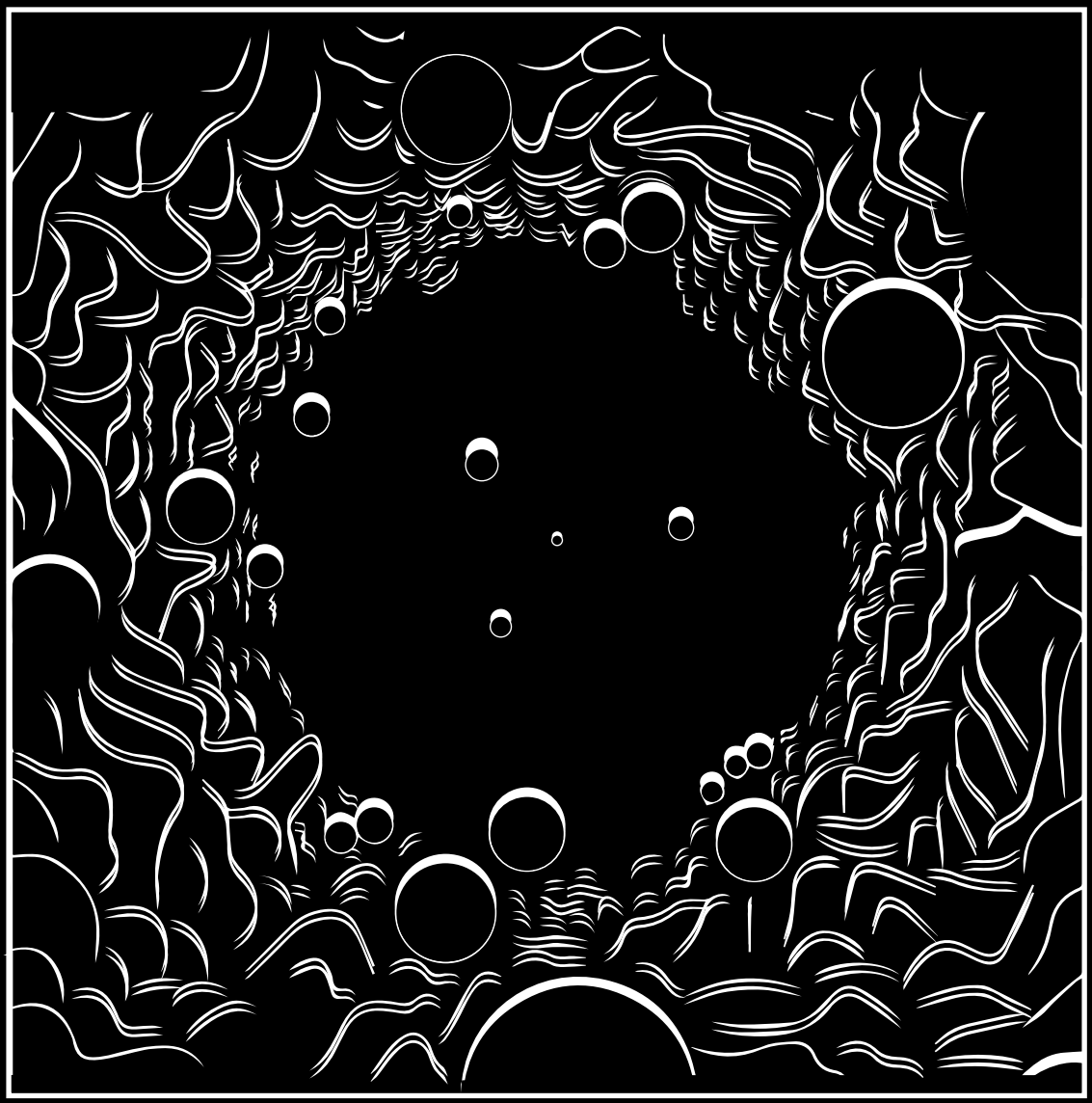
GATA2 DEPENDENT MECHANISMS IN HEMATOPOIETIC STEM CELL BIOLOGY

A voyage from ontogeny to lineage commitment and function



Cansu Koyunlar

GATA2-DEPENDENT MECHANISMS IN HEMATOPOIETIC STEM CELL BIOLOGY:

