	RPS12_rev_1	aaacCTCTGTCAATTTTACAAAGGc
RPS12_fwd_2	caccgCTGATACACGTA CTGTGCTA	RPS12_rev_2	aaacTAGACAAGTACGTGTATCAGc
RPS21_fwd_1	caccgGTACAGGTCCACGAACTCGC	RPS21_rev_1	aaacGCGAGTTCTGTTGACCTGTACc
RPS21_fwd_2	caccgGGGGGCGCTTACCATTTCCG	RPS21_rev_2	aaacCGGAAATGGTAAAGCGCCCCc
RPS5_fwd_1	caccgGTTAGTGAGGCGCTCCACAA	RPS5_rev_1	aaacTTGTGGAGCGCTCACTAACc
RPS5_fwd_2	caccgGCTGTGCACAGGCGCTCGTG	RPS5_rev_2	aaacCACGAGCGCCTGTGCACAGCc
RUVBL2_fwd_1	caccgGATGCGGGTCAGCACCGTGT	RUVBL2_rev_1	aaacACACGGTGCTGACCCGCATCc
RUVBL2_fwd_2	caccgGGAGCTCCAGAAACGCAAGG	RUVBL2_rev_2	aaacCCTTGC GTTTCTGGAGCTCCc

SAFB2_fwd_1	caccgGTCCATGCAGCTCAGTTCTG	SAFB2_rev_1	aaacCAGAAGCTGAGCTGCATGGACc
SAFB2_fwd_2	caccgGTAATTAGGACTGAAGATGG	SAFB2_rev_2	aaacCCATCTTCAGTCCTAATTACc
SDHA_fwd_1	caccgAGAGCGAGTTTGCCCCGAGG	SDHA_rev_1	aaacCCTCGGGGCAAACCTCGCTCTc
SDHA_fwd_2	caccgGCCCATGTTATAATGCACGG	SDHA_rev_2	aaacCCGTGCATTATAACATGGGCc
SERBP1_fwd_1	caccgCGACGAATCGGACCCCTTCG	SERBP1_rev_1	aaacCGAAGGGGTCCGATTCTGTCGc
SERBP1_fwd_2	caccgAATATCCGAAAACCAATGA	SERBP1_rev_2	aaacTCATTTGGTTTTTCGGATATTc
SFXN1_fwd_1	caccgGGAACAGGGCACAACACAG	SFXN1_rev_1	aaacCTGTGTTGTGCCCTGTTCCc
SFXN1_fwd_2	caccgTGAGAATGGGAACCGCTTGG	SFXN1_rev_2	aaacCCAAGCGGTTCCATTCTCAc
SHMT2_fwd_1	caccgGCTAGAGCGAGGCTACTCAC	SHMT2_rev_1	aaacGTGAGTAGCCTCGCTCTAGCc
SHMT2_fwd_2	caccgGAAAGGCGAGGGAATCACCT	SHMT2_rev_2	aaacAGGTGATTCCCTCGCCTTTCCc
SLC3A2_fwd_1	caccgAGAACCACGAGTTCTCACCC	SLC3A2_rev_1	aaacGGGTGAGAAGCTCGTGGTTCTc
SLC3A2_fwd_2	caccgGAACTCGTGGTTCTCCACTC	SLC3A2_rev_2	aaacGAGTGGAGAACCACGAGTTCCc
SND1_fwd_1	caccgGCAGCGGGGCATCATCAAGA	SND1_rev_1	aaacTCTTGATGATGCCCCGCTGCCc
SND1_fwd_2	caccgGAAGGCGAAGGCATACTCCG	SND1_rev_2	aaacCGGAGTATGCCTTCGCCTTCc
SSR4_fwd_1	caccgGACATGCAAGAACAGGGTCC	SSR4_rev_1	aaacGGACCCTGTTCTTGCATGTCc
SSR4_fwd_2	caccgTAGTAGGAAGGGGTGATCTG	SSR4_rev_2	aaacCAGATCACCCCTTCCACTAc
STOML2_fwd_1	caccgGCTGGGTGACGGCATACTCA	STOML2_rev_1	aaacTGAGTATGCCGTACCCAGCc
STOML2_fwd_2	caccgTGTTTGAGCTAGCTGGGTGA	STOML2_rev_2	aaacTCACCCAGCTAGCTCAAACAc
TKT_fwd_1	caccgCGTGGATGGACACAGCGTGG	TKT_rev_1	aaacCCACGCTGTGTCCATCCACGc
TKT_fwd_2	caccgGTACCCGGTTAACTGCCAGG	TKT_rev_2	aaacCCTGGCAGTTAACCGGGTACc
TMED9_fwd_1	caccgGAAAGAGCAGAAGTACCAGC	TMED9_rev_1	aaacGCTGGTAGTTCTGCTCTTTCCc
TMED9_fwd_2	caccgAGCGCGCTCTACTTTCACAT	TMED9_rev_2	aaacATGTGAAAGTAGAGCGCGCTc
TPR_fwd_1	caccgGGAGTGTAAGCATCTTGGG	TPR_rev_1	aaacCCCAAGATGCTTTACACTCCc
TPR_fwd_2	caccgAAGTTAAGAAAATTGGACGT	TPR_rev_2	aaacACGTCCAATTTTCTTAACCTc
TPX2_fwd_1	caccgGCAAACCAAACACCGTGAC	TPX2_rev_1	aaacGTGCACGGTGTGGTTGGTTGCc
TPX2_fwd_2	caccgGAAGTTTCTAAAAAGTACTG	TPX2_rev_2	aaacCAGTACTTTTTAGAAACTTCc
TUBA1B_fwd_1	caccgTTCAGAGGCCCGTGAAGATA	TUBA1B_rev_1	aaacTATCTTCACGGGCCTCTGAAc
TUBA1B_fwd_2	caccgCTACAGGCTGACCAGTGAC	TUBA1B_rev_2	aaacGTGCACTGGTCAGCCTGTAGc
UQCRC2_fwd_1	caccgGTACTTACACATCACCCCGC	UQCRC2_rev_1	aaacGCGGGGTGATGTGTAAGTACc
UQCRC2_fwd_2	caccgGATTGGAAAAGTGACATCAG	UQCRC2_rev_2	aaacCTGATGTCACTTTTCCAATCc
UQCRH_fwd_1	caccgTCTTTGACTTCTTGCATGCG	UQCRH_rev_1	aaacCGCATGCAAGAAGTCAAAGAc
UQCRH_fwd_2	caccgACAGAGCTCTAGCCGCTCCC	UQCRH_rev_2	aaacGGGAGCGGCTAGAGCTCTGTc
VARS_fwd_1	caccgTGCGTGGGGAGACCACCTG	VARS_rev_1	aaacCAGGGTGGTCTCCCCACGCAc
VARS_fwd_2	caccgTGCAAAGTGGCACCACCATG	VARS_rev_2	aaacCATGGTGGTGCCACTTTGCAc
XRCC5_fwd_1	caccgGTTGTGCTGTGTATGGACGT	XRCC5_rev_1	aaacACGTCCATACACAGCACAACc
XRCC5_fwd_2	caccgGAACTATATCACTTCCATAG	XRCC5_rev_2	aaacCTATGGAAGTGATATAGTTCCc
YWHAQ_fwd_1	caccgGCTAGAGATGACCCTCCAGG	YWHAQ_rev_1	aaacCCTGGAGGGTCACTCTAGCc
YWHAQ_fwd_2	caccgGCTGAGATCCATCTGCACCA	YWHAQ_rev_2	aaacTGGTGCAGATGGATCTCAGCc
ZC3H13_fwd_1	caccgTGTCGGGAGAATGTTACGC	ZC3H13_rev_1	aaacGCGTGAACATTCTCCGGACAc
ZC3H13_fwd_2	caccgCGAAAGAGAGTACTGCACAG	ZC3H13_rev_2	aaacCTGTGCAGTACTCTCTTTCCGc
ZNF326_fwd_1	caccgGGCATGGACAATCACAGTGG	ZNF326_rev_1	aaacCCACTGTGATTGTCCATGCCc
ZNF326_fwd_2	caccgTATGGCATGGACAATCACAG	ZNF326_rev_2	aaacCTGTGATTGTCCATGCCATAc

Suppl. Table 21. 24-well plate HEK 293T-LX cell transfection mix for 1 reaction

construct	700 ng
pMD2.G	175 ng
pMDLg/pPRE	280 ng
pRSV-Rev	140 ng
JetPRIME buffer	50 μ l
JetPRIME reagent	4 μ l

7.4. References

1. Hoggatt J, Pelus LM. Hematopoiesis. In: Maloy Stanley, Hughes Kelly, editors. *Brenner's Encyclopedia of Genetics (Second Edition)*. Second Edi. San Diego: Academic Press; 2013. p. 418–21.
2. Laurenti Elisa, Göttgens Berthold. From haematopoietic stem cells to complex differentiation landscapes. *Nature*. 2018;553(7689):418–26.
3. Haas Simon, Trumpp Andreas, Milsom Michael D. Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell*. 2018;22(5):627–38.
4. Hoogenkamp Maarten, Lichtinger Monika, Krysinska Hanna, Lancrin Christophe, Clarke Deborah, Williamson Andrew, Mazzarella Luca, Ingram Richard, Jorgensen Helle, Fisher Amanda, Tenen Daniel G, Kouskoff Valerie, Lacaud Georges, Bonifer Constanze. Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. *Blood*. 2009 Jul;114(2):299–309.
5. Mohrin Mary, Bourke Emer, Alexander David, Warr Matthew R, Barry-Holson Keegan, Le Beau Michelle M, Morrison Ciaran G, Passegué Emmanuelle. Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell*. 2010 Aug;7(2):174–85.
6. Cabezas-Wallscheid Nina, Klimmeck Daniel, Hansson Jenny, Lipka Daniel B, Reyes Alejandro, Wang Qi, et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell*. 2014 Oct;15(4):507–22.
7. Corces M Ryan, Buenrostro Jason D, Wu Beijing, Greenside Peyton G, Chan Steven M, Koenig Julie L, Snyder Michael P, Pritchard Jonathan K, Kundaje Anshul, Greenleaf William J, Majeti Ravindra, Chang Howard Y. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 2016 Oct;48(10):1193–203.
8. Ho Theodore T, Warr Matthew R, Adelman Emmalee R, Lansinger Olivia M, Flach Johanna, Verovskaya Evgenia V, Figueroa Maria E, Passegué Emmanuelle. Autophagy maintains the

- metabolism and function of young and old stem cells. *Nature*. 2017;543(7644):205–10.
9. Zhang Yifan, Gao Shuai, Xia Jun, Liu Feng. Hematopoietic Hierarchy – An Updated Roadmap. *Trends Cell Biol*. 2018;28(12):976–86.
 10. Yokota Takafumi. “Hierarchy” and “Holacracy”; A Paradigm of the Hematopoietic System. *Cells*. 2019 Sep;8(10).
 11. Rieger Michael A, Schroeder Timm. Hematopoiesis. *Cold Spring Harb Perspect Biol*. 2012 Dec;4(12).
 12. Medvinsky Alexander, Rybtsov Stanislav, Taoudi Samir. Embryonic origin of the adult hematopoietic system: advances and questions. *Development*. 2011 Mar 15;138(6):1017–31.
 13. Dzierzak Elaine, Philipsen Sjaak. Erythropoiesis: development and differentiation. *Cold Spring Harb Perspect Med*. 2013 Apr;3(4):a011601.
 14. Kay MM. Mechanism of removal of senescent cells by human macrophages in situ. *Proc Natl Acad Sci U S A*. 1975 Sep;72(9):3521–5.
 15. Palis James. Primitive and definitive erythropoiesis in mammals. *Front Physiol*. 2014;5:3.
 16. Granick S, Levere RD. Heme synthesis in erythroid cells. *Prog Hematol*. 1964;4:1–47.
 17. Bauer Daniel E, Orkin Stuart H. Update on fetal hemoglobin gene regulation in hemoglobinopathies. *Curr Opin Pediatr*. 2011 Feb;23(1):1–8.
 18. Risoluti Roberta, Colah Roshan, Materazzi Stefano. Editorial: Frontiers in Hemoglobinopathies: New Insights and Methods. Vol. 8, *Frontiers in Molecular Biosciences*. 2021.
 19. An Xiuli, Schulz Vincent P, Mohandas Narla, Gallagher Patrick G. Human and murine erythropoiesis. *Curr Opin Hematol*. 2015 May;22(3):206–11.
 20. Sankaran Vijay G, Xu Jian, Ragozy Tobias, Ippolito Gregory C, Walkley Carl R, Maika Shanna D, Fujiwara Yuko, Ito Masafumi, Groudine Mark, Bender MA, Tucker Philip W, Orkin Stuart H. Developmental and species-divergent globin switching are driven by BCL11A. *Nature*. 2009;460(7259):1093–7.
 21. Tirelli Valentina, Ghinassi Barbara, Migliaccio Anna Rita, Whitsett Carolyn, Masiello Francesca, Sanchez Massimo, Migliaccio Giovanni. Phenotypic Definition of the Progenitor Cells with Erythroid Differentiation Potential Present in Human Adult Blood. Sadelain Michel, editor. *Stem Cells Int*. 2011;2011:602483.
 22. Chen Ling, Gao Zhigang, Zhu Jianqiong, Rodgers Griffin P. Identification of CD13+CD36+ cells as a common progenitor for erythroid and myeloid lineages in human bone marrow. *Exp Hematol*. 2007;35(7):1047–55.
 23. Gregory CJ, Eaves AC. Three stages of erythropoietic progenitor cell differentiation distinguished by a number of physical and biologic properties. *Blood*. 1978 Mar;51(3):527–37.
 24. Varga Eszter, Hansen Marten, van den Akker Emile, von Lindern Marieke. Erythropoiesis and

- Megakaryopoiesis in a Dish. In: Mehanna Radwa Ali, editor. Cell Culture. Rijeka: IntechOpen; 2018.
25. Cantú Ileana, Philipsen Sjaak. Flicking the switch: adult hemoglobin expression in erythroid cells derived from cord blood and human induced pluripotent stem cells. Vol. 99, Haematologica. 2014. p. 1647–9.
 26. Wells Maeve, Steiner Laurie. Epigenetic and Transcriptional Control of Erythropoiesis. Front Genet. 2022;13:805265.
 27. Hattangadi Shilpa M, Wong Piu, Zhang Lingbo, Flygare Johan, Lodish Harvey F. From stem cell to red cell : regulation of erythropoiesis at multiple levels by multiple proteins , RNAs , and chromatin modifications. Blood. 2017;118(24):6258–69.
 28. Grebien Florian, Kerenyi Marc A, Kovacic Boris, Kolbe Thomas, Becker Verena, Dolznig Helmut, Pfeffer Klaus, Klingmüller Ursula, Müller Mathias, Beug Hartmut, Müllner Ernst W, Moriggl Richard. Stat5 activation enables erythropoiesis in the absence of EpoR and Jak2. Blood. 2008 May;111(9):4511–22.
 29. Socolovsky M, Nam H, Fleming MD, Haase VH, Brugnara C, Lodish HF. Ineffective erythropoiesis in Stat5a(-/-)5b(-/-) mice due to decreased survival of early erythroblasts. Blood. 2001 Dec;98(12):3261–73.
 30. Okano Masaki, Xie Shaoping, Li En. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998;19(3):219–20.
 31. Newell-Price John, Clark Adrian JL, King Peter. DNA Methylation and Silencing of Gene Expression. Trends Endocrinol Metab. 2000;11(4):142–8.
 32. Hendrich Brian, Bird Adrian. Identification and characterization of a family of mammalian methyl CpG-binding proteins. Genet Res. 1998/08/01. 1998;72(1):59–72.
 33. Yan Hongxia, Wang Yaomei, Qu Xiaoli, Li Jie, Hale John, Huang Yumin, An Chao, Papoin Julien, Guo Xinhua, Chen Lixiang, Kang Qiaozhen, Li Wei, Schulz Vincent P, Gallagher Patrick G, Hillyer Christopher D, Mohandas Narla, An Xiuli. Distinct roles for TET family proteins in regulating human erythropoiesis. Blood. 2017 Apr;129(14):2002–12.
 34. Sterner David E, Berger Shelley L. Acetylation of Histones and Transcription-Related Factors. Microbiol Mol Biol Rev. 2000;64(2):435–59.
 35. Kouzarides Tony. Chromatin Modifications and Their Function. Cell. 2007;128(4):693–705.
 36. Ji Peng, Yeh Victor, Ramirez Tzutzuy, Murata-Hori Maki, Lodish Harvey F. Histone deacetylase 2 is required for chromatin condensation and subsequent enucleation of cultured mouse fetal erythroblasts. Haematologica. 2010 Dec;95(12):2013–21.
 37. Yang Xiang Jiao, Seto Edward. The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. Nat Rev Mol Cell Biol. 2008 Mar;9(3):206–18.

38. Bannister Andrew J, Kouzarides Tony. Regulation of chromatin by histone modifications. *Cell Res.* 2011;21(3):381–95.
39. Bannister Andrew J, Kouzarides Tony. Reversing histone methylation. *Nature.* 2005;436(7054):1103–6.
40. Malcom Carrie A, Ratri Anamika, Piasecka-Srader Joanna, Borosha Shaon, Chakravarthi V Praveen, Alvarez Nehemiah S, Vivian Jay L, Fields Timothy A, Karim Rumi MA, Fields Patrick E. Primitive Erythropoiesis in the Mouse is Independent of DOT1L Methyltransferase Activity. *Front cell Dev Biol.* 2021;9:813503.
41. Breton Amandine, Theodorou Andria, Aktuna Suleyman, Sonzogni Laura, Darling David, Chan Lucas, Menzel Stephan, van der Spek Peter J, Swagemakers Sigrid MA, Grosveld Frank, Philipsen Sjaak, Thein Swee Lay. ASH1L (a histone methyltransferase protein) is a novel candidate globin gene regulator revealed by genetic study of an English family with beta-thalassaemia unlinked to the beta-globin locus. *Br J Haematol.* 2016 Nov 1;175(3):525–30.
42. Leonards Katharina, Almosailleakh Marwa, Tauchmann Samantha, Bagger Frederik Otzen, Thirant Cécile, Juge Sabine, Bock Thomas, Méreau Hélène, Bezerra Matheus F, Tzankov Alexandar, Ivanek Robert, Losson Régine, Peters Antoine HFM, Mercher Thomas, Schwaller Juerg. Nuclear interacting SET domain protein 1 inactivation impairs GATA1-regulated erythroid differentiation and causes erythroleukemia. *Nat Commun.* 2020;11(1):2807.
43. Tauchmann Samantha, Schwaller Juerg. NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer. *Life (Basel, Switzerland).* 2021 Aug;11(9).
44. Malik Jeffrey, Lillis Jacquelyn A, Couch Tyler, Getman Michael, Steiner Laurie A. The Methyltransferase Setd8 Is Essential for Erythroblast Survival and Maturation. *Cell Rep.* 2017 Nov;21(9):2376–83.
45. Ross Julie, Mavoungou Lionel, Bresnick Emery H, Milot Eric. GATA-1 Utilizes Ikaros and Polycomb Repressive Complex 2 To Suppress *Hes1* and To Promote Erythropoiesis. *Mol Cell Biol.* 2012 Sep 15;32(18):3624 LP – 3638.
46. Shi Yujiang, Lan Fei, Matson Caitlin, Mulligan Peter, Whetstine Johnathan R, Cole Philip A, Casero Robert A, Shi Yang. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell.* 2004;119(7):941–53.
47. Burda P, Laslo P, Stopka T. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia.* 2010;24(7):1249–57.
48. Zhang P, Zhang X, Iwama A, Yu C, Smith KA, Mueller BU, Narravula S, Torbett BE, Orkin SH, Tenen DG. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood.* 2000 Oct;96(8):2641–8.
49. Zhang P, Behre G, Pan J, Iwama A, Wara-Aswapati N, Radomska HS, Auron PE, Tenen DG,

- Sun Z. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc Natl Acad Sci U S A*. 1999 Jul;96(15):8705–10.
50. Gutiérrez Laura, Caballero Noemí, Fernández-Calleja Luis, Karkoulia Elena, Strouboulis John. Regulation of GATA1 levels in erythropoiesis. *IUBMB Life*. 2020 Jan;72(1):89–105.
51. Tremblay Mathieu, Sanchez-Ferras Oraly, Bouchard Maxime. GATA transcription factors in development and disease. *Development*. 2018 Oct;145(20).
52. Lowry JA, Atchley WR. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *J Mol Evol*. 2000 Feb;50(2):103–15.
53. Ferreira Rita, Ohneda Kinuko, Yamamoto Masayuki, Philipsen Sjaak. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol Cell Biol*. 2005 Feb;25(4):1215–27.
54. Lentjes Marjolein HFM, Niessen Hanneke EC, Akiyama Yoshimitsu, de Bruïne Adriaan P, Melotte Veerle, van Engeland Manon. The emerging role of GATA transcription factors in development and disease. *Expert Rev Mol Med*. 2016 Mar;18:e3.
55. Viger Robert S, Guittot Séverine Mazaud, Anttonen Mikko, Wilson David B, Heikinheimo Markku. Role of the GATA family of transcription factors in endocrine development, function, and disease. *Mol Endocrinol*. 2008 Apr;22(4):781–98.
56. Mackay Joel P, Kowalski Kasper, Fox Archa H, Czolij Robert, King Glenn F, Crossley Merlin. Involvement of the N-finger in the Self-association of GATA-1. *J Biol Chem*. 1998;273(46):30560–7.
57. Shimizu Satoshi, Sunagawa Yoichi, Hajika Naruto, Yorimitsu Natsumi, Katanasaka Yasufumi, Funamoto Masafumi, Miyazaki Yusuke, Sari Nurmila, Shimizu Kana, Hasegawa Koji, Morimoto Tatsuya. Multimerization of the GATA4 transcription factor regulates transcriptional activity and cardiomyocyte hypertrophic response. *Int J Biol Sci*. 2022;18(3):1079–95.
58. Wall L, deBoer E, Grosveld F. The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev*. 1988 Sep;2(9):1089–100.
59. Nei Yuichiro, Obata-Ninomiya Kazushige, Tsutsui Hidemitsu, Ishiwata Kenji, Miyasaka Masayuki, Matsumoto Kenji, Nakae Susumu, Kanuka Hiroataka, Inase Naohiko, Karasuyama Hajime. GATA-1 regulates the generation and function of basophils. *Proc Natl Acad Sci U S A*. 2013 Nov;110(46):18620–5.
60. Ito E, Toki T, Ishihara H, Ohtani H, Gu L, Yokoyama M, Engel JD, Yamamoto M. Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature*. 1993 Apr;362(6419):466–8.
61. Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A*. 1996 Oct;93(22):12355–8.

62. Zon LI, Tsai SF, Burgess S, Matsudaira P, Bruns GA, Orkin SH. The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. *Proc Natl Acad Sci U S A*. 1990 Jan;87(2):668–72.
63. Martin DI, Orkin SH. Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev*. 1990 Nov;4(11):1886–98.
64. Tsai Fong Ying, Keller Gordon, Kuo Frank C, Weiss Mitchell, Chen Jianzhou, Rosenblatt Margery, Alt Frederick W, Orkin Stuart H. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*. 1994;371(6494):221–6.
65. de Pater Emma, Kaimakis Polynikis, Vink Chris S, Yokomizo Tomomasa, Yamada-Inagawa Tomoko, van der Linden Reinier, Kartalaei Parham Solaimani, Camper Sally A, Speck Nancy, Dzierzak Elaine. Gata2 is required for HSC generation and survival. *J Exp Med*. 2013 Dec;210(13):2843–50.
66. Rekhtman N, Radparvar F, Evans T, Skoultschi AI. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev*. 1999 Jun;13(11):1398–411.
67. Hoppe Philipp S, Schwarzfischer Michael, Loeffler Dirk, Kokkaliaris Konstantinos D, Hilsenbeck Oliver, Moritz Nadine, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*. 2016;535(7611):299–302.
68. Strasser Michael K, Hoppe Philipp S, Loeffler Dirk, Kokkaliaris Konstantinos D, Schroeder Timm, Theis Fabian J, Marr Carsten. Lineage marker synchrony in hematopoietic genealogies refutes the PU.1/GATA1 toggle switch paradigm. *Nat Commun*. 2018 Jul;9(1):2697.
69. Gillespie Mark A, Palii Carmen G, Sanchez-Taltavull Daniel, Shannon Paul, Longabaugh William JR, Downes Damien J, Sivaraman Karthi, Espinoza Herbert M, Hughes Jim R, Price Nathan D, Perkins Theodore J, Ranish Jeffrey A, Brand Marjorie. Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. *Mol Cell*. 2020 Jun;78(5):960-974.e11.
70. Rodriguez Patrick, Bonte Edgar, Krijgsveld Jeroen, Kolodziej Katarzyna E, Guyot Boris, Heck Albert JR, Vyas Paresh, de Boer Ernie, Grosveld Frank, Strouboulis John. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J*. 2005 Jul;24(13):2354–66.
71. Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, Crossley M, Orkin SH. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*. 1997 Jul;90(1):109–19.
72. Chlon Timothy M, Crispino John D. Combinatorial regulation of tissue specification by GATA and FOG factors. *Development*. 2012 Nov;139(21):3905–16.

73. Miccio Annarita, Blobel Gerd A. Role of the GATA-1/FOG-1/NuRD pathway in the expression of human beta-like globin genes. *Mol Cell Biol.* 2010 Jul;30(14):3460–70.
74. Feng Q, Zhang Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev.* 2001 Apr;15(7):827–32.
75. Yang Tao, Jian Wei, Luo Yi, Fu Xueqi, Noguchi Constance, Bungert Jörg, Huang Suming, Qiu Yi. Acetylation of Histone Deacetylase 1 Regulates NuRD Corepressor Complex Activity*. *J Biol Chem.* 2012;287(48):40279–91.
76. Kim BowenAU - Huang, SumingAU - Qiu, YiTI - Regulating the Regulators: The Role of Histone Deacetylase 1 (HDAC1) in Erythropoiesis Min YAU Yan. *Regulating the Regulators: The Role of Histone Deacetylase 1 (HDAC1) in Erythropoiesis.* Vol. 21, International Journal of Molecular Sciences. 2020.
77. Hamlett Isla, Draper Julia, Strouboulis John, Iborra Francisco, Porcher Catherine, Vyas Pareshe. Characterization of megakaryocyte GATA1-interacting proteins: the corepressor ETO2 and GATA1 interact to regulate terminal megakaryocyte maturation. *Blood.* 2008 Oct;112(7):2738–49.
78. Love Paul E, Warzecha Claude, Li LiQi. Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends Genet.* 2014 Jan;30(1):1–9.
79. Wadman IA, Osada H, Grütz GG, Agulnick AD, Westphal H, Forster A, Rabbitts TH. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* 1997 Jun;16(11):3145–57.
80. Soler Eric, Andrieu-Soler Charlotte, de Boer Ernie, Bryne Jan Christian, Thongjuea Supat, Stadhouders Ralph, Palstra Robert Jan, Stevens Mary, Kockx Christel, van Ijcken Wilfred, Hou Jun, Steinhoff Christine, Rijkers Erikjan, Lenhard Boris, Grosveld Frank. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev.* 2010 Feb;24(3):277–89.
81. Goardon Nicolas, Lambert Julie A, Rodriguez Patrick, Nissaire Philippe, Herblot Sabine, Thibault Pierre, Dumenil Dominique, Strouboulis John, Romeo Paul Henri, Hoang Trang. ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *EMBO J.* 2006 Jan;25(2):357–66.
82. Saleque Shireen, Kim Jonghwan, Rooke Heather M, Orkin Stuart H. Epigenetic Regulation of Hematopoietic Differentiation by Gfi-1 and Gfi-1b Is Mediated by the Cofactors CoREST and LSD1. *Mol Cell.* 2007;27(4):562–72.
83. Yao Huilan, Goldman Devorah C, Fan Guang, Mandel Gail, Fleming William H. The Corepressor Rcor1 Is Essential for Normal Myeloerythroid Lineage Differentiation. *Stem Cells.* 2015 Nov;33(11):3304–14.

84. Zon LI, Youssoufian H, Mather C, Lodish HF, Orkin SH. Activation of the erythropoietin receptor promoter by transcription factor GATA-1. *Proc Natl Acad Sci*. 1991 Dec 1;88(23):10638–41.
85. Zhao Wei, Kitidis Claire, Fleming Mark D, Lodish Harvey F, Ghaffari Saghi. Erythropoietin stimulates phosphorylation and activation of GATA-1 via the PI3-kinase/AKT signaling pathway. *Blood*. 2006 Feb;107(3):907–15.
86. Deindl Philipp, Klar Martin, Drews David, Cremer Malte, Gammella Elena, Gassmann Max, Dame Christof. Mice over-expressing human erythropoietin indicate that erythropoietin enhances expression of its receptor via up-regulated Gata1 and Tal1. Vol. 99, *Haematologica*. 2014. p. e205-7.
87. Kobayashi Makoto, Yamamoto Masayuki. Regulation of GATA1 gene expression. *J Biochem*. 2007 Jul;142(1):1–10.
88. Brand Marjorie, Ranish Jeffrey A. Proteomic/transcriptomic analysis of erythropoiesis. *Curr Opin Hematol*. 2021 May;28(3):150–7.
89. Hung HL, Lau J, Kim AY, Weiss MJ, Blobel GA. CREB-Binding protein acetylates hematopoietic transcription factor GATA-1 at functionally important sites. *Mol Cell Biol*. 1999 May;19(5):3496–505.
90. Blobel GA, Nakajima T, Eckner R, Montminy M, Orkin SH. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci U S A*. 1998 Mar;95(5):2061–6.
91. Boyes J, Byfield P, Nakatani Y, Ogryzko V. Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature*. 1998 Dec;396(6711):594–8.
92. Lamonica Janine M, Vakoc Christopher R, Blobel Gerd A. Acetylation of GATA-1 is required for chromatin occupancy. *Blood*. 2006 Dec;108(12):3736–8.
93. Lamonica Janine M, Deng Wulan, Kadauke Stephan, Campbell Amy E, Gamsjaeger Roland, Wang Hongxin, Cheng Yong, Billin Andrew N, Hardison Ross C, Mackay Joel P, Blobel Gerd A. Bromodomain protein Brd3 associates with acetylated GATA1 to promote its chromatin occupancy at erythroid target genes. *Proc Natl Acad Sci U S A*. 2011 May;108(22):E159-68.
94. Crossley M, Orkin SH. Phosphorylation of the erythroid transcription factor GATA-1. *J Biol Chem*. 1994 Jun;269(24):16589–96.
95. Partington GA, Patient RK. Phosphorylation of GATA-1 increases its DNA-binding affinity and is correlated with induction of human K562 erythroleukaemia cells. *Nucleic Acids Res*. 1999 Feb;27(4):1168–75.
96. Hernandez-Hernandez Angel, Ray Pampa, Litos Gabi, Ciro Marco, Ottolenghi Sergio, Beug Hartmut, Boyes Joan. Acetylation and MAPK phosphorylation cooperate to regulate the degradation of active GATA-1. Vol. 25, *The EMBO Journal*. 2006. p. 3264–74.

97. de Thonel Aurelie, Vandekerckhove Julie, Lanneau David, Selvakumar Subramaniam, Courtois Geneviève, Hazoume Adonis, Brunet Mathilde, Maurel Sebastien, Hammann Arlette, Ribeil Jean Antoine, Zermati Yael, Gabet Anne Sophie, Boyes Joan, Solary Eric, Hermine Olivier, Garrido Carmen. HSP27 controls GATA-1 protein level during erythroid cell differentiation. *Blood*. 2010 Jul;116(1):85–96.
98. Lee Hsiang Ying, Johnson Kirby D, Fujiwara Tohru, Boyer Meghan E, Kim Shin Il, Bresnick Emery H. Controlling hematopoiesis through sumoylation-dependent regulation of a GATA factor. *Mol Cell*. 2009 Dec;36(6):984–95.
99. Collavin Licio, Gostissa Monica, Avolio Fabio, Secco Paola, Ronchi Antonella, Santoro Claudio, Del Sal Giannino. Modification of the erythroid transcription factor GATA-1 by SUMO-1. *Proc Natl Acad Sci U S A*. 2004 Jun;101(24):8870–5.
100. De Maria R, Zeuner A, Eramo A, Domenichelli C, Bonci D, Grignani F, Srinivasula SM, Alnemri ES, Testa U, Peschle C. Negative regulation of erythropoiesis by caspase-mediated cleavage of GATA-1. *Nature*. 1999 Sep;401(6752):489–93.
101. Ribeil Jean Antoine, Zermati Yael, Vandekerckhove Julie, Cathelin Severine, Kersual Joelle, Dussiot Michaël, Coulon Séverine, Cruz Moura Ivan, Zeuner Ann, Kirkegaard-Sørensen Thomas, Varet Bruno, Solary Eric, Garrido Carmen, Hermine Olivier. Hsp70 regulates erythropoiesis by preventing caspase-3-mediated cleavage of GATA-1. *Nature*. 2007;445(7123):102–5.
102. Frisan Emilie, Vandekerckhove Julie, de Thonel Aurélie, Pierre-Eugène Cécile, Sternberg Alexander, Arlet Jean Benoît, et al. Defective nuclear localization of Hsp70 is associated with dyserythropoiesis and GATA-1 cleavage in myelodysplastic syndromes. *Blood*. 2012 Feb;119(6):1532–42.
103. Pevny Larysa, Simon M Celeste, Robertson Elizabeth, Klein William H, Tsai Shih Feng, D’Agati Vivette, Orkin Stuart H, Costantini Frank. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*. 1991;349(6306):257–60.
104. Whyatt D, Lindeboom F, Karis A, Ferreira R, Milot E, Hendriks R, de Bruijn M, Langeveld A, Gribnau J, Grosveld F, Philipsen S. An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. *Nature*. 2000 Aug;406(6795):519–24.
105. Takahashi S, Onodera K, Motohashi H, Suwabe N, Hayashi N, Yanai N, Nabesima Y, Yamamoto M. Arrest in primitive erythroid cell development caused by promoter-specific disruption of the GATA-1 gene. *J Biol Chem*. 1997 May;272(19):12611–5.
106. McDevitt MA, Shivdasani RA, Fujiwara Y, Yang H, Orkin SH. A “knockdown” mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level

- of transcription factor GATA-1. *Proc Natl Acad Sci U S A*. 1997 Jun;94(13):6781–5.
107. Crispino John D, Horwitz Marshall S. GATA factor mutations in hematologic disease. *Blood*. 2017 Apr;129(15):2103–10.
 108. Ludwig Leif S, Gazda Hanna T, Eng Jennifer C, Eichhorn Stephen W, Thiru Prathapan, Ghazvinian Roxanne, George Tracy I, Gotlib Jason R, Beggs Alan H, Sieff Colin A, Lodish Harvey F, Lander Eric S, Sankaran Vijay G. Altered translation of GATA1 in Diamond-Blackfan anemia. *Nat Med*. 2014 Jul;20(7):748–53.
 109. Manwani Deepa, Bieker James J. The erythroblastic island. *Curr Top Dev Biol*. 2008;82:23–53.
 110. Yeo Jia Hao, Lam Yun Wah, Fraser Stuart T. Cellular dynamics of mammalian red blood cell production in the erythroblastic island niche. *Biophys Rev*. 2019 Dec;11(6):873–94.
 111. Papaemmanuil Elli, Gerstung Moritz, Bullinger Lars, Gaidzik Verena I, Paschka Peter, Roberts Nicola D, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med*. 2016 Jun;374(23):2209–21.
 112. De Kouchkovsky I, Abdul-Hay M. “Acute myeloid leukemia: a comprehensive review and 2016 update”. *Blood Cancer J*. 2016 Jul;6(7):e441.
 113. Khoury Joseph D, Solary Eric, Abla Oussama, Akkari Yassmine, Alaggio Rita, Apperley Jane F, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia*. 2022;36(7):1703–19.
 114. Lagunas-Rangel Francisco Alejandro, Chávez-Valencia Venice, Gómez-Guijosa Miguel Ángel, Cortes-Penagos Carlos. Acute Myeloid Leukemia-Genetic Alterations and Their Clinical Prognosis. *Int J Hematol stem cell Res*. 2017 Oct;11(4):328–39.
 115. Strickland Stephen A, Vey Norbert. Diagnosis and treatment of therapy-related acute myeloid leukemia. *Crit Rev Oncol Hematol*. 2022;171:103607.
 116. Gilliland D Gary, Griffin James D. The roles of FLT3 in hematopoiesis and leukemia. *Blood*. 2002 Sep 1;100(5):1532–42.
 117. Ley Timothy J, Miller Christopher, Ding Li, Raphael Benjamin J, Mungall Andrew J, Robertson A Gordon, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013 May;368(22):2059–74.
 118. Meyer Sara C, Levine Ross L. Translational implications of somatic genomics in acute myeloid leukaemia. *Lancet Oncol*. 2014;15(9):e382–94.
 119. Kantarjian Hagop, Kadia Tapan, DiNardo Courtney, Daver Naval, Borthakur Gautam, Jabbour Elias, Garcia-Manero Guillermo, Konopleva Marina, Ravandi Farhad. Acute myeloid leukemia: current progress and future directions. *Blood Cancer J*. 2021 Feb;11(2):41.
 120. Green Steven D, Konig Heiko. Treatment of Acute Myeloid Leukemia in the Era of Genomics—

Achievements and Persisting Challenges. Vol. 11, *Frontiers in Genetics*. 2020.

121. Issa Ghayas C, DiNardo Courtney D. Acute myeloid leukemia with IDH1 and IDH2 mutations: 2021 treatment algorithm. *Blood Cancer J*. 2021 Jun;11(6):107.
122. Samra Bachar, Konopleva Marina, Isidori Alessandro, Daver Naval, DiNardo Courtney. Venetoclax-Based Combinations in Acute Myeloid Leukemia: Current Evidence and Future Directions. *Front Oncol*. 2020;10:562558.
123. Appelbaum Frederick R, Bernstein Irwin D. Gemtuzumab ozogamicin for acute myeloid leukemia. *Blood*. 2017 Nov 30;130(22):2373–6.
124. Fagnan Alexandre, Piqué-Borràs Maria Riera, Tauchmann Samantha, Mercher Thomas, Schwaller Juerg. Molecular Landscapes and Models of Acute Erythroleukemia. *HemaSphere*. 2021 May;5(5):e558.
125. Santos FPS, Faderl S, Garcia-Manero G, Koller C, Beran M, O'Brien S, Pierce S, Freireich EJ, Huang X, Borthakur G, Bueso-Ramos C, de Lima M, Keating M, Cortes J, Kantarjian H, Ravandi F. Adult acute erythroleukemia: an analysis of 91 patients treated at a single institution. *Leukemia*. 2009 Dec;23(12):2275–80.
126. Fagnan Alexandre, Piqué-Borràs Maria Riera, Tauchmann Samantha, Mercher Thomas, Schwaller Juerg. Molecular Landscapes and Models of Acute Erythroleukemia. *HemaSphere*. 2021 May;5(5):e558.
127. SCHWARTZ STEVEN O, CRITCHLOW JOAN. Erythremic Myelosis (Di Guglielmo's Disease) : Critical Review with Report of Four Cases, and Comments on Erythroleukemia. *Blood*. 1952 Aug 1;7(8):765–93.
128. DAMESHEK WILLIAM. Editorial: Some Speculations on the Myeloproliferative Syndromes. *Blood*. 1951 Apr 1;6(4):372–5.
129. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, Sultan C. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol*. 1976;33(4):451–8.
130. Kowal-Vern Areta, Cotelingam James, Schumacher Harold R. The Prognostic Significance of Proerythroblasts in Acute Erythroleukemia. *Am J Clin Pathol*. 1992;98(1):34–40.
131. Boddu Prajwal, Benton Christopher B, Wang Wei, Borthakur Gautam, Khoury Joseph D, Pemmaraju Naveen. Erythroleukemia-historical perspectives and recent advances in diagnosis and management. *Blood Rev*. 2018;32(2):96–105.
132. Santos Fabio PS, Bueso-Ramos Carlos E, Ravandi Farhad. Acute erythroleukemia: diagnosis and management. *Expert Rev Hematol*. 2010 Dec;3(6):705–18.
133. Villeval JL, Cramer P, Lemoine F, Henri A, Bettaieb A, Bernaudin F, Beuzard Y, Berger R, Flandrin G, Breton-Gorius J. Phenotype of early erythroblastic leukemias. *Blood*. 1986 Nov

- 1;68(5):1167–74.
134. Fang Hong, Wang Sa A, You M James, Hu Shimin, Miranda Roberto N, Tang Zhenya, Lin Pei, Jorgensen Jeffrey L, Xu Jie, Thakral Beenu, Schlette Ellen J, El Hussein Siba, Bueso-Ramos Carlos, Medeiros L Jeffrey, Wang Wei. Flow cytometry immunophenotypic features of pure erythroid leukemia and the distinction from reactive erythroid precursors. *Cytometry B Clin Cytom.* 2022 Sep;
 135. HEATH JR. CLARK W, BENNETT JOHN M, WHANG-PENG JACQUELINE, BERRY ELIZABETH W, WIERNICK PETER H. Cytogenetic Findings in Erythroleukemia. *Blood.* 1969 Mar 1;33(3):453–67.
 136. Micci F, Thorsen J, Panagopoulos I, Nyquist KB, Zeller B, Tierens A, Heim S. High-throughput sequencing identifies an NFIA/CBFA2T3 fusion gene in acute erythroid leukemia with t(1;16)(p31;q24). *Leukemia.* 2013;27(4):980–2.
 137. Micci F, Thorsen J, Haugom L, Zeller B, Tierens A, Heim S. Translocation t(1;16)(p31;q24) rearranging CBFA2T3 is specific for acute erythroid leukemia. *Leukemia.* 2011;25(9):1510–2.
 138. Panagopoulos Ioannis, Micci Francesca, Thorsen Jim, Haugom Lisbeth, Buechner Jochen, Kerndrup Gitte, Tierens Anne, Zeller Bernward, Heim Sverre. Fusion of ZMYND8 and RELA Genes in Acute Erythroid Leukemia. *PLoS One.* 2013;8(5):1–6.
 139. Lessard M, Struski S, Leymarie V, Flandrin G, Lafage-Pochitaloff M, Mozziconacci MJ, et al. Cytogenetic study of 75 erythroleukemias. *Cancer Genet Cytogenet.* 2005;163(2):113–22.
 140. Fagnan Alexandre, Bagger Frederik Otzen, Piqué-Borràs Maria Riera, Ignacimoutou Cathy, Caulier Alexis, Lopez Cécile K, et al. Human erythroleukemia genetics and transcriptomes identify master transcription factors as functional disease drivers. *Blood.* 2020 Aug;136(6):698–714.
 141. Iacobucci Ilaria, Qu Chunxu, Varotto Elena, Janke Laura J, Yang Xu, Seth Aman, et al. Modeling and targeting of erythroleukemia by hematopoietic genome editing. *Blood.* 2021 Mar;137(12):1628–40.
 142. Fang Hong, Wang Sa A, Khoury Joseph D, El Hussein Siba, Kim Do Hwan, Tashakori Mehrnoosh, Tang Zhenya, Li Shaoying, Hu Zhihong, Jelloul Fatima Zahra, Patel Keyur P, McDonnell Timothy J, Kadia Tapan, Medeiros L Jeffrey, Wang Wei. Pure erythroid leukemia is characterized by biallelic TP53 inactivation and abnormal p53 expression patterns in de novo and secondary cases. *Haematologica.* 2022 May;
 143. Graf Thomas, Ade Norbert, Beug Hartmut. Temperature-sensitive mutant of avian erythroblastosis virus suggests a block of differentiation as mechanism of leukaemogenesis. *Nature.* 1978;275(5680):496–501.
 144. Beug H, Bauer A, Dolznig H, von Lindern M, Lobmayer L, Mellitzer G, Steinlein P, Wessely O,

- Müllner E. Avian erythropoiesis and erythroleukemia: towards understanding the role of the biomolecules involved. *Biochim Biophys Acta - Rev Cancer*. 1996;1288(3):M35–47.
145. Rietveld Luc EG, Caldenhoven Eric, Stunnenberg Hendrik G. Avian erythroleukemia: a model for corepressor function in cancer. *Oncogene*. 2001;20(24):3100–9.
 146. Friend Charlotte. CELL-FREE TRANSMISSION IN ADULT SWISS MICE OF A DISEASE HAVING THE CHARACTER OF A LEUKEMIA . *J Exp Med*. 1957 Apr 1;105(4):307–18.
 147. Ney Paul A, D'Andrea Alan D. Friend erythroleukemia revisited. *Blood*. 2000 Dec 1;96(12):3675–80.
 148. Orkin Stuart H. Differentiation of murine erythroleukemic (friend) cells: An in vitro model of erythropoiesis. *In Vitro*. 1978;14(1):146–54.
 149. Drexler Hans G, Matsuo Yoshinobu, MacLeod Roderick AF. Malignant hematopoietic cell lines: in vitro models for the study of erythroleukemia. Vol. 28, *Leukemia research*. England; 2004. p. 1243–51.
 150. Fan Tao, Schmidtman Anja, Xi Sichuan, Briones Victorino, Zhu Heming, Suh Hyung Chan, Gooya John, Keller Jonathan R, Xu Hong, Roayaei Jean, Anver Miriam, Ruscetti Sandra, Muegge Kathrin. DNA hypomethylation caused by Lsh deletion promotes erythroleukemia development. *Epigenetics*. 2008;3(3):134–42.
 151. Yuan Wen, Xu Mo, Huang Chang, Liu Nan, Chen She, Zhu Bing. H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation. *J Biol Chem*. 2011 Mar 11;286(10):7983–9.
 152. Brumbaugh Justin, Kim Ik Soo, Ji Fei, Huebner Aaron J, Di Stefano Bruno, Schwarz Benjamin A, Charlton Jocelyn, Coffey Amy, Choi Jiho, Walsh Ryan M, Schindler Jeffrey W, Anselmo Anthony, Meissner Alexander, Sadreyev Ruslan I, Bernstein Bradley E, Hock Hanno, Hochedlinger Konrad. Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. *Nat Cell Biol*. 2019;21(11):1449–61.
 153. Leonards Katharina, Almosaileakh Marwa, Tauchmann Samantha, Bagger Frederik Otzen, Thirant Cécile, Juge Sabine, Bock Thomas, Méreau Hélène, Bezerra Matheus F, Tzankov Alexandar, Ivanek Robert, Losson Régine, Peters Antoine HFM, Mercher Thomas, Schwaller Juerg. Nuclear interacting SET domain protein 1 inactivation impairs GATA1-regulated erythroid differentiation and causes erythroleukemia. *Nat Commun*. 2020;11(1):2807.
 154. England Samantha J, McGrath Kathleen E, Frame Jenna M, Palis James. Immature erythroblasts with extensive ex vivo self-renewal capacity emerge from the early mammalian fetus. Vol. 117, *Blood*. Washington, DC; 2011. p. 2708–17.
 155. Kumar Amit, Wempe Michael. EZH2 Inhibitor GSK126: Metabolism, drug transporter and rat pharmacokinetic studies. *Med Res Arch*. 2015;

156. Fiskus Warren, Wang Yongchao, Sreekumar Arun, Buckley Kathleen M, Shi Huidong, Jillella Anand, Ustun Celalettin, Rao Rekha, Fernandez Pravina, Chen Jianguang, Balusu Ramesh, Koul Sanjay, Atadja Peter, Marquez Victor E, Bhalla Kapil N. Combined epigenetic therapy with the histone methyltransferase EZH2 inhibitor 3-deazaneplanocin A and the histone deacetylase inhibitor panobinostat against human AML cells. *Blood*. 2009 Sep;114(13):2733–43.
157. Fujiwara Tohru, Saitoh Haruka, Inoue Ai, Kobayashi Masahiro, Okitsu Yoko, Katsuoka Yuna, Fukuhara Noriko, Onishi Yasushi, Ishizawa Kenichi, Ichinohasama Ryo, Harigae Hideo. 3-Deazaneplanocin A (DZNep), an Inhibitor of S-Adenosylmethionine-dependent Methyltransferase, Promotes Erythroid Differentiation*. *J Biol Chem*. 2014;289(12):8121–34.
158. Hoppe Philipp S, Schwarzfischer Michael, Loeffler Dirk, Kokkaliaris Konstantinos D, Hilsenbeck Oliver, Moritz Nadine, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*. 2016;535(7611):299–302.
159. Kurotaki Naohiro, Imaizumi Kiyoshi, Harada Naoki, Masuno Mitsuo, Kondoh Tatsuro, Nagai Toshiro, et al. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet*. 2002;30(4):365–6.
160. Baujat Geneviève, Cormier-Daire Valérie. Sotos syndrome. *Orphanet J Rare Dis*. 2007;2(1):36.
161. Foster Alison, Zachariou Anna, Loveday Chey, Ashraf Tazeen, Blair Edward, Clayton-Smith Jill, et al. The phenotype of Sotos syndrome in adulthood: A review of 44 individuals. *Am J Med Genet Part C Semin Med Genet*. 2019 Dec 1;181(4):502–8.
162. Jeong Yuji, Kim Taejoon, Kim Suyeun, Hong Yoon Ki, Cho Kyoung Sang, Lee Im Soon. Overexpression of histone methyltransferase NSD in *Drosophila* induces apoptotic cell death via the Jun-N-terminal kinase pathway. *Biochem Biophys Res Commun*. 2018;496(4):1134–40.
163. Choi Saeyan, Song Bokyeong, Shin Hyewon, Won Chihyun, Kim Taejoon, Yoshida Hideki, Lee Daewon, Chung Jongkyeong, Cho Kyoung Sang, Lee Im Soon. *Drosophila* NSD deletion induces developmental anomalies similar to those seen in Sotos syndrome 1 patients. *Genes Genomics*. 2021;43(7):737–48.
164. Dulmovits Brian M, Hom Jimmy, Narla Anupama, Mohandas Narla, Blanc Lionel. Characterization, regulation, and targeting of erythroid progenitors in normal and disordered human erythropoiesis. *Curr Opin Hematol*. 2017 May;24(3):159–66.
165. Barminko Jeffrey, Reinholt Brad, Baron Margaret H. Development and differentiation of the erythroid lineage in mammals. *Dev Comp Immunol*. 2016 May;58:18–29.
166. Swiers Gemma, Patient Roger, Loose Matthew. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev Biol*. 2006 Jun;294(2):525–

40.