A VOYAGE FROM ONTOGENY TO LINEAGE COMMITMENT AND FUNCTION

CANSU KOYUNLAR

Gata2-Dependent Mechanisms in Hematopoietic Stem Cell Biology: A Voyage from Ontogeny to Lineage Commitment and Function

Gata2-afhankelijke mechanismen in de hematopoetische stamcel biologie: Een reis van ontogenie naar cel differentiatie en functionaliteit

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Tuesday 6 September 2022 at 15:30 hrs
by

Cansu Koyunlar
born in Karsiyaka, Turkey

ISBN:
Layout: Egied Simons
Cover: Cansu Koyunlar and Thomas Galvan
Printing: Optima Grafische Communicatie

The work described in this thesis was performed at the Department of Hematology at the Erasmus Medical Center, Rotterdam, the Netherlands.

Printing of this thesis was financially supported by the Erasmus University Rotterdam.

Erasmus University Rotterdam

DOCTORAL COMMITTEE

Promotor: Prof.dr. I.P. Touw

Other members: Prof.dr. H.R. Delwel
Prof.dr. J.N.J. Philipsen
Prof.dr. N.A. Speck

Co-promoter: dr. E.M. de Pater

CONTENTS

Chapter 1:	General introduction	7
Chapter 2:	*Gata2*-regulated *Gfi1b* expression controls endothelial programming during endothelial-to-hematopoietic transition	31
Chapter 3:	Essential role for Gata2 in modulating lineage output from hematopoietic stem cells in zebrafish	59
Chapter 4:	Gata2b haploinsufficiency causes aberrant transcriptional signatures in HSPCs resulting in myeloid and erythroid dysplasia in zebrafish	101
Chapter 5:	Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis	127
Chapter 6:	Gata2 haploinsufficiency promotes genome instability in aged hematopoietic stem and progenitor cells upon transplantation resulting in B-cell cytopenia and monocytopenia	171
Chapter 7:	From basic biology to patient mutational spectra of GATA2 haploinsufficiencies: What are the mechanisms, hurdles and prospects of genome editing for treatment	207
Chapter 8:	General discussion	249
Addendum:	English Summary	251
	Dutch Summary	255
	List of abbreviations	259
	Curriculum Vitae	263
	List of publications	265
	PhD portfolio	267
	Word of thanks	269

1 General Introduction

1. FUNCTION AND REGULATION OF HEMATOPOIETIC STEM CELLS

Hematopoietic stem cells (HSCs) are self-renewable and multipotent precursors of the blood system. These two functional properties enable HSCs to persist while ensuring the production of all mature blood cell types throughout the organism's lifespan (*Orkin and Zon, 2008*). HSCs occupy specific stem cell niches and almost exclusively reside in the bone marrow (BM) throughout adulthood (*Mikkola and Orkin, 2006*). Upon their transplantation into irradiated recipient mice, where the hematopoietic system is ablated, HSCs can repopulate the BM and regenerate the entire blood system through their multi-lineage reconstitution and self-renewal abilities (*Zhao et al., 2000; Osawa et al., 1996*).

Adult HSCs remain quiescent under homeostatic conditions for long-term preservation by minimizing their replicative and metabolic activities (*Bakker and Passegué, 2013; Cheshier et al., 1999; Bradford et al., 1997*). HSCs can leave the quiescent state and rapidly enter the cell cycle in response to various stimuli such as blood loss, inflammation or the presence of immune insults (*Wilson et al., 2008; Morrison et al., 1997; Fleming et al., 1993*). Proliferating HSCs can undergo symmetric or asymmetric division; through symmetric division they can copy themselves (symmetric self-renewal) or produce two identical differentiated blood cells (symmetric differentiation) and through asymmetric division they can produce a differentiated cell while maintaining the HSC pool size (symmetric self-renewal) or produce two unidentical differentiated cells (asymmetric differentiation) (Figure 1) (*Murke et al., 2015*). Fine-tuning the quiescence and proliferation mechanisms is crucial for HSCs to prevent exhaustion and to sustain life-long efficient hematopoiesis (*Wilson et al., 2009*).