167. Palii Carmen G, Cheng Qian, Gillespie Mark A, Shannon Paul, Mazurczyk Michalina, Napolitani Giorgio, Price Nathan D, Ranish Jeffrey A, Morrissey Edward, Higgs Douglas R, Brand Marjorie. Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. *Cell Stem Cell*. 2019 May;24(5):812-820.e5.
168. Khoury Joseph D, Solary Eric, Ablu Oussama, Akkari Yasmine, Alaggio Rita, Apperley Jane F, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia*. 2022;36(7):1703–19.
169. Grossmann V, Bacher U, Haferlach C, Schnittger S, Pötzinger F, Weissmann S, Roller A, Eder C, Fasan A, Zenger M, Staller M, Kern W, Kohlmann A, Haferlach T. Acute erythroid leukemia (AEL) can be separated into distinct prognostic subsets based on cytogenetic and molecular genetic characteristics. *Leukemia*. 2013;27(9):1940–3.
170. Ping N, Sun A, Song Y, Wang Q, Yin J, Cheng W, Xu Y, Wen L, Yao H, Ma L, Qiu H, Ruan C, Wu D, Chen S. Exome sequencing identifies highly recurrent somatic GATA2 and CEBPA mutations in acute erythroid leukemia. *Leukemia*. 2017 Jan;31(1):195–202.
171. Cervera Nathalie, Lhoumeau Anne Catherine, Adélaïde José, Guille Arnaud, Murati Anne, Mozziconacci Marie Joëlle, Vey Norbert, Birnbaum Daniel, Gelsi-Boyer Véronique. Acute erythroid leukemias have a distinct molecular hierarchy from non-erythroid acute myeloid leukemias. Vol. 105, *Haematologica*. 2020. p. e340–2.
172. Takeda June, Yoshida Kenichi, Nakagawa Masahiro M, Nannya Yasuhito, Yoda Akinori, Saiki Ryunosuke, et al. Amplified EPOR/JAK2 Genes Define a Unique Subtype of Acute Erythroid Leukemia. *Blood cancer Discov*. 2022 Sep;3(5):410–27.
173. Brumbaugh Justin, Kim Ik Soo, Ji Fei, Huebner Aaron J, Di Stefano Bruno, Schwarz Benjamin A, Charlton Jocelyn, Coffey Amy, Choi Jiho, Walsh Ryan M, Schindler Jeffrey W, Anselmo Anthony, Meissner Alexander, Sadreyev Ruslan I, Bernstein Bradley E, Hock Hanno, Hochedlinger Konrad. Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. *Nat Cell Biol*. 2019;21(11):1449–61.
174. Strouboulis John. Erythroleukemia: all roads lead to GATA1? *Blood*. 2020 Aug;136(6):648–9.
175. Karayel Özge, Xu Peng, Bludau Isabell, Velan Bhoopalan Senthil, Yao Yu, Ana Rita Freitas Colaco, Santos Alberto, Schulman Brenda A, Alpi Arno F, Weiss Mitchell J, Mann Matthias. Integrative proteomics reveals principles of dynamic phosphosignaling networks in human erythropoiesis. *Mol Syst Biol*. 2020 Dec;16(12):e9813–e9813.
176. Gautier Emilie Fleur, Ducamp Sarah, Leduc Marjorie, Salnot Virginie, Guillonnet François,

- Dussiot Michael, Hale John, Giarratana Marie Catherine, Raimbault Anna, Douay Luc, Lacombe Catherine, Mohandas Narla, Verdier Frédérique, Zermati Yael, Mayeux Patrick. Comprehensive Proteomic Analysis of Human Erythropoiesis. *Cell Rep.* 2016/07/21. 2016 Aug 2;16(5):1470–84.
177. Heshusius Steven, Heideveld Esther, Burger Patrick, Thiel-Valkhof Marijke, Sellink Erica, Varga Eszter, Ovchynnikova Elina, Visser Anna, Martens Joost HA, von Lindern Marieke, van den Akker Emile. Large-scale in vitro production of red blood cells from human peripheral blood mononuclear cells. *Blood Adv.* 2019 Nov 12;3(21):3337–50.
178. Kurita Ryo, Suda Noriko, Sudo Kazuhiro, Miharada Kenichi, Hiroshima Takashi, Miyoshi Hiroyuki, Tani Kenzaburo, Nakamura Yukio. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS One.* 2013/03/22. 2013;8(3):e59890–e59890.
179. Ullmark Tove, Montano Giorgia, Järnström Linnea, Jernmark Nilsson Helena, Håkansson Erik, Drott Kristina, Nilsson Björn, Vidovic Karina, Gullberg Urban. Anti-apoptotic quinolinate phosphoribosyltransferase (QPRT) is a target gene of Wilms' tumor gene 1 (WT1) protein in leukemic cells. *Biochem Biophys Res Commun.* 2017 Jan;482(4):802–7.
180. Arkin Steven, Naprstek Beth, Guarini Ludovico, Ferrone Soldano, Lipton Jeffrey M. Expression of Intercellular Adhesion Molecule-1 (CD54) on Hematopoietic Progenitors. *Blood.* 1991;77(5):948–53.
181. Huang Chen Kai, Sun Ying, Lv Lei, Ping Yong. ENO1 and Cancer. *Mol Ther - Oncolytics.* 2022;24:288–98.
182. Dong Henry Y, Wilkes Steven, Yang Haisu. CD71 is selectively and ubiquitously expressed at high levels in erythroid precursors of all maturation stages: a comparative immunochemical study with glycophorin A and hemoglobin A. *Am J Surg Pathol.* 2011 May;35(5):723–32.
183. Seo Jun Young, Yaneva Rakina, Cresswell Peter. Viperin: a multifunctional, interferon-inducible protein that regulates virus replication. *Cell Host Microbe.* 2011 Dec;10(6):534–9.
184. Dorigi Kristel M, Swigut Tomek, Henriques Telmo, Bhanu Natarajan V, Scruggs Benjamin S, Nady Nataliya, Still Christopher D 2nd, Garcia Benjamin A, Adelman Karen, Wysocka Joanna. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell.* 2017 May;66(4):568-576.e4.
185. Han Heonjong, Cho Jae Won, Lee Sangyoung, Yun Ayoung, Kim Hyojin, Bae Dasom, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 2018 Jan;46(D1):D380–6.
186. Tan Nicole Y, Khachigian Levon M. Sp1 Phosphorylation and Its Regulation of Gene Transcription. *Mol Cell Biol.* 2009;29(10):2483–8.

187. Resendes Karen K, Rosmarin Alan G. Sp1 control of gene expression in myeloid cells. *Crit Rev Eukaryot Gene Expr.* 2004;14(3):171–81.
188. Bottardi Stefania, Mavoungou Lionel, Pak Helen, Daou Salima, Bourgoin Vincent, Lakehal Yahia A, Affar El Bachir, Milot Eric. The IKAROS interaction with a complex including chromatin remodeling and transcription elongation activities is required for hematopoiesis. *PLoS Genet.* 2014 Dec;10(12):e1004827.
189. Pimkin Maxim, Kossenkov Andrew V, Mishra Tejaswini, Morrissey Christopher S, Wu Weisheng, Keller Cheryl A, Blobel Gerd A, Lee Dongwon, Beer Michael A, Hardison Ross C, Weiss Mitchell J. Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. *Genome Res.* 2014/10/15. 2014 Dec;24(12):1932–44.
190. Ueki Nobuhide, Zhang Leiqing, Hayman Michael J. Ski Negatively Regulates Erythroid Differentiation through Its Interaction with GATA1. *Mol Cell Biol.* 2004 Dec 1;24(23):10118 LP – 10125.
191. Larsen J, Beug H, Hayman MJ. The v-ski oncogene cooperates with the v-sea oncogene in erythroid transformation by blocking erythroid differentiation. *Oncogene.* 1992 Oct;7(10):1903–11.
192. Cui Jie, Wang Li, Ren Xiaoyue, Zhang Yamin, Zhang Hongyi. LRP6: A Multifunctional Protein Involved in Energy Metabolism and Human Disease. Vol. 10, *Frontiers in physiology.* 2019. p. 595.
193. Chen Xiaofang, Chen Xianling, Huang Yiping, Lin Jia, Wu Yong, Chen Yuanzhong. TCP1 increases drug resistance in acute myeloid leukemia by suppressing autophagy via activating AKT/mTOR signaling. *Cell Death Dis.* 2021;12(11):1058.
194. Han Yang, Wang Xin. The emerging roles of KPNA2 in cancer. *Life Sci.* 2020;241:117140.
195. Duncan R, Bazar L, Michelotti G, Tomonaga T, Krutzsch H, Avigan M, Levens D. A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. *Genes Dev.* 1994 Feb;8(4):465–80.
196. Dijon Marilyne, Bardin Florence, Murati Anne, Batoz Michèle, Chabannon Christian, Tonnelle Cécile. The role of Ikaros in human erythroid differentiation. *Blood.* 2008 Feb;111(3):1138–46.
197. Kim J, Sif S, Jones B, Jackson A, Koipally J, Heller E, Winandy S, Viel A, Sawyer A, Ikeda T, Kingston R, Georgopoulos K. Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity.* 1999 Mar;10(3):345–55.
198. Long Jianyin, Zuo Dongmei, Park Morag. Pc2-mediated sumoylation of Smad-interacting protein 1 attenuates transcriptional repression of E-cadherin. *J Biol Chem.* 2005 Oct;280(42):35477–89.

199. Koyama Masako, Matsuura Yoshiyuki. An allosteric mechanism to displace nuclear export cargo from CRM1 and RanGTP by RanBP1. *EMBO J.* 2010 Jun;29(12):2002–13.
200. Wälde Sarah, Thakar Ketan, Hutten Saskia, Spillner Christiane, Nath Annegret, Rothbauer Ulrich, Wiemann Stefan, Kehlenbach Ralph H. The nucleoporin Nup358/RanBP2 promotes nuclear import in a cargo- and transport receptor-specific manner. *Traffic.* 2012 Feb;13(2):218–33.
201. Yazdi Parvin T, Wang Yi, Zhao Song, Patel Nimit, Lee Eva YHP, Qin Jun. SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev.* 2002 Mar;16(5):571–82.
202. Grandi P, Dang T, Pané N, Shevchenko A, Mann M, Forbes D, Hurt E. Nup93, a vertebrate homologue of yeast Nic96p, forms a complex with a novel 205-kDa protein and is required for correct nuclear pore assembly. *Mol Biol Cell.* 1997 Oct;8(10):2017–38.
203. Mello Fabiana V, Land Marcelo GP, Costa Elaine S, Teodósio Cristina, Sanchez María Luz, Bárcena Paloma, Peres Rodrigo T, Pedreira Carlos E, Alves Liliane R, Orfao Alberto. Maturation-associated gene expression profiles during normal human bone marrow erythropoiesis. *Cell death Discov.* 2019;5:69.
204. Muto Akiko, Sugihara Yoshihiko, Shibakawa Minami, Oshima Kenzi, Matsuda Tsukasa, Nadano Daita. The mRNA-binding protein Serbp1 as an auxiliary protein associated with mammalian cytoplasmic ribosomes. *Cell Biochem Funct.* 2018 Aug;36(6):312–22.
205. Fujiwara Tohru, Alqadi Yarob Wael, Okitsu Yoko, Fukuhara Noriko, Onishi Yasushi, Ishizawa Kenichi, Harigae Hideo. Role of transcriptional corepressor ETO2 in erythroid cells. *Exp Hematol.* 2013 Mar;41(3):303-15.e1.
206. Stadhouders Ralph, Cico Alba, Stephen Tharshana, Thongjuea Supat, Kolovos Petros, Baymaz H Irem, et al. Control of developmentally primed erythroid genes by combinatorial co-repressor actions. *Nat Commun.* 2015;6(1):8893.
207. Liu Hang, Wang Min, Li Min, Wang Donghai, Rao Qing, Wang Yang, Xu Zhifang, Wang Jianxiang. Expression and role of DJ-1 in leukemia. *Biochem Biophys Res Commun.* 2008 Oct;375(3):477–83.
208. Kalin Jay H, Wu Muzhou, Gomez Andrea V, Song Yun, Das Jayanta, Hayward Dawn, et al. Targeting the CoREST complex with dual histone deacetylase and demethylase inhibitors. *Nat Commun.* 2018;9(1):53.
209. Grosso Rubén, Fader Claudio M, Colombo María I. Autophagy: A necessary event during erythropoiesis. *Blood Rev.* 2017;31(5):300–5.
210. Unoki Motoko, Nishidate Toshihiko, Nakamura Yusuke. ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene.* 2004;23(46):7601–10.

211. Zhang Xianqin, Chen Shenghan, Yoo Shin, Chakrabarti Susmita, Zhang Teng, Ke Tie, Oberti Carlos, Yong Sandro L, Fang Fang, Li Lin, de la Fuente Roberto, Wang Lejin, Chen Qiuyun, Wang Qing Kenneth. Mutation in nuclear pore component NUP155 leads to atrial fibrillation and early sudden cardiac death. *Cell*. 2008 Dec;135(6):1017–27.
212. Busayavalasa Kiran, Chen Xin, Farrants Ann Kristin Östlund, Wagner Nicole, Sabri Nafiseh. The Nup155-mediated organisation of inner nuclear membrane proteins is independent of Nup155 anchoring to the metazoan nuclear pore complex. *J Cell Sci*. 2012;125(18):4214–8.
213. Choufani S, Cytrynbaum C, Chung BHY, Turinsky AL, Grafodatskaya D, Chen YA, et al. NSD1 mutations generate a genome-wide DNA methylation signature. *Nat Commun*. 2015;6(1):10207.
214. Lam Ulysses Tsz Fung, Tan Bryan Kok Yan, Poh John Jia Xin, Chen Ee Sin. Structural and functional specificity of H3K36 methylation. *Epigenetics Chromatin*. 2022;15(1):17.
215. Molenaar Thom M, Malik Muddassir, Silva Joana, Liu Ning Qing, Haarhuis Judith HI, Ambrosi Christina, Kwesi-Maliepaard Eliza Mari, van Welsem Tibor, Baubec Tuncay, Faller William J, van Leeuwen Fred. The histone methyltransferase SETD2 negatively regulates cell size. *J Cell Sci*. 2022 Oct 6;135(19):jcs259856.
216. Kudithipudi Srikanth, Lungu Cristiana, Rathert Philipp, Happel Nicole, Jeltsch Albert. Substrate Specificity Analysis and Novel Substrates of the Protein Lysine Methyltransferase NSD1. *Chem Biol*. 2014;21(2):226–37.
217. Conteduca Giuseppina, Cangelosi Davide, Coco Simona, Malacarne Michela, Baldo Chiara, Arado Alessia, Pinto Rute, Testa Barbara, Coviello Domenico A. NSD1 Mutations in Sotos Syndrome Induce Differential Expression of Long Noncoding RNAs, miR646 and Genes Controlling the G2/M Checkpoint. *Life (Basel, Switzerland)*. 2022 Jul;12(7).
218. Quintero-Rivera Fabiola, Eno Celeste C, Sutanto Christine, Jones Kelly L, Nowaczyk Małgorzata JM, Wong Derek, Earl Dawn, Mirzaa Ghayda, Beck Anita, Martinez-Agosto Julian A. 5q35 duplication presents with psychiatric and undergrowth phenotypes mediated by NSD1 overexpression and mTOR signaling downregulation. *Hum Genet*. 2021;140(4):681–90.
219. Anthony JC, Yoshizawa F, Anthony TG, Vary TC, Jefferson LS, Kimball SR. Leucine stimulates translation initiation in skeletal muscle of postabsorptive rats via a rapamycin-sensitive pathway. *J Nutr*. 2000 Oct;130(10):2413–9.
220. Weinberg Daniel N, Papillon-Cavanagh Simon, Chen Haifen, Yue Yuan, Chen Xiao, Rajagopalan Kartik N, et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature*. 2019;573(7773):281–6.
221. Streubel Gundula, Watson Ariane, Jammula Sri Ganesh, Scelfo Andrea, Fitzpatrick Darren J, Oliviero Giorgio, McCole Rachel, Conway Eric, Glancy Eleanor, Negri Gian Luca, Dillon

- Eugene, Wynne Kieran, Pasini Diego, Krogan Nevan J, Bracken Adrian P, Cagney Gerard. The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol Cell*. 2018 Apr;70(2):371-379.e5.
222. Bracken Adrian P, Brien Gerard L, Verrijzer C Peter. Dangerous liaisons: interplay between SWI/SNF, NuRD, and Polycomb in chromatin regulation and cancer. *Genes Dev*. 2019 Aug;33(15–16):936–59.
223. Brennan Kevin, Zheng Hong, Fahrner Jill A, Shin June Ho, Gentles Andrew J, Schaefer Bradley, Sunwoo John B, Bernstein Jonathan A, Gevaert Olivier. NSD1 mutations deregulate transcription and DNA methylation of bivalent developmental genes in Sotos syndrome. *Hum Mol Genet*. 2022 Jul;31(13):2164–84.
224. Tatton-Brown Katrina, Rahman Nazneen. The NSD1 and EZH2 overgrowth genes, similarities and differences. *Am J Med Genet C Semin Med Genet*. 2013 May;163C(2):86–91.
225. Griffiths Sara, Loveday Chey, Zachariou Anna, Behan Lucy Ann, Chandler Kate, Cole Trevor, D'Arrigo Stefano, Dieckmann Andrea, Foster Alison, Gibney James, Hunter Matthew, Milani Donatella, Pantaleoni Chiara, Roche Edna, Sherlock Mark, Springer Amanda, White Susan M, Tatton-Brown Katrina. EED and EZH2 constitutive variants: A study to expand the Cohen-Gibson syndrome phenotype and contrast it with Weaver syndrome. *Am J Med Genet A*. 2019 Apr;179(4):588–94.
226. Imagawa Eri, Albuquerque Edoarda VA, Isidor Bertrand, Mitsuhashi Satomi, Mizuguchi Takeshi, Miyatake Satoko, Takata Atsushi, Miyake Noriko, Boguszewski Margaret CS, Boguszewski César L, Lerario Antonio M, Funari Mariana A, Jorge Alexander AL, Matsumoto Naomichi. Novel SUZ12 mutations in Weaver-like syndrome. *Clin Genet*. 2018 Nov 1;94(5):461–6.
227. Tan Jing, Yang Xiaojing, Zhuang Li, Jiang Xia, Chen Wei, Lee Puay Leng, Karuturi RK Murthy, Tan Patrick Boon Ooi, Liu Edison T, Yu Qiang. Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev*. 2007 May;21(9):1050–63.
228. Mayr Christian, Wagner Andrej, Stoecklinger Angelika, Jakab Martin, Illig Romana, Berr Frieder, Pichler Martin, Di Fazio Pietro, Ocker Matthias, Neureiter Daniel, Kiesslich Tobias. 3-Deazaneplanocin A May Directly Target Putative Cancer Stem Cells in Biliary Tract Cancer. *Anticancer Res*. 2015 Sep;35(9):4697–705.
229. McCabe Michael T, Ott Heidi M, Ganji Gopinath, Korenchuk Susan, Thompson Christine, Van Aller Glenn S, et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature*. 2012;492(7427):108–12.
230. Kim Kimberly H, Roberts Charles WM. Targeting EZH2 in cancer. *Nat Med*. 2016

Feb;22(2):128–34.

231. Sneeringer Christopher J, Scott Margaret Porter, Kuntz Kevin W, Knutson Sarah K, Pollock Roy M, Richon Victoria M, Copeland Robert A. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci U S A*. 2010 Dec;107(49):20980–5.
232. Xu Feng, Li Xiao, Wu Lingyun, Zhang Qingxia, Yang Rui, Yang Yujuan, Zhang Zheng, He Qi, Chang Chunkang. Overexpression of the EZH2, RING1 and BMI1 genes is common in myelodysplastic syndromes: relation to adverse epigenetic alteration and poor prognostic scoring. *Ann Hematol*. 2011 Jun;90(6):643–53.
233. Rinke Jenny, Chase Andrew, Cross Nicholas CP, Hochhaus Andreas, Ernst Thomas. EZH2 in Myeloid Malignancies. *Cells*. 2020 Jul;9(7).
234. Adema Vera, Colla Simona. EZH2 Inhibitors: The Unpacking Revolution. *Cancer Res*. 2022 Feb 2;82(3):359–61.
235. Fujiwara Tohru, Saitoh Haruka, Inoue Ai, Kobayashi Masahiro, Okitsu Yoko, Katsuoka Yuna, Fukuhara Noriko, Onishi Yasushi, Ishizawa Kenichi, Ichinohasama Ryo, Harigae Hideo. 3-Deazaneplanocin A (DZNep), an Inhibitor of S-Adenosylmethionine-dependent Methyltransferase, Promotes Erythroid Differentiation. *J Biol Chem*. 2014;289(12):8121–34.
236. Gregoricchio Sebastian, Polit Lélia, Esposito Michela, Berthelet Jérémy, Delestré Laure, Evanno Emilie, Diop M'Boyba, Gallais Isabelle, Aleth Hanna, Poplineau Mathilde, Zwart Wilbert, Rosenbauer Frank, Rodrigues-Lima Fernando, Duprez Estelle, Boeva Valentina, Guillouf Christel. HDAC1 and PRC2 mediate combinatorial control in SPI1/PU.1-dependent gene repression in murine erythroleukaemia. *Nucleic Acids Res*. 2022 Aug 12;50(14).
237. Wilson Marieangela C, Trakarnsanga Kongtana, Heesom Kate J, Cogan Nicola, Green Carole, Toye Ashley M, Parsons Steve F, Anstee David J, Frayne Jan. Comparison of the Proteome of Adult and Cord Erythroid Cells, and Changes in the Proteome Following Reticulocyte Maturation. *Mol Cell Proteomics*. 2016 Jun;15(6):1938–46.
238. Khajuria Rajiv K, Munschauer Mathias, Ulirsch Jacob C, Fiorini Claudia, Ludwig Leif S, McFarland Sean K, et al. Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. *Cell*. 2018 Mar;173(1):90-103.e19.
239. Feo S, Arcuri D, Piddini E, Passantino R, Giallongo A. ENO1 gene product binds to the c-myc promoter and acts as a transcriptional repressor: relationship with Myc promoter-binding protein 1 (MBP-1). *FEBS Lett*. 2000 May;473(1):47–52.
240. Handschuh Luiza, Kaźmierczak Maciej, Milewski C. Marek, Góralski Michał, Łuczak Magdalena, Wojtaszewska Marzena, Uszczyńska-Ratajczak Barbara, Lewandowski Krzysztof, Komarnicki Mieczysław, Figlerowicz Marek. Gene expression profiling of acute myeloid

- leukemia samples from adult patients with AML-M1 and -M2 through boutique microarrays, real-time PCR and droplet digital PCR. *Int J Oncol.* 2018;52(3):656–78.
241. Beishline Kate, Azizkhan-Clifford Jane. Sp1 and the “hallmarks of cancer”. *FEBS J.* 2015 Jan;282(2):224–58.
 242. Hu Jie Hong, Navas Patrick, Cao Hua, Stamatoyannopoulos George, Song Chao Zhong. Systematic RNAi Studies on the Role of Sp/KLF Factors in Globin Gene Expression and Erythroid Differentiation. *J Mol Biol.* 2007;366(4):1064–73.
 243. Crispino John D. GATA1 in normal and malignant hematopoiesis. *Semin Cell Dev Biol.* 2005;16(1):137–47.
 244. Laugesen Anne, Helin Kristian. Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. *Cell Stem Cell.* 2014;14(6):735–51.
 245. van Bergen Maaik GJM, van der Reijden Bert A. Targeting the GFI1/1B-CoREST Complex in Acute Myeloid Leukemia. *Front Oncol.* 2019;9:1027.
 246. Maiques-Diaz Alba, Somervaille Tim Cp. LSD1: biologic roles and therapeutic targeting. *Epigenomics.* 2016 Aug;8(8):1103–16.
 247. Hu Xin, Li Xingguo, Valverde Kristell, Fu Xueqi, Noguchi Constance, Qiu Yi, Huang Suming. LSD1-mediated epigenetic modification is required for TAL1 function and hematopoiesis. *Proc Natl Acad Sci.* 2009 Jun 23;106(25):10141–6.
 248. Huang Duen Yi, Kuo Yuan Yeh, Chang Zee Fen. GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells. *Nucleic Acids Res.* 2005;33(16):5331–42.
 249. Anastas Jamie N, Zee Barry M, Kalin Jay H, Kim Mirhee, Guo Robyn, Alexandrescu Sanda, et al. Re-programing Chromatin with a Bifunctional LSD1/HDAC Inhibitor Induces Therapeutic Differentiation in DIPG. *Cancer Cell.* 2019;36(5):528-544.e10.
 250. Ishikawa Yoshinori, Gamo Kanae, Yabuki Masato, Takagi Shinji, Toyoshima Kosei, Nakayama Kazuhide, et al. A Novel LSD1 Inhibitor T-3775440 Disrupts GFI1B-Containing Complex Leading to Transdifferentiation and Impaired Growth of AML Cells. *Mol Cancer Ther.* 2017 Feb 2;16(2):273–84.
 251. Wing Casey E, Fung Ho Yee Joyce, Chook Yuh Min. Karyopherin-mediated nucleocytoplasmic transport. *Nat Rev Mol Cell Biol.* 2022;23(5):307–28.
 252. Holt JE, Ly-Huynh JD, Efthymiadis A, Hime GR, Loveland KL, Jans DA. Regulation of Nuclear Import During Differentiation; The IMP alpha Gene Family and Spermatogenesis. *Curr Genomics.* 2007 Aug;8(5):323–34.
 253. Azmi Asfar S, Uddin Mohammed H, Mohammad Ramzi M. The nuclear export protein XPO1 — from biology to targeted therapy. *Nat Rev Clin Oncol.* 2021;18(3):152–69.
 254. Tan David SP, Bedard Philippe L, Kuruvilla John, Siu Lillian L, Razak Albiruni R Abdul.

- Promising SINEs for embargoing nuclear-cytoplasmic export as an anticancer strategy. *Cancer Discov.* 2014 May;4(5):527–37.
255. Lin Kevin H, Rutter Justine C, Xie Abigail, Killarney Shane T, Vaganay Camille, Benaksas Chaima, et al. P2RY2-AKT activation is a therapeutically actionable consequence of XPO1 inhibition in acute myeloid leukemia. *Nat Cancer.* 2022;3(7):837–51.
256. Ribeil Jean Antoine, Arlet Jean Benoit, Dussiot Michael, Cruz Moura Ivan, Courtois Geneviève, Hermine Olivier. Ineffective Erythropoiesis in β -Thalassemia. Al-Tonbary Y, El-Beshlawy A, Tricta F, Badr MA, editors. *Sci World J.* 2013;2013:394295.
257. Guillem Flavia, Dussiot Michaël, Colin Elia, Suriyun Thunwarat, Arlet Jean Benoit, Goudin Nicolas, et al. XPO1 regulates erythroid differentiation and is a new target for the treatment of β -thalassemia. *Haematologica.* 2019 Nov 21;105(9 SE-Articles):2240–9.
258. Çağatay Tolga, Chook Yuh Min. Karyopherins in cancer. *Curr Opin Cell Biol.* 2018 Jun.
259. Suzuki Takuji, Ishigami Yoko, Okada Norihisa, Kaneko Akihiro, Fukutomi Ryuuta, Isemura Mamoru. Differentiation-associated alteration in gene expression of importins and exportins in human leukemia HL-60 cells. *Biomed Res.* 2008 Jun;29(3):141–5.
260. Nachmias Boaz, Khan Dilshad H, Voisin Veronique, Mer Arvind S, Thomas Geethu Emily, Segev Nadav, et al. IPO11 regulates the nuclear import of BZW1 / 2 and is necessary for AML cells and stem cells. 2022;(January).
261. Ludwig Leif S, Lareau Caleb A, Bao Erik L, Liu Nan, Utsugisawa Taiju, Tseng Alex M, et al. Congenital anemia reveals distinct targeting mechanisms for master transcription factor GATA1. *Blood.* 2022;139(16):2534–46.
262. Jans David A, Martin Alexander J, Wagstaff Kylie M. Inhibitors of nuclear transport. *Curr Opin Cell Biol.* 2019 Jun;58:50–60.
263. Yasuhara Noriko, Yamagishi Ryosuke, Arai Yoshiyuki, Mehmood Rashid, Kimoto Chihiro, Fujita Toshiharu, Touma Kenichi, Kaneko Azumi, Kamikawa Yasunao, Moriyama Tetsuji, Yanagida Toshio, Kaneko Hiroki, Yoneda Yoshihiro. Importin Alpha Subtypes Determine Differential Transcription Factor Localization in Embryonic Stem Cells Maintenance. *Dev Cell.* 2013;26(2):123–35.
264. Weiss Mitchell J, dos Santos Camila O. Chaperoning erythropoiesis. *Blood.* 2009 Mar;113(10):2136–44.
265. Matte Alessandro, De Franceschi Lucia. Oxidation and erythropoiesis. *Curr Opin Hematol.* 2019;26(3).
266. Ling Hui, He Jie, Tan Hui, Yi Lan, Liu Fang, Ji Xiaoxia, Wu Youhua, Hu Haobin, Zeng Xi, Ai Xiaohong, Jiang Hao, Su Qi. Identification of potential targets for differentiation in human leukemia cells induced by diallyl disulfide. *Int J Oncol.* 2017;50(2):697–707.

267. Li Qingye, Tang Yuxian, Qin Jing, Yi Lan, Yang Yening, Wang Juan, He Jie, Su Qi, Tan Hui. Subcellular localization of DJ-1 in human HL-60 leukemia cells in response to diallyl disulfide treatment. *Mol Med Rep.* 2016;14(5):4666–72.
268. Guest Stephen T, Kratche Zachary R, Bollig-Fischer Aliccia, Haddad Ramsi, Ethier Stephen P. Two members of the TRiC chaperonin complex, CCT2 and TCP1 are essential for survival of breast cancer cells and are linked to driving oncogenes. *Exp Cell Res.* 2015 Mar;332(2):223–35.
269. Boudiaf-Benmammar Chafika, Cresteil Thierry, Melki Ronald. The Cytosolic Chaperonin CCT/TRiC and Cancer Cell Proliferation. *PLoS One.* 2013 Apr 16;8(4):e60895.
270. Dyer Jamie O, Dutta Arnob, Gogol Madelaine, Weake Vikki M, Dialynas George, Wu Xilan, Seidel Christopher, Zhang Ying, Florens Laurence, Washburn Michael P, Abmayr Susan M, Workman Jerry L. Myeloid Leukemia Factor Acts in a Chaperone Complex to Regulate Transcription Factor Stability and Gene Expression. *J Mol Biol.* 2017 Jun;429(13):2093–107.
271. Winteringham Louise Natalie, Kobelke Simon, Williams James Howard, Ingley Evan, Klinken Svend Peter. Myeloid Leukemia Factor 1 inhibits erythropoietin-induced differentiation, cell cycle exit and p27Kip1 accumulation. *Oncogene.* 2004;23(29):5105–9.
272. Bock Christoph, Datlinger Paul, Chardon Florence, Coelho Matthew A, Dong Matthew B, Lawson Keith A, et al. High-content CRISPR screening. *Nat Rev Methods Prim.* 2022;2(1):8.
273. Morgens David W, Deans Richard M, Li Amy, Bassik Michael C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol.* 2016 Jun;34(6):634–6.
274. Chung Jennifer E, Magis Wendy, Vu Jonathan, Heo Seok Jin, Wartiovaara Kirmo, Walters Mark C, Kurita Ryo, Nakamura Yukio, Boffelli Dario, Martin David IK, Corn Jacob E, DeWitt Mark A. CRISPR-Cas9 interrogation of a putative fetal globin repressor in human erythroid cells. *PLoS One.* 2019 Jan 15;14(1):e0208237.
275. Lan Xianjiang, Khandros Eugene, Huang Peng, Peslak Scott A, Bhardwaj Saurabh K, Grevet Jeremy D, et al. The E3 ligase adaptor molecule SPOP regulates fetal hemoglobin levels in adult erythroid cells. *Blood Adv.* 2019 May;3(10):1586–97.
276. Chang Yao Jen, Kang Zhifu, Bei Jiayuan, Chou Shu Jen, Lu Mei Yeh Jade, Su Yu Lun, Lin Sheng Wei, Wang Hsin Hui, Lin Steven, Chang Ching Jin. Generation of TRIM28 Knockout K562 Cells by CRISPR/Cas9 Genome Editing and Characterization of TRIM28-Regulated Gene Expression in Cell Proliferation and Hemoglobin Beta Subunits. *Int J Mol Sci.* 2022 Jun;23(12).
277. Michlits Georg, Jude Julian, Hinterndorfer Matthias, de Almeida Melanie, Vainorius Gintautas, Hubmann Maria, Neumann Tobias, Schleiffer Alexander, Burkard Thomas Rainer, Fellner Michaela, Gijbsbertsen Max, Traunbauer Anna, Zuber Johannes, Elling Ulrich. Multilayered VBC

- score predicts sgRNAs that efficiently generate loss-of-function alleles. *Nat Methods*. 2020;17(7):708–16.
278. Lozzio CB, Lozzio BB. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood*. 1975 Mar;45(3):321–34.
279. Martin P, Papayannopoulou T. HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. *Science*. 1982 Jun;216(4551):1233–5.
280. Chiba S, Takaku F, Tange T, Shibuya K, Misawa C, Sasaki K, Miyagawa K, Yazaki Y, Hirai H. Establishment and erythroid differentiation of a cytokine-dependent human leukemic cell line F-36: a parental line requiring granulocyte-macrophage colony-stimulating factor or interleukin-3, and a subline requiring erythropoietin. *Blood*. 1991 Nov;78(9):2261–8.
281. Vinjamur Divya S, Bauer Daniel E. Growing and Genetically Manipulating Human Umbilical Cord Blood-Derived Erythroid Progenitor (HUDEP) Cell Lines. *Methods Mol Biol*. 2018;1698:275–84.

7.5. Acknowledgements

There is so much to say in this chapter and words will probably not suffice. So many have been there for me during these past 5 years and I can't really express my appreciation. I am grateful that during this time, I could surround myself with incredible people, which always helped and supported me and made my journey a lot more fun!

I would especially like to thank my PhD committee **Jürg Schwaller**, **Markus Affolter** and **Alexandre Theocharides** for reading this "book". I truly appreciated all the feedback provided during my committee meetings. I'm in debt to you all for taking the time to give me great feedback and guide me during my PhD.

Jürg, it is hard for me to find words that describe how grateful I am. Thank you for being an incredible supervisor. You are a great inspiration for all of us, I admire your tremendous knowledge and your scientific passion. I am extremely thankful that you always found time for me, no matter what, even in your free time/holidays/weekends to mentor and discuss with me. Now, every time I attend a conference and see terrible PowerPoint slides, I think back to the time I started in your lab and the hours I spent realigning and remaking graphs (I for sure missed one or two in this thesis, please don't blame me). I always felt like you were exaggerating but now I totally get your point. Your pursuit of perfection challenged me but also brought out the best in me. I would probably call it "Zuckerbrot und Peitsche", and sometimes you really pushed me to my limits, but in the end, it was all worth it. Traveling through the states to visit labs made me realize that I have learned a ton and I would not have been where I am without you. A huge thank you for giving me the opportunity to do a PhD with you and visit other labs to learn new techniques. These research stays have also helped me so much in my scientific career (everyone is always impressed when I tell them that I've been to Paris, Amsterdam, and Vienna). I know that you like to be close and always want to know what's going on, so I can imagine that it was sometimes hard not having me around. I really admire how much you care and want to be involved and I appreciate that you gave me wings and freedom and trusted me that I would bring back the best I could while being abroad. I have learned so much from you and I will always be grateful for that.

Sabine, I feel like there should be a Sabine in every workplace. Without you, the CHLK lab would fall apart. You do so much for all of us daily and I don't think it can be appreciated enough. You are the one person that knows and does it all and you are way too humble for that. We owe you Sabine! Thank you for ordering stuff last minute, cleaning, and doing lab jobs that were not supposed to be done by you. Thank you for helping with experiments, troubleshooting and always being there for everyone and anything. You are a superwoman!

So many people have passed this lab in the last few years, but I would probably not have stayed if it wasn't for **Kata**. I still think back to your amazing supervision during my master's and I have learned a lot from you. Thank you! **Angeliki, Marwa, Maria, Frederik, Dario, Amber, Amandine, Marine** it was great to have you in the lab, thank you for the fun times! **Federica, Zika, Sarah** the lab can be happy to have you, you are genuine people! Thank you for the nice talks and fun moments. **Hugues-Etienne/Huggie**, we are in this together from the start and shared many up's and down's. I'm so excited that we both made it to the end without major discrepancies (obviously we all turned a little crazy but that's part of the game;)). I wish you all the best for your future career and please invite me to your spaceship. I'm a little sad that we are not going for after-work beers with **Neil** anymore, I had so much fun with you guys! **Rathick**, call me when you want, call me when you need, call me in the morning, I'll be on the way, call me when you want, call me when you need, call me out by your name, I'll be on the way. I could write a book about the reasons you are in this paragraph, sometimes I feel like you saved me from staying in academia. Supervising you gave me so much more than you can imagine. Sharing my passion with someone that is as excited as me about science is simply wonderful. I love that you are so caring, motivated, and fiery. I could not have had a better student. I trust you with all of my heart (which really means a lot for me). I know I can always rely on you and you will give your very best, no matter how much prosecco was involved before. I think we have a very similar work ethic and that's probably why working with you was so smooth and so much fun. I was so lucky to have you with me in Vienna to push through 250 targets in such a limited time. Multipipetting until we almost died. I'm really happy you decided to do a PhD and I can't wait to see how much you have grown when we see each other again in 5 years. Vienna: Why not? Please do not lose your spark, you are an inspiration for all the youngsters that partying and being professional are not mutually exclusive. "Wer feiern kann, kann auch arbeiten!" Thank you for always listening and being such a good friend. I will miss you, please come and visit me in quirky Portland, I'm sure you'll love it, maybe you'll even meet Britney there ;). #dreamteam #gatagang

So many people outside of Basel have made my PhD more worthy and enjoyable and I'm immensely grateful to have experienced working in so many different labs and meeting so many fantastic people all across Europe.

La vie est belle. First stop, Paris. Thank you, **Thomas Mercher**, for letting me do endless ChIPs in your lab and also for your great supervision during these past years. Thank you for always having an open ear and talking about my future and giving me the strength to push through, when I couldn't see the end. I was always very happy to see you at conferences and chat with you. Thank you for your amazing supervision and feedback that definitely drove the project forward. I'm truly sad that

you are not allowed to be in my PhD committee any longer. **Cecile T** thank you for helping me with all the pipetting and daily supervision, **Cecile L, Alexandre, Anouchka**, thank you for integrating me so quickly and showing me French pubs and fun moments in and outside of the lab.

Marieke von Lindern, you have been a wonderful teacher to me, thank you for spending hours with me in cell culture to take care of the primary erythroblasts. I appreciated your great suggestions and especially your honest talks about my future. I admire your dedication and passion for teaching and sharing your knowledge. Sanquin was such a great place to be that I did not want to leave. It was fantastic being able to speak Dutch again (even if according to Roos I have a Belgian accent). Thank you, **Roos, Francesca, Greta, Han, Carlijn, Arthur, David, Natasja, Thijs** for the great lab atmosphere, all the fun borrels and clubbing nights. I genuinely enjoyed spending time with you and once again you reminded me about how much I loved living and being in the Netherlands and that a part of my heart will always be there. No matter where life will take me, I know I can always come back and be truly happy. “Spring maar achterop bij mij, achterop me fiets, en ik weet nog niet waar we naartoe gaan samen maar dat boeit ook helemaal niets. “

Last but not least, a big thank you goes to the Grebien Lab in Vienna. **Florian Grebien**, you have been a fantastic supervisor and kind mentor. I highly valued your input and you have been a huge support during my final year (especially with writing a bunch of reference letters). I would not have come so far without your tremendous help. Thank you so much for giving me the opportunity to perform the CRISPR/Cas9 screen. It was not only successful but also a lot of fun being part of your lab; the team spirit was simply mind-blowing. FG-lab, thank you for all the fun times in and outside of the lab and for always helping and always listening. Big shout out to **Gabi, Manu** for organizing, ordering, calling the customs over and over again, and being always available when something was needed. **Lisi**, you are one of the sweetest and most helpful people in the world. Thank you for your brain, thinking through experiments, always helping out, and especially supplying Rathick with money :D I'm always thinking of you when eating Ribisel'n. Thank you, **Thomas**, for all the zoom and slack discussions, analyzing the enormous amount of data, and making beautiful heatmaps and scatter plots. Being back in Vienna for EHA2022 was a special treat and felt like coming home to you again, **Meli, Ludo, Selina, Bernhard, Chloe, Paula**. Spritzer! Because no good story ever began with someone drinking a Soda Zitron. You rocked my world and I can't wait to be back! **Marketa**, the girl that always comes at 6 am. I couldn't have wished for a better roomie to share the office space. You are one of the kindest people I have met during my journey and I'm so thankful that you were there for me when I needed it the most, no matter if it was for Punsch to go at “Zum schwarzen Kameel” or just going for walks outside. On a side note, I'm delighted that you lost your virginity to “Ben and

Jerry`s” because of me. **Ecem**, du hast das Sagen, I loved having you around, you have become a real friend. With you by my side, I never have to be scared. I know you will always stand up for me and fight all the dickheads in my life. I miss your beautiful face and voice ;) “And I need you now tonight. And I need you more than ever. And if you'll only hold me tight. We'll be holding on forever. And we'll only be making it right. Cause we'll never be wrong together. We can take it to the end of the line. Forever`s gonna start tonight.”

I would like to thank my besties that have always been there for me since we were kids, **Dütschi** and **Nathalie**. You both do not really know this side of me and I`m so excited to finally show you that I can be a different and almost grown-up person in the lab :P Piep kleine ö, always and all days, thank you for always finding time to vacay with me. “Sometimes, all I think about is youuuuu, late nights in the middle of June”. You definitely helped calming my thoughts and it was always easy to switch off my brain when exploring hidden gems in foreign countries with you. I`m so thankful for these times off to recharge my battery. Schnathalia, hihi, i know you hate it aber ich kann nicht anders ;) Auch wenn unsere Leben unterschiedlicher nicht sein könnten, wir haben es geschafft uns nicht zu verlieren; mehr als 25 Jahre schon musst du mich schon ertragen :P Du bist eine wahre Freundin und wenn wir uns wiedersehen ist alles immer so wie früher. Ich weiss, dass ich in dir eine Freundin fürs Leben gefunden habe und egal wohin es mich in die Welt verschlägt, du bist immer ein Teil meines Herzens.

Beste **Mami** der Welt, es fällt mir schwer für dich nur ansatzweise die passenden Worte zu finden. Du bist immer mein grösster Fan und ich kann dir gar nicht genug Danke dafür sagen. Danke, für alles was du mir ermöglich hast und wie du mich auf diesem Weg immer unterstützt hast. Ich kann mir nicht vorstellen, wie schwer es gewesen sein muss für dich. Erst Holland, dann Basel, Paris, Amsterdam, Wien und jetzt USA. Ganz egal wo ich hin wollte, du hast mich immer bestärkt meinem Herzen zu folgen. You gave me wings and made me fly. You said no star was out of reach. You stood by me and I stood tall. Auch wenn wir oft getrennt sind, im Herzen sind wir uns immer nah. Du spürst, wenn es mir schlecht geht, du kennst mich besser als jeder andere. Du bist mein Fels in der Brandung, wenn meine Welt auseinanderbricht. Du gibst mir das Gefühl, dass ich alle meine Träume erreichen kann und hast nie aufgehört an mich zu glauben. You were my strength, when I was weak. You were my voice, when I couldn`t speak. You were my eyes, when I couldn`t see. You saw the best there was in me. Lifted me up, when I couldn`t reach. You gave me faith because you believed. I am everything I am because you love me. Thank you for everything. My world is a better place because of you. I love you to the moon and back ♥

7.6. Appendices

Appendix 1

Fagnan, A., Piqué-Borràs, MR., Tauchmann, S., Mercher, T., Schwaller, J. Molecular Landscapes and Models of Acute Erythroleukemia. HemaSphere. 2021.

Appendix 2

Leonards, K., Almosaillekh, M., Tauchmann, S., et al. Nuclear interacting SET domain protein 1 inactivation impairs GATA1-regulated erythroid differentiation and causes erythroleukemia. Nature Communications. 2020.

Appendix 3

Tauchmann, S., Schwaller, J. NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer. Life. 2021.

Molecular Landscapes and Models of Acute Erythroleukemia

Alexandre Fagnan¹, Maria-Riera Piqué-Borràs², Samantha Tauchmann², Thomas Mercher¹, Juerg Schwaller²

Correspondence: Juerg Schwaller (j.schwaller@unibas.ch)

Abstract

Malignancies of the erythroid lineage are rare but aggressive diseases. Notably, the first insights into their biology emerged over half a century ago from avian and murine tumor viruses-induced erythroleukemia models providing the rationale for several transgenic mouse models that unraveled the transforming potential of signaling effectors and transcription factors in the erythroid lineage. More recently, genetic roadmaps have fueled efforts to establish models that are based on the epigenomic lesions observed in patients with erythroid malignancies. These models, together with often unexpected erythroid phenotypes in genetically modified mice, provided further insights into the molecular mechanisms of disease initiation and maintenance. Here, we review how the increasing knowledge of human erythroleukemia genetics combined with those from various mouse models indicate that the pathogenesis of the disease is based on the interplay between signaling mutations, impaired TP53 function, and altered chromatin organization. These alterations lead to aberrant activity of erythroid transcriptional master regulators like GATA1, indicating that erythroleukemia will most likely require combinatorial targeting for efficient therapeutic interventions.