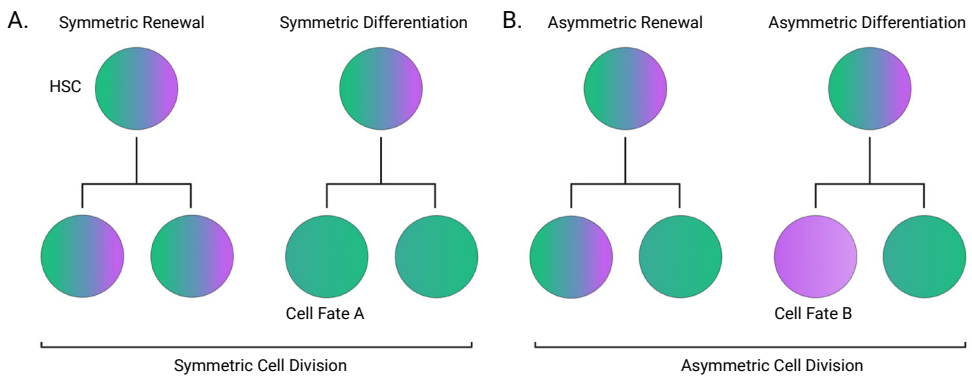


Figure 1. HSCs can undergo symmetric or asymmetric division. Adapted from Murke et al. (*Murke et al., 2015*). A) HSCs that undergo a symmetric division can produce two HSCs (symmetric self-renewal) or two progenitor cells committed to the same lineage (symmetric differentiation). B) HSCs that undergo an asymmetric division can differentiate into a lineage-committed progenitor cell while copying itself (asymmetric self-renewal) or produce two progenitor cells that are committed to different lineages (asymmetric differentiation).

Both quiescence and the cell-fate commitment in HSCs are dynamically regulated by the interactions of extrinsic and intrinsic cues. The main extrinsic factors that influence adult HSC behavior are the signals, such as cytokines and growth factors, secreted by BM microenvironment consisting of mature blood cells and non-hematopoietic niche compartments (*Wilson and Trumpp, 2006*). In cooperation with their environment, HSCs regulate their activity in a cell-autonomous manner mainly driven by the activation of transcription factor (TF) networks (*Wilkinson and Göttgens, 2013*). TFs can regulate cell-fate specific gene expression programs through their binding to the DNA and hence are the key intrinsic determinants of the HSC function.

1.1. Ontogeny and identification of hematopoietic stem cells

The first HSCs are generated in the ventral wall of the embryonic dorsal aorta, the aorta-gonad-mesonephros (AGM) region, from a specialized endothelial cell compartment (hemogenic endothelium) through endothelial-to-hematopoietic transition (EHT) events. Hemogenic endothelial cells (HECs) express hematopoietic TFs, *Runx1* and *Gata2*, which are essential for EHT in addition to pan-endothelial genes like *CD31*. However, HECs lack the expression of *bona fide* hematopoietic genes such as *cKit* and *CD41* (*Bertrand et al., 2010; Boisset et al., 2010; Kissa and Herbomel, 2010; Zovein et al., 2008; de Bruijn et al., 2002; North et al., 2002; Medvinsky and Dzierzak, 1996; Muller et al., 1994; Dieterlen-Lievre, 1975*). During EHT, transitioning HECs that acquire the hematopoietic fate break the tight-junctions to neighboring endothelial cells and gain a rounded shape so as to bulge out from the endothelial layer (*Ottersbach, 2019*). These transdifferentiation events are tightly regulated by the interplay of TFs and signaling pathways and are highly conserved across vertebrate species (*Ciau-Uitz and Patient, 2019; Ivanovs et al., 2017*).

1.1.1. Ontogeny and identification of HSCs throughout mouse development

In mouse AGM, EHT events occur between the embryonic days (E)8-E12.5 and are characterized by the formation of intra-aortic hematopoietic clusters (IAHCs). HEC-derived IAHCs contain a heterogeneous group of hematopoietic stem and progenitor cells (HSPCs) co-expressing endothelial genes such as *CD31* and hematopoietic genes like *cKit* (*Yokomizo and Dzierzak, 2010*). Around E10.5, the first HSCs are formed in IAHCs through a multistep maturation process accompanied by the sequential upregulation of cell surface markers *CD41*, *CD43* and *CD45* and the maturation is characterized as pro-HSC (*CD31*$^{+}$*cKit*$^{+}$*CD41*^{low}*CD43*$^{-}$*CD45*$^{-}$) → pre-HSC type I (pre-HSC1 or *CD31*$^{+}$*cKit*$^{+}$*CD41*^{low}*CD43*$^{+}$*CD45*$^{-}$) → pre-HSC type II (pre-HSC2) and HSCs (*CD31*$^{+}$*cKit*$^{+}$*CD41*^{low}*CD43*$^{+}$*CD45*$^{+}$) (Figure 2) *(Rybtsov et al., 2014; 2011; Taoudi et al., 2008; Sánchez et al., 1996*). Unlike the mature HSCs, the immature pro-HSCs and pre-HSCs cannot yet repopulate irradiated adult recipients. However, following an *ex vivo* cell culture step with stromal cell lines, they complete maturation and can establish long-term multilineage reconstitution upon their transplantation to adult recipients (*Rybtsov et al., 2014*).