Introduction

First described in 1917 by Giovanni Di Guglielmo, acute erythroleukemia (AEL) accounts for 1%–5% of cases with acute myeloid leukemia (AML) and is generally associated with a poor prognosis.¹ While most cases are identified in aged patients, AEL is occasionally also diagnosed in very young children. The cellular hallmark of AEL is impaired erythroid terminal differentiation and uncontrolled expansion of erythroid progenitor cells. AEL often presents with variable features that complicate the diagnosis which resulted in several changes in its classification over the years. The first French-American-British (FAB) system classified myeloid neoplasms with >30% leukemic blasts and ≥50% erythroid progenitor cells as AML-M6.² Some AEL patients present with a heterogeneous mixture of myeloid and erythroid features, while some less frequent cases present with >80% of erythroid progenitor cells considered as purely erythroid leukemia (PEL), respectively, called AML-M6a and AML-M6b in the WHO classification of 2008.³ Importantly, AEL patients may develop their disease *de novo*, but it frequently follows an antecedent myelodysplastic syndrome (MDS), myeloproliferative neoplasm (MPN), or therapeutic exposure to genotoxic agents. This suggests that the disease reflects a continuum of MDS and AML

with erythroid hyperplasia. The clinicopathological overlap and the diagnostic difficulties to distinguish MDS from AEL led the WHO in 2016 to reclassify cases previously diagnosed as AEL into either MDS or PEL.⁴ Multiple studies have shown that leukemic blasts from AEL patients often carry complex karyotypes with frequent loss of chromosomes 5 and 7.³ Only recently, deep sequencing studies have revealed a more extensive genetic landscape of AEL beyond the most prevalent mutations. However, as outlined in the following, it is important to note that the first insights into the biology of AEL emerged already over half a century ago, when researchers (accidentally) developed erythroleukemia models while studying the *in vivo* activities of avian and murine tumor viruses. Mechanistic studies aiming to understand these unusual phenotypes initiated several more rationale erythroleukemia models. More recently, the established epigenomic AEL roadmap allowed to establish models that are based on the distinct AEL-associated lesions. These models together with completely unexpected erythroleukemia-like phenotypes in genetically modified mice provide a wide experimental platform to elucidate the molecular mechanisms of AEL initiation and maintenance. Here, we discuss how the increasing knowledge of AEL genetics combined with often unexpected findings from various erythroleukemia models allowed to generate new hypotheses on the molecular mechanisms driving AEL that may provide keys for future targeted therapeutic interventions.

From complex karyotypes to epigenomic erythroleukemia landscapes

It has been recognized over 50 years ago that AEL leukemic blasts carry variable chromosomal abnormalities.^{6,7} Indeed, clonal chromosomal alterations are found in at least 75% of AEL patients and complex karyotypes were detected in at least 50% of patients.^{8,9} Complex or hypodiploid karyotypes were seen in at least 50% of cases with entire or partial monosomies of chromosome 5 and 7 being the most frequent.¹⁰

¹INSERM U1170, Equipe Labellisée Ligue Contre le Cancer, Gustave Roussy Institute, Université de Paris, Université Paris-Saclay, Villejuif, France

²University Children's Hospital beider Basel (UKBB), Department of Biomedicine, University of Basel, Basel, Switzerland.

Supplemental digital content is available for this article.

Copyright © 2021 the Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the European Hematology Association. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

HemaSphere (2021) 5:5(e558). <http://dx.doi.org/10.1097/H59.000000000000558>.

Received: 18 January 2021 / Accepted: 17 March 2021

Improved technologies have allowed the first targeted and then genome-wide sequencing of AEL patients samples, revealing recurrently mutated genes. Considering all published genetic studies, *TP53* is the most frequently mutated gene, identified in about 30% of the patients, associated with complex karyotype and poor outcome. Notably, *TP53* mutations have been identified in almost all patients with pure erythroid leukemia (PEL), thus representing a molecular hallmark of human erythroleukemia.¹¹⁻¹³ The other recurrent mutations target *NPM1*, epigenetic regulators (including *TET2* and *ASXL1* loss of function mutations, *DNMT3A*^{R882} or *IDH2*^{R140}), intermediates of signaling pathways as well as key hematopoietic transcription factors, which were also frequently identified in other AML subtypes.¹⁴⁻²¹ Signaling mutations have been reported in 25%–50% of AEL patients, including recurrent activating mutations of the JAK-STAT signaling (*JAK2*^{V617F}, *FLT3*^{ITD}, or *EPOR*) or the RAS (*NRAS*, *KRAS*, *PTPN11*, or *NF1*) pathways and are mostly associated with additional mutations such as *TP53* or *NPM1*. However, no single highly recurrent gene mutation has been reported to date supporting a high heterogeneity in the type of signaling mutation associated with AEL (Figure 1; Supplemental Digital Table 1, <http://links.lww.com/HS/A149>).

As the disease frequently develops secondary to MPN or MDS, known driver mutations of these disorders like the *BCR-ABL* fusion gene and activating mutations of *JAK2* are often associated with AEL.^{10,21,22} Notably, cytogenetic analysis of 75 AEL patients, 4 of 7 cases of pure erythroid leukemia were associated with a *BCR-ABL* fusion.²³ Single *BCR-ABL*-positive AML-M6 cases were reported to enter into long-term remission upon treatment with the imatinib tyrosine kinase inhibitor.²⁴ However, some cases of secondary AEL following a *JAK2*^{V617F}+ MPN were reported to lack the *JAK2* mutation in AEL blasts.²⁵ Of note, several of the human erythroleukemia cell lines carry one of these tyrosine kinase mutations (eg, K562: *BCR-ABL*, HEL: *JAK2*^{V617F}) suggesting that proliferative signals are essential to maintain these cells in vitro.²⁶ Interestingly, Iacobucci et al¹⁹ also highlighted age-related differences in the mutational profiles of AEL patients associated with distinct prognosis, including higher prevalence of *TP53* mutation in >60-year-old patients and higher representation of *NPM1*, *TET2*, or *DNMT3A* in 20- to 59-year-old patients.

The concept of clonal hematopoiesis of indeterminate potential (CHIP) is now well established with the observation that several mutations are particularly associated with CHIP and can predispose to the development of hematopoietic malignancies, as well as other human diseases.^{27,28} Increasing evidence suggests that AEL represents the evolution of a continuum between normal hematopoiesis, CHIP, myeloid neoplasm, and acute leukemia. Indeed, recurrent mutations identified in CHIP including *TP53*, *DNMT3A*, and *TET2* are also frequently found in human erythroleukemia. Genetic analyses of AEL shows that several mutations often present with low-allelic frequency (eg, *ASXL1*, *PTPN11*, and *WT1*) evoking their post-MPN, -MDS, -CHIP origin. In contrast, *TP53* mutations in AEL are characterized by a high-allelic frequency in leukemic cells (often with evidence of bi-allelic inactivation) and are also often identified in other cell populations than the leukemic cells (ie, T-cell compartment), supporting the idea that they arise at an early time point during disease development in an immature multipotent hematopoietic progenitor. Together, these results support an early acquisition of genetic lesions (including those associated with CHIP) and acquisition of additional sometimes subclonal alterations compatible with the emergence of AEL directly or indirectly from a CHIP situation. However, this hypothesis will need to be formally proven.

Although rare, AEL can affect pediatric patients. Interestingly, some recurrent translocations have been described in pediatric AEL, which so far have never been associated with other AML forms including t(1;16)(p31;q24) and t(11;20)(p11;q11) leading to expression of *NFIA-ETO2* and *ZMYND8-RELA* fusions, respectively.²⁹⁻³³ Very recently, a novel t(1;8)(p31;q21) translocation leading to the expression of an *NFIA-ETO* fusion was described in an infant with PEL presenting as erythroblastic sarcoma.³⁴ Fusions targeting the nucleopore components like *NUP98* or *NUP214*, like *NUP98-NSD1*, *NUP98-KDM5A*, or *DEK-NUP214* described in other AML types have also been found in pediatric AEL.³⁵ In addition, several less frequent fusions were identified, involving erythroid-associated factors (*MYB-GATA1* and *APLP2-EPOR*), epigenetic factors (*ZEB1-KDM4C* and *SMARCA4-CBS*), or signaling pathways like *ASNS-PTPN1*, *SRC-VWC2*, *RUNX2-STAT3* and *PRKAR2B-PIK3C*, or *PCM1-JAK2*.^{15,19} Of note, most of these fusions are individually very rare, suggesting that many different genetic

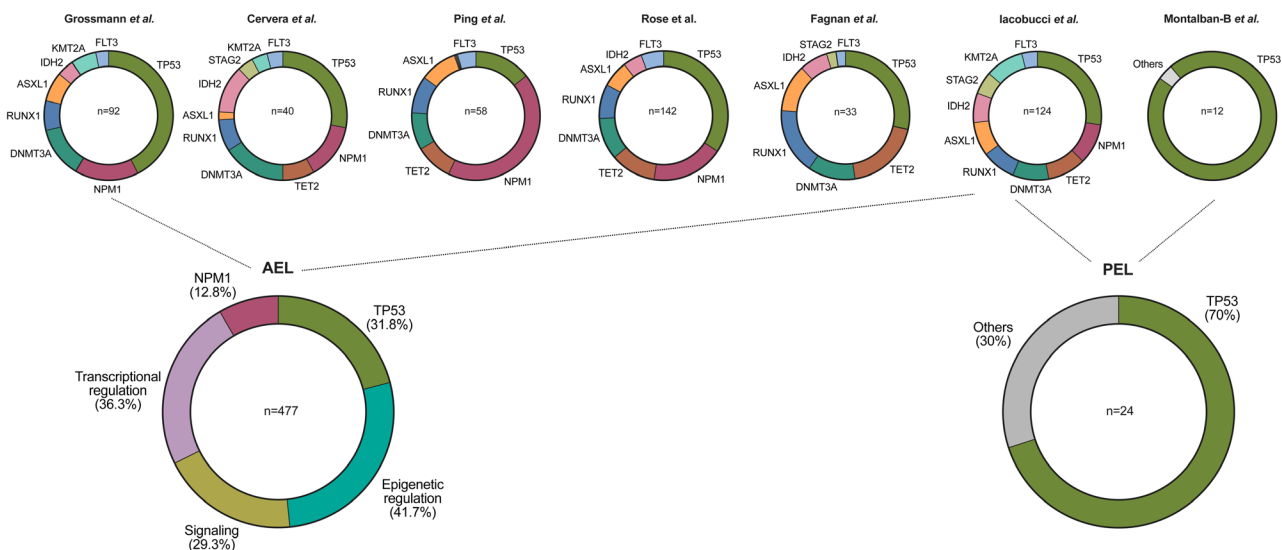


Figure 1. Genetic landscape of human AEL. Schematic representation of the most prevalent genetic lesions reported in human AEL patients.^{14,15,17,19,21} Alterations of the *TP53* tumor suppressor were found in $\geq 30\%$ of AEL and in the majority PEL patients. Note: only the studies by Iacobucci et al and Montalban-Bravo et al included patients with leukemia diagnosed as PEL (according to the WHO 2016 classification). AEL = acute erythroleukemia; PEL = purely erythroid leukemia.

alterations can lead to childhood AEL. Taken together, these studies proposed a molecular classification of AEL patients according to genetic and transcriptomic landscapes that could be associated with age at diagnosis and distinct clinical outcomes.

From tumor viruses to rationale erythroleukemia models

The first erythroleukemia models emerged over half a century ago mostly from phenotypes induced by tumor viruses. These models set the stage for more rational models exploring the transforming potential of signaling mediators and transcription factors in cells of the erythroid lineage. In addition to more recent models that explored the transforming activity of genetic alterations found in AEL patients, unexpected erythroleukemia phenotypes in genetically modified mice provided insight into epigenetic regulation of erythroid differentiation (Figure 2).

Tumor virus-induced erythroleukemia models

Avian leukemia viruses

The first hematopoietic malignancies-inducing oncogenes have been discovered by studying the *in vivo* activity of avian viruses. The avian erythroblastosis retrovirus (AEV), encoding for viral oncogenic variants *v-ErbA* and *v-ErbB* of cellular genes (respectively, the thyroid hormone receptor alpha [TR α] and a mutated epithelial growth receptor) was found to induce fatal erythroleukemia in young chicken. AEV blocks terminal differentiation of committed erythroid progenitor cells. Further functional studies led to the hypothesis that *v-ErbA* cooperates with activated cellular stem cell factor receptor Kit and *v-ErbB* to efficiently arrest terminal erythroid maturation.^{36,37} Molecular studies delineated a more general mechanism of oncogenesis based on the inability of altered nuclear receptors to efficiently respond to physiological concentrations of ligands, which was also shown to be the driving force of other AML forms such as

acute promyelocytic leukemia mediated by retinoid acid receptor alpha (RARA) fusion proteins.³⁸

The E26 avian retrovirus, which induces massive erythroblastosis in newborn chicken, encodes for a fusion between a portion of the viral *gag* sequences to truncated mutated forms of the transcription factors MYB and ETS1.³⁹ Functional studies revealed that *v-ets* is required for the E26-mediated block erythroid differentiation. Interestingly, the *Myb-Ets* fusion protein seems to inhibit *v-ErbA* and RARA, indicating overlapping pathways of malignant transformation by E26 and AEV.^{40,41}

In 1957, Friend⁴² reported that intraperitoneal injection of cell-free extracts prepared from ascites of mice inoculated with Ehrlich's carcinoma cells induced a leukemia-like disease. Electron microscopic analysis indicated that Ehrlich's cells, derived from a spontaneous mouse mammary adenocarcinoma, contained particles similar to what has been seen in virus-infected cells.⁴² Intraperitoneal injections of spleen cell suspensions or filtrates into Swiss albino mice resulted in signs of disease in >80% of the recipients. Affected mice had significant infiltrations in hematopoietic organs by cells that looked like erythroid progenitor cells. Subsequent studies suggested that erythropoietin (EPO)-sensitive erythroid progenitors and in particular late burst forming unit or colony forming unit-erythroid (CFU-E) are the targets of the Friend leukemia virus. Permanent cell lines could be established called Friend tumor cells or murine erythroleukemia cells (MELs). Notably, the observation that particular chemicals (eg, dimethyl sulfoxide [DMSO]) are able to induce partial terminal erythroid differentiation made these cells one of the most widely used *in vitro* platform to study erythroid maturation.⁴³ Similar to Friend's, other viruses like the Rauscher or Graffi MuLV were shown to induce an erythroblastosis that phenocopied human erythroleukemia in mice.^{44,45}

Friend virus contains 2 components, the replication-competent Friend murine leukemia virus acting as a helper for the replication-defective spleen focus forming virus (SFFV), which is the erythroblastosis-inducing component. The pathogenic activity is mediated by an *env*-derived viral 55 kDa glycoprotein (gp55) that directly interacts with and activates the EPO

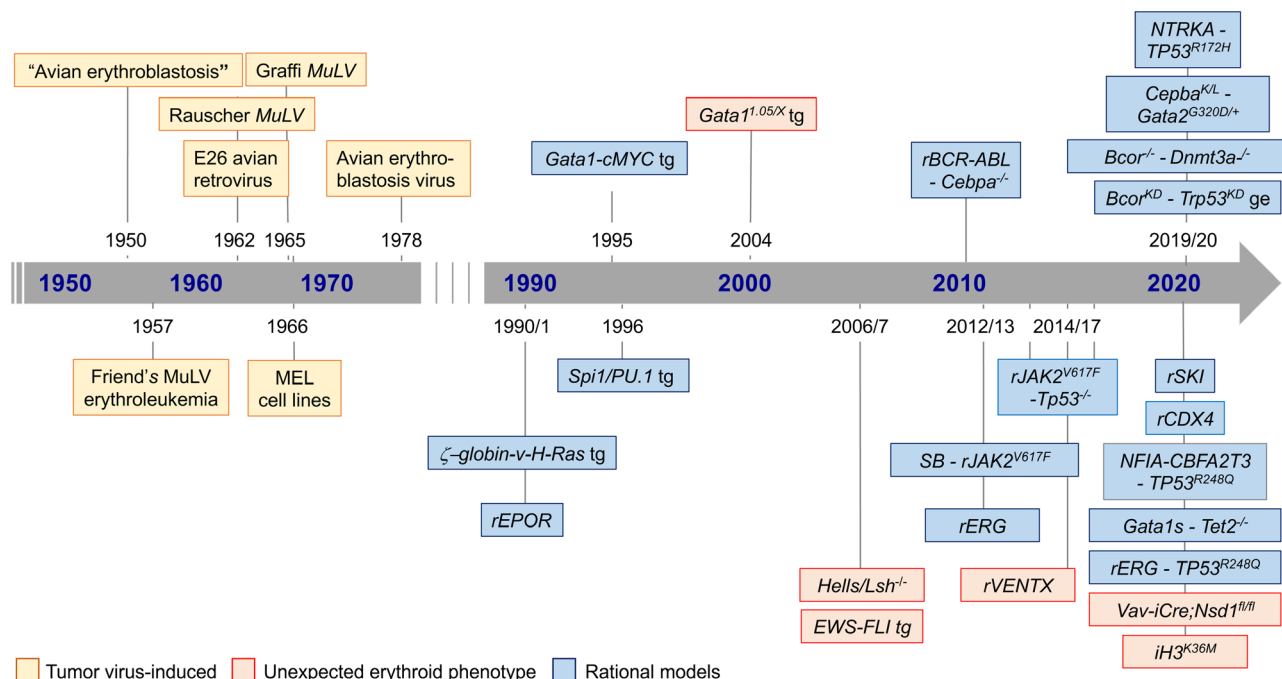


Figure 2. Chronology of AEL mouse models. Schematic timeline of tumor virus-induced, unexpected, and rational erythroleukemia mouse models. r = viral overexpression; tg = transgenic; ge = genome editing; kd = knock-down; MuLV = murine leukemia virus; AEL = acute erythroleukemia.

receptor (EPOR) promoting EPO-independent proliferation and differentiation. Recruitment of a cellular receptor tyrosine kinase receptor *stk/RON* by *gp55* results in activation of downstream signaling effectors including signal activators of transcription (STATs), PI3K/AKT, or MAP kinases.⁴⁶ Proviral integration cloning revealed that Friend virus integrated almost exclusively in a site called *SFFV-proviral integration site-1 (Spi-1)*, which resulted in transcriptional activation of the *Spi-1* gene locus by the viral LTR enhancers and in overexpression of the *Spi-1* mRNA.⁴⁷

Rationale erythroleukemia mouse models

Erythroleukemia by overexpression of a *Spi-1* transgene

To model the biological activity of aberrant *Spi-1* expression, Françoise Moreau-Gachelin and co-workers established a transgenic mouse model in which a *Spi-1* mini-gene was expressed under the control of the *SFFV-LTR* (Table 1). During an observation period of 12 months, 50% of *Spi-1* transgenic mice developed hepatosplenomegaly with extensive erythroblast infiltration and occasional tumor cells on the peripheral blood smears. Cells from diseased mice could be grown *ex vivo* as EPO-dependent cell lines but did not induce the disease upon transplantation, indicating that ectopic *Spi-1* expression blocks

erythroid differentiation but does not overcome growth factor requirement for survival.⁵⁰ Under hypertransfusion stress, tumor cells (referred as “HS-2 cells”) emerged that were able to proliferate independent of EPO and induce the disease in immunodeficient mice. HS-2 cells carried mutations of the Kit receptor tyrosine kinase leading to constitutive activation of PI3K/AKT and MAPK signaling,⁵⁸ representing a good example of acquired mutation in signaling molecules as a cooperative mechanism contributing to differentiation blockage in cancer. Notably, erythroblasts from the late stage FLV-induced disease harbored allelic losses or missense *Trp53* mutations.⁵⁹ Loss of *Trp53* alleles also increased penetrance and reduced the latency of erythroleukemia in *Spi-1* transgenic mice.⁶⁰

Erythroleukemia by constitutive EPOR activation

SFFV-encoded *gp55* glycoprotein binds and activates the EPOR bypassing the cellular requirement for EPO and supporting proliferation and survival of infected erythroid cells.⁶¹ To address the leukemogenic potential of a constitutively active EPOR, Longmore and Lodish⁴⁸ generated an SFFV in which they replaced *gp55* with a mutated EPOR. Injection of this modified SFFV-induced polycythemia and splenomegaly in mice.⁴⁸ Transfer of growth-factor-independent erythroblasts isolated from the spleens of these mice rapidly induced an

Table 1.

Rational Genetic AEL Mouse Models

Year	Gene	Model	Phenotype	Major Findings	Surface Markers on Leukemic Cells	References
1990	<i>EPOR</i>	<i>SFFV</i> ORF transduction.	Erythroleukemia	Polycythemia, splenomegaly	n.d.	48
1995	<i>c-MYC</i>	Transgene under control of <i>Gata1</i> enhancer/promoter	Early onset erythroleukemia, mice died before reaching sexual maturity	Splenomegaly, erythroid progenitors in peripheral blood, tumor cell infiltration, severe anemia, and moderate thrombocytopenia	n.d.	49
1996	<i>Spi-1</i>	Classical transgene (“mini gene”)	50% homozygous mice developed a multi-step erythroleukemia within 1.5 to 6 mo of birth. Transplantable into nude recipients	Hepatosplenomegaly with erythroblast infiltration	Increase Ter119 ⁺ , reduced B220 ⁺ , CD4 ⁺ , CD8 ⁺ , Mac1 ⁺ , Gr1 ⁺ , Sca1 ⁺	50
1998	<i>H-Ras</i>	Transgene under control of <i>Zeta-globin</i> enhancer/promoter mouse	Mesenchymal and epithelial neoplasms, <5% erythroleukemia	Mesenchymal and epithelial neoplasms, <5% showed hepatosplenomegaly with erythroblast infiltration	Ter119 ⁺	51
2004	<i>GATA1^{1.05/x}</i>	Transgenic mouse, insertion of <i>Neo</i> cassette before <i>Gata1</i> -GIE region	50% of <i>GATA1^{1.05/x}</i> mice developed disease: 2 phenotypes: myeloid disease after 143 d and lymphoid disease after median latency of 387 d. Transplantable into nude mice	Anemia, thrombocytopenia, erythroblasts, and megakaryocytes in spleens of mice with myeloid disease	Myeloid disease: Kit ⁺ , CD71 ⁺ , Ter119 ^{-/dull} , CD19 Lymphoid disease: Kit ⁻ , Sca1 ⁺ /CD43 ⁺ /CD19 ⁺	52
2007	<i>EWS-FLI</i>	Inducible transgene in <i>Rosa26</i> , activated by <i>Mx1-iCre</i> (plpC)	Rapid, highly penetrant (90%–100%) erythroleukemia (+plpC: 19 d, –plpC: 95 d). Transplantable in 19/26 sublethally irradiated wild-type and 7/7 NOD/SCID recipients	Anemia, peripheral blood blasts, no thrombocytopenia, hepatosplenomegaly with infiltration of tumor cells	Kit ⁺ , CD43 ⁺ , CD71 ⁺ . Many cells Gata1 ⁺ with increased c-Myc expression	53
2012	<i>ERG</i>	<i>MSCV</i> viral overexpression and BM reconstitution	Erythro-megakaryoblastic leukemia, T-cell acute lymphoblastic leukemia. Transplantable	Hematopoietic malignancies affecting erythroid, megakaryocytic, and T-cell lineage	CD71 ⁺ /Ter119 ⁺ , CD4 ⁺ /CD8 ⁺	54
2012	<i>ERG</i>	<i>MSCV</i> viral overexpression and BM reconstitution	Lymphoid leukemia, erythroid-megakaryocytic leukemia. Transplantable (30/32)	Accumulation of immature erythroblasts <i>in vivo</i> , cell clones exhibited both erythroid and megakaryocytic differentiation <i>in vitro</i>	CD4 ⁺ /CD8 ⁺ CD71 ⁺ /Ter119 ⁺ , CD71 ⁺ /CD41 ⁺	55
2016	<i>VENTX</i>	<i>MSCV</i> viral overexpression and BM reconstitution	Erythroleukemia. Transplantable (13/19)	Infiltrations of BM and spleen by erythroblasts after long latency	Primary: Mac-1 ⁺ , Mac-1/Gr-1 ⁺ , Kit ^{low} , CD71 ⁺ , Ter119 ⁻ Secondary: Ter119 ⁺ , partly CD71 ⁺	56
2019	<i>CDX4</i>	<i>MSCV</i> viral overexpression and BM reconstitution	Erythroleukemia-like diseases after a long latency. Transplantable	Anemia, splenomegaly, infiltration of erythroid progenitors. Tumor cells expressed low levels of Gata1	CD71 ⁺ /Ter119 ^{+/-}	57
2020	<i>SKI</i>	<i>MSCV</i> viral overexpression and BM reconstitution	Leukemia of erythroid and myeloid phenotype	Pan-cytopenia associated with accumulation of erythroid and myeloid progenitors in BM, spleen and liver	CD71 ⁺ /Ter119 ^{+/-} CD11b ⁺ /Gr-1 ⁺	21

AEL = acute erythroleukemia; BM = bone marrow; MSCV = murine stem cell virus.

erythroleukemia-like disease in the recipients. Of note, tumor cells did neither secrete a pathogenic virus nor did they integrated into *Spi-1*, but they carried inactivating *Trp53* rearrangements. Aberrant EPO expression was also found in serially propagated FLV erythroleukemia cell lines, due to genomic rearrangements independent of retroviral integration.⁶²

Erythroleukemia by targeted expression of master oncogenes

In vitro studies showed that DMSO-induced MEL cell erythroid differentiation is associated with reduced expression of the *MYC* proto-oncogene and that its overexpression inhibited differentiation.^{63,64} To study the in vivo transforming potential of *MYC* in the erythroid lineage, Phil Leder and co-workers used regulatory sequences potentially controlling the expression of the erythroid master regulator GATA1 in transgenic mice.⁴⁹ Diseased mice presented with splenomegaly with significant tumor cell infiltration and erythroid progenitors in the peripheral blood. Tumor cells showed clonogenic activity in methylcellulose (MC) without growth factors and induced the same disease phenotype when transplanted. Furthermore, tumor cells expressed erythroid genes (*EPOR*, *globin*), but, unlike MEL cells, exposure to DMSO did not induce terminal differentiation. These observations suggest that aberrant *MYC* activation at a particular vulnerable phase of erythroid differentiation is most likely sufficient to induce erythroleukemia.

In addition to *MYC*, increased levels of the *H-Ras* and *K-Ras* oncogenes were found in Friend's murine erythroleukemia.⁶⁵ Leder and co-workers established another series of transgenic mice in which the embryonic alpha-like zeta-globin gene was driving expression of an activated *H-Ras* oncogene. Unexpectedly, these transgenic mice ("Tg.AC") developed multiple mesenchymal and epithelial neoplasms and only few mice (<5%) showed hepatosplenomegaly with erythroblast infiltration.⁵¹ Impaired EPO-induced differentiation of an FLV-induced erythroleukemia cell line (SKT6) by a constitutively active *H-Ras*^{G12V} mutant also suggested that aberrant RAS activation can enhance erythroid transformation.⁶⁶

Erythroleukemia by aberrant activity of erythroid transcriptional regulators

GATA binding protein 1

The *GATA1* gene on chromosome X encodes for a zinc-finger transcription factor expressed in erythroid, megakaryocytic, eosinophilic, and mast hematopoietic cells as well as in Sertoli cells of the testes.⁶⁷ As the generation of germline *Gata1*-null alleles resulted in embryonic lethality, Yamamoto and colleagues aimed to alter *Gata1* expression by inserting a *neomycin* selection cassette in the promoter region between the so-called double GATA sequence and the erythroid-specific exon 1 (IE). Whereas hemizygous mutant male embryos died in utero due to impaired primitive erythropoiesis, heterozygous female mice survived due to random inactivation of the X-chromosome. As these mice expressed about 5% of *Gata1* transcripts, this targeted mutation was referred as the *Gata1*^{1.05X} allele. While the decreased number of erythroid cells and CFU-E formation was observed in fetal livers, accumulation of primitive erythroid progenitors was observed as early as E9.5 in mutant mice.⁶⁸ Heterozygous *Gata1*^{1.05X} female mice developed signs of distress at the age of 5 months presenting with anemia, thrombocytopenia with massive accumulation of erythroblasts and megakaryocytes in their spleens.⁶⁹ Detailed analysis of a larger cohort of heterozygous female *Gata1*^{1.05X} mice revealed a high incidence (~50% penetrance) of leukemia composed of either Kit⁺ erythroid blasts (starting at 143 d) or CD19⁺ lymphoid blasts (starting at 387 d). Tumor cells were Kit⁺/CD71⁺/Ter119^{-dull}/CD19⁻ proerythroblasts. Tracking of the cells with a GATA1-controlled fluorescent reporter suggested that immature erythroid cells were

already expanding in the hematopoietic organs of *GATA1*^{1.05X} mice at the late embryonic stages.⁵²

Several models indicate that GATA1 activity tightly controls the balance between proliferative erythroid progenitors and maturing cells. First, in the *Gata1*^{1.05X} model, leukemia development is completely abolished by transgenic expression of wild-type *Gata1*. A cell line ("GAK-14") was established from *Gata1*^{1.05X} diseased mice that maintained an immature erythroblastic phenotype (CD71⁺/Kit⁺/Ter119⁻) when grown on OP9 stroma cells in the presence of EPO and stem cell factor. Retroviral overexpression of *Gata1* resulted in GAK-14 differentiation into mature erythroid cells when cultured on fetal liver-derived stroma cells.⁷⁰ Similarly, through expression of the apoptosis inhibitor BCL2 into *Gata1*-deficient embryonic stem cells followed by in vitro erythroid differentiation, Weiss et al⁷¹ generated a stable erythroblastic cell line ("G1E"). Exogenous *Gata1* expression in G1E cells restored erythroid maturation⁷¹ and allowed to functionally dissect *Gata1* critical domains, post-translational modifications and target genes.⁷²⁻⁷⁴ Collectively, the observations suggested that impaired GATA1 activity is an important feature for induction and most likely also maintenance of transformed murine erythroid progenitor cells.⁷⁵

ETS transcription factor ERG

ERG is a member of the E26 transformation-specific family of transcription factors that contain a highly conserved ETS DNA binding domain that interacts together with other transcription factors to enhancer elements.⁷⁶ Functional studies in mice revealed that *ERG* expression promotes HSC maintenance but also controls erythromegakaryocytic differentiation.⁷⁷⁻⁷⁹ ERG was found to bind together with GATA1 to regulatory elements of key hematopoietic transcription factors like *SCL/TAL1*.⁸⁰ The *ERG* gene is targeted by several chromosomal translocations associated with AML but also solid tumors.⁸¹ High *ERG* expression levels have been associated with poor prognosis in cytogenetically normal AML.⁸² In addition, increased *ERG* gene dosage seems to cooperate with the N-terminal GATA1 mutation (GATA1s) in the transient myeloproliferative disorder associated with Down's syndrome.⁸³ *ERG* is not only highly expressed in trisomy 21-related but also in sporadic cases of acute megakaryoblastic leukemia (AMKL). Increased *ERG* expression was shown to promote in vitro megakaryopoiesis and synergize with GATA1 to immortalize hematopoietic progenitor cells.⁸⁴

Several groups explored the oncogenic potential of increased *ERG* expression levels in the hematopoietic system of the mouse. Brady and colleagues reported that transplantation of fetal liver-derived murine hematopoietic stem and progenitor cells (HSPC) retrovirally overexpressing a human *ERG* ORF into sublethally irradiated mice resulted in fully penetrant megakaryoblastic leukemia.⁵⁴ Tsuzuki and Seto⁵⁵ found that transplantation of adult bone marrow (BM) cells from 5-FU-stimulated donor mice retrovirally expressing a human *ERG* ORF into lethally irradiated mice induced a leukemia-like disease characterized by accumulation of CD71⁺/Ter119⁺ erythroblasts and expansion of CD4⁺/CD8⁺ double positive T cells. Kile and colleagues reported that transplantation of fetal liver-derived or adult mouse bone marrow (BM) (from 5-FU treated donors) retrovirally expressing a murine *Erg* ORF into irradiated mice induced a leukemia-like disease. Similar to the observation by Seto, some mice developed CD4⁺/CD8⁺ T-cell leukemia, others developed nonlymphoid disease composed of CD71⁺/Ter119^{+/−} cells in some, but also CD71⁺/CD41⁺ cells in other mice.⁸⁵ Collectively, these studies suggest that abnormally high *ERG* expression contributes to hematopoietic malignancies affecting the erythromegakaryoblastic and T-cell lineage.

The Sleeping Beauty (SB) transposon-based mutagenesis system allows to identify potentially cooperating genetic lesions for cancer development. Targeting a conditional SB allele to

the hematopoietic system in mice expressing the constitutively active *JAK2^{V617F}* resulted in a strong phenotypic selection for an erythroleukemia-like disease.⁸⁶ The vast majority of *SB/JAK2^{V617F}* mice developed an aggressive erythroleukemia occasionally coincident with CD4⁺/CD8⁺ T-cell ALL. Interestingly, the most prevalent common transposon insertion sites were the genes encoding for the transcription factors ERG and ETS1. Notably, transplantation of fetal liver-derived HSPC retrovirally expressing an AML-associated *TLS-ERG* fusion also induced a very similar erythroleukemia as observed in the *SB/JAK2^{V617F}* mice. Expression of *TLS-ERG* in SB mice resulted in acceleration of the disease. Interestingly, the *Jak2* gene locus was among the most common CIS in this model furthermore underlining cooperation of ERG and constitutively active JAK2 in murine erythroleukemia.⁸⁶

Caudal-type homeobox 4

The caudal-type homeobox family comprises CDX1, CDX2, and CDX4 known as developmental regulators of the clustered HOX homeobox genes.⁸⁷ In normal hematopoiesis, CDX4 mirrors *HOX* gene expression with a peak in hematopoietic stem cell and decreasing upon differentiation.⁸⁸ Ectopic *Cdx4* expression in mouse embryonic stem (ES) increased the hematopoietic colony output associated with upregulation of a *Hox* gene expression.⁸⁹ However, *Cdx4* gene inactivation only minimally affected adult hematopoiesis in mice.⁹⁰ Retroviral *Cdx4* overexpression induced aberrant self-renewal potential in mouse hematopoietic cells in vitro and transplantation induced an AML-like disease in about 50% of mice.⁹¹

Feuring-Buske and colleagues recently reported *Cdx4* mRNA expression in a small cohort of AEL patients and in 3 established AML cell lines with an erythroid phenotype. Similar to previous studies, retroviral *Cdx4* overexpression provided aberrant serial replating potential to BM-derived hematopoietic cells.^{57,91} Notably, mice transplanted with *Cdx4* virally transduced BM-derived HSPC developed a transplantable erythroleukemia-like disease after a long latency, characterized by anemia, splenomegaly with infiltration of CD71⁺Ter119⁺ erythroid progenitors, occasionally erythroid progenitors in the periphery, and multiorgan infiltrations upon propagation into secondary recipients. Tumor cells were characterized by low expression levels of genes associated with erythroid specification or differentiation including *Gata1*. Interestingly, leukemic blasts from diseased mice carried some additional mutations in erythroid transcription factors like GATA1 or GATA2. These observations indicate that aberrant *Cdx4* expression levels in a permissive progenitor may induce a transcriptional program that interferes with normal erythroid development. However, the direct relevance for the human disease remains unclear, as the transcriptome analysis of large AEL patient cohorts did not highlight *Cdx4* alterations.^{19,21}

EWS-FLI fusion

The *Fli-1* gene encoding for an ETS-transcription factor was identified as an additional common integration site in FLV-induced erythroleukemia.^{92,93} In addition to FLV-induced mouse erythroleukemia, *Fli-1* mRNA expression was found in some human AML cells lines with erythroid phenotypes.⁹⁴ Experimental *Fli-1* overexpression was shown to reduce the expression of GATA1 and to impair induced erythroid differentiation in human and mouse cell lines.⁹⁵ Apart from FLV-driven mouse erythroleukemia, *FLI-1* is better known as fusion partner to *EWSR1* (EWS) as consequence of a t(11;22)(q24;q12) chromosomal translocation found in Ewing's sarcoma and other neuroectodermal tumors.⁹⁶ Interestingly, *Mx1-iCre*-controlled hematopoietic expression of a transgenic *EWS-FLI-1* ORF in the *Rosa26* murine gene locus rapidly resulted in a highly penetrant aggressive transplantable erythroleukemia with tumor cells expressing Kit, CD71, CD43, and Gata1, but no Ter-119

or other lineage markers. Leukemic cells expressed high levels of *Myc* but did not harbor any gross chromosomal or *Trp53* alterations.⁵³ Although the erythroleukemia phenotype does not match the human disease associated with this fusion, these transgenic mice provided an in vivo platform to study strategies for therapeutic targeting of *EWS-FLI1*-driven tumors.⁹⁷ Notably, several compounds were found that inhibit Fli-1 transcriptional activities and impaired *EWS-FLI1*-driven erythroleukemia in mice. Their detailed mode of action and clinical value for human erythroleukemia remains to be elucidated.^{97,98}

Erythroleukemia by cooperating genetic lesions *BCR/ABL* and loss of *C/EBP α*

Earlier work suggested that AML is the product of signaling mutations (eg, in tyrosine kinases) supporting proliferation and survival that functionally cooperate with mutations in hematopoietic transcription factor mutations blocking differentiation (Table 2).¹⁰⁴ BCR-ABL is a constitutively active tyrosine kinase fusion associated with chronic myeloid leukemia (CML) and is also recurrently found in AEL.¹⁰ Tenen and colleagues developed a model of progression from chronic to acute disease by retrovirally expressing BCR-ABL in fetal liver-derived HSPCs lacking the myeloid transcription factor CCAAT-enhancer binding protein (*C/EBP α*).⁹⁹ Mice transplanted with *BCR-ABL*-expressing *Cebpa*^{-/-} cells developed acute erythroleukemia with infiltration of BM and spleens and erythroblasts on peripheral blood smears. Furthermore, tumor cells expressed erythroid regulators such as SCL/TAL1 and GATA1. Notably, similar to tumor cells from diseased *BCR-ABL* transduced *Cebpa*^{-/-} mice, the human erythroleukemia cell line K562 (established from a patient with CML in blast crisis) also expresses BCR-ABL and GATA1 and lacks *C/EBP α* . Functional studies with *Cebpa*^{-/-} fetal liver progenitors revealed that *C/EBP α* functions in hematopoietic cell fate decisions by the dual actions of inhibiting erythroid and inducing myeloid gene expression.¹⁰⁵

GATA2 and *C/EBP α* mutations

Genetic alterations of the *C/EBP α* and GATA2 transcription factors regulating myeloid differentiation and HSC self-renewal have been reported in some AEL patients.^{17,19} A recent mouse model demonstrated that bi-allelic *Cebpa* mutations led to myeloid leukemia development and that an additional *Gata2* mutation enhanced leukemogenesis with a subset of triple transgenic mice (40%) developing leukemia with erythroid and myeloid features.¹⁰¹ Interestingly, the identified leukemia-initiating cells in both models were neutrophil-monocyte progenitors and molecular characterization of this population revealed distinct function of each cooperating mutations. While bi-allelic *Cebpa* mutations increased expression of erythroid genes, the *Gata2* mutation increased chromatin accessibility at erythroid TF motifs (eg, GATA1, ZFP1, and KLF1) and decreased it at myeloid TF motifs. These findings suggested that the erythroid phenotype of this leukemia model is driven by the aberrant chromatin accessibility at key erythroid TF-regulated loci, controlled by aberrant GATA2 activity.

JAK2^{V617F} and loss of *Trp53*

The majority of BCR-ABL-negative MPNs are driven by somatic activating mutations of the JAK2 tyrosine kinase of which V617F is the most prevalent. A subset of MPN patients progress to AML that is associated with recurrent somatic alterations affecting epigenetic regulators, splicing-related factors and/or the TP53 tumor suppressor.¹⁰⁶ To demonstrate potential cooperation, researchers retrovirally overexpressed *JAK2^{V617F}* in either wildtype or *Trp53*^{-/-} BM-derived HSPCs and transplanted them into irradiated wildtype recipients.^{23,100} Mice developed a serially transplantable leukemic phenotype with hepatosplenomegaly with infiltration of CD71⁺/Ter119⁻ erythroid progenitor cells. Kurokawa and colleagues

Table 2.

Rational AEL Mouse Models Based on Genetic Cooperation

Year	Gene	Model	Phenotype	Major Findings	Surface Markers on Leukemic Cells	Reference
2010	<i>BCR-ABL—Cebpa</i> ^{-/-}	MSCV viral overexpression and BM reconstitution	Erythroleukemia after 26–157 d, 100% penetrance. Transplantable (5/6 mice)	Splenomegaly, erythroblasts on PB smears. Tumor cells expressed SCL/TAL1	CD71 ⁺ , Ter119 ⁻	99
2013	<i>Sleeping beauty—JAK2</i> ^{V617F}	SB insertion tagging, MSCV viral overexpression and BM reconstitution	Erythroleukemia (75%), after a median latency of 50 d. Transplantable	Erythroblast infiltration in BM and spleen. ERG and ETS1 were the most frequent integration site	CD71 ⁺ , Ter119 ^{-/lo} , some CD41 ⁺ cells	86
2014	<i>JAK2</i> ^{V617F} — <i>Trp53</i> ^{-/-}	MSCV viral overexpression and BM reconstitution	Erythroleukemia -like, after 14–100 d. Transplantable	Hepatosplenomegaly, normal platelet counts	Kit ⁺ , CD71 ⁺ , Ter119 ^{+/-}	23
2017	<i>JAK2</i> ^{V617F} — <i>Trp53</i> ^{-/-}	MSCV viral overexpression and BM reconstitution	Erythroleukemia, 100% penetrance, median latency 46.5 d. Transplantable	Anemia, hepatosplenomegaly, pulmonary hemorrhage, and expansion of dysplastic erythroid progenitors	CD71 ⁺ , Ter119 ⁻	100
2019	<i>NTRK1</i> ^{H498R} — <i>Trp53</i> ^{R173H}	MSCV viral overexpression and BM reconstitution	Erythroleukemia, 100% penetrance, median latency 25 d. Transplantable	Hepatosplenomegaly, infiltration of GATA1 ⁺ , RUNX1 ⁺ , Ter119 ^{+/-} tumor cells. Tumor cells were sensitive to Larotrectinib	CD71 ⁺ , Ter119 ^{+/-}	19
2019	<i>Gata2—Cebpa</i> ^{K/L}	<i>Cebpa</i> ^{K/L} ; <i>Gata2</i> ^{G320D/+} knock-in alleles, fetal liver hematopoietic cell transplants	Bi-lineage acute erythroleukemia almost 100% penetrance, after 8–10 mo	Anemia, thrombocytopenia, and splenomegaly	Kit ⁺ , CD71 ^{high} , Ter119 ^{low} , but also some Mac1 ⁺ cells	101
2020	<i>Gata1s—Tet2</i> ^{-/-}	Erythroblasts from <i>Tet2</i> ^{-/-} + <i>Gata1s</i> transgenic mice grown in vitro and transplanted	Erythroleukemia, 100% penetrance	Anemia, Hepato-splenomegaly, Infiltration of erythroblast GATA1 ⁺ in BM, spleen, and liver	CD71 ^{high} , Ter119 ^{+/-}	21
2020	<i>ERG—TP53</i> ^{R248Q}	MSCV viral overexpression and BM reconstitution (HUPKI). EB sorting from the first degree recipients and injected into the second degree recipients	Erythroleukemia, 100% penetrance with median latency 60 d (2 nd recipients)	Anemia, Hepatosplenomegaly, Infiltration of erythroblast GATA1 ⁺ in BM, spleen, and liver	CD71 ^{high} , Ter119 ^{+/-}	21
2020	<i>NFIA-ETO2—TP53</i> ^{R248Q}	MSCV viral overexpression and BM reconstitution (HUPKI)	Pure erythroid leukemia 80%–100% penetrance. Fully transplantable	Anemia, hepatosplenomegaly, organ infiltration, erythroblasts on blood smears	Kit ⁺ , CD71 ⁺ , Ter119 ^{+/-}	102
2021	<i>TP53/Bcor/Dnmt3a and TP53/Bcor/Rb1/Nfix</i>	Lentiviral vectors with different of sgRNA for multiplex genome editing in Cas9-eGFP- lineage-negative HSCs and BM reconstitution	Erythroleukemia	AEL cells sensitive to CDK9 inhibitor (LY2857785)	Not yet access	103

AEL = acute erythroleukemia; BM = bone marrow; MSCV = murine stem cell virus.

found abnormal karyotypes, such as hyperdiploidy, suggesting increased genomic instability upon *Trp53* loss.¹⁰⁰ Viral expression of wildtype *Trp53* significantly reduced clonogenic activity of in vitro and in vivo leukemia induction by *Jak2*^{V617F}; *Trp53*^{-/-} erythroblasts suggesting an active role in aberrant self-renewal. Both groups explored therapeutic approaches in their models. Kurokawa and colleagues reported that a potent JAK1/2 inhibitor (INCB18424) reduced the spleen weight but did not affect tumor cells in the BM.¹⁰⁰ Levine and colleagues found that treatment with the ruxolitinib JAK2 inhibitor somehow prolonged the survival of secondary recipients of *Jak2*^{V617F}; *Trp53*^{-/-} erythroblasts; however, treatment with an HSP90 inhibitor appeared more efficient in reducing tumor cell load and restoring normal myelopoiesis.²³ Although these studies demonstrated cooperation of *Jak2*^{V617F} and loss of *Trp53* in mice, post-MPN erythroleukemia patients mostly develop *TP53* DNA-binding domain point mutations rather than loss of both alleles that seemed to be essential for the observed mouse phenotype in these experiments.

NTRK1 and mutated TP53

NTRK1 is a member of the neurotrophic receptor tyrosine kinase gene family including the TRK-A, TRK-B, and TRK-C receptors, respectively. NTRKs are target of gene fusions, deletions/truncations and point mutations not only in hematological malignancies but also in various solid cancers that mostly lead to constitutive activity.^{107,108} Iacobucci et al¹⁹ found NTRK1

tyrosine kinase domain (H498R, G6117D, and H766R) mutations in tumor cells from 3 AEL patients carrying *TP53* mutations. While activation of downstream signaling pathways is likely dependent on both the cell context and the type of NTRK mutation, NTRK fusion expression in hematopoietic progenitors led to increased phosphorylation of AKT and PLCγ1.¹⁰⁸ Together with the observation that PI3K-AKT mutations have been found in about 7% of AEL (primarily older adults), aberrant activation of this pathway may be a driver of some human AEL.¹⁹ NTRK1 mutations in AEL gained particular attraction as selective small molecule TRK inhibitors (TRKi) such as larotrectinib, which has been shown to have clinical activity in cancer patients.¹⁰⁹ To model functional cooperation, Iacobucci and colleagues transplanted wildtype and *Trp53*^{R172H} BM-derived HSPCs retrovirally expressing wildtype or mutant *NTRK1* into irradiated syngeneic recipient mice. *Trp53*^{R172H} is the murine homolog to human *TP53*^{R175H} one of the most frequent cancer-associated *TP53* DNA-binding domain mutations. Notably, expression of wildtype or mutant *NTRK1* in *Trp53*^{R172H} cells resulted in a transplantable erythroleukemia-like disease. Expression of the *NTRK1*^{H498R} did result in significantly earlier disease than overexpression of wildtype *NTRK1*. Notably, *NTRK1/Trp53* comutant tumor cells appeared very sensitive to in vivo TRKi therapy. Whereas vehicle-treated mice rapidly succumbed to the disease, larotrectinib-treated mice did not develop any disease >100 days posttransplant. Although NTRK mutations are relatively rare, these observations suggest that,

like other cancers carrying NTRK alterations, treatment with selective TRK inhibitors could be of therapeutic benefit for AEL patients carrying these alterations.

NFIA-ETO2 and mutated TP53

The t(1;16)(p31;q24) chromosomal translocation found in PEL from very young children leads to fusion of the nuclear factor 1A (NFIA) to ETO2 (also known as CBFA2/RUNX1 Partner Transcriptional Co-Repressor 3, CBFA2T3).^{29–33} The transcription factor NFIA has previously been shown to control erythroid fate of hematopoietic progenitors, while ETO2 is as transcriptional cofactor controlling HSC and differentiation of erythroid progenitor cells.^{110,111} We observed that retroviral overexpression of *NFIA-ETO2* fusion blocked in vitro erythroid differentiation of MEL cells and primary murine erythroblasts. However, *NFIA-ETO2*-expressing cells could not be serially propagated in growth-factor containing methylcellulose and transplantation of *NFIA-ETO2*-expressing BM or fetal liver erythroid progenitor cells into irradiated mice did not result in any disease. In contrast, *NFIA-ETO2*-expressing erythroblasts harboring one of the most frequent cancer and AEL-associated TP53 mutation, *TP53*^{R248Q}, could be serially plated in MC and when transplanted into irradiated recipients, induced a fully penetrant transplantable lethal erythroleukemia-like disease characterized by hepatosplenomegaly, anemia, thrombocytopenia, and the presence of erythroid progenitor cells on peripheral blood smear.¹⁰² Molecular studies suggested that *NFIA-ETO2* primarily blocks erythroid differentiation by repressing NFIA as well as GATA1 target genes, and that the *TP53*^{R248Q} mutation endowed cells with aberrant stemness and aberrant activity of the polycomb complex 2 (PRC2). In addition, similar to other ETO-protein containing fusions, *NFIA-ETO2* immortalized cells may also be sensitive to small peptides that disrupt ETO-NHR-domain-mediated protein/protein interactions, suggesting a potential therapeutic vulnerability.^{102,112}

Alterations of BCOR collaborating with TP53 and DNMT3A mutations

To functionally demonstrate oncogenic cooperation, Iacobucci and colleagues used multiplexed CRISPR/Cas9-mediated genome editing of HSPCs followed by BM reconstitution in irradiated mice.¹⁰³ They established 14 genetically different leukemia mouse models in which induction or an AEL phenotype was associated with inactivation of the *Bcl-6* co-repressor (*Bcor*) and *Trp53* either alone or co-mutated with *Dnmt3A*, *Retinoblastoma 1* (*Rb1*) or *Nuclear factor 1 X* (*Nfix1*). Erythroleukemia in the mice was characterized by aberrant expression of erythroid developmental regulators such as *Gata1*, *Kruppel-like factor 1* (*Klf1*), or *Nuclear factor erythroid-2* (*Nfe2l3*), driven by the interaction of mutations of the epigenetic modifiers *Dnmt3a* and *Tet2* that perturbed methylation and thus expression of lineage-specific transcription factors. Putative loss of function BCOR mutations represent a substantial fraction of cytogenetically normal (CN)-AML patients, are frequently associated with DNMT3A mutations and were proposed to be associated with an inferior outcome.¹¹³ Sportoletti et al¹¹⁴ explored functional cooperation with a conditional mouse model mimicking AML-associated BCOR truncating mutations. They observed expansion of erythromegakaryocytic progenitors, anemia, and thrombocytosis but no overt leukemia in *Bcor*^{-/-} mice. In contrast, all *Bcor/Dnmt3A* double knockout mice developed AEL-like leukemia characterized by the expansion of Kit⁺/Ter119⁺ cells. Interestingly, the gene expression signatures of the leukemic cells suggested functional interference with GATA1-regulated erythroid differentiation.¹¹⁴ Both of these studies demonstrated functional cooperation of *Bcor* alterations with *Trp53* and/or *Dnmt3A* in mice; however, BCOR mutations seemed to be very rare events in human AEL.^{14–21} Interestingly, murine *Bcor*/

Trp53-mutated AEL cells were sensitive to the PARP inhibitor talazoparib and the demethylating agent decitabine, and combined *Trp53/Bcor/Dnmt3a* mutation conferred sensitivity of AEL cells to CDK7/9 inhibitors.¹⁰³

Unexpected erythroleukemia phenotypes in genetically modified mice

Inactivation of the HELLS/lymphoid specific helicase chromatin remodeler

HELLS (also known as lymphoid specific helicase or SMARCA6) is a member of the SNF2 subfamily of helicases mostly known for their chromatin remodeling activity. It is thought that the HELLS protein controls the access of de novo methyltransferases DNMT3A/B mostly at DNA repeat elements but also at selected targets including the *Hox* gene loci (Table 3).¹¹⁸ HELLS also seems to modulates chromatin binding of lineage-specific transcription factors including GATA3, SCL/TAL1 or E2A.¹¹⁵ *Hells*^{-/-} mice have multiple defects (growth retardation, premature aging phenotype) and mostly die prematurely.¹¹⁹ To address the role of HELLS for normal hematopoiesis, Muegge and colleagues transplanted fetal liver-derived cells from *Hells*^{-/-} mice into irradiated recipients. Interestingly, recipient mice developed hematologic malignancies including lymphoma or an erythroleukemia-like disease. Although only about 10% of the mice developed this disease, over 50% of them had signs of abnormal erythropoiesis. Although no detailed characterization was reported, the tumor cells appeared to express erythroid CD71 and Ter119 markers. Molecular analysis revealed global DNA hypomethylation and de-repression of endogenous retroviral repeats. In addition, increased expression of *Spi1* mRNA and protein was found in *Hells*^{-/-} fetal livers. The loss of *Hells* was associated with reduced Dnmt3b binding to retroviral elements within the PU.1 gene suggesting that HELLS generally controls Dnmt3B binding to chromatin.¹²⁰ The human HELLS homolog (also known as proliferation-associated SNF2-like gene product) is widely expressed in AML and ALL cell lines and primary samples. Interestingly, a 25-bp mRNA deletion lacking a region critical for transcriptional activation of the yeast homolog was detected in about half of the leukemia samples tested but not in other cancers; however, its functional significance remains unclear.¹²¹

Inactivation of the nuclear receptor interacting SET domain 1 methyltransferase

Posttranslational modification of the histone tails is one of the key events of epigenetic gene regulation. Hereby, trimethylation of lysine 4 and 36 of histone 3 are generally associated with active transcribed regions, whereas trimethylation of lysine 9 or 27 correlates with repression of a given gene locus. The so-called epigenetic code is based on the interplay of histone lysine methyltransferases (HMT) that set these marks (referred as “writers”) and demethylases (HDM) that remove them (referred as “erasers”). Multiple genes encoding for these epigenetic regulators are targets of recurrent somatic gene alterations that have been shown to contribute to AML initiation and maintenance.¹²² H3K36me3 is the product of mono- and dimethylation by the nuclear receptor interacting SET domain (NSD)1–3 family, ASH1L, or SETMAR followed by SETD2-mediated trimethylation.¹²³ NSD1 is the target of recurrent genomic alterations in human cancers. Highly prevalent putative loss of function *NSD1* mutations have been found in head and neck and other solid cancers, and *NSD1* expression was reported epigenetically silenced in renal carcinomas.¹²⁴ In contrast to solid cancers, *NSD1* mutations are rare in hematological malignancies; however, *NSD1* was found to be target of a recurrent t(5;11) chromosomal translocation found in childhood AML that results in a fusion with the *NUP98* gene.¹²⁵ To better understand its function in hematopoiesis, we inactivated *NSD1* in human and mouse hematopoietic cells.¹¹⁷

Table 3.

Unexpected AEL mouse model

Year	Gene	Model	Phenotype	Major Findings	Surface markers on leukemic cells	Reference
2008	<i>Hells</i> ^{-/-}	Constitutive gene knockout	<i>Hell</i> ^{-/-} fetal liver hematopoietic cell transplant. 7% erythroleukemia, some lymphoma. 11% erythroleukemia upon co-deletion of <i>Trp53</i>	Anemia, splenomegaly, infiltration by erythroblasts	Relative increase in CD71 ⁺ /Ter119 ⁺ cells	115
2019	<i>iH3K36M</i>	Dox-inducible (<i>rtTA</i> , <i>Rosa26</i>), transgene in <i>Col1A1</i> locus	Lethal hematologic disorder after median latency of 50 d	Anemia, thrombocytopenia, splenomegaly, accumulation of erythroid progenitors. Relative BM hypocellularity	Increase in colony forming unit-erythroid progenitors and CD71 ⁺ proerythroblasts, expressing low levels of c-Kit. Decrease in Ter119 ⁺ EBs	116
2020	<i>Nsd1</i> ^{-/-}	Targeted gene knockout (<i>Vav-iCre</i> ; <i>Nsd1</i> ^{fl/m})	Pure erythroleukemia-like disease, transplantable, 100% penetrance	Anemia, thrombocytopenia, splenomegaly, erythroblasts on peripheral smears. Multiorgan infiltration and relative BM hypocellularity	CD71 ^{low} , Kit ^{+/+} , FcγRIII/III ^{+/+} , CD34 ⁻ , B220 ⁻ and Sca-1 ⁻	117

AEL = acute erythroleukemia; BM = bone marrow.

Hematopoietic (*Vav1-iCre*-mediated) inactivation of *Nsd1* induced a fully penetrant erythroleukemia-like disease characterized by anemia, thrombocytopenia, splenomegaly, and multiorgan infiltrations with occasional erythroblasts on peripheral blood smears. *Nsd1*^{-/-} erythroblasts formed abnormal serially replating burst forming unit-erythroid in EPO-containing MC. Transplantation of BM cells from diseased mice propagated the disease in wild-type recipients, alone or in competition with normal cells. Despite constitutive expression of the erythroid master regulator GATA1, *in vitro* erythroid terminal maturation of *Nsd1*^{-/-} erythroblasts was significantly impaired. Expression of known positively regulated GATA1 targets was decreased, while the regulation of GATA1-repressed target genes was less affected. Retroviral overexpression of *Gata1* was able to overcome the terminal differentiation block *in vitro*. Similarly, retroviral expression of wildtype, but not a catalytically inactive *Nsd1*^{N1918Q} mutant, was also able to rescue the terminal maturation block associated with upregulation of erythroid differentiation-associated genes on the mRNA and global proteome level. Despite very similar *Gata1* mRNA and protein levels, only *Nsd1*^{-/-} erythroblasts expressing wildtype *Nsd1* showed significantly increased binding of Gata1 to many known target genes. These observations suggested that the catalytic activity of NSD1 is an essential permissive factor for proper transactivation of GATA1 targets for productive terminal erythroid maturation. Importantly, knockdown of *NSD1* mRNA significantly altered the clonogenic growth of human CD34⁺ HSCP leading to accumulation of immature erythroid progenitor cells strongly suggesting that independent of the species, NSD1 activity controls terminal erythroid maturation.

Hematopoietic expression of an inducible H3^{K36M} oncohistone transgene

H3K36 is target of multiple aberrations in human cancer including mutations miswriting the marks or aberrant expression of the respective HMT, but also by mutations of the. ¹²⁶ Originally identified in the histone 3.3. (H3.3.) variant in chondrosarcoma, H3 lysine (K) to methionine (M) mutations were later also found in H3.1 in several human cancers (also known as oncohistones) including pediatric soft-tissue sarcomas and some head and neck squamous cell carcinoma. The H3^{K36M} mutant protein seems to sequester the SETD2 HMT resulting in globally reduced H3K36me3 marks but also inhibits other HMTs active on H3K36 such as NSD1 and NSD2 resulting in reduced H3K36me1/2. ¹²⁷ To address the impact of a H3^{K36M} mutation in primary cells *in vivo*, Hochedlinger and colleagues established ES cells and mice with Doxycycline (DOX)-induced overexpression of a H3^{K36M} transgene integrated in the *Col1A1* gene locus. ¹¹⁶ Adult *iH3K36M* transgenic mice developed symptoms of disease after 4–7 weeks on DOX. In addition to nonhematopoietic

aberrations (testicular atrophy, lack of Paneth cells in the intestine), the mice developed thymic atrophy, splenomegaly, and reduced BM cellularity. They presented with anemia, thrombocytopenia, and increased white blood counts with erythroid progenitors in the periphery mimicking acute erythroleukemia. Gene expression profiling revealed an increased expression of regulators of the erythroid lineage and downregulation of genes known to control HSC and/or myelopoiesis. Notably, the H3K36me3 mark was depleted in downregulated genes of which some showed increased H3K27me3 marks at promoter and adjacent to gene bodies. Overall, there was a correlation between loss of H3K36me and decreased chromatin accessibility, but only modest decreases in DNA methylation were observed over gene bodies of some hematopoietic regulators. The erythroleukemia phenotype was strikingly similar between *iH3K36M* and *Nsd1*^{-/-} mice, which underlines that appropriately regulated H3K36 methylation is critical for terminal erythroid differentiation.

Emerging molecular mechanism of erythroleukemia

Is GATA1 a core player in the pathogenesis of human erythroleukemia?

GATA1 is a master regulator of normal erythropoiesis acting in transcriptionally active complexes with TAL1, LMO2, LDB1, RUNX1, ETO-, and ETS-family proteins. ^{67,128} GATA1 undergoes multiple posttranslational modifications including phosphorylation, acetylation, sumoylation, and ubiquitination that in part control its transcriptional activity. ¹²⁹ Mutational and transcriptomic analysis of primary human AEL cells supports the idea that several alterations may converge on the functional alteration of GATA1 through various mechanisms. First, some transcriptionally active proteins of the GATA1 complexes, including GATA1 itself, are targeted by mutations or part of fusion genes in AEL (eg, GATA1s, NFIA-ETO2, or MYB-GATA1). ^{19,21,29–33} However, these alterations are rare and therefore do not account for the erythroid phenotype of the majority of AEL. Second, mouse models have shown that altering the expression of several factors, including ectopic expression of ERG, SPI1, or FLI1 or reduced expression of GATA1 can induce AEL phenotypes, indicating that additional mechanisms converging on GATA1 lead to the development of erythroid leukemia. ^{50,52,85,93} In the line of these observations in mice, we found that aberrantly high expression of several proteins that impact GATA1 function are recurrent alterations in human primary AEL cells. This includes high expression of factors like ERG, CBFA2T3, or SKI that functionally antagonize GATA1-dependent differentiation

in G1E cells. Notably, ectopic expression of these transcription factors in mouse erythroblasts resulted in their immortalization associated with decreased chromatin accessibility at GATA1 binding sites.²¹

In the 2 AEL cohorts for which transcriptome data are available, the subset of AEL samples presenting alterations of expression of either transcription factors or signaling intermediates impacting GATA1 activity represents up to 25% of cases.²¹ Notably, aberrantly high expression levels of ERG or EPOR were also reported in another independent AEL cohort resulting from *bona fide* genomic amplifications including ERG and EPOR genes.^{21,130,131} Overall, these findings strongly support a functional convergence on aberrant GATA1 activity in human AEL either through direct genetic alterations (eg, in regulatory regions to be identified) or as a result of an epigenetic drift as outlined in the following section.

Mechanistically, aberrant GATA1 activity may result in at least 2 cellular consequences. First, inhibition of GATA1 activity may derive from distinct mechanisms involving aberrant maintenance of GATA1 transcriptional repressor (eg, ETS-associated or ETO2 transcriptional complexes), GATA1 destabilization at the protein levels through alterations of GATA1 posttranslational modifications or aberrant protein-protein interaction, or by alterations affecting GATA1 chromatin binding. These mechanisms would contribute to prevent erythroid progenitor differentiation progression toward fully mature erythroid cells. Second, GATA1 activity could be aberrantly activated or maintained, including through constitutive activation of signaling factors (eg, through JAK2^{V617F} or high EPOR expression), leading to the abnormal commitment of early immature progenitors toward the erythroid lineage. This idea is supported by the recent model combining bi-allelic *Cebpa* and *Gata2* mutations.¹⁰¹ Hereby, *Cebpa* and *Gata2* mutations synergize by increasing erythroid transcription factor (*Gata1*, *Klf1*, and *Zfmp1*) expression and erythroid chromatin access, respectively, thereby installing ectopic erythroid potential. In addition, while bi-allelic *Cebpa* mutation resulted in increased erythroid transcription factors expression, it also led to increased expression of several erythroid repressors including *Gata2*, *Erg*, or *Cbfa2t3*. These findings collectively suggest that a combination of these 2 antagonistic mechanisms on GATA1 activity in the same progenitor could explain both, the erythroid bias and the blockage of differentiation. The degree of erythroid commitment may depend on both the place of the targeted progenitor in the hematopoietic hierarchy and the cooperating mutations. Of note, GATA1 is also mutated in myeloid leukemia of Down syndrome (ML-DS) that frequently present with erythroid cell marker expression, and is generally considered at the frontier between megakaryoblastic and erythroid leukemia.¹³² While numerous studies have addressed GATA1 function, it is likely that further analyses are required to fully understand how the activity of GATA1 is fine-tuned and how these alterations contribute to erythroid transformation.

Impaired TP53 activity and malignant erythropoiesis

TP53 regulates HSPC quiescence and self-renewal; thus, impaired function of TP53 promotes HSPC proliferation that likely leads to additional DNA damage and hematopoietic malignancies.^{133,134} Although TP53 mutations represent by far the most frequent mutated genes in AEL and particularly in PEL, no model clearly demonstrated yet a link with the erythroid phenotype. Impaired erythropoiesis in MDS carrying a deletion on 5q has been linked to activation of TP53 upon inactivation of the ribosomal protein small subunits (RPS)-14 or -19.¹³⁵ More recently, activation of TP53 during ribosomal biogenesis have been proposed to regulated normal erythroid differentiation.¹³⁶

Most TP53 alterations identified in AEL and other AML subtypes are missense mutations in the DNA binding domain but their functional consequences (ie, inactivating, gain-of-function or dominant negative) remain a matter of debate. Functional studies in mice suggested that TP53 mutations are drivers of clonal hematopoiesis.¹³⁷ Several TP53 DNA-binding domain mutations have been reported to disrupt the structure and activity of the protein, allowing neomorphic interactions with several tumor suppressive factors, including TP73 that abrogate its function.¹³⁸ In addition, such potentially TP53-gain-of-function mutants appear to be stabilized by binding to HSP90 leading to the inactivation of MDM2 and CHIP E3 ligase-mediated degradation.¹³⁹ Of importance, these molecular mechanisms could be pharmacologically blocked by small-molecules interfering with TP73 and HSP90.^{134,140} However, more recent functional studies suggested that the most prevalent AML-associated TP53 mutations act in a dominant-negative manner rather than gain-of-function.¹⁴¹

TP53 interferes with the activity of multiple transcription factors. For example, some TP53 mutants have been shown to interact with the ETS1 transcription factor leading to increased expression of multidrug-resistance 1 (MDR1) associated with a poor outcome in AML.¹⁴² Experimental evidence linked TP53 to GATA1 activity in normal erythropoiesis. TP53 and GATA1 may interact, through their transactivation domain and DNA-binding domain, respectively, leading to mutual inhibition of their transcriptional activities.¹⁴³

TP53-DNA-binding domain mutations were shown to interact with epigenetic factors, including EZH2, KMT2A, KMT2D, and KAT6A. Notably, TP53 mutants bind to and enhance EZH2 chromatin association, resulting in an increased level of H3K27me3 at essential regulator of HSC function and differentiation.¹³⁸ We recently found that the PEL-associated *NFIA-ETO2* fusion gene functionally cooperated with one of the most prevalent AEL-associated TP53^{R248Q} mutation most likely also through functional interference with the PRC2 complex.¹¹² TP53-DNA-binding domain mutations may also interact and enhance the activity of the methyltransferase KMT2A/2D and the acetyl transferase KAT6A leading to increase genome-wide methylation and acetylation.¹⁴⁴ Consistently, KMT2A/2D and KAT6A are upregulated in mutant TP53 AML and are often mutated in AML, including AEL.^{19,145} While additional work is necessary to dissect the link between TP53 mutation and epigenetic gene regulation, it could open the avenue for novel therapeutic avenues. If TP53 mutants indeed enhance KMT2A activity through KMT2A partners such as MEN1, TP53-mutant leukemia including AEL may benefit from the recent development of highly potent and selective small molecules blocking the functionally critical interaction of KMT2A and Menin.¹⁴⁶

In MDS, TP53 mutations are generally associated with high-risk disease, rapid transformation to AML, therapy resistance, and poor outcome. Studying a large MDS patient cohort, Papemmanuil and colleagues recently found that two-thirds of the patients had multiple hits indicating biallelic targeting which was predictive for leukemic transformation and early death. Interestingly, monoallelic patients die not differ from TP53 wildtype patients in outcomes and therapy response which would not really support a dominant-negative activity of these mutations (at least in the context of MDS).¹⁴⁷ However, in at least 80% of TP53-mutated AML patients, more than 1 genetic alteration is present, reflecting the requirement for different oncogenic cooperation mechanisms.¹³⁵ Indeed, TP53 mutations were shown to cooperate with multiple cellular signaling pathways. Loss of TP53 activity has been shown to cooperate with the KRAS^{G12D} activating mutation, inducing an aggressive AML in mice and with NRAS^{G12D} to promote megakaryocytic-erythroid progenitor (MEP) transformation leading to AML.^{148,149} The RAS signaling pathway is target of recurrent

mutations in AEL.^{19,21} Similarly, as outlined before, mouse models have shown that genetic TP53 inactivation cooperates with several mutations, such as loss of function alterations of CEBPA or BCOR or constitutively active mutated tyrosine kinases such as JAK2^{V617F} or NTRK1^{H498R} to induce erythroleukemia in mice.^{19,23,100–103,114} Hence, treatment of xenografted AEL patient cells carrying either JAK2 or EPOR amplification with a JAK2 inhibitor significantly suppressed cell growth and prolonged overall survival.¹³⁰ Together, these observations strongly suggest that TP53 mutations are essential players not only in AEL development but also crucial for the maintenance of the disease. However, it remains to be elucidated how TP53 mutations functionally interfere with transcriptional control of erythroid differentiation.

Aberrant chromatin organization and erythroleukemia

AEL mouse models as well as sequencing of human samples strongly suggest that several proteins that control chromatin architecture by DNA methylation or histone modification play a central role in erythroid malignancies. DNA methylation is regulated by several factors including TET2, DNMT3A/B, or HELLS/Lsh.¹⁵⁰ While DNMT3A/B is involved in de novo methylation by transferring a methyl group S-adenosyl-L-methionine to carbon position 5 of the nucleotide cytosine (5-mC), TET2 controls demethylation through the oxidation of 5-mC in 5-hmC. Mutational inactivation of these factors is recurrently observed in patients with MDS and AML resulting in enhanced HSC self-renewal and a decline in the output of differentiated progeny, thus predisposing to leukemic transformation in mice.¹⁵¹ Inactivation of *Tet2* or *Dnmt3a* in HSC resulted in increased myeloid and decreased erythroid gene expression signature resulting in aberrant accumulation of erythroid progenitors in mice. Notably, DNMT3A and TET2 were suggested to regulate hematopoietic differentiation by controlling accessible binding sites for distinct transcription factors including

GATA1.^{152–158} In addition, very recent work has shown that precise DNA methylation patterning can control binding and regulation of GATA1 activity.¹⁵⁹ Interestingly, a novel signaling pathway has been characterized that links TET2 activation to JAK2-mediated phosphorylation resulting not only in increased cytosine hydroxy-methylation and genome-wide loss of cytosine methylation but also enhanced interaction with the erythroid transcription factor KLF1.¹⁶⁰ Interestingly, combined TET2 and DNMT3A inactivation, as frequently found in AEL, was also reported to increase of KLF1 and EPOR expression.¹⁶¹ As outlined above, genetic inactivation of *Dnmt3A* in mice lacking the transcriptional coregulator *Bcor* shifted the phenotype from macrocytic anemia to an erythroleukemia-like disease.¹¹⁴ Based on these observations, we speculate that the erythroid phenotype in some AEL cases is based on aberrant DNA methylation that impairs the erythroid transcriptional program. Aberrant DNA methylation is therapeutically targeted by cytidine analogs (eg, 5-Azacytidine or Decitabine) that incorporate DNA instead of deoxycytidine, covalently bind the enzyme and lead to DNMT degradation. Although a rather low specificity, positive clinical effects resulted into FDA approval to treat MDS and AML patients.¹⁶² Notably, earlier case reports of positive therapeutic responses to azacytidine were supported in a larger retrospective study including over 80 AEL patients treated with hypomethylating agents.¹⁶³ Although a deeper molecular understanding of these effects would be important, it appears that similar to other AML forms HMAs as single agents are not curative for AEL.

Unexpected erythroleukemia models emerged from the inactivation or mutations of several chromatin-associated factors. Inactivation of the histone H3K36me1/2 methyl-transferase NSD1 leads to uncontrolled accumulation of erythroid progenitor cells resulting in a fully penetrant erythroleukemia-like disease in mice. Molecular analyses revealed a decrease of GATA1 target genes expression without significant expression changes in known GATA1-repressors suggesting that NSD1 is an essential epigenetic modulator of GATA1 target genes.¹¹⁷ Of importance, NSD1 has been identified being mutated in some cases of pediatric AEL either through an NUP98-NSD1 fusion or a NSD1 loss-of-function mutation found in a single patient.¹⁹ Notably, the *NSD1* gene is located on the long arm of chromosome 5, a region that is most frequently target of cytogenetic alteration in human erythroleukemia.^{5,10} As outlined above, inducible hematopoietic overexpression of an *H3K36M* transgene resulting in reduced H3K36me1/2 methylation induced a very similar erythroid phenotype in mice as observed upon NSD1 inactivation. Molecular characterization of *H3K36M* overexpressing HSPC revealed an expression of an aberrant erythroid signature, but the putative relation to GATA1 target activation remains unknown.¹¹⁶ Extensive biochemical in vitro experiments revealed that NSD1-mediated H3K36me2 marks are required for the recruitment of DNMT3A and maintenance of DNA methylation.¹⁶⁴ Genetic ablation of *Nsd1* and its paralog *Nsd2* in murine cells resulted in a redistribution of DNMT3A to H3K36me3-modified gene bodies and a reduction in the methylation of intergenic regions. Notably, both, blood samples from individual with SOTOS overgrowth syndrome (carrying germline *NSD1* loss of function mutations) as well as NSD1-mutant cancer cells exhibited hypomethylation of intergenic DNA. This suggests that reduced H3K36 methylation connects human cancers and developmental overgrowth through aberrant intergenic CpG methylation.

Another chromatin modifier that has also been recently found to interact with GATA1 and other components of the GATA1-transcriptional complex is the KDM5A histone H3K4me demethylase.¹⁶⁵ Notably, NUP98-KDM5A fusions have been found in pediatric AMKL and AEL patients.^{19,35} Although the ectopic expression of this fusion in human cord blood HSPC lead to immortalization and multilineage leukemia when

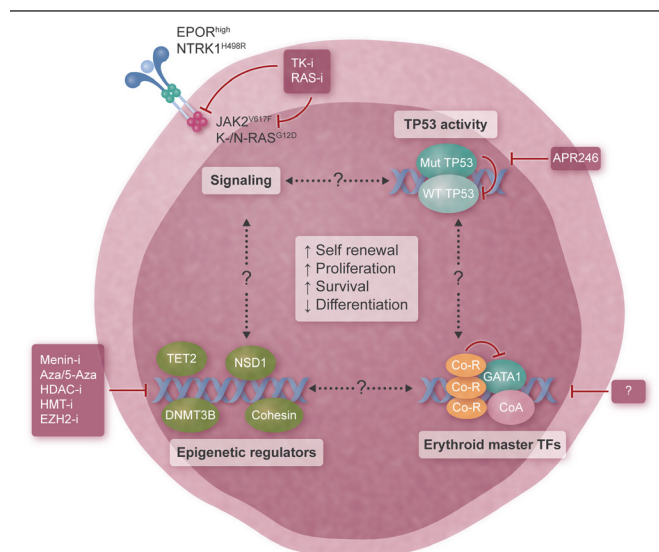


Figure 3. Targeted therapeutic strategies emerging from the epigenomic landscapes and AEL disease models. Schematic representation of major players identified to date in human AEL and the potential strategies for targeted interference, including (1) blocking aberrant activation of the JAK/STAT signaling axis (upper left); (2) restoration of the tumor suppressive TP53 activity (upper right); (3) inhibition of epigenetic regulators (lower left); and (4) reactivation of impaired activity of erythroid master transcription factors (like GATA1) to boost terminal differentiation of malignant erythroblasts (lower right). AEL = acute erythroleukemia.

injected in mice its impact on the chromatin and on GATA1 activity remains unknown.¹⁶⁶

Chromatin organization is also orchestrated by the cohesin protein family, frequently mutated in myeloid malignancies including AEL.^{19,21,130,167} Mutations in cohesin-encoding genes are strongly associated with GATA1 mutations in ML-DS.^{168,169} Overall, cohesin mutations are associated with a global decrease in chromatin accessibility, but a relative enhancement of chromatin accessibility at binding sequences for some master hematopoietic stem-cell transcription factors such as GATA2, RUNX1, and ERG involved in myeloid transformation.¹⁷⁰ Importantly, cohesin deficiency severely impaired erythroid differentiation of a multipotent cell line and enhanced self-renewal programs.¹⁷¹ These observations suggest that altered cohesin function may support erythroid transformation; however, their role for induction or maintenance of AEL remains to be elucidated. Collectively, these data support the idea that erythroid differentiation is tightly linked to chromatin organization and that aberrant expression or mutations of the regulatory proteins may lead to erythroid malignancies. While relatively few direct erythroid regulatory genes have been identified mutated in AEL, aberrant activity of epigenomic factors that control erythroid differentiation on the chromatin level are therefore valuable candidates to explain the erythroid phenotype of the disease.

Taken together, insights from multiple mouse models as well as the epigenomic landscape recently defined by deep sequencing and transcriptomics studies suggest that AEL reflects a disease continuum between MDS and AML that is characterized by a unique erythroid identity. While the molecular links between genetic/transcriptomic/epigenetic AEL alterations network remains to be elucidated (Figure 3), several therapeutic opportunities can be envisioned. Interference with the most prevalent genetic lesions like TP53 DNA binding mutations but also inhibition of classical signaling pathways, including the JAK/STAT pathway supports growth and survival, appears of primary interest and will require further preclinical assessments. In addition to these signaling nodes that are drivers of several myeloid malignancies, AEL is characterized by different degrees of aberrant erythroid maturation. Recent insights from rationale and unexpected mouse models indicate that the erythroid identity in a significant fraction of AEL is based on the impaired activity of transcriptional master regulators like GATA1. Therapeutic restoring of the GATA1 activity could lead to terminal differentiation of aberrantly accumulated erythroblasts, which may resolve PEL and reduce the cellular burden of AEL. However, more research is needed to dissect the critical molecular mechanisms to translate this strategy into clinically effective therapies.

Disclosures

The authors have no conflicts of interest to disclose.

Sources of funding

The work on erythroleukemia by J.S. is supported by the Swiss Cancer Research KFS-3487-08-2014 and KFS-4258-08-2017; the Swiss National Science Foundation (31000A_173224), the Novartis Biomedical Research Foundation, Basel, the San Salvatore Foundation, Lugano (201525), and the Wilhelm-Sander Foundation, Munich (2017-035.1). AF was supported by Ligue Contre le Cancer & Fondation pour la Recherche Médicale. TM is supported by Institut National du Cancer (PLBIO-2018-169), PAIR-Pédiatrie/CONNECT-AML (Collaborative Network for Children and Teenagers with Acute Myeloblastic Leukemia: INCa-ARC-LIGUE_11905 and Association Laurette Fugain), Société Française des Cancers de l'Enfant, SIRIC-SOCRATE (INCa-DGOS-INSERM_12551).

References










- Bain BJ. Di Guglielmo and his syndromes. *Br J Haematol.* 2003;120:939–943.
- Bennett JM, Catovsky D, Daniel MT, et al. Proposals for the classification of the acute leukaemias. French-American-British (FAB) Co-operative Group. *Br J Haematol.* 1976;33:451–458.
- Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood.* 2009;114:937–951.
- Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood.* 2016;127:2391–2405.
- Hasserjian RP, Zuo Z, Garcia C, et al. Acute erythroid leukemia: a reassessment using criteria refined in the 2008 WHO classification. *Blood.* 2010;115:1985–1992.
- Kiossoglou KA, Mitus WJ, Dameshek W. Chromosomal aberrations in acute leukemia. *Blood.* 1965;26:610–641.
- Castoldi G, Mitus WJ, Yam LT, et al. Chromosomal studies in erythro-leukemia and chronic erythremic myelosis. *Blood.* 1968;31:202–215.
- Mazzella FM, Kowal-Vern A, Shrit MA, et al. Acute erythro-leukemia: evaluation of 48 cases with reference to classification, cell proliferation, cytogenetics, and prognosis. *Am J Clin Pathol.* 1998;110:590–598.
- Santos FP, Faderl S, Garcia-Manero G, et al. Adult acute erythro-leukemia: an analysis of 91 patients treated at a single institution. *Leukemia.* 2009;23:2275–2280.
- Lessard M, Struski S, Leymarie V, et al. Cytogenetic study of 75 erythro-leukemias. *Cancer Genet Cytogenet.* 2005;163:113–122.
- Hou HA, Chou WC, Kuo YY, et al. TP53 mutations in de novo acute myeloid leukemia patients: longitudinal follow-ups show the mutation is stable during disease evolution. *Blood Cancer J.* 2015;5:e331.
- Wang W, Wang SA, Medeiros LJ, et al. Pure erythroid leukemia. *Am J Hematol.* 2017;92:292–296.
- Montalban-Bravo G, Benton CB, Wang SA, et al. More than 1 TP53 abnormality is a dominant characteristic of pure erythroid leukemia. *Blood.* 2017;129:2584–2587.
- Grossmann V, Bacher U, Haferlach C, et al. Acute erythroid leukemia (AEL) can be separated into distinct prognostic subsets based on cytogenetic and molecular genetic characteristics. *Leukemia.* 2013;27:1940–1943.
- Cervera N, Carbuccioni N, Garnier S, et al. Molecular characterization of acute erythroid leukemia (M6-AML) using targeted next-generation sequencing. *Leukemia.* 2016;30:966–970.
- Cervera N, Carbuccioni N, Mozziconacci MJ, et al. Revisiting gene mutations and prognosis of ex-M6a-acute erythroid leukemia with regard to the new WHO classification. *Blood Cancer J.* 2017;7:e594.
- Ping N, Sun A, Song Y, et al. Exome sequencing identifies highly recurrent somatic GATA2 and CEBPA mutations in acute erythroid leukemia. *Leukemia.* 2017;31:195–202.
- Rose D, Haferlach T, Schnittger S, et al. Subtype-specific patterns of molecular mutations in acute myeloid leukemia. *Leukemia.* 2017;31:11–17.
- Iacobucci I, Wen J, Meggendorfer M, et al. Genomic subtyping and therapeutic targeting of acute erythro-leukemia. *Nat Genet.* 2019;51:694–704.
- Cervera N, Lhoumeau AC, Adélaïde J, et al. Acute erythroid leukemias have a distinct molecular hierarchy from non-erythroid acute myeloid leukemias. *Haematologica.* 2020;105:e340–e342.
- Fagnan A, Bagger FO, Piqué-Borràs MR, et al. Human erythro-leukemia genetics and transcriptomes identify master transcription factors as functional disease drivers. *Blood.* 2020;136:698–714.
- Ware AD, Birkness J, Duffield AS, et al. Molecular evidence of JAK2 p.V617F mutated pure erythroid leukemia arising from polycythemia vera. *Virchows Arch.* 2018;473:131–135.
- Rampal R, Ahn J, Abdel-Wahab O, et al. Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms. *Proc Natl Acad Sci U S A.* 2014;111:E5401–E5410.
- Kreft A, Springer E, Lipka DB, et al. Wild-type JAK2 secondary acute erythro-leukemia developing after JAK2-V617F-mutated primary myelofibrosis. *Acta Haematol.* 2009;122:36–38.
- Pompetti F, Spadano A, Sau A, et al. Long-term remission in BCR/ABL-positive AML-M6 patient treated with Imatinib Mesylate. *Leuk Res.* 2007;31:563–567.
- Zhao W, Du Y, Ho WT, et al. JAK2V617F and p53 mutations coexist in erythro-leukemia and megakaryoblastic leukemic cell lines. *Exp Hematol Oncol.* 2012;1:15.

27. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126:9–16.
28. Bowman RL, Busque L, Levine RL. Clonal hematopoiesis and evolution to hematopoietic malignancies. *Cell Stem Cell*. 2018;22:157–170.
29. Micci F, Thorsen J, Haugom L, et al. Translocation t(1;16)(p31;q24) rearranging CBFA2T3 is specific for acute erythroid leukemia. *Leukemia*. 2011;25:1510–1512.
30. Panagopoulos I, Micci F, Thorsen J, et al. Fusion of ZMYND8 and RELA genes in acute erythroid leukemia. *PLoS One*. 2013;8:e63663.
31. Micci F, Thorsen J, Panagopoulos I, et al. High-throughput sequencing identifies an NFIA/CBFA2T3 fusion gene in acute erythroid leukemia with t(1;16)(p31;q24). *Leukemia*. 2013;27:980–982.
32. Liu H, Guinipero TL, Schieffer KM, et al. De novo primary central nervous system pure erythroid leukemia/sarcoma with t(1;16)(p31;q24) NFIA/CBFA2T3 translocation. *Haematologica*. 2020;105:e194–e197.
33. Linnik Y, Pastakia D, Dryden I, et al. Primary central nervous system erythroid sarcoma with NFIA-CBFA2T3 translocation: a rare but distinct clinicopathologic entity. *Am J Hematol*. 2020;95:E299–E301.
34. King RL, Siaghani PJ, Wong K, et al. Novel t(1;8)(p31.3;q21.3) NFIA-RUNX1T1 translocation in an infant erythroblastic sarcoma. *Am J Clin Pathol*. 2020:aqaa216. [Epub ahead of print].
35. Chisholm KM, Heerema-McKenney AE, Choi JK, et al. Acute erythroid leukemia is enriched in NUP98 fusions: a report from the Children's Oncology Group. *Blood Adv*. 2020;4:6000–6008.
36. Graf T, Ade N, Beug H. Temperature-sensitive mutant of avian erythroblastosis virus suggests a block of differentiation as mechanism of leukaemogenesis. *Nature*. 1978;275:496–501.
37. Beug H, Bauer A, Dolznig H, et al. Avian erythropoiesis and erythroleukemia: towards understanding the role of the biomolecules involved. *Biochim Biophys Acta*. 1996;1288:M35–M47.
38. Rietveld LE, Caldenhoven E, Stunnenberg HG. Avian erythroleukemia: a model for corepressor function in cancer. *Oncogene*. 2001;20:3100–3109.
39. Ivanov X, Mladenov Z, Nedyalk S, Todorov TG. Experimental investigation into avian leukoses. I. Transmission experiments of certain diseases of the avian leukosis complex in Bulgaria. *Bulgarian Acad Sci Bull Inst Pathol Comp Animaux Domestiques*. 1962;9:5–36.
40. Rasclé A, Ferrand N, Gandrillon O, et al. Myb-Ets fusion oncoprotein inhibits thyroid hormone receptor/c-ErbA and retinoic acid receptor functions: a novel mechanism of action for leukemogenic transformation by E26 avian retrovirus. *Mol Cell Biol*. 1996;16:6338–6351.
41. Blair DG, Athanasiou M. Ets and retroviruses—transduction and activation of members of the Ets oncogene family in viral oncogenesis. *Oncogene*. 2000;19:6472–6481.
42. Friend C. Cell-free transmission in adult Swiss mice of a disease having the character of a leukemia. *J Exp Med*. 1957;105:307–318.
43. Orkin SH. Differentiation of murine erythroleukemic (Friend) cells: an in vitro model of erythropoiesis. *In Vitro*. 1978;14:146–154.
44. Fey F, Graffi A. Erythroblast-Leukämie nach Injektion von Virus der myeloischen Leukämie der Maus. *Z Krebsforsch*. 1965;67:145–151.
45. Siegel BV, Weaver WJ, Koler RD. Mouse erythroleukemia of viral etiology. *Nature*. 1964;201:1042–1043.
46. Ney PA, D'Andrea AD. Friend erythroleukemia revisited. *Blood*. 2000;96:3675–3680.
47. Moreau-Gachelin F, Tavitian A, Tambourin P. Spi-1 is a putative oncogene in virally induced murine erythroleukaemias. *Nature*. 1988;331:277–280.
48. Longmore GD, Lodish HF. An activating mutation in the murine erythropoietin receptor induces erythroleukemia in mice: a cytokine receptor superfamily oncogene. *Cell*. 1991;67:1089–1102.
49. Skoda RC, Tsai SF, Orkin SH, et al. Expression of c-MYC under the control of GATA-1 regulatory sequences causes erythroleukemia in transgenic mice. *J Exp Med*. 1995;181:1603–1613.
50. Moreau-Gachelin F, Wendling F, Molina T, et al. Spi-1/PU.1 transgenic mice develop multistep erythroleukemias. *Mol Cell Biol*. 1996;16:2453–2463.
51. Trempus CS, Ward S, Farris G, et al. Association of v-Ha-ras transgene expression with development of erythroleukemia in Tg.AC transgenic mice. *Am J Pathol*. 1998;153:247–254.
52. Shimizu R, Kuroha T, Ohneda O, et al. Leukemogenesis caused by incapacitated GATA-1 function. *Mol Cell Biol*. 2004;24:10814–10825.
53. Torchia EC, Boyd K, Rehg JE, et al. EWS/FLI-1 induces rapid onset of myeloid/erythroid leukemia in mice. *Mol Cell Biol*. 2007;27:7918–7934.
54. Salek-Ardakani S, Smooha G, de Boer J, et al. ERG is a megakaryocytic oncogene. *Cancer Res*. 2009;69:4665–4673.
55. Tsuzuki S, Seto M. Expansion of functionally defined mouse hematopoietic stem and progenitor cells by a short isoform of RUNX1/AML1. *Blood*. 2012;119:727–735.
56. Gentner E, Vegi NM, Mulaw MA, et al. VENTX induces expansion of primitive erythroid cells and contributes to the development of acute myeloid leukemia in mice. *Oncotarget*. 2016;7:86889–86901.
57. Thoene S, Mandal T, Vegi NM, et al. The ParaHox gene Cdx4 induces acute erythroid leukemia in mice. *Blood Adv*. 2019;3:3729–3739.
58. Kosmider O, Denis N, Lacout C, et al. Kit-activating mutations cooperate with Spi-1/PU.1 overexpression to promote tumorigenic progression during erythroleukemia in mice. *Cancer Cell*. 2005;8:467–478.
59. Munroe DG, Peacock JW, Benchimol S. Inactivation of the cellular p53 gene is a common feature of Friend virus-induced erythroleukemia: relationship of inactivation to dominant transforming alleles. *Mol Cell Biol*. 1990;10:3307–3313.
60. Lavigueur A, Bernstein A. p53 transgenic mice: accelerated erythroleukemia induction by Friend virus. *Oncogene*. 1991;6:2197–2201.
61. Li JP, D'Andrea AD, Lodish HF, et al. Activation of cell growth by binding of Friend spleen focus-forming virus gp55 glycoprotein to the erythropoietin receptor. *Nature*. 1990;343:762–764.
62. Howard JC, Berger L, Bani MR, et al. Activation of the erythropoietin gene in the majority of F-MuLV-induced erythroleukemias results in growth factor independence and enhanced tumorigenicity. *Oncogene*. 1996;12:1405–1415.
63. Lachman HM, Skoultchi AI. Expression of c-myc changes during differentiation of mouse erythroleukaemia cells. *Nature*. 1984;310:592–594.
64. Dmitrovsky E, Kuehl WM, Hollis GF, et al. Expression of a transfected human c-myc oncogene inhibits differentiation of a mouse erythroleukaemia cell line. *Nature*. 1986;322:748–750.
65. Robert-Lézénès J, Meneceur P, Ray D, et al. Protooncogene expression in normal, preleukemic, and leukemic murine erythroid cells and its relationship to differentiation and proliferation. *Cancer Res*. 1988;48:3972–3976.
66. Matsuzaki T, Aisaki Ki, Yamamura Y, et al. Induction of erythroid differentiation by inhibition of Ras/ERK pathway in a friend murine leukemia cell line. *Oncogene*. 2000;19:1500–1508.
67. Ferreira R, Ohneda K, Yamamoto M, et al. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol Cell Biol*. 2005;25:1215–1227.
68. Takahashi S, Onodera K, Motohashi H, et al. Arrest in primitive erythroid cell development caused by promoter-specific disruption of the GATA-1 gene. *J Biol Chem*. 1997;272:12611–12615.
69. Takahashi S, Komeno T, Suwabe N, et al. Role of GATA-1 in proliferation and differentiation of definitive erythroid and megakaryocytic cells in vivo. *Blood*. 1998;92:434–442.
70. Mukai HY, Suzuki M, Nagano M, et al. Establishment of erythroleukemic GAK14 cells and characterization of GATA1 N-terminal domain. *Genes Cells*. 2013;18:886–898.
71. Weiss MJ, Yu C, Orkin SH. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol*. 1997;17:1642–1651.
72. Nichols KE, Crispino JD, Poncz M, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet*. 2000;24:266–270.
73. Letting DL, Rakowski C, Weiss MJ, et al. Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. *Mol Cell Biol*. 2003;23:1334–1340.
74. Welch JJ, Watts JA, Vakoc CR, et al. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood*. 2004;104:3136–3147.
75. Shimizu R, Engel JD, Yamamoto M. GATA1-related leukaemias. *Nat Rev Cancer*. 2008;8:279–287.
76. Beck D, Thoms JA, Perera D, et al. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood*. 2013;122:e12–e22.
77. Loughran SJ, Kruse EA, Hacking DF, et al. The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat Immunol*. 2008;9:810–819.
78. Taoudi S, Bee T, Hilton A, et al. ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification. *Genes Dev*. 2011;25:251–262.
79. Knudsen KJ, Rehn M, Hasemann MS, et al. ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. *Genes Dev*. 2015;29:1915–1929.
80. Wilson NK, Foster SD, Wang X, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide

- analysis of ten major transcriptional regulators. *Cell Stem Cell*. 2010;7:532–544.
81. Martens JH. Acute myeloid leukemia: a central role for the ETS factor ERG. *Int J Biochem Cell Biol*. 2011;43:1413–1416.
 82. Baldus CD, Burmeister T, Martus P, et al. High expression of the ETS transcription factor ERG predicts adverse outcome in acute T-lymphoblastic leukemia in adults. *J Clin Oncol*. 2006;24:4714–4720.
 83. Birger Y, Goldberg L, Chlon TM, et al. Perturbation of fetal hematopoiesis in a mouse model of Down syndrome's transient myeloproliferative disorder. *Blood*. 2013;122:988–998.
 84. Stankiewicz MJ, Crispino JD. ETS2 and ERG promote megakaryopoiesis and synergize with alterations in GATA-1 to immortalize hematopoietic progenitor cells. *Blood*. 2009;113:3337–3347.
 85. Carmichael CL, Metcalf D, Henley KJ, et al. Hematopoietic over-expression of the transcription factor Erg induces lymphoid and erythro-megakaryocytic leukemia. *Proc Natl Acad Sci U S A*. 2012;109:15437–15442.
 86. Tang JZ, Carmichael CL, Shi W, et al. Transposon mutagenesis reveals cooperation of ETS family transcription factors with signaling pathways in erythro-megakaryocytic leukemia. *Proc Natl Acad Sci U S A*. 2013;110:6091–6096.
 87. Lengerke C, Daley GQ. Caudal genes in blood development and leukemia. *Ann N Y Acad Sci*. 2012;1266:47–54.
 88. Pineault N, Helgason CD, Lawrence HJ, et al. Differential expression of Hox, Meis1, and Pbx1 genes in primitive cells throughout murine hematopoietic ontogeny. *Exp Hematol*. 2002;30:49–57.
 89. Davidson AJ, Ernst P, Wang Y, et al. *Cdx4* mutants fail to specify blood progenitors and can be rescued by multiple *hox* genes. *Nature*. 2003;425:300–306.
 90. Koo S, Huntly BJ, Wang Y, et al. *Cdx4* is dispensable for murine adult hematopoietic stem cells but promotes MLL-AF9-mediated leukemogenesis. *Haematologica*. 2010;95:1642–1650.
 91. Bansal D, Scholl C, Fröhling S, et al. *Cdx4* dysregulates Hox gene expression and generates acute myeloid leukemia alone and in cooperation with *Meis1a* in a murine model. *Proc Natl Acad Sci U S A*. 2006;103:16924–16929.
 92. Ben-David Y, Giddens EB, Letwin K, et al. Erythroleukemia induction by Friend murine leukemia virus: insertional activation of a new member of the *ets* gene family, *Fli-1*, closely linked to *c-ets-1*. *Genes Dev*. 1991;5:908–918.
 93. Li Y, Luo H, Liu T, et al. The *ets* transcription factor *Fli-1* in development, cancer and disease. *Oncogene*. 2015;34:2022–2031.
 94. Klemsz MJ, Maki RA, Papayannopoulou T, et al. Characterization of the *ets* oncogene family member, *fli-1*. *J Biol Chem*. 1993;268:5769–5773.
 95. Athanasiou M, Mavrothalassitis G, Sun-Hoffman L, et al. *Fli-1* is a suppressor of erythroid differentiation in human hematopoietic cells. *Leukemia*. 2000;14:439–445.
 96. Delattre O, Zucman J, Plougastel B, et al. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature*. 1992;359:162–165.
 97. Li YJ, Zhao X, Vecchiarelli-Federico LM, et al. Drug-mediated inhibition of *Fli-1* for the treatment of leukemia. *Blood Cancer J*. 2012;2:e54.
 98. Liu T, Xia L, Yao Y, et al. Identification of diterpenoid compounds that interfere with *Fli-1* DNA binding to suppress leukemogenesis. *Cell Death Dis*. 2019;10:117.
 99. Wagner K, Zhang P, Rosenbauer F, et al. Absence of the transcription factor CCAAT enhancer binding protein alpha results in loss of myeloid identity in *bcr/abl*-induced malignancy. *Proc Natl Acad Sci U S A*. 2006;103:6338–6343.
 100. Tsuruta-Kishino T, Koya J, Kataoka K, et al. Loss of p53 induces leukemic transformation in a murine model of *Jak2 V617F*-driven polycythemia vera. *Oncogene*. 2017;36:3300–3311.
 101. Di Genua C, Valletta S, Buono M, et al. *C/EBPα* and *GATA-2* mutations induce bilineage acute erythroid leukemia through transformation of a neomorphic neutrophil-erythroid progenitor. *Cancer Cell*. 2020;37:690–704.e8.
 102. Pique-Borras MR, Otzen Bagger F, Filgueira Bezerra M, et al. Transformation mechanisms of the *Nfia-ETO2* fusion gene associated with pediatric pure acute erythroleukemia. Paper presented at: 61st ASH Annual Meeting & Exposition. Orlando, FL: Blood. 2020
 103. Iacobucci I, Qu, C, Varotto E, et al. Modeling and targeting of erythroleukemia by hematopoietic genome editing. *Blood*. 2021 January 19. [Epub ahead of print]. doi: 10.1182/blood.202009103
 104. Kelly LM, Gilliland DG. Genetics of myeloid leukemias. *Annu Rev Genomics Hum Genet*. 2002;3:179–198.
 105. Suh HC, Gooya J, Renn K, et al. *C/EBPα* determines hematopoietic cell fate in multipotential progenitor cells by inhibiting erythroid differentiation and inducing myeloid differentiation. *Blood*. 2006;107:4308–4316.
 106. Abdel-Wahab O, Manshouri T, Patel J, et al. Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Res*. 2010;70:447–452.
 107. Joshi SK, Davare MA, Druker BJ, et al. Revisiting NTRKs as an emerging oncogene in hematological malignancies. *Leukemia*. 2019;33:2563–2574.
 108. Taylor J, Pavlick D, Yoshimi A, et al. Oncogenic TRK fusions are amenable to inhibition in hematologic malignancies. *J Clin Invest*. 2018;128:3819–3825.
 109. Cocco E, Scaltriti M, Drilon A. NTRK fusion-positive cancers and TRK inhibitor therapy. *Nat Rev Clin Oncol*. 2018;15:731–747.
 110. Starnes LM, Sorrentino A, Pelosi E, et al. *NFI-A* directs the fate of hematopoietic progenitors to the erythroid or granulocytic lineage and controls beta-globin and G-CSF receptor expression. *Blood*. 2009;114:1753–1763.
 111. Steinauer N, Guo C, Zhang J. Emerging roles of *MTG16* in cell-fate control of hematopoietic stem cells and cancer. *Stem Cells Int*. 2017;2017:12.
 112. Thirant C, Ignacimoutou C, Lopez CK, et al. *ETO2-GLIS2* hijacks transcriptional complexes to drive cellular identity and self-renewal in pediatric acute megakaryoblastic leukemia. *Cancer Cell*. 2017;31:452–465.
 113. Grossmann V, Tiacci E, Holmes AB, et al. Whole-exome sequencing identifies somatic mutations of *BCOR* in acute myeloid leukemia with normal karyotype. *Blood*. 2011;118:6153–6163.
 114. Sportoletti P, Sorcini D, Guzman AG, et al. *Bcor* deficiency perturbs erythro-megakaryopoiesis and cooperates with *Dnmt3a* loss in acute erythroid leukemia onset in mice. *Leukemia (Advanced Online)*. 2020 November 06. [Epub ahead of print]. doi: 10.1038/s41375-020-01075-3
 115. Ren J, Finney R, Ni K, et al. The chromatin remodeling protein *Lsh* alters nucleosome occupancy at putative enhancers and modulates binding of lineage specific transcription factors. *Epigenetics*. 2019;14:277–293.
 116. Brumbaugh J, Kim IS, Ji F, et al. Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. *Nat Cell Biol*. 2019;21:1449–1461.
 117. Leonards K, Almosaillekh M, Tauchmann S, et al. Nuclear interacting SET domain protein 1 inactivation impairs *GATA1*-regulated erythroid differentiation and causes erythroleukemia. *Nat Commun*. 2020;11:2807.
 118. Briones V, Muegge K. The ghosts in the machine: DNA methylation and the mystery of differentiation. *Biochim Biophys Acta*. 2012;1819:757–762.
 119. Sun LQ, Lee DW, Zhang Q, et al. Growth retardation and premature aging phenotypes in mice with disruption of the *SNF2*-like gene, *PASG*. *Genes Dev*. 2004;18:1035–1046.
 120. Fan T, Schmidtman A, Xi S, et al. DNA hypomethylation caused by *Lsh* deletion promotes erythroleukemia development. *Epigenetics*. 2008;3:134–142.
 121. Lee DW, Zhang K, Ning ZQ, et al. Proliferation-associated *SNF2*-like gene (*PASG*): a *SNF2* family member altered in leukemia. *Cancer Res*. 2000;60:3612–3622.
 122. Pastore F, Levine RL. Epigenetic regulators and their impact on therapy in acute myeloid leukemia. *Haematologica*. 2016;101:269–278.
 123. Wagner EJ, Carpenter PB. Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol*. 2012;13:115–126.
 124. Bennett RL, Swaroop A, Troche C, Licht JD. The role of nuclear receptor-binding SET domain family histone lysine methyltransferases in cancer. *Cold Spring Harb Perspect Med*. 2017;7:a026708.
 125. Hollink IH, van den Heuvel-Eibrink MM, Arentsen-Peters ST, et al. *NUP98/NSD1* characterizes a novel poor prognostic group in acute myeloid leukemia with a distinct *HOX* gene expression pattern. *Blood*. 2011;118:3645–3656.
 126. Mohammad F, Helin K. Oncohistones: drivers of pediatric cancers. *Genes Dev*. 2017;31:2313–2324.
 127. Lu C, Jain SU, Hoelper D, et al. Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science*. 2016;352:844–849.
 128. Gutiérrez L, Caballero N, Fernández-Calleja L, et al. Regulation of *GATA1* levels in erythropoiesis. *IUBMB Life*. 2020;72:89–105.