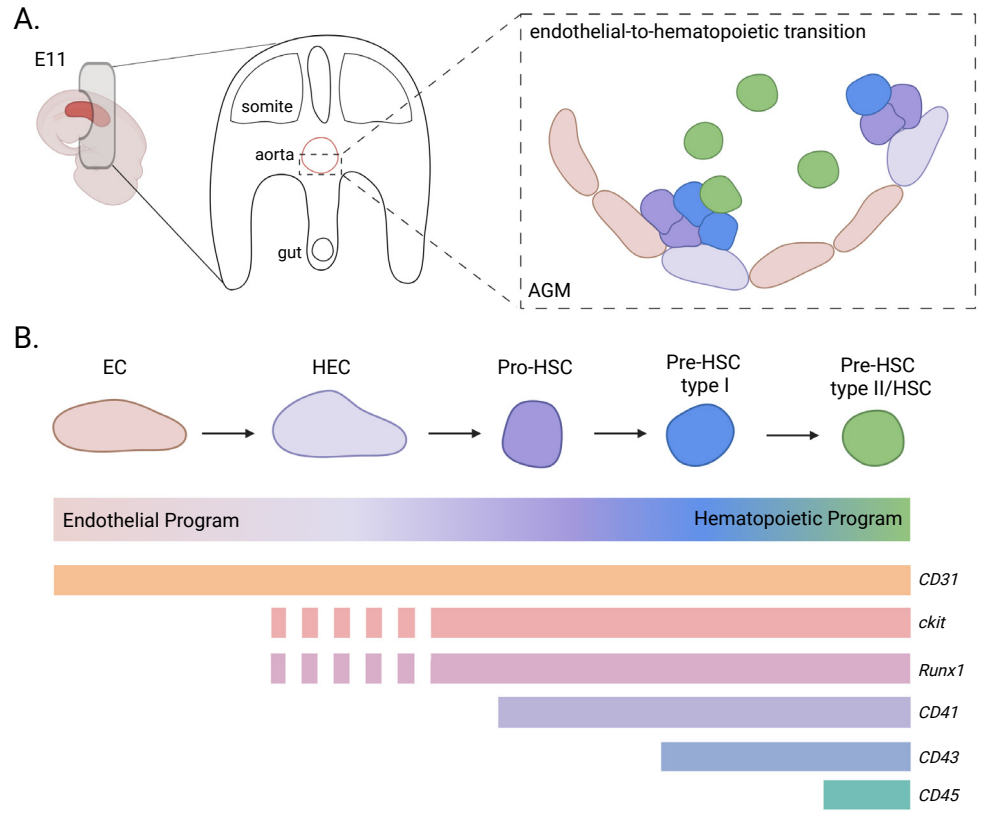


Figure 2. Illustration of the endothelial-to-hematopoietic transition events in E11 mouse embryo. A) Endothelial-to-hematopoietic transition (EHT) events occur in the aorta-gonad-mesonephros (AGM) region by the formation of intra-aortic hematopoietic clusters (IAHCs) toward the lumen of the aorta. B) EHT is characterized as endothelial cell (EC) → hemogenic endothelial cell (HEC) → Pro-HSC → Pre-HSC type I → Pre-HSC type II/HSC and is accompanied by the progressive downregulation of endothelial transcriptional programming and upregulation of hematopoietic transcriptional programming.