129. DeVilbiss AW, Tanimura N, McIver SC, et al. Navigating transcriptional coregulator ensembles to establish genetic networks: a GATA factor perspective. *Curr Top Dev Biol.* 2016;118:205–244.
130. Takeda J, Yoshida K, Nannya Y, et al. Novel molecular pathogenesis and therapeutic target in acute erythroid leukemia. Paper presented at: Annual Meeting of the American Society of Hematology (ASH); December 7–10, 2019; Orlando: Blood.
131. Adélaïde J, Cervera N, Guille A, et al. Gains of EPOR and ERG genes in adult erythroleukaemia. *Br J Haematol.* 2020;189:e174–e177.
132. Garnett C, Cruz Hernandez D, Vyas P. GATA1 and cooperating mutations in myeloid leukaemia of Down syndrome. *IUBMB Life.* 2020;72:119–130.
133. Asai T, Liu Y, Bae N, et al. The p53 tumor suppressor protein regulates hematopoietic stem cell fate. *J Cell Physiol.* 2011;226:2215–2221.
134. Prokocimer M, Molchadsky A, Rotter V. Dysfunctional diversity of p53 proteins in adult acute myeloid leukemia: projections on diagnostic workup and therapy. *Blood.* 2017;130:699–712.
135. Schneider RK, Schenone M, Ferreira MV, et al. Rps14 haploinsufficiency causes a block in erythroid differentiation mediated by S100A8 and S100A9. *Nat Med.* 2016;22:288–297.
136. Le Goff S, Boussaid I, Floquet C, et al. p53 activation during ribosome biogenesis regulates normal erythroid differentiation. *Blood.* 2021;137:89–102.
137. Chen S, Wang Q, Yu H, et al. Mutant p53 drives clonal hematopoiesis through modulating epigenetic pathway. *Nat Commun.* 2019;10:5649.
138. Stein Y, Rotter V, Aloni-Grinstein R. Gain-of-function mutant p53: all the roads lead to tumorigenesis. *Int J Mol Sci.* 2019;20:6197.
139. Li D, Marchenko ND, Schulz R, et al. Functional inactivation of endogenous MDM2 and CHIP by HSP90 causes aberrant stabilization of mutant p53 in human cancer cells. *Mol Cancer Res.* 2011;9:577–588.
140. Schulz-Heddergott R, Moll UM. Gain-of-function (GOF) mutant p53 as actionable therapeutic target. *Cancers (Basel).* 2018;10:188.
141. Boettcher S, Miller PG, Sharma R, et al. A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. *Science.* 2019;365:599–604.
142. Yamamoto S, Iwakuma T. Regulators of oncogenic mutant TP53 gain of function. *Cancers (Basel).* 2018;11:4.
143. Trainor CD, Mas C, Archambault P, et al. GATA-1 associates with and inhibits p53. *Blood.* 2009;114:165–173.
144. Zhu J, Sammons MA, Donahue G, et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature.* 2015;525:206–211.
145. Barbosa K, Li S, Adams PD, et al. The role of TP53 in acute myeloid leukemia: Challenges and opportunities. *Genes Chromosomes Cancer.* 2019;58:875–888.
146. Krivtsov AV, Evans K, Gadrey JY, et al. A menin-MLL inhibitor induces specific chromatin changes and eradicates disease in models of MLL-rearranged leukemia. *Cancer Cell.* 2019;36:660–673.e11.
147. Bernard E, Nannya Y, Hasserjian RP, et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nat Med.* 2020;26:1549–1556.
148. Zhang J, Kong G, Rajagopalan A, et al. p53^{-/-} synergizes with enhanced NrasG12D signaling to transform megakaryocyte-erythroid progenitors in acute myeloid leukemia. *Blood.* 2017;129:358–370.
149. Zhao Z, Zuber J, Diaz-Flores E, et al. p53 loss promotes acute myeloid leukemia by enabling aberrant self-renewal. *Genes Dev.* 2010;24:1389–1402.
150. Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet.* 2018;19:81–92.
151. Lio CJ, Yuita H, Rao A. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. *Blood.* 2019;134:1487–1497.
152. Li Z, Cai X, Cai CL, et al. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood.* 2011;118:4509–4518.
153. Moran-Crusio K, Reavie L, Shih A, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell.* 2011;20:11–24.
154. Challen GA, Sun D, Jeong M, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet.* 2011;44:23–31.
155. Izzo F, Lee SC, Poran A, et al. DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat Genet.* 2020;52:378–387.
156. Yan H, Wang Y, Qu X, et al. Distinct roles for TET family proteins in regulating human erythropoiesis. *Blood.* 2017;129:2002–2012.
157. Qu X, Zhang S, Wang S, et al. TET2 deficiency leads to stem cell factor-dependent clonal expansion of dysfunctional erythroid progenitors. *Blood.* 2018;132:2406–2417.
158. Ketkar S, Verdoni AM, Smith AM, et al. Remethylation of Dnmt3a^{-/-} hematopoietic cells is associated with partial correction of gene dysregulation and reduced myeloid skewing. *Proc Natl Acad Sci U S A.* 2020;117:3123–3134.
159. Yang L, Chen Z, Stout ES, et al. Methylation of a CGATA element inhibits binding and regulation by GATA-1. *Nat Commun.* 2020;11:2560.
160. Jeong JJ, Gu X, Nie J, et al. Cytokine-regulated phosphorylation and activation of TET2 by JAK2 in hematopoiesis. *Cancer Discov.* 2019;9:778–795.
161. Zhang X, Su J, Jeong M, et al. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet.* 2016;48:1014–1023.
162. Castillo-Aguilera O, Depreux P, Halby L, Arimondo PB, Goossens L. DNA methylation targeting: the DNMT/HMT crosstalk challenge. *Biomolecules.* 2017;7:3.
163. Almeida AM, Prebet T, Itzykson R, et al. Clinical outcomes of 217 patients with acute erythroleukemia according to treatment type and line: a retrospective multinational study. *Int J Mol Sci.* 2017;18:837.
164. Weinberg DN, Papillon-Cavanagh S, Chen H, et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature.* 2019;573:281–286.
165. Karia D, Gilbert RCG, Biasutto AJ, et al. The histone H3K4 demethylase JARID1A directly interacts with hematopoietic transcription factor GATA1 in erythroid cells through its second PHD domain. *R Soc Open Sci.* 2020;7:191048.
166. Cardin S, Bilodeau M, Roussy M, et al. Human models of NUP98-KDM5A megakaryocytic leukemia in mice contribute to uncovering new biomarkers and therapeutic vulnerabilities. *Blood Adv.* 2019;3:3307–3321.
167. Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368:2059–2074.
168. Yoshida K, Toki T, Okuno Y, et al. The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat Genet.* 2013;45:1293–1299.
169. Labuhn M, Perkins K, Matzk S, et al. Mechanisms of progression of myeloid preleukemia to transformed myeloid leukemia in children with down syndrome. *Cancer Cell.* 2019;36:340.
170. Mazumdar C, Shen Y, Xavy S, et al. Leukemia-associated cohesin mutants dominantly enforce stem cell programs and impair human hematopoietic progenitor differentiation. *Cell Stem Cell.* 2015;17:675–688.
171. Sasca D, Yun H, Giotopoulos G, et al. Cohesin-dependent regulation of gene expression during differentiation is lost in cohesin-mutated myeloid malignancies. *Blood.* 2019;134:2195–2208.

Nuclear interacting SET domain protein 1 inactivation impairs GATA1-regulated erythroid differentiation and causes erythroleukemia

Katharina Leonards ^{1,2,12}, Marwa Almosailleakh ^{1,2,12}, Samantha Tauchmann^{1,2,12}, Frederik Otzen Bagger^{1,2,3,4,12}, Cécile Thirant ⁵, Sabine Juge^{1,2}, Thomas Bock ⁶, Hélène Méreau², Matheus F. Bezerra ^{1,2,7}, Alexandar Tzankov ⁸, Robert Ivanek ^{2,3}, Régine Losson^{9,13}, Antoine H. F. M. Peters^{10,11}, Thomas Mercher⁵ & Juerg Schwaller ^{1,2} 

The nuclear receptor binding SET domain protein 1 (NSD1) is recurrently mutated in human cancers including acute leukemia. We show that NSD1 knockdown alters erythroid clonogenic growth of human CD34⁺ hematopoietic cells. Ablation of *Nsd1* in the hematopoietic system of mice induces a transplantable erythroleukemia. In vitro differentiation of *Nsd1*^{-/-} erythroblasts is majorly impaired despite abundant expression of GATA1, the transcriptional master regulator of erythropoiesis, and associated with an impaired activation of GATA1-induced targets. Retroviral expression of wildtype NSD1, but not a catalytically-inactive NSD1^{N1918Q} SET-domain mutant induces terminal maturation of *Nsd1*^{-/-} erythroblasts. Despite similar GATA1 protein levels, exogenous NSD1 but not NSD1^{N1918Q} significantly increases the occupancy of GATA1 at target genes and their expression. Notably, exogenous NSD1 reduces the association of GATA1 with the co-repressor SKI, and knockdown of SKI induces differentiation of *Nsd1*^{-/-} erythroblasts. Collectively, we identify the NSD1 methyltransferase as a regulator of GATA1-controlled erythroid differentiation and leukemogenesis.

¹University Children's Hospital Basel, Basel, Switzerland. ²Department of Biomedicine, University of Basel, 4031 Basel, Switzerland. ³Swiss Institute of Bioinformatics, 4031 Basel, Switzerland. ⁴Genomic Medicine, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark. ⁵INSERM U1170, Equipe Labellisée Ligue Contre le Cancer, Gustave Roussy Institute, Université Paris Diderot, Université Paris-Sud, Villejuif 94800, France. ⁶Proteomics Core Facility, Biozentrum University of Basel, Basel, Switzerland. ⁷Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, Brazil. ⁸Institute for Pathology, University Hospital Basel, 4031 Basel, Switzerland. ⁹Institute de Génétique et de Biologie Moléculaire et Cellulaire (I.G.B.M.C.), CNRS/INSERM Université de Strasbourg, BP10142, 67404 Illkirch Cedex, France. ¹⁰Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland. ¹¹Faculty of Sciences, University of Basel, 4056 Basel, Switzerland. ¹²These authors contributed equally: Katharina Leonards, Marwa Almosailleakh, Samantha Tauchmann, Frederik Otzen Bagger. ¹³Deceased: Régine Losson. ✉email: J.Schwaller@unibas.ch

Steady-state erythropoiesis is primarily controlled by erythropoietin (EPO) and other hormones including stem cell factor and glucocorticoids. Different pathways translate external signals to the activation of transcription factors and co-regulators that drive expression programs that define erythroid identity¹. Erythroid differentiation is mainly regulated by a relatively small number of transcriptional regulators, including GATA-1, SCL/TAL1, LMO2, LDB1, KLF1, and GFI1b, that dynamically form multiprotein complexes. However, it remains poorly understood how distinct complexes interact and activate or repress specific gene expression programs².

The best studied erythroid transcription factor is the GATA1 zinc-finger protein. GATA1 was shown to activate its target genes by complexing with SCL/TAL1, the bHLH protein E2A, and the LIM domain containing factors LMO2 and LDB1. GATA1-mediated repression was proposed to be executed by complexes containing FOG1, GFI1b, and/or Polycomb repressive complex 2 (PRC2) proteins^{2,3}. Inactivation studies in mice revealed that GATA1 is an essential master regulator of erythropoiesis as *Gata1*-null embryos died in utero from anemia⁴. Moreover, some adult female mice that are heterozygous for the targeted disruption of the X chromosome-linked *Gata1* promoter region displayed reduced *Gata1* gene expression (*Gata1*^{1.05/X} allele) and developed an early onset erythroleukemia-like disease⁵. This mouse model suggested that reduced *Gata1* activity contributes to leukemogenesis by preventing proper erythroid differentiation. Acute erythroleukemia is a rare form of human acute myeloid leukemia (AML) generally associated with poor outcome⁶. Recent studies started to unravel the genetic AEL landscape but the molecular mechanisms that control the erythroid identity of the tumor cells remain poorly understood⁷.

The nuclear receptor SET domain protein 1 (NSD1) histone methyltransferase was identified as a protein interacting with several nuclear receptors^{8,9}. Mono- and di-methylation of histone 3 lysine 36 (H3K36) and lysine 168 of linker histone 1.5 have been proposed to be the major cellular NSD1 substrates^{10,11}. Multiple studies suggest that *NSD1* can act as a tumor suppressor gene. First, the *NSD1* gene locus is subject to recurrent putative loss-of-function mutations in hematological malignancies and solid cancers^{12–16}. Second, the CpG island promoter of the *NSD1* locus has also been reported to be frequently hyper-methylated in certain human cancers, thereby epigenetically silencing the allele^{17,18}. Third, heterozygous germline point mutations in *NSD1* are the molecular correlate for SOTOS, an overgrowth syndrome with learning disabilities and increased cancer risk^{19,20}. Finally, *NSD1* was identified as putative cancer predisposition gene mediated by rare germline variants and somatic loss-of-heterozygosity (LOH)²¹. However, the mechanism of how NSD1 protects different cell types from malignant transformation remains unknown.

We study the role of NSD1 in steady-state hematopoiesis and leukemia. We observe that reduced *NSD1* expression alters the clonogenic growth of erythroid progenitor cells derived from human CD34⁺ hematopoietic cells. Targeted *Nsd1* gene inactivation during late fetal hematopoiesis in mice leads to malignant accumulation of erythroblasts phenocopying human acute erythroleukemia. Complementation experiments reveal that the NSD1-SET domain is critical for in vitro erythroblast terminal differentiation. In addition, our work suggests that NSD1 controls target gene activation by the erythroid master regulator GATA1, most likely through regulated association with the transcriptional co-repressor SKI. Collectively, we identify NSD1 as a co-regulator of GATA1-controlled terminal erythroid maturation and leukemogenesis.

Results

NSD1 knockdown in human CD34⁺ hematopoietic cells. To address the role of NSD1 in hematopoiesis, we first optimized lentiviral shRNA-mediated knockdown in human CD34⁺ hematopoietic cells (Supplementary Fig. 1a–d). We identified three NSD1 shRNA that reduced the numbers of colonies grown in methylcellulose (MC) containing growth factors including EPO (Fig. 1a, b, Supplementary Fig. 1a, b). Interestingly, whereas very few colonies were generated upon replating of *Ctrl*-shRNA-transduced cells, cells transduced with *NSD1*-shRNA “372” or “353” formed abundant relatively dense reddish colonies (Fig. 1b, c, Supplementary Fig. 1c). These colonies were mostly composed of CD45^{low} cells expressing the transferrin receptor (CD71) and glycophorin-A (GPA) presenting with a proerythroblast-like morphology (Fig. 1d, e, Supplementary Fig. 1d). The cells could however not be further expanded in MC or in liquid cultures. Very similar results were obtained with human cord blood-derived cells (Supplementary Fig. 1e–h). Collectively, these data suggest that NSD1 regulates clonogenic erythroid differentiation of fetal and adult human CD34⁺ hematopoietic cells in vitro.

Ablation of *Nsd1* induces erythroleukemia in mice. To address its role in steady-state hematopoiesis, we inactivated *Nsd1* in mice²². *Nsd1*^{f/f}; *Vav1-iCre*^{tg/tg} transgenic mice (here referred as *Nsd1*^{-/-}) efficiently excised both alleles in cells from different lineages leading to almost undetectable levels of *Nsd1* exon 5 mRNA and protein expression (Supplementary Fig. 2a–g). At the age of 6–25 weeks (median 91 days, *n* = 24) all *Nsd1*^{-/-} mice developed signs of distress, significant hepatosplenomegaly with extensive cellular multi-organ infiltrations, reduced red blood cell (RBC) counts and hemoglobin levels, reticulocytosis, and severe thrombocytopenia (Fig. 2a–h, Supplementary Fig. 2h, Supplementary Table 1). White blood cell (WBC) counts were mostly within the normal range but “unclassified leukocytes” were detected and erythroblast-like cells were seen on peripheral blood smears (Fig. 2i, Supplementary Fig. 2i–k). Transplantation of BM cells from symptomatic *Nsd1*^{-/-} mice (alone or 1:1 in competition with normal cells) rapidly induced the same disease in lethally irradiated wild-type recipients, after a latency of 33 and 42 days, respectively, characterized by hepatosplenomegaly, multi-organ infiltration, anemia, thrombocytopenia, and erythroblasts in the periphery (Fig. 2j, Supplementary Fig. 2j, Supplementary Table 2).

BM and spleen cells from diseased mice expressed modest levels of the transferrin receptor (CD71) and variable amounts of Kit and FcγRII/III, but were negative for CD34, B220, and Sca-1 (Fig. 3a, Supplementary Fig. 3a, b). Erythroid differentiation was defined by staining of CD71 and Ter119 progressing from immature CD71^{low}Ter119^{low} (“R0”) to CD71^{low}Ter119^{high} (“R4”) cells (Supplementary Fig. 3c)²³. Whereas a decrease of the R4 fraction that was mostly evident in the spleens, all diseased *Nsd1*^{-/-} mice significantly accumulated CD71^{dim}/Ter119^{low} cells in BM and spleen (Fig. 3b, Supplementary Fig. 3d). BM cells of diseased *Nsd1*^{-/-} mice formed reduced numbers of colonies in MC with significant reduction of CFU-GM and BFU-E colonies accompanied with sometimes large and abnormally dense, reddish and benzidine-staining positive “BFU-E-like” serially platable colonies, composed of myeloid and erythroid progenitors (Fig. 3c–e).

Diseased *Nsd1*^{-/-} BM contained reduced numbers of lineage marker negative, Kit⁺/Sca-1⁺ (LSK), long-term- (LT-HSC), and short-term repopulating hematopoietic stem cells (ST-HSC) (Fig. 3f). The number of multi-potent progenitors (MPP) and granulocytic-macrophage progenitors (GMP) was also reduced, common myeloid progenitors (CMP) were less affected, and the

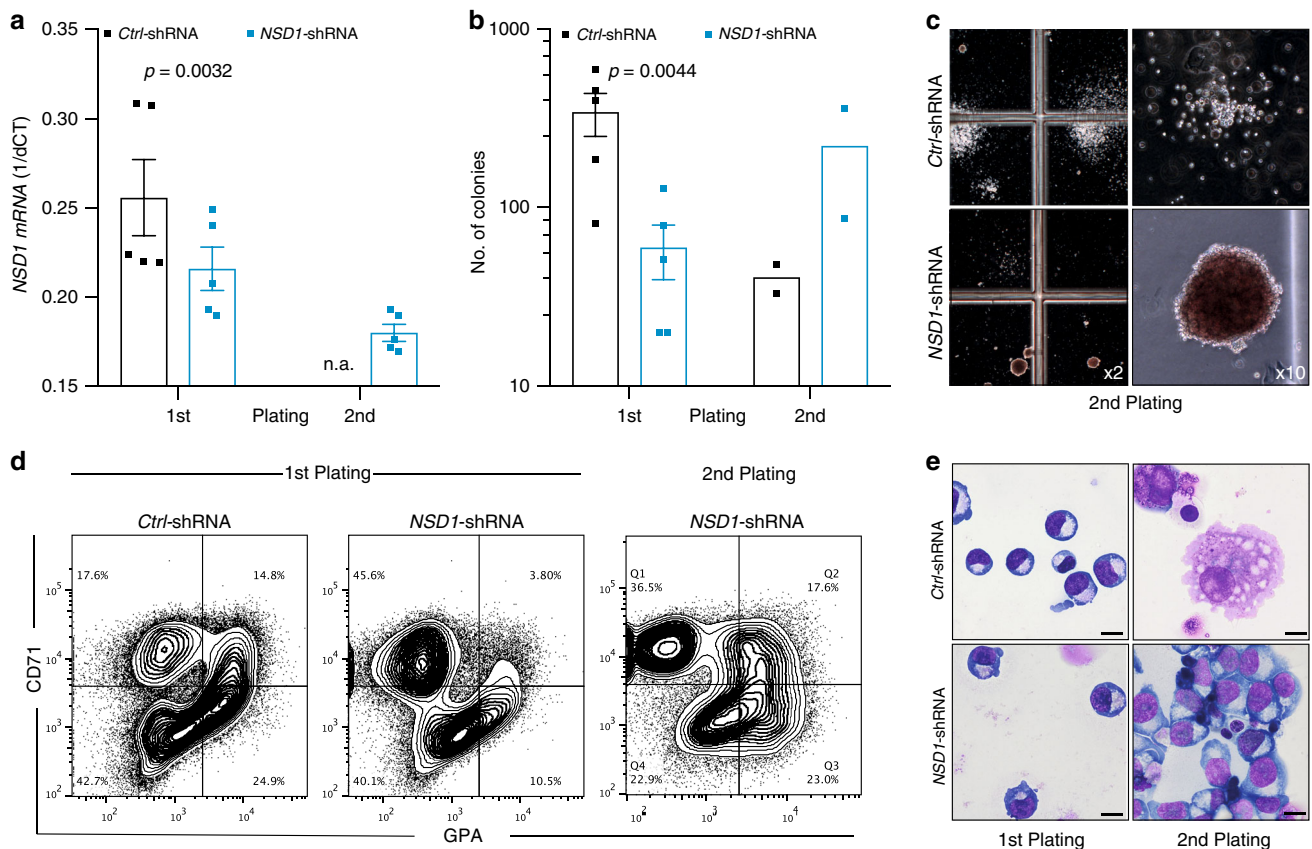


Fig. 1 *NSD1* knockdown alters clonogenic erythroid differentiation of human CD34⁺ hematopoietic cells. **a** Relative *NSD1* mRNA expression (1/dCt) in peripheral blood CD34⁺ cells transduced with *pLKO.1* expressing control shRNA (*Ctrl*) or *NSD1* shRNA (#372) harvested from the first and second plating in growth-factor-containing MC (H4434). Bars represent average relative expression normalized to ($n = 5$ per group). **b** Numbers of colonies formed by 4×10^4 peripheral CD34⁺ cells transduced with *pLKO.1* expressing control shRNA (*Ctrl*) or *NSD1* shRNA in the first plate ($n = 5$) and upon replating ($n = 2$) in growth-factor-containing MC (H4434). **c** Representative images of colonies formed in MC (H4434) by 4×10^4 peripheral CD34⁺ cells transduced with *pLKO.1* expressing control or *NSD1* shRNA. **d** Flow cytometric analysis of cells harvested from the first and second plating in MC (H4434) revealed accumulation of CD71^{high} and glycoprotein A (GPA)⁻ cells upon replating. The plots represent one out of three independent experiments. **e** Representative images of Wright Giemsa-stained cytopsin preparations from control shRNA (*Ctrl*) or *NSD1* shRNA-expressing CD34⁺ cells harvested from the MC (H4434) cultures after the first and second plating, illustrating the overall predominance of cells with erythroblast morphology upon replating (one out of three experiments) ($\times 1000$, the size bar = 10 μm). 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**, **b** was tested with paired two-tailed *t*-test.

number of other progenitors (pre-GM, pre-MegE, MkP) did not significantly differ from littermate controls (Supplementary Fig. 3e, f).

As *Vav1*-promoter driven Cre expression resulted in significant reduction of *Nsd1* expression as early as at E13.5 of development, we also analyzed the impact of *Nsd1* inactivation during fetal liver hematopoiesis (Supplementary Fig. 3g, h)²⁴. Hereby, we observed clusters of large cells with a dark-blue cytoplasm on E19.5 fetal liver sections (Fig. 3g). MC cultures did not display any significant differences in total colony number; however, E19.5 *Nsd1*^{-/-} fetal liver cells formed dense colonies of mostly CD71⁺ cells (Fig. 3h, i) (Supplementary Fig. 3i) resembling those formed by diseased adult BM (Fig. 3d).

Comparison of the transcriptomes of BM cells from symptomatic *Nsd1*^{-/-} mice ($n = 5$) with littermate controls ($n = 3$) (Supplementary Fig. 3j, Supplementary Data 1) revealed significant upregulation of 1705 (of 18301 genes, 9.3%) and downregulation of 1558 (8.5%) genes (FDR < 0.05). Gene set enrichment analysis (GSEA) revealed positive correlations between differentially expressed genes (DEGs) with a previously characterized signature of MYC targets and negative correlation with a gene signature of murine terminal erythroid differentiation (Fig. 3j)²⁵.

Collectively, these data show that inactivation of *Nsd1* in the hematopoietic system induces an erythroleukemia-like disease in mice²⁶.

Aberrant regulation of GATA1 in *Nsd1*^{-/-} erythroblasts. To elucidate the role of *Nsd1* in erythroleukemia, we first established culture conditions for primary erythroblasts that maintain cytokine dependency as well as differentiation potential towards enucleated erythrocytes (Fig. 4a)²⁷. Growth of fetal liver (FL)-derived *Nsd1*^{-/-} erythroblasts did not significantly differ from littermate controls in maintenance medium (“MM”, containing dexamethasone, hIGF1, cholesterol, and hEPO). In contrast, differentiation of *Nsd1*^{-/-} cells was significantly impaired while control cells completely matured in mSCF and hEPO containing differentiation-inducing medium (“DM”) (Fig. 4b–d, Supplementary Fig. 4a–d).

Erythropoiesis is controlled by the transcriptional master regulator GATA1²⁸. While *Nsd1*^{-/-} erythroblasts expressed reduced *Gata1* mRNA levels in DM, GATA1 protein expression remained abundant in maintenance medium and during induced differentiation (Fig. 5a, b, Supplementary Fig. 4e, f). Interestingly, retroviral expression of a full-length murine *Gata1* cDNA resulting

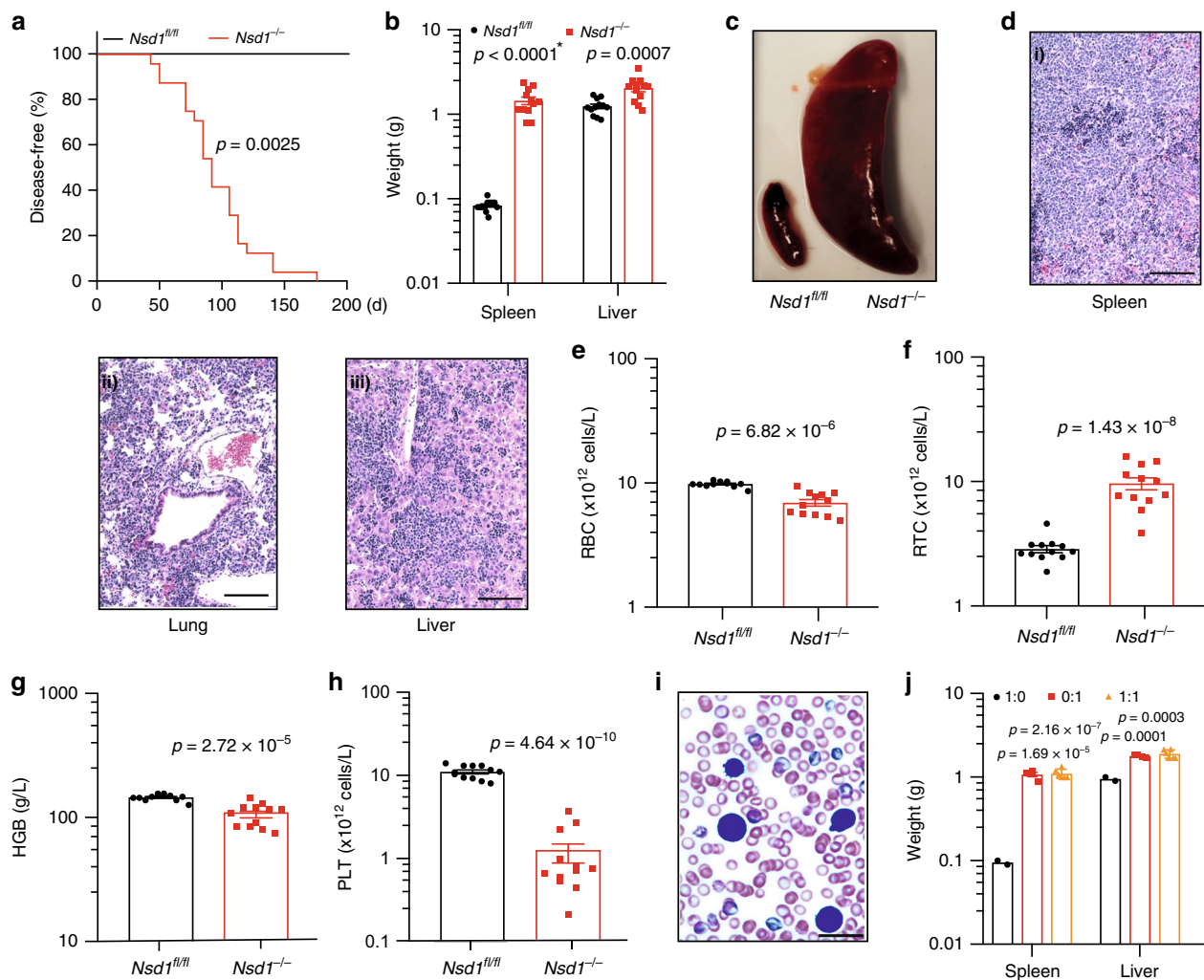


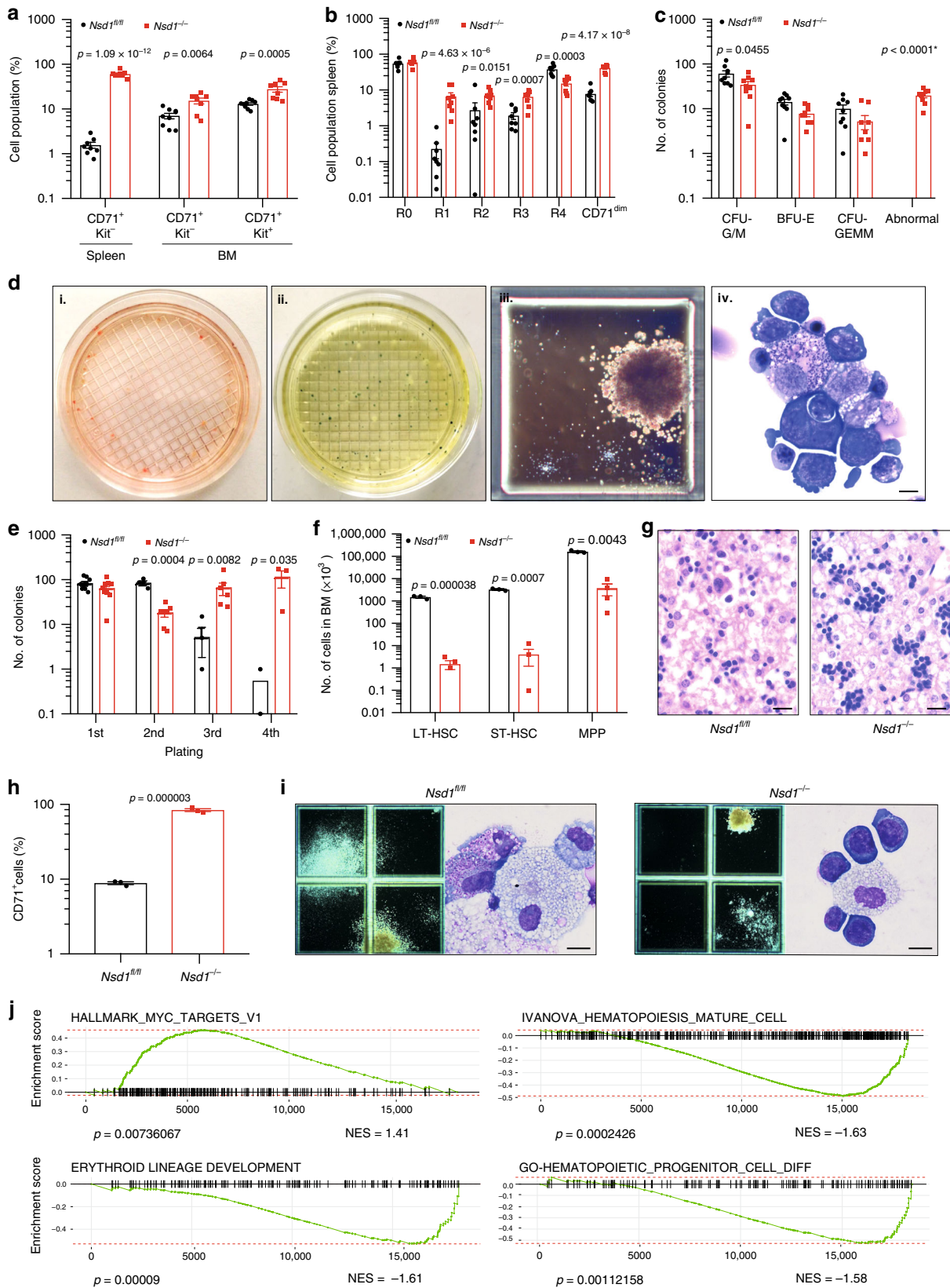
Fig. 2 Hematopoietic ablation of *Nsd1* leads to a fully penetrant and transplantable leukemia-like disease in mice. **a** Kaplan Meier plot of disease-free survival of *Nsd1^{fl/fl}* (n = 12, black line) and *Nsd1^{-/-}* (n = 24; red line) mice. Median survival of *Nsd1^{-/-}* mice was 91 days; *Nsd1^{fl/fl}* mice did not develop any disease. **b** Spleen and liver weight of *Nsd1^{fl/fl}* and symptomatic *Nsd1^{-/-}* mice in grams (n = 12 per group) (* indicates a p value smaller than 1×10^{-15}). **c** Representative image of spleens of *Nsd1^{fl/fl}* (left) and diseased *Nsd1^{-/-}* (right) mice. **d** Representative images (from 1 out of 24 mice) of HE-stained sections ($\times 200$, size bars = 50 μm) of (i) spleen, (ii) lung, and (iii) liver of diseased *Nsd1^{-/-}* mice showing significant cell infiltrations in all organs. **e** Peripheral red blood cell counts (RBC, $\times 10^{12}$ cells/L), **f** reticulocyte counts (RTC, $\times 10^{12}$ cells/L), **g** hemoglobin levels (HGB, g/L), and **h** platelet counts (PLT, $\times 10^{12}$ cells/L) in diseased *Nsd1^{-/-}* compared to *Nsd1^{fl/fl}* littermate controls (n = 12 per group). **i** Representative image of a Wright Giemsa-stained peripheral blood smear of a symptomatic *Nsd1^{-/-}* mouse with the presence of an erythroblast (from 1 out of 24 mice, $\times 600$, the size bar = 10 μm). **j** Spleen and liver weight in grams of WT mice transplanted with whole BM from diseased *Nsd1^{-/-}* mice (red bars) (n = 4) or *Nsd1^{fl/fl}* littermate controls (black bars) (n = 2) or in a 1:1 mixture of *Nsd1^{-/-}* (CD45.2) and B6.SJL (CD45.1) cells (orange bars) (n = 6). 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** was tested with log-rank Mantel Cox test, and in **b**, **e-h**, **j** with an unpaired two-tailed t-test.

in 2.6–3.2-fold increased level of exogenous protein not only reduced the number of aberrant colonies in MC but also induced terminal maturation of *Nsd1^{-/-}* erythroblasts (Fig. 5c–e, Supplementary Fig. 4g).

To study the impact of NSD1 on GATA1 transcription factor activity we compared the expression of previously proposed GATA1 target genes. Overexpression of GATA1 promoted induced expression of several genes in *Nsd1^{-/-}* cells, including *Hba-A*, *Hbb-B*, and *Bcl2l1*, which are normally activated by GATA1 during differentiation (Fig. 5f, Supplementary Fig. 4h, i). In contrast, exogenous *Gata1* did not affect expression of *Spi1* but further reduced expression of *Kit* and *Gata2* known to be downregulated by GATA1 during normal erythroid differentiation (Fig. 5g, Supplementary Fig. 4j, k). Together, these data

suggest that the activation of GATA1-controlled target genes during erythroid differentiation is modulated by NSD1²⁹.

NSD1-SET is essential for in vitro erythroblast maturation. To address how NSD1 controls erythroid differentiation we compared the effects of expression of wild type (*Nsd1*) or a catalytically inactive SET domain mutant (*Nsd1^{N1918Q}*) in *Nsd1^{-/-}* erythroblasts (Fig. 6a)³⁰. Expression of *Nsd1* but not *Nsd1^{N1918Q}* significantly rescued terminal maturation as illustrated by cellular morphology, a shift of CD71 and Ter119 surface expression, formation of reddish cell pellets, reduced proliferation, and reduced colony formation in MC (Fig. 6b–f, Supplementary Fig. 5a–c).



To address the molecular mechanisms, we measured DEGs and total proteome expression in BM-derived $Nsd1^{-/-}$ erythroblasts retrovirally expressing $Nsd1$ or $Nsd1^{N1918Q}$ expanded in MM and kept for 24h in DM (Fig. 7a). After 24 h in DM, expression of

about 2% of the genes significantly ($p < 0.05$) increased (270 of 15804) or decreased (318 of 15804) in cells expressing $Nsd1$ compared to $Nsd1^{N1918Q}$ (Fig. 7b, Supplementary Data 2). Among more highly expressed genes we found the cell cycle

Fig. 3 Cellular and molecular characterization of the erythroleukemia-like disease of *Nsd1*^{-/-} mice. **a** CD71⁺ and Kit⁺ cell populations (given in %) in single-cell suspensions of spleen and BM of healthy *Nsd1*^{fl/fl} (spleen, *n* = 8, BM, *n* = 9, black bars) and diseased *Nsd1*^{-/-} mice (spleen, *n* = 8; BM, *n* = 8, red bars). **b** Comparative flow cytometric analysis of erythroid maturation (R1-R4) of single-cell suspensions of total BM of healthy *Nsd1*^{fl/fl} (*n* = 9, black bars) and diseased *Nsd1*^{-/-} mice (*n* = 9, red bars). **c** Colony types formed by 4×10^4 BM cells of *Nsd1*^{fl/fl} (*n* = 9, black bars) and *Nsd1*^{-/-} mice (*n* = 9, red bars) in growth-factor-containing MC (M3434). * indicating a *p* value smaller than 1×10^{-15} . **d** Representative images (illustrating one out of four experiments) of MC (M3434) cultures of *Nsd1*^{-/-} BM cells demonstrating (i) abnormal large red colonies, (ii) partially benzidine-staining-positive colonies, (iii) a large dense isolated colony, and (iv) Wright Giemsa-stained cytospin of an isolated colony ($\times 4$ and $\times 1000$, size bar = $10 \mu\text{m}$). **e** Number of colonies in four consecutive rounds of plating in MC (M3434) formed by 4×10^4 BM cells of *Nsd1*^{fl/fl} (black dots; 1st plating: *n* = 9, 2nd plating: *n* = 4, 3rd plating: *n* = 3, 4th plating: *n* = 2) and *Nsd1*^{-/-} mice (red squares; 1st plating: *n* = 9, 2nd plating: *n* = 7, 3rd plating: *n* = 6, 4th plating: *n* = 3). **f** Number of LT-HSC, ST-HSC, and MPP ($\times 10^4$) in lineage-marker-depleted single-cell BM suspensions of *Nsd1*^{fl/fl} (*n* = 3, black bars) and *Nsd1*^{-/-} mice (*n* = 4, red bars) relative to the total number of lineage-depleted cells obtained during each procedure. **g** Representative HE-stained biopsies of E19.5 fetal livers from a *Nsd1*^{fl/fl} (left panel, illustrating one out of two experiments) and *Nsd1*^{-/-} (right panel, illustrating one out of four experiments) mouse ($\times 400$, size bar = $10 \mu\text{m}$). **h** CD71⁺ cells (%) in E19.5 fetal livers of *Nsd1*^{fl/fl} (*n* = 3, black bar) and *Nsd1*^{-/-} (*n* = 3, red bar) mice. **i** Representative images of colonies in MC cultures (M3434) and Wright Giemsa-stained cytospin preparations from 4×10^4 E19.5 fetal liver-derived hematopoietic cells of *Nsd1*^{fl/fl} (left panels, illustrating one out of three experiments) and *Nsd1*^{-/-} (right panels, illustrating one out of three experiments) mice ($\times 2$ and $\times 1000$, size bars = $10 \mu\text{m}$). **j** Gene set enrichment analysis (GSEA) (weighted Kolmogorov-Smirnov-like statistics, two-sided, with adjustment for multiple comparisons) of selected signatures of differentially expressed gene between *Nsd1*^{-/-} mice (*n* = 5) and littermate controls (*n* = 3). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in **a-c**, **e**, **f**, **h** was tested with unpaired two-tailed *t*-test.

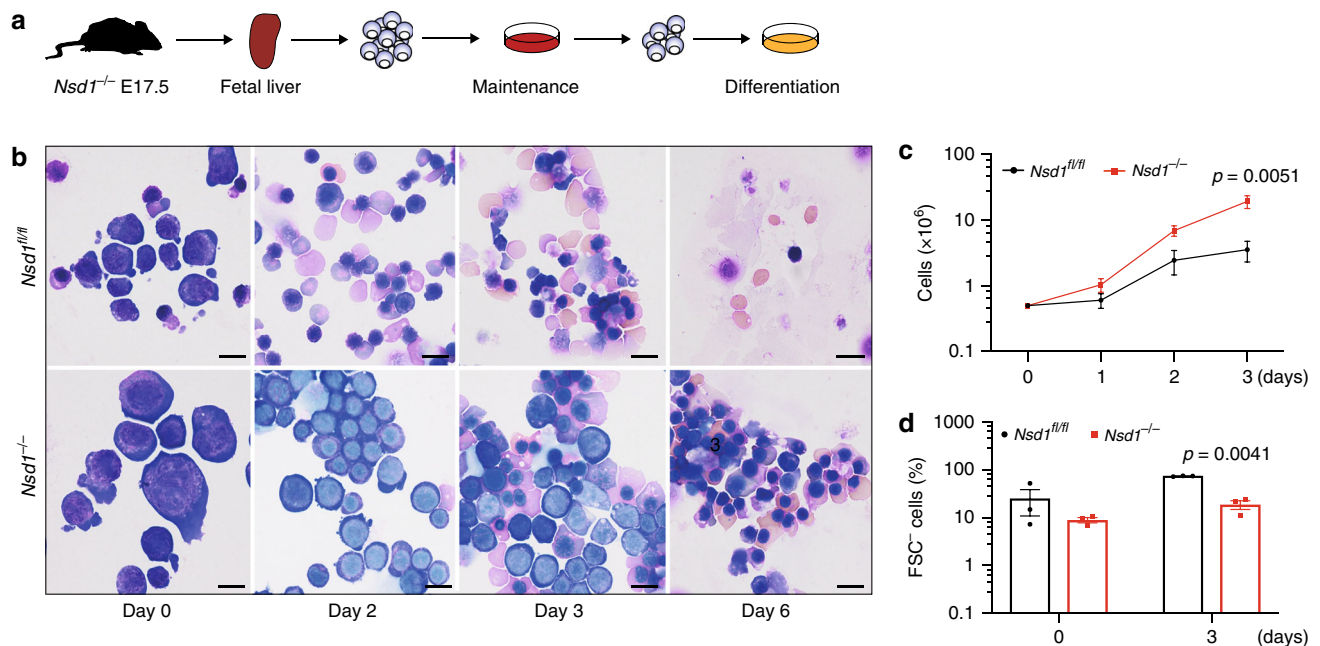


Fig. 4 Impaired erythroid maturation of *Nsd1*^{-/-} erythroblasts. **a** Experimental setup: lineage-depleted E17.5 fetal liver cells of *Nsd1*^{fl/fl} and *Nsd1*^{-/-} mice were grown in maintenance medium (>6 days) before induction of maturation in differentiation cultures. **b** Representative images (illustrating one out of two maturation experiments) of Wright Giemsa-stained cytospin preparations of E17.5 *Nsd1*^{fl/fl} and *Nsd1*^{-/-} fetal liver-derived erythroblasts expanded in maintenance medium (day 0) and induced to mature in differentiation medium, shown at days 2, 3, and 6 ($\times 1000$, the size bar = $10 \mu\text{m}$). **c** Growth of fetal liver-derived *Nsd1*^{fl/fl} (*n* = 6, black line) and *Nsd1*^{-/-} (*n* = 4, red line) erythroblasts in differentiation medium. Living cells were counted using trypan blue exclusion. **d** Forward scatter-negative (FSC⁻) *Nsd1*^{fl/fl} (black bars) and *Nsd1*^{-/-} (red bars) living cells (%) before (day 0) and after 3 days in differentiation medium (*n* = 3). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in **c**, **d** was tested with unpaired two-tailed *t*-test.

regulator *Cdk2* and the epigenetic regulator *Kmt5a* (*Setd8*) previously shown to be important for erythroid differentiation³¹⁻³⁴. Among the lower expressed genes, we found the transcription factor *Gata2* and the RNA-binding protein *Zfp36l2* known for their role in regulating self-renewal of hematopoietic stem and erythroid progenitor cells^{35,36}. GSEA revealed significant (*p* < 0.0001) positive correlations between DEGs of *Nsd1*^{-/-} cells expressing *Nsd1* for 24 h in DM with signatures linked to heme metabolism, erythroid differentiation, and putative GATA1 target genes, and inverse correlation with a negative regulatory differentiation signature (Fig. 7c, Supplementary

Data 3). DEG between cells expressing *Nsd1* or the *Nsd1*^{N1918Q} mutant kept 24 h in DM did not only positively correlate with the expression signature of murine terminal erythroid differentiation but also with signatures related to heme metabolism and cell cycle checkpoints, and negatively with signatures related to hematopoietic stemness (Fig. 7d, Supplementary Data 4). In parallel to differential mRNA expression we also determined changes of the global proteome. *Nsd1*^{-/-} erythroblasts expressing wild-type *Nsd1* for 24 h in DM expressed significant higher protein levels of several proposed GATA1 targets like hemoglobin (HBA, HBB1, HBE), exportin 7 (XPO7), or mitoferrin (MFRN1) (Fig. 7e,

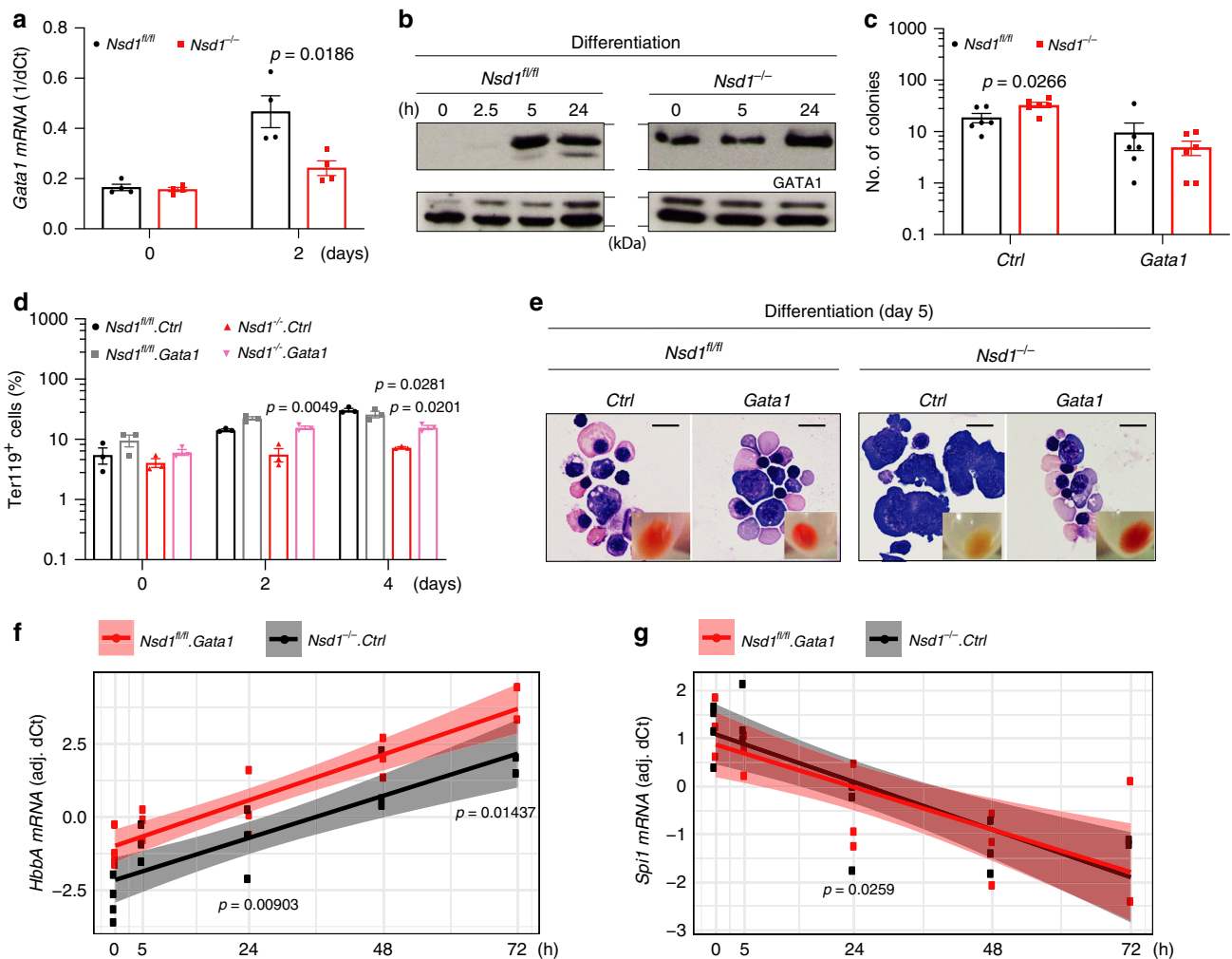


Fig. 5 Aberrant regulation of GATA1 expression in *Nsd1*^{-/-} erythroblasts. **a** Relative *Gata1* mRNA expression levels (1/dCt) in BM-derived erythroblasts from *Nsd1*^{fl/fl} (*n* = 4, black bars) and *Nsd1*^{-/-} mice (*n* = 4, red bars) in maintenance medium (day 0) and after 2 days in differentiation medium. Ct values were normalized to *Gapdh* expression. **b** GATA1 protein levels in *Nsd1*^{fl/fl} (left panels) and *Nsd1*^{-/-} (right panels) BM-derived erythroblasts expanded in maintenance medium (0 h) and in differentiation medium (2.5, 5, and 24 h). LAMIN-A/C was used as immunoblot loading control for nuclear proteins (one out of two experiments). **c** Number of colonies formed by 5 × 10³ lineage-marker-depleted BM-derived erythroblasts in MC (M3434) from *Nsd1*^{fl/fl} (black bars) and *Nsd1*^{-/-} mice (red bars) transduced with *pMSCV-puro* (*Ctrl*) or *pMSCV-mGata1-puro* (*Gata1*) (*n* = 6 per group). **d** Ter119 expression (Ter119⁺, in %) in maintenance medium (0 h) and after 2 and 4 days in differentiation medium of *Nsd1*^{fl/fl} (black and gray bars) and *Nsd1*^{-/-} (red and pink bars) BM-derived erythroblasts transduced with control virus (*Ctrl*, black and red bars) or *Gata1*-expressing virus (*Gata1*, gray or pink bars) (*n* = 3 per group). **e** Representative images (one out of two experiments) of Wright Giemsa-stained cytopsin preparations and cell pellets (small insets) of *Nsd1*^{fl/fl} and *Nsd1*^{-/-} BM-derived erythroblasts transduced with control virus (*Ctrl*) or *Gata1*-expressing virus (*Gata1*) after 5 days in differentiation medium (×600, size bars = 10 μm). **f** *HbbA* and **g** *Spi1* mRNA levels in BM-derived *Nsd1*^{-/-} BM-derived erythroblasts transduced with control virus (*Ctrl*, black dots) or *Gata1*-expressing virus (*Gata1*, red dots) measured 0, 5, 24, 48, and 72 h in differentiation medium. Values are residual ΔCT relative to *Gapdh*, after adjustment for effect of individual mouse (Tukey test for difference in expression between transductions at 24 h in linear model with interaction between time *a* and transduction, adjusting for effect of mouse, two-sided, adjusted for multiple comparisons). 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**, **c** was tested with unpaired two-tailed *t*-test. Statistical significance in **d** was tested with either unpaired (*Nsd1*^{fl/fl}.*Gata1* vs. *Nsd1*^{-/-}.*Gata1*) or paired (*Nsd1*^{-/-}.*Ctrl* vs. *Nsd1*^{-/-}.*Gata1*) two-tailed *t*-test. Statistical significance in **f**, **g** was tested using Tukey test for difference in expression between transductions at 24 h in linear model with interaction between time *a* and transduction, adjusting for effect of mouse.

Supplementary Fig. 5d, Supplementary Data 5, and Supplementary Table 3^{37,38}. Collectively, these observations suggest that the catalytic activity of NSD1 is essential for terminal erythroid maturation and regulation of GATA1 targets.

***Nsd1* regulates GATA1 chromatin binding and protein interactions.** As expression of wild type or mutant *Nsd1* did not overtly change GATA1 protein levels in *Nsd1*^{-/-} erythroblasts kept for 2 days in DM, we compared chromatin binding and putative protein interactions of GATA1 by ChIP and IP-MS after

24 h of induced differentiation (Fig. 8a–c). Hereby, we found increased occupancy of GATA1 at over 3000 sites in the genome overlapping with 1362 genes (*p* < 0.01) in cells expressing *Nsd1* in comparison to the catalytically inactive *Nsd1*^{N1918Q} mutant (Fig. 8d). Of genes with significantly increased binding of GATA1, 731 of them had the promoter regions decorated by H3K27^{ac} while H3K36^{me3} marks overlapped with 1179 gene bodies (Supplementary Data 6). Hence, while global levels of GATA1 remains constant, reintroduction of *Nsd1* resulted in increased DNA binding to available GATA1 sites in promoter

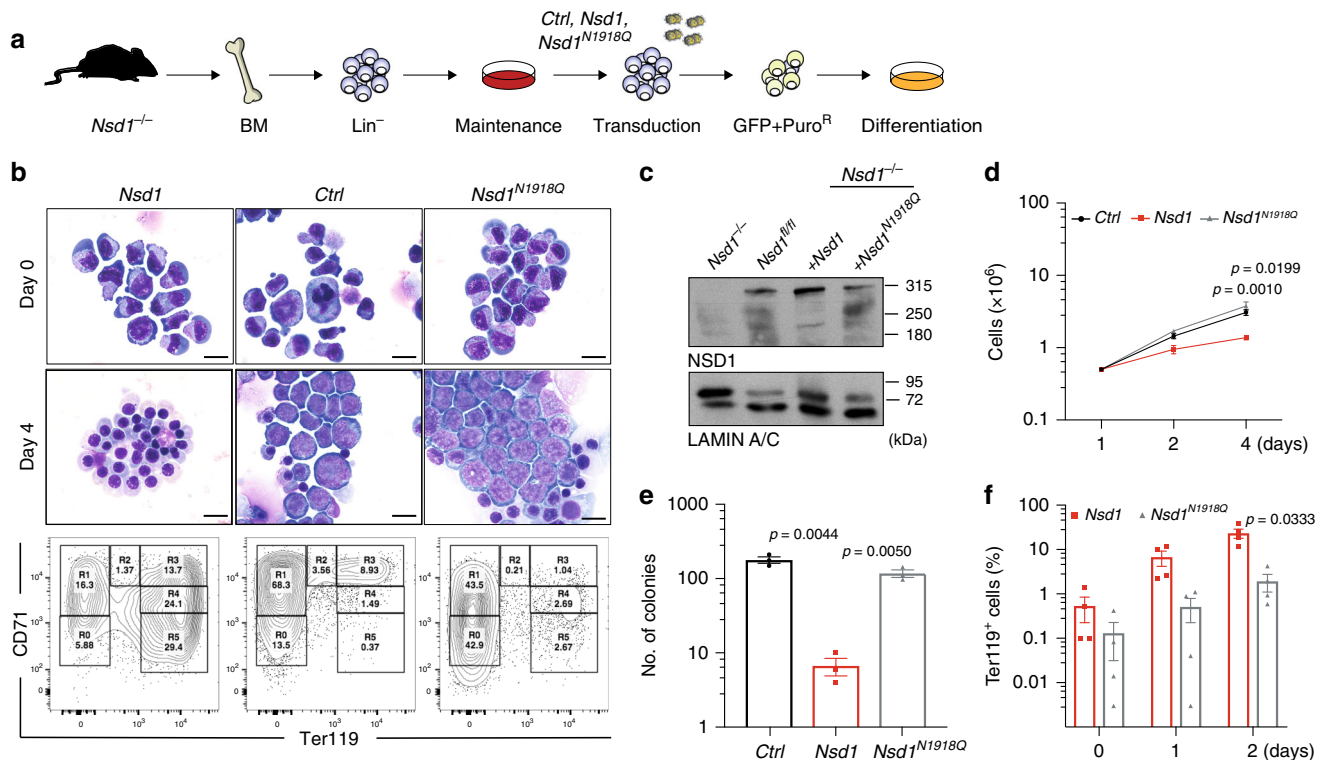


Fig. 6 NSD1-SET is essential for in vitro erythroblast maturation. **a** Experimental setup: BM-derived *Nsd1*^{fl/fl} and *Nsd1*^{-/-} BM-derived erythroblasts were transduced with either *pMSCV-GFP-Puro* (Ctrl), *pMSCV-Nsd1-GFP-Puro* (*Nsd1*), or *pMSCV-Nsd1*^{N1918Q}-*GFP-Puro* (*Nsd1*^{N1918Q}) in maintenance medium, GFP⁺ cells were expanded in the presence of Puromycin before induced differentiation and analysis. **b** Representative pictures of Wright Giemsa-stained cytopsin preparations transduced with *Nsd1*, *Nsd1*^{N1918Q}, or control virus in maintenance medium (day 0, top panels) and after 4 days in differentiation medium (middle panels). The lower panels show flow cytometric analysis of CD71 and Ter119 expression of the cells after 4 days in differentiation medium. These data represent one of four independent experiments ($\times 1000$, size bars = 10 μm). **c** Western blot analysis showing NSD1 protein expression in 1×10^6 *Nsd1*^{fl/fl} and *Nsd1*^{-/-} (untransduced), and *Nsd1*^{-/-} transduced erythroblasts either expressing *Nsd1* or *Nsd1*^{N1918Q} in maintenance medium. LAMIN-A/C was used as a loading control. These data represent one out of two experiments. **d** Growth of *Nsd1* (red line), *Nsd1*^{N1918Q} (gray line), or the control (Ctrl, black line) virus transduced *Nsd1*^{-/-} BM-derived erythroblasts in differentiation medium (1–4 days). Nucleated living cells were counted by the Trypan blue exclusion ($n = 3$ per group). **e** Number of colonies formed by 1×10^4 *Nsd1* (red bar), *Nsd1*^{N1918Q} (gray bar) or the control virus (Ctrl, black bar) transduced *Nsd1*^{-/-} BM-derived erythroblast in MC (M3434) after 11 days ($n = 3$ per group). **f** Ter119⁺ stained *Nsd1*^{-/-} BM-derived erythroblasts transduced with *Nsd1* (red bars) and *Nsd1*^{N1918Q} (gray bars) in maintenance medium (day 0) and after 1 and 2 days in differentiation medium ($n = 4$). Values are presented as individual points, bar graphs represent the mean value of biological replicates, error bars as standard error of the mean. Statistical significances in **d–f** tested with a paired two-tailed *t*-test.

regions, similarly reflected in changes in H3K36^{me3} and H3K27^{ac} at the genomic coordinates. Interestingly, changes in gene expression aligned with H3K27^{ac} around TSS, confirming that these epigenetic marks are directly regulating the down-stream transcriptional programming (Fig. 8e). However, we could not detect any gene loci with statistically significant increase of all three, GATA1, H3K36^{me3}, and H3K27^{ac} (Supplementary Data 7 and 8), which could be a matter of temporal distance along the activation pathway.

Nevertheless, *Nsd1*-induced regulation of several erythroid regulators was associated with simultaneously changed GATA1 binding, H3K27^{ac} and H3K36^{me3} marks. The *Pklr* gene locus, encoding for the liver-red cell pyruvate kinase linked to erythroid differentiation and *Art4*, encoding for the developmentally regulated Dombrock blood group glycoprotein, were both higher expressed in *Nsd1*^{-/-} cells expressing wild-type *Nsd1* associated with a narrow GATA1 peak in the promoter region within a broader decoration of H3K27^{ac}, followed by gene body-wide H3K36^{me3} marks (Fig. 8f)^{39,40}. The opposite was observed for the gene encoding for *Fgf2* (fibroblast growth factor 2) associated with inhibition of efficient erythroid differentiation that appeared higher expressed in *Nsd1*^{N1918Q} than in *Nsd1*-expressing cells

(Supplementary Fig. 6a)⁴¹. Immunoblot and mass spectrometry analysis revealed globally reduced mono-, di-, and tri-methylated H3K36 in *Nsd1*^{-/-} erythroblasts expressing the inactive *Nsd1*^{N1918Q} mutant compared to those expressing wild-type *Nsd1* (Supplementary Fig. 6b, c).

To address whether impaired chromatin binding and transactivation of GATA1 in the absence of *Nsd1* might be associated with altered protein interactions we immunoprecipitated GATA1 followed by mass spectrometry in *Nsd1*^{-/-} cells either expressing wild-type *Nsd1* of the inactive *Nsd1*^{N1918Q} mutant kept for 24 h in DM. We identified 413 differentially expressed proteins ($p < 0.05$) (Supplementary Data 9), of which the most significant ones included known interactors of GATA1 such as MBD2, RBBP4, ZFPM1, RUNX1, and TAL1 suggesting functionality of the assay (Fig. 8g)⁴². Interestingly, mass spectrometry analysis revealed that differentiation of *Nsd1*-expressing *Nsd1*^{-/-} erythroblasts was associated with a highly significant reduction ($\log\text{FC} = -1.96$; $p < 1.08 \times 10^{-7}$) of the transcriptional repressor protein SKI previously proposed to interact with and inhibit GATA1 activation, most likely in cooperation with the nuclear co-repressor (NCoR) complex (Fig. 8g, h, Supplementary Fig. 6d)^{43,44}. Notably, several members of the NCoR complex (NCOR1, HDAC3, TBLXR1) co-

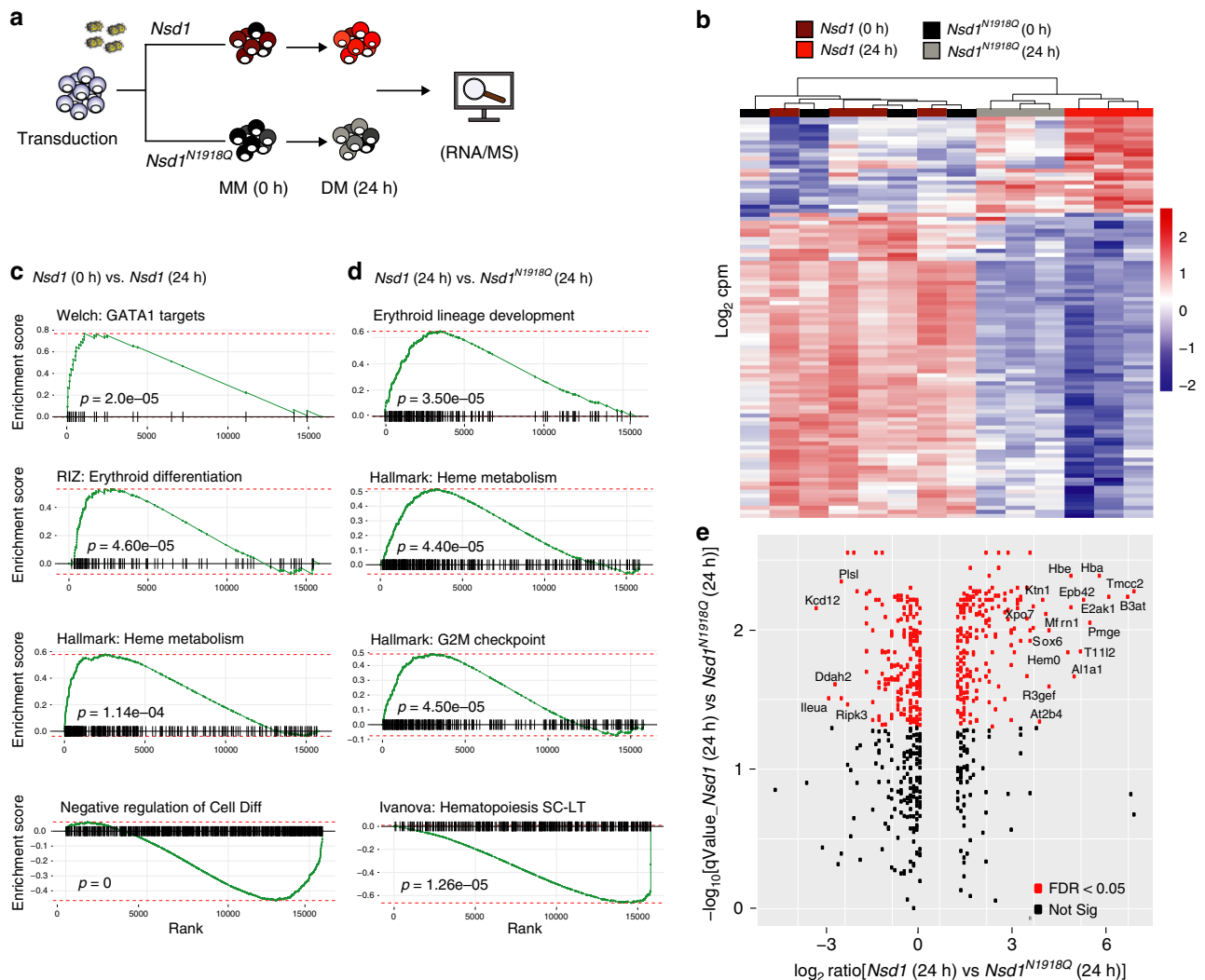


Fig. 7 *Nsd1* expression induces an erythroid gene and protein signature. **a** Experimental setup: *Nsd1*^{-/-} BM-derived erythroblasts expressing either *Nsd1* or *Nsd1*^{N1918Q} were analyzed during expansion in maintenance medium (0 h) and after 24 h in differentiation medium by RNA-seq and global proteome analysis. **b** Heatmap of the top 100 differentially expressed genes (corresponding to FDR < 1.06 × 10⁹) of *Nsd1*^{-/-} BM-derived erythroblasts expressing *Nsd1* (brown squares) and *Nsd1*^{N1918Q} (black squares) in maintenance medium, and after 24 h in differentiation medium (*Nsd1*, red squares; *Nsd1*^{N1918Q}, gray square). Columns clustering was done by Wards linkage on correlations. **c** Gene set enrichment analysis (GSEA) (weighted Kolmogorov-Smirnov-like statistics, two-sided, with adjustment for multiple comparisons) of differential expression between *Nsd1*^{-/-} BM-derived erythroblast expressing *Nsd1* before and after 24 h in differentiation medium. **d** GSEA (weighted Kolmogorov-Smirnov-like statistics, two-sided, with adjustment for multiple comparisons) of differential expression between *Nsd1*^{-/-} BM-derived erythroblasts expressing either *Nsd1* or *Nsd1*^{N1918Q} kept for 24 h in differentiation medium. **e** Differential protein expression of *Nsd1*^{-/-} erythroblasts expressing *Nsd1* or *Nsd1*^{N1918Q} kept for 24 h in differentiation medium (n = 3 per group, FDR < 0.05, p value < 0.05). Labels are shown for proteins with both logFC > 3 and FDR < 0.05.

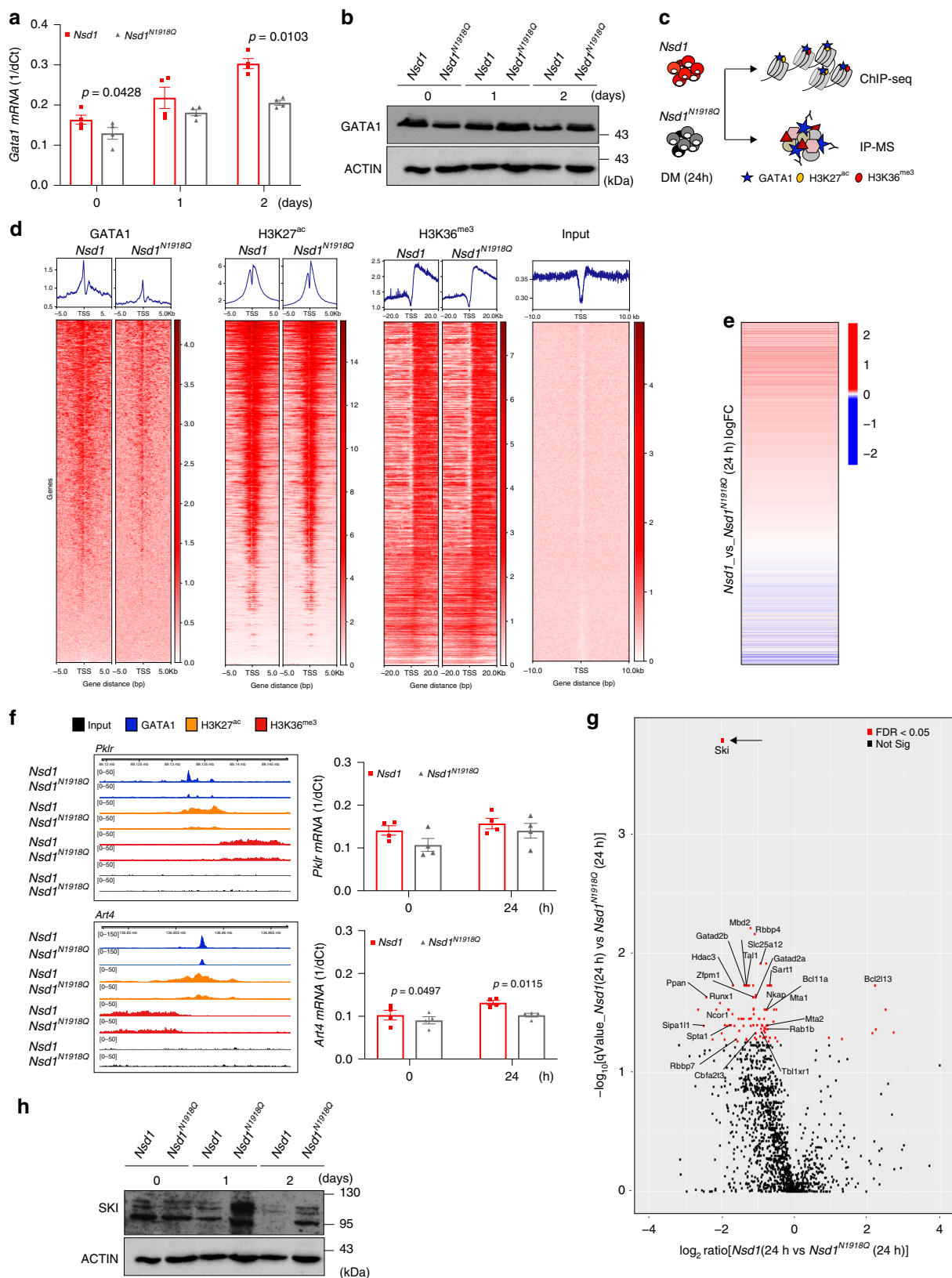
appeared with SKI, as differentially regulated (Fig. 8g, Supplementary Data 9).

SKI knockdown differentiates *Nsd1*^{-/-} erythroblasts. To functionally explore reduced GATA1-SKI association upon *Nsd1* expression, we asked whether experimental shRNA-mediated reduction of SKI might be sufficient to initiate maturation of *Nsd1*^{-/-} erythroblasts (Fig. 9a). We found that SKI knockdown significantly increased in vitro induced terminal maturation of erythroblasts from three independent *Nsd1*^{-/-} mice, as shown by cellular morphology, flow cytometry (CD71/Ter119/Kit), and proliferation (Fig. 9b–d, Supplementary Fig. 7a). SKI knockdown did not alter GATA1 protein levels (Fig. 9e). Notably, prolonged culture in DM was associated with a general reduction of SKI levels suggesting a role for SKI during initiation rather than

terminal differentiation. SKI knockdown also significantly reduced clonogenic growth, total number of cells, and Kit⁺ expression of the cells in MC (Fig. 9f–h, Supplementary Fig. 7b). Collectively, these data suggest that in the absence of *Nsd1*, terminal erythroid maturation is blocked as a consequence of impaired GATA1 transactivation dependent on its association with the transcriptional repressor SKI.

Discussion

The observations that reduced expression of NSD1 altered erythroid clonogenic growth of human CD34⁺ cells and significantly impaired terminal erythroid maturation leading to an erythroleukemia-like disease in mice characterizes NSD1 as a regulator of erythroid differentiation. Mechanistically, we found



that NSD1 activity regulated chromatin binding and target gene activation by the erythroid master regulator GATA1.

***Nsd1* inactivation phenocopies hallmarks of *Gata1* deficiencies in mice.** Earlier studies showed that constitutive *Gata1* gene

inactivation in mice resulted in embryonic death at approximately E11.5 due to arrested maturation of primitive erythroid cells⁴. In contrast, some mice with 95% reduced *Gata1* mRNA expression due to a knockdown mutation (*Gata1*^{+1.05}) developed a late-onset B-cell lymphoproliferative disease or an earlier erythroleukemia-like disease⁵. Similar to *Nsd1*^{-/-} erythroblasts, *Gata1*^{+1.05/+}

Fig. 8 *Nsd1* expression increases GATA1 chromatin binding and changes GATA1 protein interaction partners during induced differentiation of *Nsd1*^{-/-} cells. **a** Relative *Gata1* mRNA expression levels (1/dCt) in *Nsd1*^{-/-} BM-derived erythroblasts virally expressing *Nsd1* (red bars) or *Nsd1*^{N1918Q} (gray bars) expanded in maintenance medium (day 0) or after 1 and 2 days in differentiation medium. Values were normalized to *Gapdh* ($n = 4$ per group). **b** Western blot showing GATA1 protein expression in 1×10^6 *Nsd1*^{-/-} BM-derived erythroblasts expressing *Nsd1* or *Nsd1*^{N1918Q} upon expansion in maintenance medium (day 0), and after 1 and 2 days in differentiation medium. Actin was used as a loading control. This data represents one of two experiments. **c** Experimental setup of the ChIP-seq and IP-MS experiment. **d** Heatmaps of genome-wide ChIP-seq signals in *Nsd1*^{-/-} BM-derived erythroblasts expressing *Nsd1* (left column) or *Nsd1*^{N1918Q} (right column) after 24 h in differentiation medium for GATA1, H3K27^{ac}, and H3K36^{me3}. 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. Density plots above each heatmap depicts corresponding averaged binding around TSS. **e** One-dimensional heatmap of logFC between gene expression of *Nsd1*^{-/-} BM-derived erythroblasts expressing *Nsd1* or *Nsd1*^{N1918Q} after 24 h in differentiation medium (as presented in Fig. 5h, j) sorted according to read coverage around TSS for H3K27^{ac} ChIP (data as shown in panel c, sorted independently. Only overlapping genes are displayed). **f** Integrated genome viewer (IGV) representation of GATA1, H3K27^{ac}, and H3K36^{me3} ChIP peaks in the *Pklr* (top panel) and *Art4* gene locus (lower panel) from *Nsd1*^{-/-} BM-derived erythroblasts either expressing *Nsd1* or *Nsd1*^{N1918Q} after 24 h in differentiation medium. Right panels show *Pklr* and *Art4* mRNA relative expression levels (1/dCt) in *Nsd1*^{-/-} BM-derived erythroblasts expressing *Nsd1* or *Nsd1*^{N1918Q} in maintenance medium (day 0) and after 24 h differentiation medium. Values are shown as relative expression normalized to *Gapdh* ($n = 4$). **g** Volcano plot of differential protein enrichments by GATA1 immunoprecipitation (GATA1-IP) in *Nsd1*^{-/-} BM-derived erythroblasts either expressing *Nsd1* or *Nsd1*^{N1918Q} kept for 24 h in differentiation medium, each group is normalized to IgG control ($n = 2$). Significantly reduced GATA1-SKI association (indicated by a black arrow) was observed upon expression of *Nsd1* compared to *Nsd1*^{N1918Q} (FDR < 0.05). **h** Western blot analysis showing SKI protein expression in 1×10^6 BM-derived *Nsd1*^{-/-} erythroblasts either expressing *Nsd1* or *Nsd1*^{N1918Q} during expansion in maintenance medium (day 0), and after 1 and 2 days in differentiation medium. Actin was used as a loading control. This data represent one of two experiments. 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** tested with a paired two-tailed t-test.

erythroleukemia cells were able to differentiate into mature erythrocytes when complemented with full-length *Gata1* (ref. 45). In contrast to *Gata1*^{1.05/+} mice, *Nsd1*^{-/-} mice developed a fully penetrant erythroleukemia-like phenotype after a shorter latency (Fig. 2).

The best-studied in vivo erythroleukemia model is *Friend's virus complex* induced erythroblastosis in which viral integration results in aberrant expression of the *Spi1* gene encoding for PU.1 (ref. 46). Similar to *Nsd1*^{-/-}, *Spi1* transgenic mice develop anemia, thrombocytopenia, and multi-organ infiltration of erythroblasts progressing from an EPO-dependent stage to EPO-independence by acquisition of activating mutations in the *c-kit* receptor tyrosine kinase^{47,48}. However, very similar to *Nsd1*^{-/-} erythroblasts, *Friend's virus* erythroblastosis-derived MEL cells constitutively expressed GATA1 protein that could not be explained by the interaction with PU.1 (Fig. 5)⁴⁹. In addition, conditional activation of exogenous *Gata1* was also reported to induce erythroid differentiation in some MEL cell lines^{49,50}.

***Nsd1* controls GATA1 protein interaction and activation of erythroid regulators.** To study the mechanism of *Nsd1*-controlled erythroid differentiation we faced the problem that primary erythroblast cultures can contain significant fractions of myeloid cells, which are not present in *Nsd1*^{-/-} cultures. Therefore, we chose to virally express WT or a previously reported catalytically inactive *Nsd1*^{N1918Q} SET-mutant in *Nsd1*^{-/-} erythroblasts³⁰. However, not only the large size of the *Nsd1* ORF drastically impaired the gene transfer efficacy, transduced cells also did not tolerate high levels of exogenous *Nsd1*, which limited generation of stably expressing cells in time and numbers. Nevertheless, low-level mRNA expression resulted in detectable *Nsd1* protein expression sufficient to restore terminal maturation of *Nsd1*^{-/-} erythroblasts in a methyltransferase activity-dependent manner (Fig. 6). Interestingly, *Nsd1* expression was associated with increased binding to and transactivation of a large number of previously proposed GATA1 target genes associated with changes in H3K27^{ac} and H3K36^{me3} marks (Fig. 8). These observations led us to speculate that in the absence of *Nsd1*, GATA1 might be functionally trapped in some saturated interactions that may limit its transactivation potential, which can be overcome by expression of additional “free” GATA1.

Characterization of putative GATA1 interactions by immunoprecipitation and mass spectrometry suggested that in the absence of *Nsd1*, GATA1 associates with potent transcriptional repressors (Fig. 8). Notably, expression of wild type but not the inactive SET *Nsd1* mutant resulted in a highly significant reduced association of GATA1 with the transcriptional co-repressor SKI. SKI is well known for its role as a regulator of the TGF-beta/Smad signaling pathway^{51,52}. SKI was also found to be overexpressed in AML and proposed to repress retinoic acid receptor and RUNX1-mediated signaling.^{53–55} In addition, SKI was reported to control HSC fitness in myelodysplastic syndromes (MDS)⁵⁶. Most importantly, SKI was shown to physically interact, to repress GATA1-mediated transactivation, and to block erythroid differentiation by blocking its interaction with DNA⁴³. In addition, SKI-mediated repression seemed to be NCoR dependent, and several NCoR complex proteins were altered in GATA1 pull-downs upon expression of *Nsd1* (Fig. 8)⁴⁴. These observations suggest that the methyltransferase activity of NSD1 controls the interaction of GATA1 with SKI or other, yet to be defined mediators.

Very recent studies using quantitative proteomics revealed that co-repressors are dramatically more abundant than co-activators in erythroblasts⁵⁷. How the lack of *Nsd1* directly regulates differential interaction of GATA1 co-activators and co-repressors remains to be elucidated. One can hypothesize that *Nsd1*-mediated H3K36 methylation provides the anchors for ultimate accumulation of sufficient co-activators on critical target gene loci.

Recent work suggested that H3K36 methylation is critical for normal erythroid differentiation. A conditional H3K36M mutation severely affected the murine hematopoietic system resulting in defects that partially phenocopy those observed in the *Nsd1*^{-/-} mice. *H3K36M* transgenic mice also developed anemia, thrombocytopenia, and splenomegaly. Most notably, these mice also showed a dramatic increase in early Ter119⁻ erythroid progenitor cells in the BM but also in the periphery⁵⁸. Another study found that reduced H3K36^{me2} by *Nsd1* inactivation in ES cells resulted in re-localization of the DNMT3A DNA methyltransferase, which interacts with the H3K36^{me3} through its PWWP domain. This led to hypomethylation of euchromatic intergenic regions as observed in SOTOS patients having *NSD1* loss of function mutations⁵⁹. Interestingly, normal erythroid maturation and particularly at

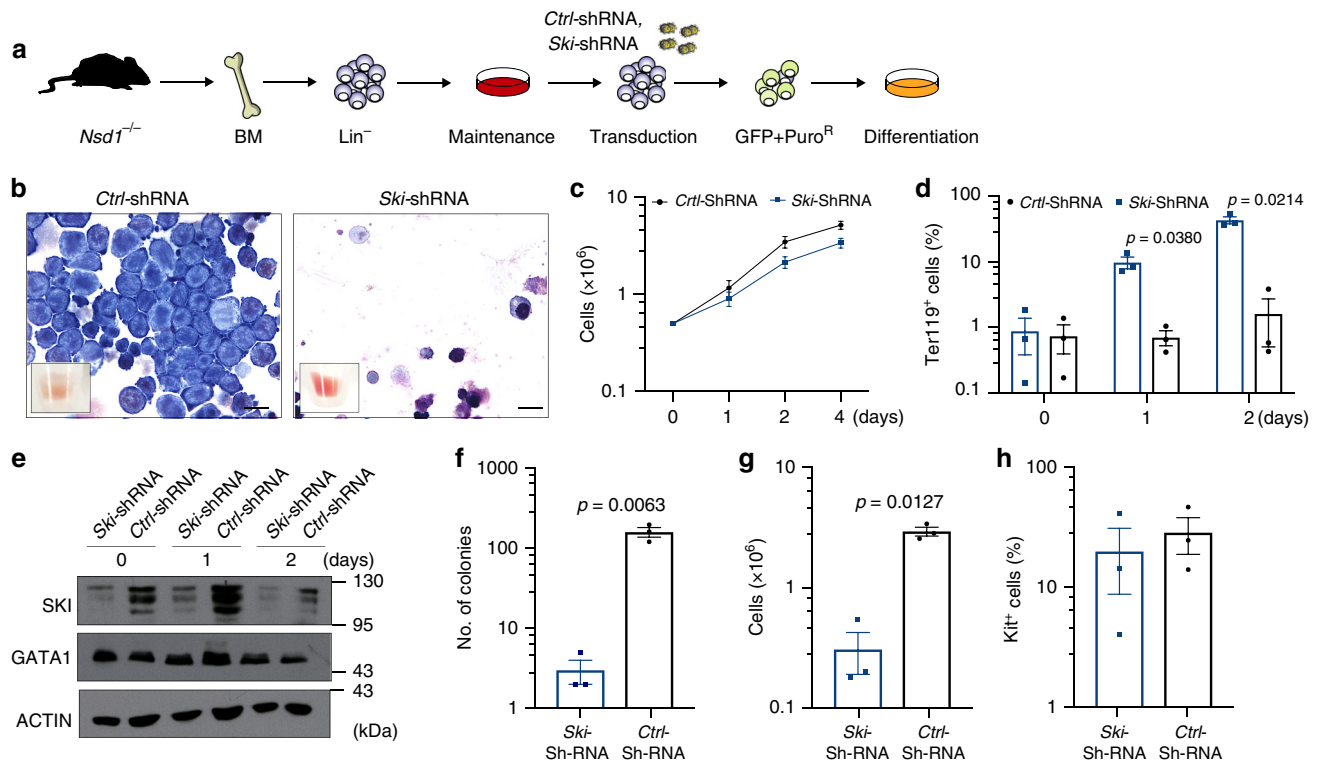


Fig. 9 SKI knockdown results in terminal differentiation of *Nsd1*^{-/-} erythroblasts. **a** Experimental setup: BM-derived *Nsd1*^{-/-} erythroblasts were transduced with either *pLMP-empty-shRNA-GFP-Puro* (Ctrl-shRNA) or *pLMP-Ski-shRNA-GFP-Puro* (Ski-shRNA) in maintenance medium, sorted for GFP, and selected with Puromycin for 2 days before induced differentiation and analysis. **b** Representative images of Wright Giemsa-stained cytospin preparations of control (Ctrl-shRNA, left panel) and Ski shRNA (Ski-shRNA, right panel) transduced *Nsd1*^{-/-} BM-derived erythroblasts after 2 days in differentiation medium. The small insets show the cell pellets before analysis. These data illustrate one of three experiments ($\times 1000$, size bar = 10 μm). **c** Growth of *Nsd1*^{-/-} BM-derived erythroblasts transduced with Ski- (Ski-shRNA, blue line) or control shRNA (Ctrl-shRNA, black line) grown for 4 days in differentiation medium. Nucleated living cells were counted by Trypan blue exclusion ($n = 3$ per group). P value > 0.05 for all time points. **d** Fraction of Ter119⁺ cells (%) of *Nsd1*^{-/-} BM-derived erythroblasts transduced with Ski- (Ski-shRNA, blue bars) or control (Ctrl-shRNA, black bars) virus grown for 2 days in differentiation medium ($n = 3$ per group). **e** Western blot showing SKI and GATA1 protein expression in *Nsd1*^{-/-} BM-derived erythroblasts transduced with Ski (Ski-shRNA) or control (Ctrl-shRNA) virus during expansion in maintenance medium (day 0) and following 1-2 days in differentiation medium. Actin was used as a loading control. These data represent one of three experiments. **f** Total number of colonies counted at day 11 after plating of 1×10^4 *Nsd1*^{-/-} BM-derived erythroblasts expressing either Ski shRNA (Ski-shRNA, blue bar) or control (Ctrl-shRNA, black bar) in MC (M3434) ($n = 3$ per group). **g** Total number of cells obtained from 1×10^4 *Nsd1*^{-/-} BM-derived erythroblasts expressing either Ski shRNA (Ski-shRNA, blue bar) or control (Ctrl-shRNA, black bar) after 11 days in MC (M3434) ($n = 3$ per group). **h** Percentage of Kit⁺ living cells obtained from 1×10^4 *Nsd1*^{-/-} BM-derived erythroblasts expressing either Ski shRNA (Ski-shRNA, blue bar) or control (Ctrl-shRNA, black bar) after 11 days in MC (M3434) ($n = 3$ per group). P value > 0.05 for all time points. 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**, **f**, **g**, **h** tested with a paired two-tailed *t*-test.

transition from CFU-E to proerythroblasts was found to correlate with activation of a significant number of genes associated with gained DNA methylation on selective genes including numerous GATA1 targets⁶⁰. Together, these findings suggest that the loss of H3K36 methylation and redistribution of DNMT3A could be directly responsible for impaired binding of GATA1 and its co-activators. The fact that we pulled down DNMT3A by immunoprecipitation of GATA1 in *Nsd1*^{-/-} cells (Supplementary Datas 10 and 11) suggests that GATA1 binding could indeed not only be dependent on H3K36me but also on DNMT3A-mediated DNA methylation.

NSD1, SKI, and human erythroleukemia. *Nsd1* gene inactivation during late stage fetal liver hematopoiesis induced a fully penetrant lethal disease that phenocopied several aspects of acute erythroleukemia, a rare form of human AML⁶. Putative loss of function missense or frameshift *NSD1* mutations have been found in various human cancers including AML (<https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=NSD1>). Interrogation of the cancer cell line encyclopedia (CCLE) revealed that very few human

cancer cell lines express even negligible levels of NSD1 mRNA and protein, including F-36P, a cell line established from a patient with acute erythroleukemia (<https://portals.broadinstitute.org/ccle/page?gene=NSD1>)⁶¹. Notably, very recent work revealed several cases of childhood acute erythroleukemia that harbored fusion genes involving *NSD1*⁷. Based on our findings one can speculate that in such cases the fusion may either act in a dominant-negative manner to *NSD1* expressed from the non-arranged allele, or the presence of LOH is reducing *NSD1* activity as recently reported for a significant number of solid cancers²¹. Interestingly, we also found aberrantly high SKI expression levels in tumor cells of some erythroleukemia patients, and that in vivo overexpression of a SKI ORF in BM-derived HSPC resulted in an erythroleukemia-like disease in mice, suggesting that SKI expression may not only be critical for impaired erythroid differentiation in *Nsd1*^{-/-} mice but also a driver of the human disease⁶². Collectively, our observations suggest that impaired *NSD1* activity functionally interferes with lineage-associated transcriptional master regulators such as GATA1 resulting in impaired cellular differentiation as a first step to malignant transformation.

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 \pm SEM). Statistical significance was tested with unpaired (*Nsd1^{fl/fl}* vs. *Nsd1^{-/-}*) or paired (viral transduction in *Nsd1^{-/-}* cells) two-tailed *t*-test, assuming equal variance, unless otherwise specified. Statistical test was performed in log₁₀ space, or for qPCR kept in log₂ space.

shRNA-mediated knockdown. Human CD34⁺ cells were obtained by enrichment using the CD34 MicroBead Kit (Miltenyi Biotec, Bergisch Gladbach, Germany) from peripheral blood or cord blood from healthy donors and kept in StemLine II medium (Sigma Aldrich, Buchs, Switzerland), supplemented with human cytokines such as 50 ng/ml hTPO (Peprotech, London, UK), 50 ng/ml hFLT3 ligand (Peprotech, London, UK), 50 ng/ml hSCF (Peprotech, London, UK), and 1 U/ml hEPO (Eprex 4000, Pharmacy of University Hospital Basel). shRNAs were expressed from lentiviral vectors (*pLKO.1*). For transduction lentiviral stock was produced by transient co-transfection of packaging vectors (*pMD2G*, *pMLDG/PRE*, *pRSV/Rev*) and respective lentiviral shRNA plasmid (shRNA Ctrl and shRNA *NSD1* #353, #369 and #372) 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 48 and 72 h after transfection, snap frozen in liquid nitrogen, and stored in -80°C until usage. Cells were spin-infected in the presence of 5 $\mu\text{g}/\text{ml}$ polybrene (Sigma Aldrich, Buchs, Switzerland) with virus for 90 min, 2500 r.p.m. at 30°C . Six hours after spin infection, cells were washed with PBS and plated in StemLine II cytokine enriched medium. Two days after spin infection cells were selected with 2 $\mu\text{g}/\text{ml}$ puromycin (Gibco, Thermo Fisher Scientific, Reinach, Switzerland).

Transgenic mice. Mice carrying a *Nsd1^{L3}* allele were previously described²². The floxed *pgk-neomycin* selection cassette was removed by viral *Cre* expression in ES cells, leaving two *loxP* sites flanking the largest coding exon 5, 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 hematopoiesis⁶³. All mice in this study were kept under specific pathogen-free conditions. Mice were genotyped using the KAPA Mouse Genotyping Kit Hot Start Kit (Cat. KK7352; KapaBiosystems, Wilmington, USA) following the manufacturer's instructions. PCR reaction program was 5 min 95°C , 40 cycles 95°C 15 s, 60°C 15 s, 72°C 30 s followed by 5 min 72°C and 1 min 4°C . PCR amplicons were visualized on 2% agarose gels containing ethidium bromide. Used oligonucleotide primers can be found in Supplementary Table 4.