Throughout EHT, HSPCs undergo transcriptional changes associated with the progressive downregulation of endothelial transcriptional programming and upregulation of hematopoietic transcriptional programming (*Oatley et al., 2020; Baron et al., 2018; Zhou et al., 2016; Swiers et al., 2013*). The endothelial-to-hematopoietic transcriptional switch during EHT is regulated by the activity of TF networks; whilst early hematopoietic commitment is established by the activation of *Runx1* and the downstream activity of *Gata2* in HECs and IAHCs, activation of *Gfi1* and *Gfi1b* is crucial for the repression of endothelial identity throughout EHT. During the early phases of EHT, HECs undergoing EHT are marked by the co-expression of *Runx1* and *Gfi1*. At later stages of EHT, *Gfi1* expression is gradually replaced by *Gfi1b,* mainly in pre-HSCs, indicative of the exclusive yet complementary roles of these TFs on the repression of endothelial identity during EHT (*Yzaguirre et al., 2018;*

Kang et al., 2018; Thambyrajah et al., 2016a; 2016b; Liakhovitskaia et al., 2014). The role of *Gata2* in this interplay of TFs regulating endothelial-to-hematopoietic transcriptional switch, moreover, how the expression of *Gfi1b* take off in Pre-HSCs remain unknown and these mechanisms are explored in **chapter 2**.

Notch signaling is essential for the specification of HECs and therefore is necessary for EHT. The earliest hematopoietic TFs detected in HECs, *Runx1* and *Gata2*, are downstream targets of the Notch signaling pathway (*Gama-Norton et al., 2016; Robert-Monero et al., 2005*). Furthermore, *Notch1*-deficient embryos fail to produce long-term definitive HSCs (*Hadland et al., 2014*). Conversely, downregulation of Notch activity is required during HSC maturation (*Porcheri et al., 2020; Souilhol et al., 2016*), indicating fine-tuning the Notch signaling is crucial for the generation of HSCs.

Despite the abundance of IAHC cells in the AGM (~700 IAHC cells/AGM at E11) (*Ganuza et al., 2017*), only a small proportion of these cells matures into HSCs during EHT (~1-3 HSCs/AGM at E11) (*Kumaravelu et al., 2002*). Following their maturation, HSCs detach from IAHCs and via the bloodstream and populate the fetal liver (FL) from E11. Here, HSCs rapidly expand in numbers and transform into adult HSCs (lineage$^-$*Sca1*$^+$*cKit*$^+$*CD48*$^-$*CD150*$^+$ or LSK SLAM) before settling in the adult niche BM around birth (E18-E21) (Figure 3) (*Zovein et al., 2008; Ema and Nakauchi et al., 2000*) .

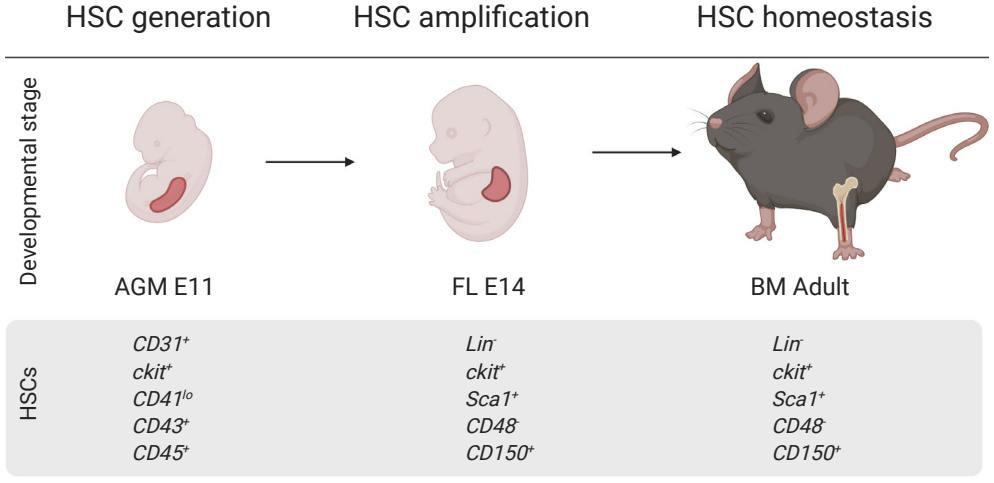


Figure 3. Hematopoietic organs and identification of HSCs throughout mouse development. HSCs occupy distinct stem-cell niches during their generation (aorta-gonad-mesonephros or AGM), amplification (fetal liver or FL) and homeostasis (bone marrow or BM) and are characterized by the corresponding cell surface markers throughout the mouse development.