Analysis of the mouse phenotype. Mice were sacrificed by CO_2 asphyxia, organs removed and fixed in buffered 4% formalin solution. Paraffin-embedded tissue sections were stained with hematoxylin and eosin (H&E). Blood was collected from the tail vein or by vena cava inferior puncture (terminal) and counts were determined using an Advia120 Hematology Analyzer using Multispecies Version 5.9.0-MS software (Bayer, Leverkusen, Germany). Differential blood counts were analyzed on smears stained using Wright-Giemsa staining (Hematology, University Hospital Basel). Sections were analyzed on an Olympus BX61 microscope (Tokyo, Japan) or Nikon TI (Tokyo, Japan).

PCR analysis of cleavage of floxed *Nsd1* exon 5. Genomic DNA was extracted from total BM or flow sorting enriched myeloid, erythroid, T and B cells using the Genra Puregene Cell kit (Qiagen/158767) according to the manufacturer's protocol. Fifty nanograms of gDNA were amplified using the GoTaq G2 Hot Start Polymerase (Promega/M7405) under the following conditions: 94°C 1 min, 60°C 1 min, 72°C 1 min for 25 cycles. PCR products were separated on a 1.5% agarose gel containing ethidium bromide. PCR primers are given in Supplementary Table 4.

BM transplantation. Transplantations were performed using whole BM of *Nsd1^{-/-}* mice at indicated age. For competitive transplantation, 1×10^6 total BM cells of symptomatic or asymptomatic *Nsd1^{-/-}* mice (CD45.2) was mixed in a 1:1 ratio with supporting BM of B6.SJL (CD45.1) donor mice and transplanted into lethally irradiated (2×600 cGy) B6.SJL (CD45.1) recipients via tail vein.

Analysis of mouse hematopoiesis. To obtain fetal liver cells, one male mouse was placed with two female mice during the light period and left overnight. At the indicated dates, pregnant females were sacrificed by CO_2 asphyxia, and fetal liver cells from individual embryos were surgically isolated, minced, and passed through 70 μm cell strainer (Cat. 352350; BD, New Jersey, USA). For adult mice, total BM was harvested by crushing long bones and spine, while spleens were dissected and single-cell suspensions obtained by pressing through a 70 μm cell strainer. Red blood cells were lysed with ammonium-chloride potassium (ACK) lysis buffer (150 mM NH_4Cl , 10 mM KHCO_3 , and 0.1 mM EDTA, pH 8.0) for 10 min on ice. Lineage depletion was achieved according to the manufacturers' protocol of mouse hematopoietic lineage depletion kit (Cat. 130-090-858; Miltenyi Biotec, Bergisch

Gladbach, Germany). Cytospin preparations of approximately 10^5 cells were made by centrifugation for 3 min at 300 r.p.m. using a Shandon Cytospin 3 centrifuge using cytofunnel disposable sample chambers (Cat. 5991040; Thermo Fisher Scientific, Reinach, Switzerland) and non-coated cytoslides (Cat. 5991051; Thermo Fisher Scientific, Reinach, Switzerland). Cytospots were stained with Wright Giemsa solution.

Flow cytometry. Cells in suspension were washed with FACS buffer (0.5% BSA, 1 mM EDTA in PBS) and incubated with indicated antibodies for 45 min on ice, washed, and stained with 1 $\mu\text{g}/\text{ml}$ DAPI (Life Technologies, Paisley, UK) in PBS. Stained cells were analyzed on a CyAn ADP analyzer (Beckman-Coulter) or LSR Fortessa (BD, New Jersey, USA). Data were analyzed with FlowJo software (Tree Star). For CD71/Ter119 staining, the preparation still contained red blood cells, for CD71/Kit/Sca-1/FcγRII/III stem and progenitor staining the red blood cells were depleted. For stem and myeloid progenitor staining, lineage-positive cells were depleted as described before. All antibodies used in this study are indicated in Supplementary Table 5. For calculating number of stem and progenitor cells in BM, lineage-marker-depleted cells were counted and absolute numbers of cells adjusted to this number. For differentiation analysis of mouse or human cells in vitro, cells were filtered, washed twice with PBS, and stained in 100 μL FACS buffer. Detailed information about the FACS gating strategies can be found in Supplementary Fig. 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 *Gapdh* expression and relative expression was quantified using $1/\text{dCt}$ or the $2^{-(\text{ddCt})}$ method⁶⁴. Primers are given in Supplementary Table 6. For multivariate analysis of RT-qPCR the multicom R package was used to model dCt values kept log space, extracting coefficients at each timepoint using Tukey's test. Where specified the model included adjustment of the effect of individual mouse and plots depicts the residuals following this regression.

Colony forming assay. For whole BM analysis, approximately 4×10^4 cells were plated in methylcellulose M3434 (Methocult, StemCell Technologies, Vancouver, Canada). Colonies were scored after 8–10 days. Pictures were taken on Olympus IX50 microscope with $\times 2$, $\times 4$, and $\times 10$ magnification. Cells were washed, resuspended, counted with trypan blue and if applicable replated into fresh methylcellulose. For colony formation analysis of human CD34⁺ cells, 5×10^3 cells were plated into H4434 (Methocult; StemCell Technologies, Vancouver, Canada). After scoring of M3434 plates, when indicated, dishes were incubated with a mix of two volumes of 0.3% hydrogen peroxide and five volumes of 0.2% di-hydrochloride benzidine (Sigma Aldrich, Buchs, Switzerland) in 0.5 M acetic acid/1 \times PBS for 5 min at 37°C .

Western blotting. For protein detection, total cell extracts were isolated from freshly cultured 1×10^6 cells using 60 μL 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 was placed in a new tube. Nuclear protein lysates were prepared by resuspending cells in hypotonic lysis buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT) for 15 min on ice, followed by treatment with 0.1% NP-40 and 15 s vortexing. Nuclei were spun down at 14,000 r.p.m. for 2 min at 4°C and supernatant containing cytoplasmic fraction kept for analysis. Pellets were resuspended in nuclear lysis buffer (20 mM HEPES pH 7.9, 0.4 M NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT). In addition, pellets were sonicated for three cycles (30 s sonication, 30 s pause) on a Bioruptor p10 sonicator (Diagenode, Seraing, Belgium) and left for 20 min on ice before spinning down at 14,000 r.p.m. for min at 4°C . Lysates were kept for analysis of nuclear proteins and remaining pellets used for histone extraction in 0.2 N HCl and beta-mercaptoethanol. Lysis buffers were supplemented with Complete Mini protease inhibitors (Cat. 11836153001; Roche). Proteins were quantified by Bradford assay (Biorad, München, Germany) and loading adjusted. Samples were prepared in 4 \times Laemmli buffer (Biorad, München, Germany) and boiled for 10 min at 95°C before loading on pre-cast (BioRad) or hand casted gels of different percentages. For NSD1 blot, 50 μg of nuclear extract was loaded on a 5% running gel. Wet transfer was done overnight at 4°C in 5% methanol/0.1% SDS/tris-base-bicine buffer on 0.45 μm nitrocellulose membranes. For blotting GATA1, 10 μg nuclear extract was loaded on 10% gels and semi-dry transfer was done for 30 min on nitrocellulose 0.2 μm (Biorad, München, Germany). For SKI, whole lysate from 1 MIO cells was boiled and loaded on 6–7.5% gels. Wet transfer was carried out for 3 h at 4°C . Membranes were blocked in 5% non-fatty milk (NFM) in PBS–1% Tween for 2 h at room temperature. Blots were probed overnight with antibody at 4°C in 2.5% NFM/PBS–1% Tween, washed three times for 15 min 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 min 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). Uncropped original scans of the western blots membranes as shown in Figs. 5b, 6c, 8b, h are provided in Supplementary Fig. 9. Information regarding the used antibodies can be found in Supplementary Table 7.

In vitro erythroid differentiation assay. Fetal liver and adult BM-derived erythroblasts cells were obtained following a previously published protocol²⁷. Erythroblast cultures from adult mice were established after lineage depletion of BM cells. Cells were cultured for more than one week in maintenance medium composed of StemSpan SFEM (StemCell Technologies, Vancouver, Canada), supplemented with 1% Pen/Strep, 0.4% cholesterol (Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 2 U/ml hEpo (Eprex 4000, 9096976, Pharmacy of University Hospital Basel), 100 ng/ml mScf (PeproTech, London, UK), 10^{-6} M dexamethasone (Calbiochem, Sigma Aldrich, Buchs, Switzerland), and 40 ng/ml hIGF-1 (PeproTech, London, UK). Cells were split every second day and presence of proerythroblasts was verified by flow cytometry (DAPI⁻/FSC⁺/CD71⁺/Ter119⁻) and cytospins. Erythroblasts were subjected to terminal maturation in differentiation medium composed of IMDM (Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 1% P/S, 10% FCS, 10% PFHMII (Gibco, Thermo Fisher Scientific, Reinach, Switzerland), 5% hPDS (0.45 μ M filtered, Blood donation service, University Hospital Basel), monothio glycerol (Sigma Aldrich, Buchs, Switzerland), 100 ng/ml mSCF and 2 U/ml hEpo.

Retroviral gene transfer. Full-length cDNAs for murine *Nsd1* (*pSG5*) was obtained from R. Losson (Strasbourg). Wild type (*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 an SKI-specific mir-shRNA was a gift from M. Hayman (Buffalo, NY). A full-length cDNA for murine *Gata1* was obtained from T. Mercher (Paris) and cloned into *pMSCV* and sequenced verified. Retroviral stocks were produced by transient co-transfection of packaging vectors (*pPAK6*) and respective plasmids using Turbofect or Jetprime transfection reagent (Life Technologies, Paisley, UK) in HEK293T-LX cells kept in DMEM (Gibco, Lubio, Thermo Fisher Scientific, Reinach, Switzerland) with 10% FCS and 1% P/S. Viral supernatants were harvested 48 and 72 h after transfection, 10 \times Vivaspin 20 (Sartorius, Göttingen, Germany) concentrated at 4000 r.p.m. for 40 min at 4 °C and snap frozen in liquid nitrogen and stored in -80 °C until usage. Cells were spin-infected either in StemSpan SFEM, supplemented with 50 ng/ml hTPO (PeproTech, London, UK) and 50 ng/ml mSCF or in maintenance medium used for erythroblast culture as described above, in the presence of 5 μ g/ml polybrene (Sigma Aldrich, Buchs, Switzerland) with virus for 90 min, 2500 rpm at 30 °C. Four hours after spin infection, the cells were washed with PBS and plated in maintenance medium. Two days after spin infection, the cells were selected with 2 μ g/ml puromycin (Gibco, Thermo Fisher Scientific, Reinach, Switzerland) or EGFP⁺ cells were FACS enriched as described before. Information regarding the used plasmids can be found in Supplementary Table 8.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The RNA raw expression data are accessible with the following number [GSE136811](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE136811). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD017657 (ref. 65).

Received: 8 August 2019; Accepted: 17 April 2020;
Published online: 12 June 2020

References

- Hattangadi, S. M., Wong, P., Zhang, L., Flygare, J. & Lodish, H. F. From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood* **118**, 6258–6268 (2011).
- Kerenyi, M. A. & Orkin, S. H. Networking erythropoiesis. *J. Exp. Med.* **207**, 2537–2541 (2010).
- Yu, M. et al. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol. Cell* **36**, 682–695 (2009).
- Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA* **93**, 12355–12358 (1996).
- Shimizu, R. et al. Leukemogenesis caused by incapacitated GATA-1 function. *Mol. Cell Biol.* **24**, 10814–10825 (2004).
- Boddu, P. et al. Erythroleukemia-historical perspectives and recent advances in diagnosis and management. *Blood Rev.* **32**, 96–105 (2018).
- Iacobucci, I. et al. Genomic subtyping and therapeutic targeting of acute erythroleukemia. *Nat. Genet.* **51**, 694–704 (2019).
- 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.* **17**, 3398–3412 (1998).
- Wang, X. et al. Identification and characterization of a novel androgen receptor coregulator ARA267- α in prostate cancer cells. *J. Biol. Chem.* **276**, 40417–40423 (2001).
- Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126 (2012).
- Kudithipudi, S., Lungu, C., Rathert, P., Happel, N. & Jeltsch, A. Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. *Chem. Biol.* **21**, 226–237 (2014).
- Dolnik, A. et al. Commonly altered genomic regions in acute myeloid leukemia are enriched for somatic mutations involved in chromatin remodeling and splicing. *Blood* **120**, e83–e92 (2012).
- Garg, M. et al. Profiling of somatic mutations in acute myeloid leukemia with FLT3-ITD at diagnosis and relapse. *Blood* **126**, 2491–2501 (2015).
- Papillon-Cavanagh, S. et al. Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nat. Genet.* **49**, 180–185 (2017).
- Su, X. et al. NSD1 inactivation and SETD2 mutation drive a convergence toward loss-of-function of H3K36 writers in clear-cell renal cell carcinomas. *Cancer Res.* **77**, 4835–4845 (2017).
- Peri, S. et al. NSD1- and NSD2-damaging mutations define a subset of laryngeal tumors with favorable prognosis. *Nat. Commun.* **8**, 1772 (2017).
- Berdasco, M. et al. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc. Natl. Acad. Sci. USA* **106**, 21830–21835 (2009).
- Lee, S. T. & Wiemels, J. L. Genome-wide CpG island methylation and intergenic demethylation propensities vary among different tumor sites. *Nucleic Acids Res.* **44**, 1105–1117 (2016).
- Kurotaki, N. et al. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* **30**, 365–366 (2002).
- Baujart, G. & Cormier-Daire, V. Sotos syndrome. *Orphanet J. Rare Dis.* **2**, 36 (2007).
- Park, S., Supek, F. & Lehner, B. Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits. *Nat. Commun.* **9**, 2601 (2018).
- Rayasam, G. V. et al. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* **22**, 3153–3163 (2003).
- Koulnis, M. et al. Identification and analysis of mouse erythroid progenitors using the CD71/TER119 flow-cytometric assay. *J. Visualized Exp.* **54**, 2809 (2011).
- Ogilvy, S. et al. Promoter elements of *vav* drive transgene expression in vivo throughout the hematopoietic compartment. *Blood* **94**, 1855–1863 (1999).
- Kingsley, P. D. et al. Ontogeny of erythroid gene expression. *Blood* **121**, e5–e13 (2013).
- Kogan, S. C. et al. Bethesda proposals for classification of nonlymphoid hematopoietic neoplasms in mice. *Blood* **100**, 238–245 (2002).
- England, S. J., McGrath, K. E., Frame, J. M. & Palis, J. Immature erythroblasts with extensive ex vivo self-renewal capacity emerge from the early mammalian fetus. *Blood* **117**, 2708–2717 (2011).
- Ferreira, R., Ohneda, K., Yamamoto, M. & Philipsen, S. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol. Cell Biol.* **25**, 1215–1227 (2005).
- Welch, J. J. et al. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104**, 3136–3147 (2004).
- Wang, G. G., Cai, L., Pasillas, M. P. & Kamps, M. P. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat. Cell Biol.* **9**, 804–812 (2007).
- Dai, M. S., Mantel, C. R., Xia, Z. B., Broxmeyer, H. E. & Lu, L. An expansion phase precedes terminal erythroid differentiation of hematopoietic progenitor cells from cord blood in vitro and is associated with up-regulation of cyclin E and cyclin-dependent kinase 2. *Blood* **96**, 3985–3987 (2000).
- DeVilbiss, A. W., Boyer, M. E. & Bresnick, E. H. Establishing a hematopoietic genetic network through locus-specific integration of chromatin regulators. *Proc. Natl. Acad. Sci. USA* **110**, E3398–E3407 (2013).
- DeVilbiss, A. W. et al. Epigenetic determinants of erythropoiesis: role of the histone methyltransferase SetD8 in promoting erythroid cell maturation and survival. *Mol. Cell Biol.* **35**, 2073–2087 (2015).
- Malik, J., Getman, M. & Steiner, L. A. Histone methyltransferase Setd8 represses Gata2 expression and regulates erythroid maturation. *Mol. Cell Biol.* **35**, 2059–2072 (2015).
- Zhang, L. et al. ZFP36L2 is required for self-renewal of early burst-forming unit erythroid progenitors. *Nature* **499**, 92–96 (2013).
- Wu, W. et al. Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. *Genome Res.* **21**, 1659–1671 (2011).

37. Hattangadi, S. M. et al. Histones to the cytosol: exportin 7 is essential for normal terminal erythroid nuclear maturation. *Blood* **124**, 1931–1940 (2014).
38. Shaw, G. C. et al. Mitoferrin is essential for erythroid iron assimilation. *Nature* **440**, 96–100 (2006).
39. Aizawa, S. et al. Ineffective erythropoiesis in mutant mice with deficient pyruvate kinase activity. *Exp. Hematol.* **33**, 1292–1298 (2005).
40. Gubin, A. N. et al. Identification of the dombrock blood group glycoprotein as a polymorphic member of the ADP-ribosyltransferase gene family. *Blood* **96**, 2621–2627 (2000).
41. Bartunek, P. et al. bFGF signaling and v-Myb cooperate in sustained growth of primitive erythroid progenitors. *Oncogene* **21**, 400–410 (2002).
42. Rodriguez, P. et al. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J.* **24**, 2354–2366 (2005).
43. Ueki, N., Zhang, L. & Hayman, M. J. Ski negatively regulates erythroid differentiation through its interaction with GATA1. *Mol. Cell Biol.* **24**, 10118–10125 (2004).
44. Ueki, N. & Hayman, M. J. Signal-dependent N-CoR requirement for repression by the Ski oncoprotein. *J. Biol. Chem.* **278**, 24858–24864 (2003).
45. Mukai, H. Y. et al. Establishment of erythroleukemic GAK14 cells and characterization of GATA1 N-terminal domain. *Genes Cells* **18**, 886–898 (2013).
46. Moreau-Gachelin, F. Lessons from models of murine erythroleukemia to acute myeloid leukemia (AML): proof-of-principle of co-operativity in AML. *Haematologica* **91**, 1644–1652 (2006).
47. Moreau-Gachelin, F. et al. Spi-1/PU.1 transgenic mice develop multistep erythroleukemias. *Mol. Cell Biol.* **16**, 2453–2463 (1996).
48. Kosmider, O. et al. Kit-activating mutations cooperate with Spi-1/PU.1 overexpression to promote tumorigenic progression during erythroleukemia in mice. *Cancer cell* **8**, 467–478 (2005).
49. Rekhtman, N., Radparvar, F., Evans, T. & Skoultschi, A. I. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Development* **13**, 1398–1411 (1999).
50. Choe, K. S. et al. Reversal of tumorigenicity and the block to differentiation in erythroleukemia cells by GATA-1. *Cancer Res.* **63**, 6363–6369 (2003).
51. Tecalco-Cruz, A. C., Rios-Lopez, D. G., Vazquez-Victorio, G., Rosales-Alvarez, R. E. & Macias-Silva, M. Transcriptional cofactors Ski and SnoN are major regulators of the TGF-beta/Smad signaling pathway in health and disease. *Signal Transduct. Target Ther.* **3**, 15 (2018).
52. Bonnon, C. & Atanasoski, S. c-Ski in health and disease. *Cell Tissue Res.* **347**, 51–64 (2012).
53. Ritter, M. et al. Inhibition of retinoic acid receptor signaling by Ski in acute myeloid leukemia. *Leukemia* **20**, 437–443 (2006).
54. Teichler, S. et al. MicroRNA29a regulates the expression of the nuclear oncogene Ski. *Blood* **118**, 1899–1902 (2011).
55. 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.* **46**, 3412–3428 (2018).
56. Muench, D. E. et al. SKI controls MDS-associated chronic TGF-beta signaling, aberrant splicing, and stem cell fitness. *Blood* **132**, e24–e34 (2018).
57. Gillespie, M. A. et al. Absolute quantification of transcription factors reveals principles of gene regulation in erythropoiesis. *Mol Cell*, <https://doi.org/10.1016/j.molcel.2020.03.031> (2020).
58. Brumbaugh, J. et al. Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. *Nat. Cell Biol.*, <https://doi.org/10.1038/s41556-019-0403-5> (2019).
59. Weinberg, D. N. et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281–286 (2019).
60. Schulz, V. P. et al. A unique epigenomic landscape defines human erythropoiesis. *Cell Rep.* **28**, 2996–3009 e2997 (2019).
61. Chiba, S. et al. Establishment and erythroid differentiation of a cytokine-dependent human leukemic cell line F-36: a parental line requiring granulocyte-macrophage colony-stimulating factor or interleukin-3, and a subline requiring erythropoietin. *Blood* **78**, 2261–2268 (1991).
62. Fagman, A. et al. Human erythroleukemia genetics and transcriptomes identify master transcription factors as functional disease drivers. *Blood* **003062**, <https://doi.org/10.1182/blood.2019003062> (2020).
63. Georgiades, P. et al. VavCre transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages. *Genesis* **34**, 251–256 (2002).
64. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402–408 (2001).
65. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

Acknowledgements

We thank P. Schär, D. Schürmann, S. Kaspar, M. Florescu, D. Jankovic, E. Paraskevopoulou, B. Schlegelberger, P. Lundberg, R. Skoda, M. Kristiansen, C. Porcher, P. Vyas, P. Fillipakopoulos, A. Jeltsch, S. Kudithipudi, L. Steiner, S. Atanasoski, A. Schmidt, D. Labes, E. Traunecker, N. Meier, B. Fahrenkrog, L. Brault, T. Pabst, R. Hasserjian, and M. Hayman for reagents and/or other support of this study. We also appreciate the service of the flow- and animal experimentation facilities of the DBM & FMI, and we particularly thank Tianke Wang, Fred Zilbermann, Patrick Kopp, and Jean-Francois Spetz for the generation of the *Nsd^{fl}* mouse strain from the *Nsd1^{+/-L3}* allele. This work was supported by funding from the Swiss Cancer League (SKL): KFS-3487-08-2014 and KFS-4258-08-2017; the Swiss National Science Foundation (SNF), 31003_A_173224/1), the Gertrude Von Meissner Foundation Basel, the Novartis Biomedical Research Foundation, Basel, the San Salvatore Foundation, Lugano (201525), and the Wilhelm-Sander Foundation, München (2017-035.1).

Author contributions

K.L., M.A., S.T., and F.O.B. planned, performed, analyzed research, and wrote the manuscript; C.T., S.J., T.B., H.M., and M.F.B. performed and analyzed research; A.T., R.I., and R.L. provided services and/or critical material; A.H.F.M.P. and T.M. supervised research and wrote the manuscript; J.S. planned, performed, analyzed research, supervised the project, and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-16179-8>.

Correspondence and requests for materials should be addressed to J.S.

Peer review information *Nature Communications* thanks Emery Bresnick and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Review

NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer

Samantha Tauchmann  and Juerg Schwaller * 

University Children's Hospital, Department of Biomedicine, University of Basel, 4031 Basel, Switzerland; Samantha.Tauchmann@unibas.ch

* Correspondence: J.Schwaller@unibas.ch; Tel.: +41-61-265-3517 or +41-61-265-3504

Abstract: Recurrent epigenomic alterations associated with multiple human pathologies have increased the interest in the nuclear receptor binding SET domain protein 1 (NSD1) lysine methyltransferase. Here, we review the current knowledge about the biochemistry, cellular function and role of NSD1 in human diseases. Several studies have shown that NSD1 controls gene expression by methylation of lysine 36 of histone 3 (H3K36me1/2) in a complex crosstalk with de novo DNA methylation. Inactivation in flies and mice revealed that NSD1 is essential for normal development and that it regulates multiple cell type-specific functions by interfering with transcriptional master regulators. In humans, putative loss of function NSD1 mutations characterize developmental syndromes, such as SOTOS, as well as cancer from different organs. In pediatric hematological malignancies, a recurrent chromosomal translocation forms a NUP98-NSD1 fusion with SET-dependent leukemogenic activity, which seems targetable by small molecule inhibitors. To treat or prevent diseases driven by aberrant NSD1 activity, future research will need to pinpoint the mechanistic correlation between the NSD1 gene dosage and/or mutational status with development, homeostasis, and malignant transformation.



Citation: Tauchmann, S.; Schwaller, J. NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer. *Life* **2021**, *11*, 877. <https://doi.org/10.3390/life11090877>

Academic Editors: Albert Jeltsch and Arunkumar Dhayalan

Received: 14 July 2021

Accepted: 23 August 2021

Published: 25 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: NSD1; H3K36; SOTOS; cancer; NUP98-NSD1; AML

1. Introduction

Gene expression is controlled by temporarily and spatially coordinated modification of chromatin. Hereby, the N-terminal tails of the histone octamers formed by H2A, H2B, H3, and H4 undergo post-translational modifications including methylation, phosphorylation, acetylation, ubiquitylation and sumoylation executed by proteins acting as “writers” of an epigenetic code [1]. Histone lysine methyltransferases (KMTs) have been characterized as critical regulators of multiple cellular processes including DNA replication, DNA damage response, cell cycle progression or cytokinesis. Genetic lesions (mutations, translocations) as well as altered gene expression functionally affecting KMTs are recurrently found in various human malignancies but also in developmental disorders [2]. An increasing number of compounds that selectively target aberrantly activated KMTs have been developed and underwent clinical trials as novel cancer therapeutics [3]. In this review, we summarize the current knowledge on the nuclear receptor binding SET domain protein 1 (NSD1, aka KMT3B), a H3 lysine 36 (H3K36) methyltransferase that has recently gained attention because of its critical role in several human pathologies, such as germline developmental syndromes and cancers.

2. Identification and Structure of NSD1

NSD1 was discovered in a yeast two hybrid screen for proteins associated with the ligand-binding domain (LBD) of the retinoic acid receptor alpha (RARα). NSD1 was shown to interact directly with the LBD of several nuclear receptors, including the retinoic acid (RAR), thyroid (TR), retinoid X (RXR), and estrogen (ER) receptors. These interactions are mediated by two distinct nuclear receptor interaction domains (NID) in NSD1, NID^{-L} and

NID^{+L} . NID^{-L} interacts with RAR and TR when a ligand is absent, whereas NID^{+L} binds RAR, TR, RXR and ER when a ligand is present, indicating that NSD1 controls repression or activation of target genes by distinct binding to nuclear receptors [4]. Similarly, a yeast-two-hybrid screen, using the LBD of the androgen receptor (AR) and the orphan receptor TR4 as baits, allowed for the detection of a human androgen receptor-associated protein of 267 Kd (ARA267) that showed the highest homology to mouse NSD1. ARA-267 (which turned out to be NSD1) was shown to be widely expressed in different tissues, with highest levels in lymph nodes. Functional studies have suggested its primary role as co-activator of AR controlled transcription [5].

The NSD1 gene maps to human chromosome 5q35.3, close to the telomere, with an 8088 bp open reading frame (ORF) [6]. Interrogation of ensembl.org indicates the existence of three NSD1 isoforms produced by alternative splicing, one long isoform and two shorter ones, with additional potential smaller isoforms that have been computationally mapped [7]. NSD1 isoform 1 (NSD1(204), Q96L73-1; ARA267-beta) with an ORF starting at exon 2 and ending at exon 23 has been chosen as the canonical sequence and is 2696 amino acids (aa) long, resulting in 296 kDa [6,7]. NSD1 isoform 2 (NSD1(202), Q96L73-2; ARA267-alpha) is 2427aa and 267 kDa and differs at the 5'UTR, compared to isoform 1 where 1-269aa are missing [8]. Furthermore, through an mRNA splicing event, a 740 bp long intron within exon 2 is removed, leading to an additional exon with 90 bp (exon 3), resulting in a total length of 24 exons. NSD1 isoform 3 (NSD1(201), Q96L73-3) is similar to isoform 2 with a 740 bp spliced intron; however, it differs by lacking 310-412aa, thereby resulting in a smaller intron between exon 1 and 2, with 841 bp. [7] Furthermore, exon 24 has a length of 1931 bp, which is smaller compared to isoform 2 that has a 6379 bp long exon 24. However, the ORF for both, isoform 2 and 3, starts at exon 2 and ends at exon 24, resulting in the same length and size of the protein (Figure 1A). Notably, the three isoforms (204, 202, 201) encode for proteins that contain all of the functionally characterized NSD1 domains, suggesting that variations close to the 5' end of the ORF may be linked to regulation of gene expression.

Interrogating public databases suggests ubiquitous NSD1 expression (or its related homologs) in most tissues from various organisms. Somehow higher NSD1 mRNA levels seem to be expressed in normal brain, pancreas, male reproductive tract, and hematopoietic organs such as the bone marrow and lymphoid tissues [9]. Significant NSD1 mRNA expression in bone marrow polymorphonuclear cells, CD4, CD8 and NK cells is also supported by genevisible.com [10]. Integrated expression analysis in normal tissues and cell lines indicates abundant NSD1 protein expression in B-lymphocytes, CD8 T cells, platelets, fetal brain, retina, fetal gut, rectum, liver, adipocytes, pancreas, placenta and ovaries [11]. However, there seems to be an overall low tissue specificity for NSD1 protein expression.

The NSD1 protein contains two NIDs, two proline-tryptophan-tryptophan-proline (PWWP) domains, five plant homeodomains (PHD), an atypical (C5HCH) plant homeo-domain (PHD) finger and a catalytic domain (CD) composed of a pre-SET (AWS), Su(var)3-9, Enhancer-of-zeste, Trithorax (SET) and post-SET domain [6]. The aa sequences from both the PWWP-I and PHD-II domains are 100% identical between mouse and human NSD1, while the SET domain is 99% identical. A 97% homology between human and mouse was found for PHD-I and PHD-III. PWWP-II was 95% conserved whereas the NID^{-L} and NID^{+L} showed the least identity, with 88 and 83%, respectively [6] (Figure 1B).

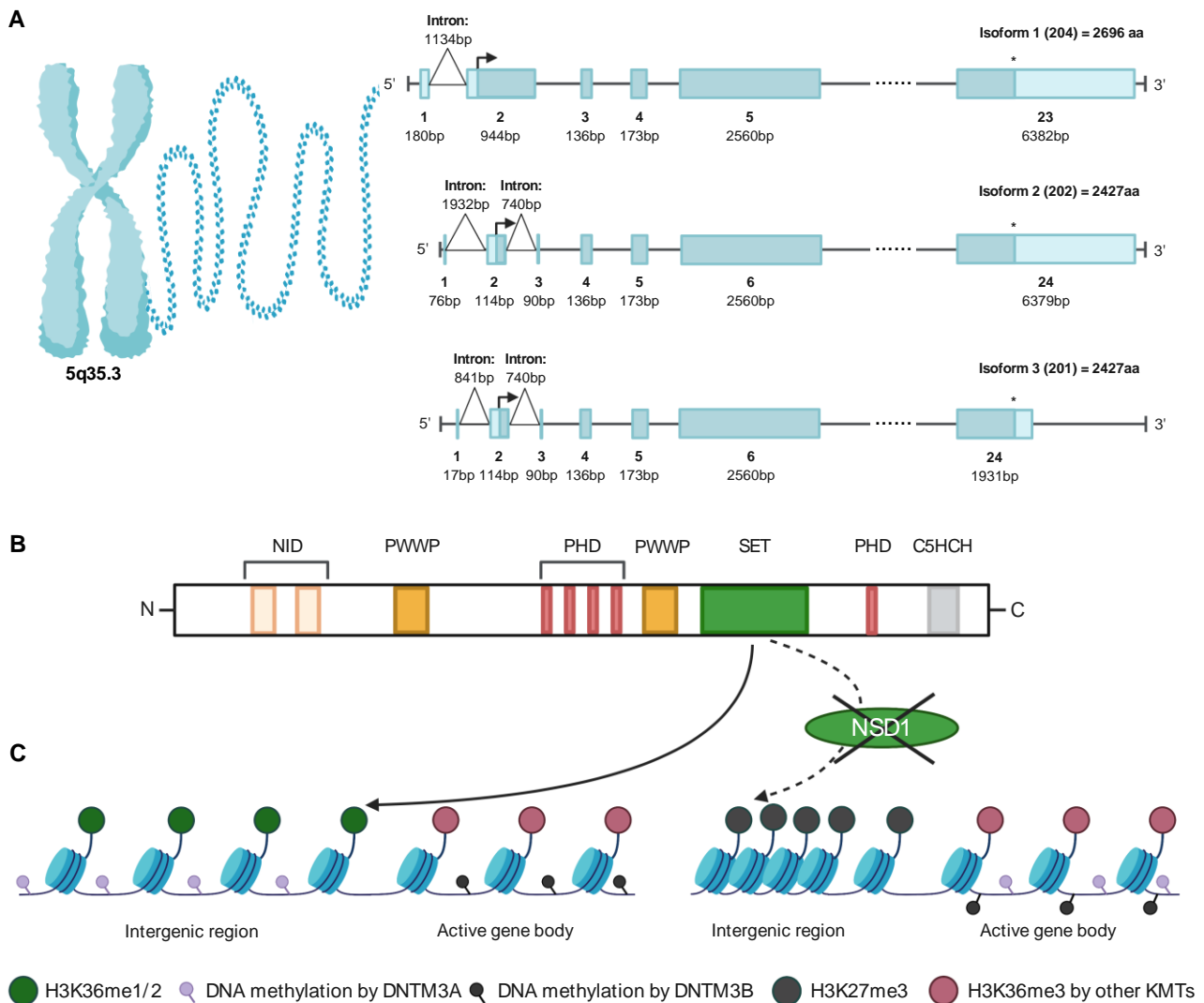


Figure 1. NSD1 gene and protein architecture and function. (A) Exon structure of the three different *NSD1* isoforms. Isoform 1 (204) contains 23 exons, whereas isoform 2 (202) and 3 (201) contain 24 exons. Open reading frame is shown by an arrow as start and asterisk at the end. (B) All three major *NSD1* isoforms contain two nuclear receptor interacting domains (NID), two proline-tryptophan-tryptophan-proline (PWWP), five plant homodomain zinc fingers (PHD), the catalytic Su(var)3-9, enhancer-of-zeste, Trithorax (SET) and the C-terminal C5HCH (Cys-His) domain. (C) The *NSD1* SET domain methylates H3K36me1/2 predominantly at intergenic regions allowing recruitment of DNMT3A and facilitating H3K36me3 by other KMTs allowing recruitment of DNMT3B to active gene bodies. Reduced *NSD1* catalytic activity results in loss of H3K36me2 marks, which allows spreading of PRC2-mediated H3K27me3 marks at intergenic regions and redistribution of DNMT3A-mediated DNA methylation to active gene bodies.

NSD1 is a member of a SET-containing methyltransferase protein family, which contains two additional members, *NSD2* and *NSD3*. Both are significantly smaller than *NSD1* due to the absence of the NID^{-L} and NID^{+L} in the N-terminus. *NSD2*, also called Wolf-Hirschhorn Syndrome Candidate 1 (WHSC1) or Multiple Myeloma SET domain protein (MMSET) is located on the short arm of chromosome 4 (4p16.3), a locus targeted by a recurrent t(4;14)(p16;q32) translocation found in up to 20% of patients with multiple myeloma. *NSD2* contains a PWWP domain, a SET domain, PHD zinc fingers and a high mobility group (HMG) box with 75% homology to *NSD1* [6]. Similarly, *NSD3*, also called WHSC1L1, contains a PWWP, SET and PHD zinc finger domains but lacks the HMG box and is therefore only 68% identical to *NSD1*. *NSD3* was mapped to the short arm of chromosome 8 (8p11.2), a locus involved in cancer-associated amplifications and translocations, such as

t(8;11)(p11;p15) associated with myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML) [12].

3. NSD1 Is an Epigenetic Regulator Writing and Reading Chromatin Marks

3.1. The SET Domain Mediates the Catalytic Activity

NSD1-3 have been functionally characterized as histone methyltransferases (HMT) due to its conserved catalytic SET domain involved in methylation of histone 3-K4, -K9, -K27, -K36, and -K79, and methylation of histone 4-K20 [13]. Members of the NSD family seem to differ from other protein lysine methyltransferases (PKMTs) as in the absence of a ligand, the SET histone binding site is closed, preventing any access to the catalytic groove [14]. In general, SET domains are approximately 130 aa long and contain binding sites for the lysine ligand and the co-factor S-adenosylmethionine (SAM), which donates methyl groups. The C-terminal post-SET domain can form a loop, thereby regulating substrate binding by forming one side of the SAM binding pocket [15]. The PWWP domains of NSD1 are critical for binding to H3K36me marks but also to DNA, whereas the PHD zinc fingers are needed for interactions with other methylated histones, such as H3K4 and H3K9 [16].

Several, mostly in vitro studies, reported other histones (H4K20) and non-histone proteins as potential NSD1 substrates. Berdasco et al. found that loss of NSD1 by 5'-CpG island DNA hypermethylation interferes with histone lysine methylation not only by decreasing the levels of H3K36me3 but also of H4K20me3 [17]. Lu et al. suggested that NSD1 acts (in tandem with the F-box and leucine-rich repeat protein 11 (FBXL11) demethylase) as a regulator of the NF κ B signaling pathway indicated by reversible methylation of K218 and K221 of NF κ B-p65. However, these observations were based on associations upon NSD1 overexpression or knockdown, and not validated in biochemical assays [18]. Using a biochemical approach, others were unable to validate NSD1-SET mediated methylation of H4K20 and NF κ B-p65. However, they found in addition to H3K36, H1 linker histones, in particular H1.5 (K168) but also H1.2 (K168) and H1.3 (K169) as well as H4 (K44), as potential NSD1 substrates. Furthermore, they identified peptides of 50 non-histone proteins recognized by NSD1-SET. NSD1 methylation on two of those non-histone proteins, the chromatin remodeler ATRX (K1033) and the small nuclear RNA-binding protein U3 (K189), could be validated in vitro [19].

3.2. NSD1 Chromatin Modification and Regulation

Methylation of histone H3K36 occurs in three states mono-, di- and trimethylation and is primarily described as a hallmark of active transcription. Several KMTs were shown to be recruited by RNA polymerase II and deposit H3K36me3 over gene bodies essential for transcriptional elongation, whereas H3K36me2 is enriched at intergenic regions or promoters [20,21].

Functional studies have shown that NSD1 catalyzes mono- and dimethylation of H3K36 specifically. NSD2 leads to mono- and dimethylation of H3K36, whereas it prefers to catalyze dimethylation compared to monomethylation. Interestingly, H3K36me2 marks are not only set by NSD1-3 but also by ASH1L (ASH1 Like Histone Lysine Methyltransferase), whereas SETD2 (SET Domain Containing 2, Histone Lysine Methyltransferase) is the only enzyme able to introduce K36 methylation up to the trimethylation stage (H3K36me3) [22]. Previous studies have shown that NSD1 exhibits an autoinhibitory state that is relieved by binding to nucleosomes enabling dimethylation of histone H3 at Lys36 (H3K36) [23]. To better understand H3K36 recognition by NSD proteins, Li et al. recently solved the cryo-electron microscopy structures of mononucleosome-bound NSD2 and NSD3 [24]. They observed that binding of NSD2 and NSD3 causes DNA near the linker region to unwrap, facilitating insertion of the catalytic core between the histone octamer and the unwrapped DNA segment. Multiple DNA- and histone-specific contacts between NSD and the nucleosome precisely defined the position of the enzyme on the nucleosome.

Yuan et al. suggested that H2A mono ubiquitination (ubH2A) impairs the enzymatic activity of HMTs including NSD1, indicating another layer of complexity in NSD1 regulation [25]. Notably, ubH2A can recruit the Polycomb Repressive Complex 2 (PRC2). PRC2 regulates gene expression by methylation of lysine 27 of histone 3 (H3K27) marks through its enzymatic component EZH2 (Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit). The different degrees of H3K27 methylation (H3K27me1/me2/me3) have distinct genomic distributions: H3K27me1 is enriched within gene bodies of actively transcribed genes; H3K27me2 is abundant, marking 50–70% of total histone H3 and covering inter- and intragenic regions. H3K27me3 (present on 5–10% of histone H3) is strongly enriched at sites overlapping with PRC2 binding and is considered the hallmark of PRC2-mediated gene repression [26]. Streubel et al. found that genetic inactivation of *Nsd1* leads to genome-wide expansion of H3K27me3 not only at PRC2 target genes but also as de novo accumulation within broad H3K27me2 marked domains. Thus, NSD1-mediated H3K36me2 seems crucial to restrict PRC2 activity by preventing uncontrolled deposition of H3K27me3 [27].

3.3. Functional Interaction with DNA Methyltransferases

In addition to PRC2, epigenomic regulation by NSD1 also involves DNA methyltransferases (DNMTs), which methylate CpG dinucleotides. In total, there are five different DNMTs, of which three play a role in DNA methylation. DNMT1 is important to maintain methylation during DNA replication and acts in response to DNA damage, while DNMT3A and DNMT3B are responsible for de novo methylation [28]. DNMT3 enzymes are recruited through their PWWP domain to methylated H3K36 [29]. DNMT3B colocalizes selectively with H3K36me3 and methylates active gene bodies to enhance gene expression. DNMT3A binds more strongly to H3K36me2 than to H3K36me3 and preferentially methylates intergenic chromatin, which often co-occurs with PRC2-mediated H3K27me2 as well as NSD1-mediated-H3K36me2 [30,31]. Functional studies in ES cells revealed that ablation of NSD1 results in redistribution of DNMT3A to H3K36me gene bodies and reduced methylation of intergenic DNA [32]. Likewise, expression of a H3K36M mutant (not recognized by NSD1), resulted in an increase in H3K27me3 at intergenic regions and redistribution of PRC2 resulting in aberrant gene expression [31] (Figure 1C).

3.4. Regulation of Gene Expression

Depletion of NSD1 leads to both up- and down-regulation of gene expression, indicating NSD1 functions as transcriptional co-activator and co-repressor. In earlier studies, distinct stretches of the NSD1 ORF sequence were tested for their transcriptional activity by fusing them to a GAL4 DNA binding domain, which identified a region (1084–1400 aa) with a significant repressive activity in vitro. This suggested that NSD1 has a silencing domain that functions autonomously, which might act as corepressor for unliganded TR and RAR [4]. Although the mechanisms of gene repression by NSD1 are not fully understood, experimental work suggested that transcription is impaired through binding of the NSD1 C5HCH domain (adjacent to the C-terminus of PHD-V) to the C2HR zinc finger motif of ZNF496 (aka NSD1 interacting zinc finger protein 1, NIZP1) tethered on RNA polymerase II promoters [33,34]. In contrast, only expression of an N-terminal stretch of NSD1 (1–731) fused to the estrogen-receptor alpha DNA binding domain showed strong transcriptional activation in yeast but not mammalian cells [4]. More recent work demonstrated that loss of NSD1 increases H3K27ac associated with active enhancers in mESCs. NSD1 was shown to recruit the histone deacetylase 1 (HDAC1), which can deacetylate H3K27ac. Hence, inactivation of HDAC1 recapitulated increased H3K27ac similar to loss of NSD1 [35]. Overall, although these studies provided some insights into the role of NSD1 as a transcriptional co-repressor, its function as a co-activator, particularly in the context of specific nuclear receptors remains poorly understood.

4. Cellular Functions of NSD1

Earlier *in vitro* studies showed that NSD1 overexpression allowed NIH-3T3 fibroblasts to grow in reduced serum levels, whereas vector-transfected control cells did not. Overexpression of *Schizosaccharomyces pombe* SET2, which contains a SET domain but no PHD or PWWP domains, conferred reduced serum dependence, indicating that the catalytic NSD1 activity is able to modulate serum dependence [36].

4.1. Modeling NSD1 Activity in the Fly

To better understand the function of NSD1 *in vivo*, gain- and loss-of-function studies in various organisms have been performed. Ubiquitous NSD (the fly NSD1 homolog) overexpression in *Drosophila melanogaster* caused developmental delay and reduced body size at the larval stage, resulting in pupal lethality. Targeted overexpression in various tissues led to significant alterations that rescued RNAi-based NSD knockdown. NSD overexpression enhanced the transcription of pro-apoptotic genes and led to caspase activation. Notably, NSD-overexpression associated wing atrophy was reduced by a loss-of-function mutation in Jun N-terminal (JNK) kinase [37]. NSD1 overexpression in *Drosophila* imaginal discs induced organ atrophy. Interestingly, ectopic expression of the DNA replication-related element-binding factor (DREF) resulted in increased NSD expression [38]. DREF proteins are central regulators of cell proliferation; however, whether the human homolog ZBED1 (zinc finger BED-type-containing 1) regulates NSD1 expression remains unknown. Pan-glial, but not pan-neuronal NSD overexpression induced apoptosis in *Drosophila* larval brain cells. However, pan-glial NSD overexpression also induced caspase-3 cleavage in neuronal cells. Among the various glial cell types, NSD overexpression in only astrocytic glia induced apoptosis and abnormal learning defects in the larval stage. These observations in *Drosophila* suggested that aberrant NSD expression may result in neurodevelopmental disorders through functional interference with astrocytes [39]. In contrast, NSD deletion by CRISPR/Cas9-mediated knock-out resulted in an increase in the body size of *Drosophila* larvae. Although the NSD mutant flies survived to adulthood, their fecundity was dramatically decreased. NSD lacking flies also showed neurological dysfunctions, such as lower memory performance and motor defects, and a diminished extracellular signal-regulated kinase activity [40]. Collectively, these functional studies in the fly suggested that NSD is a central regulator of proliferation and, cell and/or body size.

4.2. Modeling NSD1 Activity in the Mouse

To gain insight into the biological functions of NSD1 in mammals, Losson and colleagues have generated mice carrying a floxed *Nsd1* exon 5 containing the nuclear factor interaction domain. Ubiquitous inactivation (*Actin-iCre;Nsd1^{ff}*) embryos displayed a high incidence of apoptosis and failed to complete gastrulation, indicating that NSD1 is essential for early post-implantation development [41]. More recent work, using the same *Nsd1^{ff}* allele, showed that conditional targeted ablation in primordial germ cells (*Tnap-iCre;Nsd1^{ff}*) resulted in male sterility associated with absence of mature spermatozoa and loss of testicular germ cells in adult testis and epididymis. A similar effect was seen when DNMT3A was conditionally ablated in germ cells. Male mutant mice presented with impaired spermatogenesis due to loss of methylation at two out of three paternally imprinted loci in spermatogonia [42]. Molecular studies confirmed previous findings that NSD1 safeguards a subset of genes against H3K27me3-associated transcriptional silencing. In contrast, H3K36me2 in oocytes is predominantly dependent on the SETD2 HMT coinciding with H3K36me3. Hence, in contrast to males, *Nsd1^{-/-}* females are fertile. These studies showed that NSD1 plays a critical role in the maturation of mouse gametes by regulating distinct profiles of H3K36 methylation [43]. A third study using the floxed *Nsd1* mouse allele generated by Losson et al. inactivated the gene in the hematopoietic system. Unexpectedly homozygous ablation during late fetal liver hematopoiesis (*Vav-iCre;Nsd1^{ff}*) resulted in a fully penetrant hematological malignancy phenocopying many aspects of human acute erythroleukemia. Functional studies revealed that lack of *Nsd1*

impairs terminal differentiation of erythroblasts, which could be rescued by expression of wildtype, but not a catalytically inactive SET-domain NSD1^{N1918Q} mutant. Interestingly, NSD1, but not the inactive mutant, significantly increased the occupancy of the erythroid transcriptional master regulator GATA1 at target genes and their expression. These studies identified NSD1 as a novel regulator of GATA1-controlled erythroid differentiation [44]. Very recently, Zou and coworkers used the same floxed murine *Nsd1* allele for targeted activation of the gene in mesenchymal progenitor cells (*Prx1-iCre;Nsd1^{fl/fl}*). Ablation of *Nsd1* in mesenchymal progenitors resulted in impaired cartilage development, skeletal growth defects, and impaired fracture healing. Chondrogenic differentiation was impaired, which was associated with reduced H3K36me2 marks and lower expression of critical mediators including the SRY-box transcription factor 9 (SOX9). Interestingly, in chondrocytes NSD1 seems to bind the promoter and to control expression of the hypoxia-inducible factor 1alpha (HIF1alpha), a well-known regulator of SOX9 [45]. Importantly, *Sox9* overexpression rescued the chondrogenic differentiation effects of *Nsd1*^{-/-} cells. Collectively, these data suggest that NSD1 controls chondrogenic differentiation by direct (H3K36me2) and indirect (HIF1A) regulation of SOX9 [46].

Piper and colleagues used a CRISPR/Cas9 strategy to inactivate exon 3 of *Nsd1*. Although they did not find any major morphologic defects in *Nsd1*^{+/-} brains, the animals exhibited deficits in social behavior without significant learning or memory deficits. *Nsd1*^{-/-} E9.5 embryos had a smaller prosencephalon compared to heterozygous and wild-type animals, with abnormal morphology and aberrant formation of the luminal cavity of the brain [47]. Taken together, NSD1 inactivation studies in *Drosophila* and mice showed that NSD1 is essential for normal development and that it regulates a wide variety of cellular functions, of which many seem to be cell type-specific, most likely by controlling the activity of distinct transcriptional master regulators.

5. Role of NSD1 in Human Diseases

5.1. Aberrant NSD1 Activity Is a Hallmark of Developmental Syndromes

Germline lesions (including missense, truncating and splice-site mutations and sub-microscopic deletions) potentially resulting in loss-of-function of the NSD1 protein have been linked to a developmental syndrome called SOTOS [48]. SOTOS is a childhood overgrowth syndrome characterized by a distinctive facial appearance, physical overgrowth with height and head circumference >97th percentile, advanced bone age and learning disabilities [49]. Interestingly, microduplications of 5q35.2–q35.3 encompassing the *NSD1* gene locus have been reported in rare patients with a clinically reversed SOTOS syndrome. These individuals are characterized by short stature, microcephaly, learning disability or mild to moderate intellectual disability, and distinctive facial features. These observations suggest that the NSD1 gene dosage determines the phenotype of these developmental syndromes [50].

Analysis of a cohort of >700 individuals with overgrowth and intellectual disability revealed a putative causal mutation in less than 15 genes in almost half of the individuals [51]. Notably, epigenetic regulation was a prominent biological process not only represented by NSD1 but also by five additional genes including PRC2 complex proteins (EZH2, EED), H1.5 linker histone (HIST1H1E), the de novo DNMT3A methyltransferase, and the chromatin remodeler CHD8. Other patients had mutations in genes controlling cellular growth (PTEN, AKT3, PIK3CA, MTOR, PPP2R5D). The PI3K/AKT pathway is a central regulator of growth by increased cell metabolism, survival, and turnover, as well as protein synthesis. As deregulated cellular growth is a hallmark of cancer, and certain human overgrowth syndromes are associated with increased cancer risk, it is not unexpected that the majority of the mutated genes in overgrowth syndromes including NSD1, EZH2, DNMT3A, PTEN, CHD8, HIST1H1E, MTOR, PIK3CA are also frequently altered in human cancers [51]. Interestingly, overgrowth-related PIK3CA mutations were shown to exhibit a striking allele dose-dependent stemness phenotype in human pluripotent stem cells (PSC) [52,53]. Whether NSD1 mutations affect PSC stemness remains unknown.

Analysis of genome-wide DNA methylation of SOTOS syndrome patients revealed a highly specific signature able to differentiate patients with pathogenic NSD1 mutations from controls, benign NSD1 variants and clinically overlapping syndromes. This NSD1^{+/-} DNA methylation signature encompasses genes that function in cellular morphogenesis and neuronal differentiation reflecting cardinal features of SOTOS syndrome [54]. SOTOS-related DNA methylation signatures were used to model epigenetic clocks that predict biological age. The so-called Horvath epigenetic clock model revealed that NSD1 loss-of-function mutations substantially accelerate epigenetic aging [55].

5.2. Aberrant NSD1 in Human Cancers

The first evidence linking NSD1 genetic aberrations to cancer came from cloning of a cytogenetically silent t(5;11)(q35;115) chromosomal translocation associated with pediatric de novo MDS or aggressive AML that leads to fusion of the N-terminal domains of the nucleoporin 98 (NUP98) protein to the C-terminal part (including the SET) of NSD1 [56]. Importantly, in most patients, additional genetic lesions are found in NUP98-NSD1⁺ AML cells of which activating FLT3-ITD mutations are by far the most prevalent, present in about 80% of the cases [57]. Reconstitution of lethally irradiated mice with bone marrow retrovirally overexpressing the NUP98-NSD1 fusion (in presence or absence of a functionally cooperating FLT3-ITD mutation) was reported to induce an AML-like disease in mice [58,59]. Functional studies suggested that the NUP98-NSD1 fusion binds genomic elements adjacent to the *HoxA7* and *HoxA9* loci and maintains histone H3K36 methylation and histone acetylation, preventing transcriptional repression of the *HoxA* gene cluster during differentiation. Structure functional analysis indicated that the phenylalanine-glycine (FG) repeats of the NUP98 moiety as well as the NSD1-SET domain are necessary for its transforming activity [58]. Targeted sequencing of a large number of genes associated with hematologic malignancies revealed rare and potentially deleterious *NSD1* mutations in AML patients suggesting that not only gain but also loss of NSD1 can contribute to transformation of hematopoietic cells [60].

Analysis of cancer-associated aberrant CpG promoter methylation revealed epigenetic silencing of NSD1 in human brain tumor cell lines associated with reduced H3K36 methylation [17]. While NSD1 overexpression impaired colony growth in semi-solid medium and proliferation of cancer cells, RNAi-mediated knock-down increased proliferation, suggesting a role of a tumor suppressor [17]. Frequent NSD1 epigenetic silencing was also found in human clear cell renal cell carcinoma (ccRCC). Notably, tumors harboring NSD1 promoter methylation were of higher grade and stage, and NSD1 promoter methylation correlated with somatic mutations in the SETD2 H3K36me3 HMT. Interestingly, ccRCC with epigenetic NSD1 silencing displayed a specific genome-wide methylome signature consistent with the NSD1 mutation methylome signature observed in SOTOS syndrome [61]. Comprehensive genomic characterization of human head and neck squamous cell carcinomas (HNSCC) identified inactivating NSD1 mutations and focal homozygous deletions in up to 10% of the patients [62]. Further studies revealed recurrent mutations including a K36M oncomutation in multiple H3 histone genes. Interestingly, direct in vitro inhibition of NSD2 and SETD2 by H3K36M has been described, whereas inhibition of NSD1 was only found in steady-state kinetic analysis using inhibitory H3 (27–43) peptide containing K36M [63,64]. Notably, along with previously described NSD1 mutations, they corresponded to a specific DNA methylation cluster. In addition, the K36M substitution and NSD1 defects converged on altering methylation of H3K36, subsequently blocking cellular differentiation and promoting oncogenesis [62]. Extensive genetic analysis of HNSCCs revealed that, similar to what has been experimentally observed in ES cells, loss of function NSD1 mutations are responsible for reduced intergenic H3K36me2 marks, followed by loss of DNA methylation and gain of H3K27me3 in the affected genomic regions. Those regions seem enriched in cis-regulatory elements, and subsequent loss of H3K27ac correlated with reduced expression of putative target genes [65]. In addition to HNSCC, H3.3 K36M mutations are recurrently found in several rare human cancers including chondroblastomas and poorly

differentiated sarcomas. Comparison of the epigenomic and transcriptomic landscape of mesenchymal cells experimentally depleted of H3K36me2 indicated recapitulation of H3K36M's effect on H3K27me3 redistribution and gene expression [66]. Notably, transgenic mice overexpressing H3.3K36M in the hematopoietic system developed a lethal phenotype characterized by blocked erythroid differentiation that was very similar to that reported upon conditional *Nsd1* inactivation again supporting the converting consequences on epigenomic regulation [44,67].

A similar hypomethylated tumor subtype enriched for inactivating NSD1 mutations and deletions was also found in lung squamous cell carcinoma (LUSC). NSD1-altered HNSCC and LUSC correlated at the DNA methylation and gene expression levels, featuring ectopic expression of developmental transcription factors and genes that are also hypomethylated in SOTOS syndrome. Reduced expression of NSD1 was also reported to be part of an epigenetic gene signature able to distinguish non-malignant tumor from tissue of prostate cancer. Surprisingly, metastatic lesions appeared to express significantly higher NSD1 levels than primary tumors [68]. Highly prevalent NSD1 mutations were also found in testicular germ cell tumors, and low NSD1 expression was associated with resistance to cisplatin [69]. However, the functional significance of NSD1 alterations in human urogenital cancers remains to be investigated.

Comprehensive genomic analysis of 21 tumor types originating from >6000 samples revealed that the degrees of overall methylation in CpG island and demethylation in intergenic regions, defined as the 'backbone', are highly variable between different tumors [70]. Interestingly, NSD1 mutations showed the most significant association with backbone DNA demethylation not only in HNSCC but also in other cancers. In fact, bi-allelic NSD1 aberrations by mutation or gene copy loss showed the highest backbone demethylation [70]. A computational search for cancer predisposition genes based on the Knudson's two-hit hypothesis using genome data of ~10,000 tumors identified genes including *NSD1* that may contribute to cancer through a combination of rare germline variants and somatic loss-of-heterozygosity (LOH). Interestingly, rare germline variants in such genes may contribute substantially to cancer risk, particularly of ovarian carcinomas, but also other cancers [71].

Researchers also explored the correlation between allele frequency of somatic variants and total gene expression of the affected gene using matched tumor and normal RNA and DNA sequencing data from almost 400 individuals across 10 cancer types. They defined higher allele frequency of somatic variants in cancer-implicated genes. This study revealed that somatic alleles bearing premature terminating variants (PTVs) in cancer implicated genes seemed to be less degraded via nonsense-mediated mRNA decay, possibly favoring truncated proteins. Notably, NSD1 appeared as a gene with more than five somatic variants and PTVs with high allele frequency [72].

Collectively, increased NSD1-SET activity drives a particular hematological cancer, whereas loss-of function mutation or impaired expression characterize a wide variety of mostly solid human cancers (Figure 2).

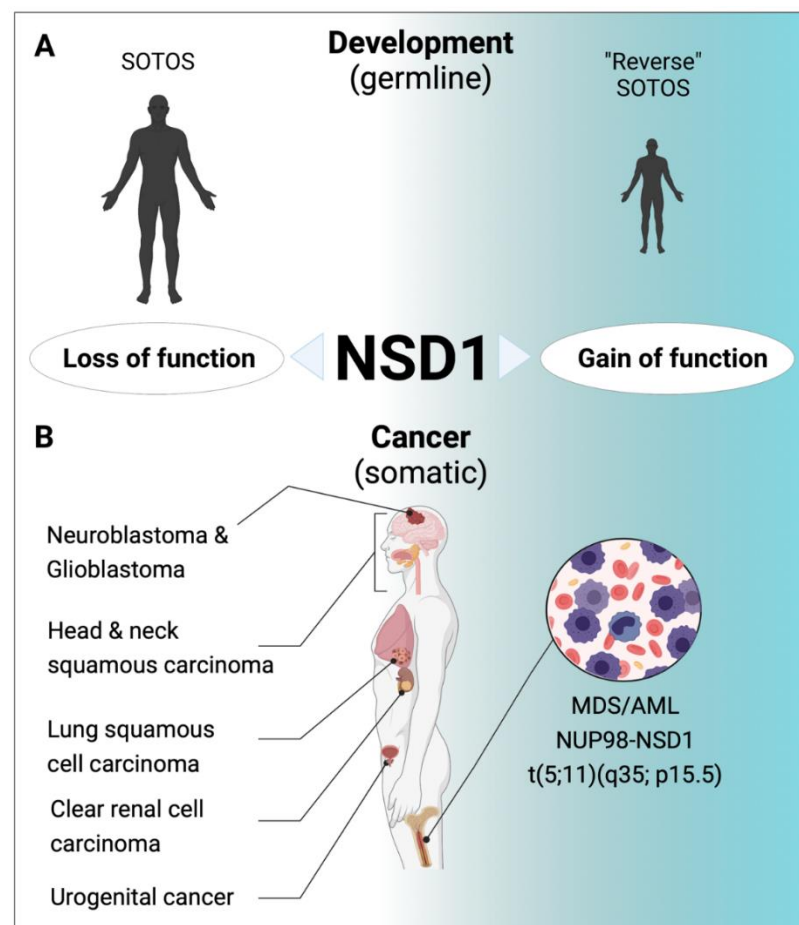


Figure 2. Role of NSD1 in human diseases. **(A)** Inactivating mutations in NSD1 are the molecular hallmark of SOTOS syndrome, a developmental disorder characterized by a distinctive facial appearance, physical overgrowth advanced bone age and learning disabilities [48,49]. “Reverse” SOTOS Syndrome is characterized by a short stature, microcephaly, and learning disability, and is associated with microdeletions of 5q35 carrying NSD1 [50]. **(B)** Putative loss of function mutations of NSD1 are among the most prevalent lesions in human head and neck and lung squamous cell carcinomas, neuroblastomas and glioblastomas [17,61,62,65,66,68]. NSD1 gene silencing was found in human clear cell renal cell carcinoma, and urogenital cancers [69,71]. In pediatric myeloid malignancies (de novo MDS and AML) the chromosomal translocation $t(5;11)(q35;115)$ results in expression of a NUP98-NSD1 fusion gene with SET-dependent leukemogenic activity [56–59].

6. Therapeutic Interference with NSD1

Several strategies have been explored to selectively interfere with NSD1 activity. Earlier work characterized a small molecule methyltransferase inhibitor (BIX-01294) able to modulate H3K9 methylation. BIX.01294 was characterized as a G9a inhibitor by binding to the histone-tail groove in the SET domain [73]. Notably, BIX-01294 was also found to differentially inhibit NSD1, NDS2 and NSD3 in vitro based on the structural conserved catalytic SET domain but the molecule clearly lacks any NSD1 specificity [74].

NSD1 contains several PHD zinc fingers, whereby the PHD-V C5HCH domain serves as a binding site for protein–protein interactions. This region is particularly interesting as it has been shown to be involved in dysregulated Hox gene activation in AML and occurrence of point mutations in SOTOS Syndrome [16]. PHD-V C5HCH recruits a transcriptional repressor, resulting in a direct finger–finger interaction with the C2HR domain of Nizp1 [34]. The consequences of this binding are not clear; therefore, interfering with this interaction would be interesting to elucidate the biological and pathological relevance. Targeting PHD fingers has been considered to pharmacologically interfere with protein function; however,

the affinity of compounds to specifically target a particular region is not advanced enough to be implemented *in vivo*. Berardi et al. designed a computational and experimental pipeline to investigate the druggability by using a 3D model of the PHD-V C5HCH domain of NSD1 with the C2HR domain of Nizp1 [34]. Applying a structure-based *in silico* screening following NMR validation, they found three structurally related molecules that were able to bind to the PHD-V C5HCH domain of NSD1: type II topoisomerase inhibitor mitoxantrone, chloroquine and quinacrine. Even if these compounds are interesting to target the NSD1/Nizp1 interaction, the consequences of derepressing transcription and selective inhibition are not clear and more functional studies have to be performed before this can be translated into the clinic.

Using a luminescence screening platform that quantifies S-adenosyl homocysteine (which is produced during methyl transfer from S-adenosylmethionine used by NSD1 and other HMTs), researchers identified suramin and other scaffolds as potential inhibitors of the enzymatic NSD1 HMT activity [75]. A computational strategy incorporating ligand contact information into classical alignment-based comparisons applied to SET containing proteins revealed additional scaffolds that inhibited NSD1 activity [76].

More recently, Grembecka, Cierpicki and colleagues employed a fragment-based screening strategy to identify and optimize first-in-class irreversible small-molecule inhibitors of the NSD1 SET domain [77]. Structural analysis revealed that NSD1 in complex with covalently bound ligands results in a conformational change in the autoinhibitory loop of the SET domain and formation of a channel-like pocket suitable for targeting with small molecules. Importantly, their lead-compound (“BT5”) demonstrated on-target activity in NUP98-NSD1 immortalized cells associated with reduction of H3K36me2 and downregulation of critical target genes, such as the *HOXA* gene cluster and *MEIS1*. Notably, BT5 also impaired the clonogenic growth of primary NUP98-NSD1⁺ AML cells but not leukemic cells carrying an *KMT2A-MLL1* fusion or normal human CD34⁺ hematopoietic stem and progenitor cells. The discovery of this compound provides a platform for the development of potent and selective NSD1-SET inhibitors [77].

7. Outlook

It is well established that NSD1 regulates gene expression programs through H3K36 methylation in a complex crosstalk between activating and repressive histone marks, as well as DNA methylation. In addition, NSD1 is a target of recurrent germline or somatically acquired loss and gain-of-function alterations associated with developmental syndromes (e.g., SOTOS) and various human cancers. However, many open questions remain; in particular, it is currently poorly understood how a putative loss of function mutation or reduced expression results in the observed developmental and cancer phenotype.

Molecular characterization of SOTOS patient-derived DNA confirmed the connection between NSD1 loss-of-function mutations and aberrant DNA CpG methylation [54]. However, it seems unclear whether SOTOS is based on simple NSD1 haploinsufficiency, or whether particular mutants eventually have dominant-negative activity, functionally impairing the protein expressed from the unmutated allele. Some studies have suggested an increased risk for SOTOS patients to develop cancer, raising the question about the role of NSD1 mutations in this context. As the cancer risk in SOTOS patients is small, one wonders whether further reduction of the NSD1 gene dosage (by, e.g., epigenetically silencing of the wildtype allele) could be involved [78]. A better characterization of the gene dosage and protein activity relationship in developmental syndromes is necessary to explore whether the presence of NSD1 mutations can serve more than as a diagnostic marker but eventually also provide some translational opportunities [79].

As outlined before, predicted loss-of-function mutations or epigenetic silencing of NSD1 have been described in a variety of human cancers. Investigating HNSCC or lung squamous cell carcinomas revealed that mutations did not abrogate NSD1 expression in most samples. Notably, interrogation of the cancer cell line encyclopedia (CCLE) indicates that only a very small number (5/1457) of human cancer cell lines completely lost NSD1

expression at the mRNA level [80,81]. Currently, it remains unclear how a single NSD1 point mutation will contribute to malignant transformation. Is further reduction of the NSD1 gene dosage, e.g., by loss of heterozygosity (LOH), necessary to significantly enhance malignant transformation? Notably, heterozygous *Nsd1*^{+/-} mice do not develop any pathologies and express normal *Nsd1* mRNA and protein levels [44]. In addition, we observed that shRNA-mediated knockdown experiments only significantly affected growth of various human and mouse cells after reduction of NSD1 mRNA levels over 50% (unpublished data). Hence, a systematic analysis of the functional NSD1 gene dose in malignant transformation is necessary. It also remains unresolved whether genetic alterations are early or late events in cancer development.

In addition, the critical downstream effectors of the tumor suppressive activity of NSD1 remain unknown. Although recent molecular analysis of human HNSCC cancer cell lines with and without NSD1 mutations (generated by CRISPR/Cas9 genome editing) revealed aberrant regulation of genes related to oxidative phosphorylation, MYC, mTORC1 or RAS signaling and other pathways, the impact on the cell biology has not been addressed and no particular transformation effector genes have been validated [65]. In addition, further studies are necessary to show whether the disease phenotypes with aberrant functional NSD1 dose are purely the consequence of its chromatin regulatory role or whether yet to be identified non-chromatin NSD1 substrate proteins are critically involved [3].

Notably, one of the most significantly down-regulated pathways in HNSCC cells, carrying engineered NSD1 mutations, was interferon alpha/gamma signaling [65]. Earlier studies identified human HNSCC and lung squamous cell carcinoma enriched for NSD1 inactivating mutations and deletions that displayed an immune-cold phenotype characterized by low degree of infiltration by tumor-associated leukocytes (macrophages, CD8⁺ T cells) as well as low expression of immune checkpoint ligands and receptors (PD1, PDL1, PDCD1LG2) [82]. Interestingly, tumors formed by lung cancer cell lines with shRNA-mediated reduced NSD1 expression in immunodeficient mice contained also less tumor-infiltrating T cells and were associated with reduced expression of various cytokines and chemokines [82]. Another study proposed that a chemokine expression signature allows classification of HNSCC into high and low CD8⁺ T cell-infiltrated tumor phenotypes (TCIP-H vs. TCIP-L) associated with different clinical outcome. Notably about 20% of TCIP-L tumors carried loss of function NSD1 mutations [83]. These observations suggest that human cancers may escape the immune system through acquisition of NSD1 mutations. Further work is necessary to dissect the cellular and molecular circuits of cell-autonomous from non-autonomous consequences of aberrant NSD1 activity in human diseases. Interestingly, in vitro functional studies performed with human brain and breast cancer cells lines found a potential link of reduced expression or mutations of NSD1 to drug resistance; however, its general significance for cancer therapy remains to be validated [84,85].

When fused to the N-terminus of NUP98, the NSD1-SET gains transforming activities in hematopoietic cells, resulting in myelodysplasia and AML, and the presence of a NUP98-NSD1 (and other NUP98-fusions) is often associated with primary resistance to chemotherapy [86,87]. Functional studies suggested that transformation by these fusions involves the NUP98-GFLG repeats recruiting a large WDR82-SET1A/B-COMPASS protein complex to promote H3K4 trimethylation and favor active transcription [88]. The fusions may also directly interact with KMT2A (aka MLL1) to reach critical target gene loci such as the *HOX-A* gene cluster regulated by the fusion partner like NSD1 that favors transcription by H3K36 methylation [58,89]. These findings strongly suggest that targeted inactivation of the NSD1-SET domain shows anti-leukemic activity in NUP98-NSD1⁺ hematological malignancies.

Although selective NSD1-SET inhibitors are highly relevant for aggressive NUP98-NSD1⁺ pediatric AML, one has to take into consideration that loss-of-function mutations of NSD1 are much more prevalent in human cancers. Will such NSD1-SET inhibitors also block NSD1's role as a tumor suppressor? Significantly reduced NSD1 activity may

functionally affect transcription factors controlling maturation of hematopoietic cells (and eventually also cells from other tissues). In the best-case scenario, some reduction of the NSD1-SET might be sufficient to induce differentiation of NUP98-NSD1-transformed myeloid cells, whereas significant side effects (as observed in gene targeted mice) may only develop upon complete inactivation over a longer time period that will most likely never be reached by such compounds.

The future NSD1 research agenda should aim to (i) mechanistically determine the gene dosage–phenotype correlation in germline syndromes with aberrant NSD1 activity, (ii) identify the cellular and molecular mechanisms of malignant transformation by altered NSD1 activity (mutations, epigenetic silencing), and (iii) optimize and validate small molecule NSD1-SET inhibitors for therapy of pediatric AML, driven by the NUP98-NSD1 fusion gene, and research for strategies to selectively interfere in situations when reduced NSD1 activity is the driving force.

Author Contributions: S.T. and J.S. both conceptualized and wrote the manuscript. Both authors have read and agreed to the published version of the manuscript.

Funding: Research on NSD1 was supported by Cancer Research Switzerland (KFS-4258-08-2017).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors thank Albert Jeltsch and Jonathan Séguin for critical reading the manuscript. All figures were created with BioRender.com (accessed on 25 August 2021).

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

References

- Jenuwein, T.; Allis, C.D. Translating the histone code. *Science* **2001**, *293*, 1074–1080. [[CrossRef](#)]
- Husmann, D.; Gozani, O. Histone lysine methyltransferases in biology and disease. *Nat. Struct. Mol. Biol.* **2019**, *26*, 880–889. [[CrossRef](#)]
- Bhat, K.P.; Kaniskan, H.Ü.; Jin, J.; Gozani, O. Epigenetics and beyond: Targeting writers of protein lysine methylation to treat disease. *Nat. Rev. Drug Discov.* **2021**, *20*, 265–286. [[CrossRef](#)]
- Huang, N.; Baur, E.V.; Garnier, J.; Lerouge, T.; Vonesch, J.; Lutz, Y.; Chambon, P.; Losson, R. Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *EMBO J.* **1998**, *17*, 3398–3412. [[CrossRef](#)] [[PubMed](#)]
- Sampson, E.R.; Yeh, S.Y.; Chang, H.-C.; Tsai, M.Y.; Wang, X.; Ting, H.J.; Chang, C. Identification and characterization of androgen receptor associated coregulators in prostate cancer cells. *J. Biol. Regul. Homeost. Agents* **2001**, *15*, 123–129.
- Kurotaki, N.; Harada, N.; Yoshiura, K.-I.; Sugano, S.; Niikawa, N.; Matsumoto, N. Molecular characterization of NSD1, a human homologue of the mouse *Nsd1* gene. *Gene* **2001**, *279*, 197–204. [[CrossRef](#)]
- Ensembl Gene: NSD1 ENSG00000165671. Available online: http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000165671;r=5:177133025-177300213;t=ENST00000347982 (accessed on 17 June 2021).
- Wang, X.; Yeh, S.; Wu, G.; Hsu, C.-L.; Wang, L.; Chiang, T.; Yang, Y.; Guo, Y.; Chang, C. Identification and Characterization of a Novel Androgen Receptor Coregulator ARA267- α in Prostate Cancer Cells. *J. Biol. Chem.* **2001**, *276*, 40417–40423. [[CrossRef](#)]
- The Human Protein Atlas. Available online: <https://www.proteinatlas.org/ENSG00000165671-NSD1/tissue> (accessed on 13 July 2021).
- GeneCards: The Human Gene Database: NSD1. Available online: https://www.genecards.org/cgi-bin/carddisp.pl?gene=NSD1#protein_expression (accessed on 13 July 2021).
- GeneCards the Human Gene Database. Available online: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=NSD1> (accessed on 17 June 2021).
- Rosati, R.; La Starza, R.; Veronese, A.; Aventin, A.; Schwienbacher, C.; Vallespi, T.; Negrini, M.; Martelli, M.F.; Mecucci, C. NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15). *Blood* **2002**, *99*, 3857–3860. [[CrossRef](#)]
- Murn, J.; Shi, Y. The winding path of protein methylation research: Milestones and new frontiers. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 517–527. [[CrossRef](#)]
- Morishita, M.; di Luccio, E. Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1. *Biochem. Biophys. Res. Commun.* **2011**, *412*, 214–219. [[CrossRef](#)]
- Amin, N.; Nietlispach, D.; Qamar, S.; Coyle, J.; Chiarparin, E.; Williams, G. NMR backbone resonance assignment and solution secondary structure determination of human NSD1 and NSD2. *Biomol. NMR Assign.* **2016**, *10*, 315–320. [[CrossRef](#)]

16. Pasillas, M.P.; Shah, M.; Kamps, M.P. NSD1 PHD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human sotos syndrome. *Hum. Mutat.* **2011**, *32*, 292–298. [[CrossRef](#)] [[PubMed](#)]
17. Berdasco, M.; Ropero, S.; Setien, F.; Fraga, M.; Lapunzina, P.; Losson, R.; Alaminos, M.; Cheung, N.-K.; Rahman, N.; Esteller, M. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21830–21835. [[CrossRef](#)]
18. Lu, T.; Jackson, M.W.; Wang, B.; Yang, M.; Chance, M.R.; Miyagi, M.; Gudkov, A.V.; Stark, G.R. Regulation of NF κ B by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Cytokine* **2009**, *48*, 19–20. [[CrossRef](#)]
19. Kudithipudi, S.; Lungu, C.; Rathert, P.; Happel, N.; Jeltsch, A. Substrate Specificity Analysis and Novel Substrates of the Protein Lysine Methyltransferase NSD1. *Chem. Biol.* **2014**, *21*, 226–237. [[CrossRef](#)]
20. Li, J.; Ahn, J.H.; Wang, G.G. Understanding histone H3 lysine 36 methylation and its deregulation in disease. *Cell. Mol. Life Sci.* **2019**, *76*, 2899–2916. [[CrossRef](#)] [[PubMed](#)]
21. Huang, C.; Zhu, B. Roles of H3K36-specific histone methyltransferases in transcription: Antagonizing silencing and safeguarding transcription fidelity. *Biophys. Rep.* **2018**, *4*, 170–177. [[CrossRef](#)]
22. Edmunds, J.W.; Mahadevan, L.C.; Clayton, A.L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **2007**, *27*, 406–420. [[CrossRef](#)]
23. Qiao, Q.; Li, Y.; Chen, Z.; Wang, M.; Reinberg, D.; Xu, R.-M. The Structure of NSD1 Reveals an Autoregulatory Mechanism Underlying Histone H3K36 Methylation. *J. Biol. Chem.* **2011**, *286*, 8361–8368. [[CrossRef](#)]
24. Li, W.; Tian, W.; Yuan, G.; Deng, P.; Sengupta, D.; Cheng, Z.; Cao, Y.; Ren, J.; Qin, Y.; Zhou, Y.; et al. Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases. *Nat. Cell Biol.* **2021**, *590*, 498–503.
25. Yuan, G.; Ma, B.; Yuan, W.; Zhang, Z.; Chen, P.; Ding, X.; Feng, L.; Shen, X.; Chen, S.; Li, G.; et al. Histone H2A Ubiquitination Inhibits the Enzymatic Activity of H3 Lysine 36 Methyltransferases. *J. Biol. Chem.* **2013**, *288*, 30832–30842. [[CrossRef](#)] [[PubMed](#)]
26. Laugesen, A.; Højfeldt, J.W.; Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell* **2019**, *74*, 8–18. [[CrossRef](#)]
27. Streubel, G.; Watson, A.; Jammula, S.G.; Scelfo, A.; Fitzpatrick, D.J.; Oliviero, G.; McCole, R.; Conway, E.; Glancy, E.; Negri, G.L.; et al. The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol. Cell* **2018**, *70*, 371–379.e5. [[CrossRef](#)]
28. Hoang, N.-M.; Rui, L. DNA methyltransferases in hematological malignancies. *J. Genet. Genom.* **2020**, *47*, 361–372. [[CrossRef](#)] [[PubMed](#)]
29. Dhayalan, A.; Rajavelu, A.; Rathert, P.; Tamas, R.; Jurkowska, R.Z.; Ragozin, S.; Jeltsch, A. The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. *J. Biol. Chem.* **2010**, *285*, 26114–26120. [[CrossRef](#)]
30. Deevy, O.; Bracken, A.P. PRC2 functions in development and congenital disorders. *Development* **2019**, *146*, dev181354. [[CrossRef](#)]
31. Li, Y.; Chen, X.; Lu, C. The interplay between DNA and histone methylation: Molecular mechanisms and disease implications. *EMBO Rep.* **2021**, *22*, e51803. [[CrossRef](#)] [[PubMed](#)]
32. Weinberg, D.; Papillon-Cavanagh, S.; Chen, H.; Yue, Y.; Chen, X.; Rajagopalan, K.N.; Horth, C.; McGuire, J.T.; Xu, X.; Nikbakht, H.; et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nat. Cell Biol.* **2019**, *573*, 281–286. [[CrossRef](#)]
33. Nielsen, A.L.; Jørgensen, P.; Lerouge, T.; Cerviño, M.; Chambon, P.; Losson, R. 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*, 5184–5196. [[CrossRef](#)]
34. Berardi, A.; Ghitti, M.; Quilici, G.; Musco, G. In silico derived small molecules targeting the finger-finger interaction between the histone lysine methyltransferase NSD1 and Nizp1 repressor. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 4082–4092. [[CrossRef](#)]
35. Fang, Y.; Tang, Y.; Zhang, Y.; Pan, Y.; Jia, J.; Sun, Z.; Zeng, W.; Chen, J.; Yuan, Y.; Fang, D. The H3K36me2 methyltransferase NSD1 modulates H3K27ac at active enhancers to safeguard gene expression. *Nucleic Acids Res.* **2021**, *49*, 6281–6295. [[CrossRef](#)] [[PubMed](#)]
36. Yamada-Okabe, T.; Matsumoto, N. Decreased serum dependence in the growth of NIH3T3 cells from the overexpression of human nuclear receptor-binding SET-domain-containing protein 1 (NSD1) or fission yeast su(var)3-9, enhancer-of-zeste, trithorax 2 (SET2). *Cell Biochem. Funct.* **2007**, *26*, 146–150. [[CrossRef](#)] [[PubMed](#)]
37. Jeong, Y.; Kim, T.; Kim, S.; Hong, Y.-K.; Cho, K.S.; Lee, I.-S. Overexpression of histone methyltransferase NSD in Drosophila induces apoptotic cell death via the Jun-N-terminal kinase pathway. *Biochem. Biophys. Res. Commun.* **2018**, *496*, 1134–1140. [[CrossRef](#)] [[PubMed](#)]
38. Kim, S.; Kim, T.; Jeong, Y.; Choi, S.; Yamaguchi, M.; Lee, I.-S. The Drosophila histone methyltransferase NSD is positively regulated by the DRE/DREF system. *Genes Genom.* **2018**, *40*, 475–484. [[CrossRef](#)] [[PubMed](#)]
39. Kim, T.; Shin, H.; Song, B.; Won, C.; Yoshida, H.; Yamaguchi, M.; Cho, K.S.; Lee, I. Overexpression of H3K36 methyltransferase NSD in glial cells affects brain development in Drosophila. *Glia* **2020**, *68*, 2503–2516. [[CrossRef](#)]
40. Choi, S.; Song, B.; Shin, H.; Won, C.; Kim, T.; Yoshida, H.; Lee, D.; Chung, J.; Cho, K.S.; Lee, I.-S. Drosophila NSD deletion induces developmental anomalies similar to those seen in Sotos syndrome 1 patients. *Genes Genom.* **2021**, *43*, 737–748. [[CrossRef](#)]
41. Rayasam, G.V.; Wendling, O.; Angrand, P.-O.; Mark, M.; Niederreither, K.; Song, L.; Lerouge, T.; Hager, G.L.; Chambon, P.; Losson, R. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* **2003**, *22*, 3153–3163. [[CrossRef](#)] [[PubMed](#)]

42. Kaneda, M.; Okano, M.; Hata, K.; Sado, T.; Tsujimoto, N.; Li, E.; Sasaki, H. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nat. Cell Biol.* **2004**, *429*, 900–903. [[CrossRef](#)]
43. Shirane, K.; Miura, F.; Ito, T.; Lorincz, M.C. NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. *Nat. Genet.* **2020**, *52*, 1088–1098. [[CrossRef](#)]
44. Leonards, K.; Almosaileakh, M.; Tauchmann, S.; Bagger, F.O.; Thirant, C.; Juge, S.; Bock, T.; Méreau, H.; Bezerra, M.F.; Tzankov, A.; et al. Nuclear interacting SET domain protein 1 inactivation impairs GATA1-regulated erythroid differentiation and causes erythroleukemia. *Nat. Commun.* **2020**, *11*, 1–15. [[CrossRef](#)]
45. Amarilio, R.; Viukov, S.V.; Sharir, A.; Eshkar-Oren, I.; Johnson, R.; Zelzer, E. HIF1 α regulation of Sox9 is necessary to maintain differentiation of hypoxic prechondrogenic cells during early skeletogenesis. *Development* **2007**, *134*, 3917–3928. [[CrossRef](#)]
46. Shao, R.; Zhang, Z.; Xu, Z.; Ouyang, H.; Wang, L.; Ouyang, H.; Zou, W. H3K36 methyltransferase NSD1 regulates chondrocyte differentiation for skeletal development and fracture repair. *Bone Res.* **2021**, *9*, 1–11.
47. Oishi, S.; Zalucki, O.; Vega, M.S.; Harkins, D.; Harvey, T.J.; Kasherman, M.; Davila, R.A.; Hale, L.; White, M.; Piltz, S.; et al. Investigating cortical features of Sotos syndrome using mice heterozygous for Nsd1. *Genes Brain Behav.* **2020**, *19*, e12637. [[CrossRef](#)] [[PubMed](#)]
48. Kurotaki, N.; Imaizumi, K.; Harada, N.; Masuno, M.; Kondoh, T.; Nagai, T.; Ohashi, H.; Naritomi, K.; Tsukahara, M.; Makita, Y.; et al. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* **2002**, *30*, 365–366. [[CrossRef](#)] [[PubMed](#)]
49. Tatton-Brown, K.; Rahman, N. Sotos syndrome. *Eur. J. Hum. Genet.* **2006**, *15*, 264–271. [[CrossRef](#)] [[PubMed](#)]
50. Dikow, N.; Maas, B.; Gaspar, H.; Kreiss-Nachtsheim, M.; Engels, H.; Kuechler, A.; Garbes, L.; Netzer, C.; Neuhann, T.M.; Koehler, U.; et al. The phenotypic spectrum of duplication 5q35.2-q35.3 encompassing NSD1: Is it really a reversed sotos syndrome? *Am. J. Med Genet. Part A* **2013**, *161*, 2158–2166. [[CrossRef](#)]
51. Tatton-Brown, K.; Loveday, C.; Yost, S.; Clarke, M.; Ramsay, E.; Zachariou, A.; Elliott, A.; Wylie, H.; Ardisson, A.; Rittinger, O.; et al. Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability. *Am. J. Hum. Genet.* **2017**, *100*, 725–736. [[CrossRef](#)]
52. Madsen, R.R.; Knox, R.G.; Pearce, W.; Lopez, S.; Mahler-Araujo, B.; McGranahan, N.; Semple, R.K. Oncogenic PIK3CA promotes cellular stemness in an allele dose-dependent manner. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 8380–8389. [[CrossRef](#)]
53. Madsen, R.R. PI3K in stemness regulation: From development to cancer. *Biochem. Soc. Trans.* **2020**, *48*, 301–315. [[CrossRef](#)]
54. Choufani, S.; Cytrynbaum, C.; Chung, B.H.Y.; Turinsky, A.L.; Grafodatskaya, D.; Chen, Y.A.; Cohen, A.S.A.; Dupuis, L.; Butcher, D.T.; Siu, M.T.; et al. NSD1 mutations generate a genome-wide DNA methylation signature. *Nat. Commun.* **2015**, *6*, 10207. [[CrossRef](#)]
55. Martin-Herranz, D.E.; Aref-Eshghi, E.; Bonder, M.J.; Stubbs, T.M.; Choufani, S.; Weksberg, R.; Stegle, O.; Sadikovic, B.; Reik, W.; Thornton, J.M. Screening for genes that accelerate the epigenetic aging clock in humans reveals a role for the H3K36 methyltransferase NSD1. *Genome Biol.* **2019**, *20*, 1–19. [[CrossRef](#)] [[PubMed](#)]
56. Jaju, R.J. 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*, 1264–1267. [[CrossRef](#)]
57. Hollink, I.H.I.M.; Heuvel-Eibrink, M.M.V.D.; Arentsen-Peters, S.T.C.J.M.; Pratcorona, M.; Abbas, S.; Kuipers, J.E.; Van Galen, J.F.; Beverloo, H.B.; Sonneveld, E.; Kaspers, G.-J.J.L.; et al. NUP98/NSD1 characterizes a novel poor prognostic group in acute myeloid leukemia with a distinct HOX gene expression pattern. *Blood* **2011**, *118*, 3645–3656. [[CrossRef](#)]
58. Wang, G.G.; Cai, L.; Pasillas, M.P.; Kamps, M.P. NUP98–NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat. Cell Biol.* **2007**, *9*, 804–812. [[CrossRef](#)] [[PubMed](#)]
59. Thanasopoulou, A.; Tzankov, A.; Schwaller, J. Potent co-operation between the NUP98-NSD1 fusion and the FLT3-ITD mutation in acute myeloid leukemia induction. *Haematologica* **2014**, *99*, 1465–1471. [[CrossRef](#)]
60. Dolnik, A.; Engelmann, J.C.; Scharfenberger-Schmeer, M.; Mauch, J.; Kelkenberg-Schade, S.; Haldemann, B.; Fries, T.; Krönke, J.; Kühn, M.W.M.; Paschka, P.; et al. Commonly altered genomic regions in acute myeloid leukemia are enriched for somatic mutations involved in chromatin remodeling and splicing. *Blood* **2012**, *120*, e83–e92. [[CrossRef](#)] [[PubMed](#)]
61. Su, X.; Zhang, J.; Mouawad, R.; Compérat, E.; Rouprêt, M.; Allanic, F.; Malouf, G.G. NSD1 Inactivation and SETD2 Mutation Drive a Convergence toward Loss of Function of H3K36 Writers in Clear Cell Renal Cell Carcinomas. *Cancer Res.* **2017**, *77*, 4835–4845. [[CrossRef](#)]
62. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **2015**, *517*, 576–582. [[CrossRef](#)]
63. Mohammad, F.; Helin, K. Oncohistones: Drivers of pediatric cancers. *Genes Dev.* **2017**, *31*, 2313–2324. [[CrossRef](#)]
64. Schuhmacher, M.; Kusevic, D.; Kudithipudi, S.; Jeltsch, A. Kinetic Analysis of the Inhibition of the NSD1, NSD2 and SETD2 Protein Lysine Methyltransferases by a K36M Oncohistone Peptide. *ChemistrySelect* **2017**, *2*, 9532–9536. [[CrossRef](#)]
65. Farhangdoost, N.; Horth, C.; Hu, B.; Bareke, E.; Chen, X.; Li, Y.; Coradin, M.; Garcia, B.A.; Lu, C.; Majewski, J. Chromatin dysregulation associated with NSD1 mutation in head and neck squamous cell carcinoma. *Cell Rep.* **2021**, *34*, 108769. [[CrossRef](#)]
66. Rajagopalan, K.N.; Chen, X.; Weinberg, D.N.; Chen, H.; Majewski, J.; Allis, C.D.; Lu, C. Depletion of H3K36me2 recapitulates epigenomic and phenotypic changes induced by the H3.3K36M oncohistone mutation. *Proc. Natl. Acad. Sci. USA* **2021**, *118*. [[CrossRef](#)]

67. Brumbaugh, J.; Kim, I.S.; Ji, F.; Huebner, A.J.; Di Stefano, B.; Schwarz, B.A.; Charlton, J.; Coffey, A.; Choi, J.; Walsh, R.M.; et al. Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. *Nat. Cell Biol.* **2019**, *21*, 1449–1461. [[CrossRef](#)]
68. Bianco-Miotto, T.; Chiam, K.; Buchanan, G.; Jindal, S.; Day, T.K.; Thomas, M.; Pickering, M.A.; O'Loughlin, M.A.; Ryan, N.K.; Raymond, W.A.; et al. Global Levels of Specific Histone Modifications and an Epigenetic Gene Signature Predict Prostate Cancer Progression and Development. *Cancer Epidemiol. Biomark. Prev.* **2010**, *19*, 2611–2622. [[CrossRef](#)] [[PubMed](#)]
69. Bakardjieva-Mihaylova, V.; Kramarzova, K.S.; Slamova, M.; Svaton, M.; Rejlova, K.; Zaliova, M.; Dobiasova, A.; Fiser, K.; Stuchly, J.; Grega, M.; et al. Molecular Basis of Cisplatin Resistance in Testicular Germ Cell Tumors. *Cancers* **2019**, *11*, 1316. [[CrossRef](#)]
70. Lee, S.-T.; Wiemels, J.L. Genome-wide CpG island methylation and intergenic demethylation propensities vary among different tumor sites. *Nucleic Acids Res.* **2015**, *44*, 1105–1117. [[CrossRef](#)] [[PubMed](#)]
71. Park, S.; Supek, F.; Lehner, B. Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits. *Nat. Commun.* **2018**, *9*, 1–13. [[CrossRef](#)]
72. Spurr, L.; Li, M.; Alomran, N.; Zhang, Q.; Restrepo, P.; Movassagh, M.; Trenkov, C.; Tunnessen, N.; Apanasovich, T.; Crandall, K.A.; et al. Systematic pan-cancer analysis of somatic allele frequency. *Sci. Rep.* **2018**, *8*, 7735. [[CrossRef](#)]
73. Chang, Y.; Zhang, X.; Horton, J.; Upadhyay, A.K.; Spannhoff, A.; Liu, J.; Snyder, J.P.; Bedford, M.T.; Cheng, X. Structural basis for G9a-like protein lysine methyltransferase inhibition by BIX-01294. *Nat. Struct. Mol. Biol.* **2009**, *16*, 312–317. [[CrossRef](#)] [[PubMed](#)]
74. Morishita, M.; Mevius, D.E.H.F.; Shen, Y.; Zhao, S.; di Luccio, E. BIX-01294 inhibits oncoproteins NSD1, NSD2 and NSD3. *Med. Chem. Res.* **2017**, *26*, 2038–2047. [[CrossRef](#)]
75. Drake, K.M.; Watson, V.G.; Kisielewski, A.; Glynn, R.; Napper, A.D. A Sensitive Luminescent Assay for the Histone Methyltransferase NSD1 and Other SAM-Dependent Enzymes. *ASSAY Drug Dev. Technol.* **2014**, *12*, 258–271. [[CrossRef](#)] [[PubMed](#)]
76. Rabal, O.; Castellar, A.; Oyarzabal, J. Novel pharmacological maps of protein lysine methyltransferases: Key for target deorphanization. *J. Chem.* **2018**, *10*, 1–19. [[CrossRef](#)] [[PubMed](#)]
77. Huang, H.; Howard, C.A.; Zari, S.; Cho, H.J.; Shukla, S.; Li, H.; Ndoj, J.; González-Alonso, P.; Nikolaidis, C.; Abbott, J.; et al. Covalent inhibition of NSD1 histone methyltransferase. *Nat. Chem. Biol.* **2020**, *16*, 1403–1410. [[CrossRef](#)]
78. Rahman, N. Mechanisms predisposing to childhood overgrowth and cancer. *Curr. Opin. Genet. Dev.* **2005**, *15*, 227–233. [[CrossRef](#)]
79. Cytrynbaum, C.; Choufani, S.; Weksberg, R. Epigenetic signatures in overgrowth syndromes: Translational opportunities. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **2019**, *181*, 491–501. [[CrossRef](#)]
80. Ghandi, M.; Huang, F.W.; Jané-Valbuena, J.; Kryukov, G.; Lo, C.C.; McDonald, E.R., III; Barretina, J.; Gelfand, E.T.; Bielski, C.M.; Li, H.; et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **2019**, *569*, 503–508. [[CrossRef](#)]
81. Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A.A.; Kim, S.; Garraway, L.A. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **2012**, *483*, 603–607. [[CrossRef](#)]
82. Brennan, K.; Shin, J.H.; Tay, J.K.; Prunello, M.; Gentles, A.J.; Sunwoo, J.B.; Gevaert, O. NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. *Sci. Rep.* **2017**, *7*, 1–12. [[CrossRef](#)]
83. Saloura, V.; Izumchenko, E.; Zuo, Z.; Bao, R.; Korzinkin, M.; Ozerov, I.; Zhavoronkov, A.; Sidransky, D.; Bedi, A.; Hoque, M.O.; et al. Immune profiles in primary squamous cell carcinoma of the head and neck. *Oral Oncol.* **2019**, *96*, 77–88. [[CrossRef](#)]
84. Han, J.; Jun, Y.; Kim, S.H.; Hoang, H.-H.; Jung, Y.; Kim, S.; Kim, J.; Austin, R.H.; Lee, S.; Park, S. Rapid emergence and mechanisms of resistance by U87 glioblastoma cells to doxorubicin in an in vitro tumor microfluidic ecology. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 14283–14288. [[CrossRef](#)] [[PubMed](#)]
85. Pereira, A.M.M.; Sims, D.; Dexter, T.; Fenwick, K.; Assiotis, I.; Kozarewa, I.; Mitsopoulos, C.; Hakas, J.; Zvelebil, M.; Lord, C.; et al. Genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2730–2735. [[CrossRef](#)]
86. McNeer, N.; Philip, J.; Geiger, H.; Ries, R.E.; Lavalley, V.-P.; Walsh, M.; Shah, M.; Arora, K.; Emde, A.-K.; Robine, N.; et al. Abstract 2870: Genetic mechanisms of primary chemotherapy resistance in pediatric acute myeloid leukemia. *Tumor Biol.* **2019**, *33*, 1934–1943.
87. Barresi, V.; Di Bella, V.; Andriano, N.; Privitera, A.; Bonaccorso, P.; La Rosa, M.; Iachelli, V.; Spampinato, G.; Pulvirenti, G.; Scuderi, C.; et al. NUP-98 Rearrangements Led to the Identification of Candidate Biomarkers for Primary Induction Failure in Pediatric Acute Myeloid Leukemia. *Int. J. Mol. Sci.* **2021**, *22*, 4575. [[CrossRef](#)] [[PubMed](#)]
88. Franks, T.M.; McCloskey, A.; Shokhirev, M.N.; Benner, C.; Rathore, A.; Hetzer, M.W. Nup98 recruits the Wdr82-Set1A/COMPASS complex to promoters to regulate H3K4 trimethylation in hematopoietic progenitor cells. *Genes Dev.* **2017**, *31*, 2222–2234. [[CrossRef](#)] [[PubMed](#)]
89. Xu, H.; Valerio, D.G.; Eisold, M.E.; Sinha, A.; Koche, R.; Hu, W.; Chen, C.-W.; Chu, S.H.; Brien, G.; Park, C.Y.; et al. NUP98 Fusion Proteins Interact with the NSL and MLL1 Complexes to Drive Leukemogenesis. *Cancer Cell* **2016**, *30*, 863–878. [[CrossRef](#)]