1.1.2. Ontogeny and identification of HSCs throughout zebrafish development

In zebrafish AGM, EHT events occur between 30 hours post fertilization (hpf) and 56 hpf through the egress of HECs from the aortic floor into the subaortic space (*Kissa and Herbomel, 2010*). Instead of developing from aortic clusters as in mice, zebrafish HSCs are

formed as single cells during EHT and are characterized by the co-expression of endothelial genes like *flt1* and hematopoietic genes like *CD41* (Figure 4). Although *CD41* is the *bona fide* marker for HSCs, transplantable HSCs are identified as CD41 intermediate-expressing cells (CD41^{int}) as the high expression of *CD41* (CD41^{hi}) marks the thrombocyte (platelet) population in zebrafish (*de Pater and Trompouki, 2018; Ma et al., 2011; Lin et al., 2005*).

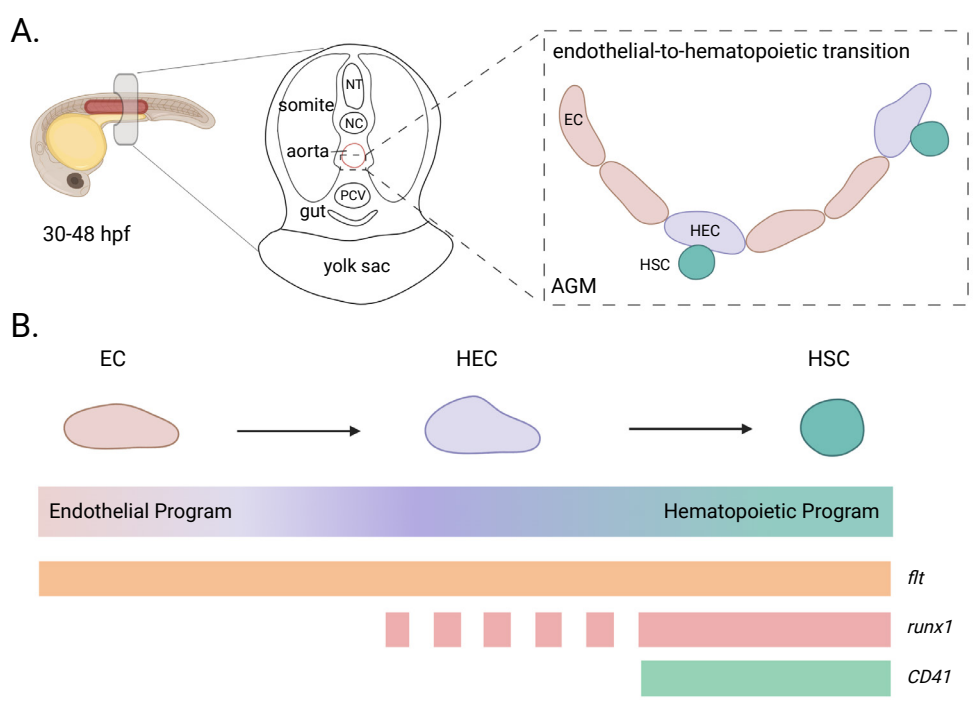


Figure 4. Illustration of endothelial-to-hematopoietic transition events in 30-48 hpf zebrafish. A) Endothelial-to-hematopoietic transition (EHT) events in the aorta-gonad-mesonephros (AGM) region are characterized by the formation of single HSCs toward the subaortic space. B) EHT is characterized as endothelial cell (EC) → hemogenic endothelial cell (HEC) → HSC and is accompanied by the progressive downregulation of endothelial transcriptional programming and upregulation of hematopoietic transcriptional programming. NT, neural tube ; NC, notochord ; PCV, posterior cardinal vein.

Despite the anatomical differences, TFs and signaling pathways regulating EHT are highly conserved between zebrafish and mammals; e.g., the activation of *runx1* in the HECs and the requirement of Notch signaling during EHT (*Bonkhofer et al., 2019; Kim et al., 2014*). Once detached from the aortic floor, HSCs enter the bloodstream through the axial vein and populate to-, and expand in, the caudal hematopoietic tissue (CHT), the equivalent of mammalian FL (*Tamplin et al., 2015*). At around 5 days post fertilization (dpf), CD41^{int} HSCs colonize the adult hematopoietic tissue in the kidney marrow (KM), the analogue of BM in mammals, where they reside throughout adulthood (Figure 5) (*Traver et al., 2003*).

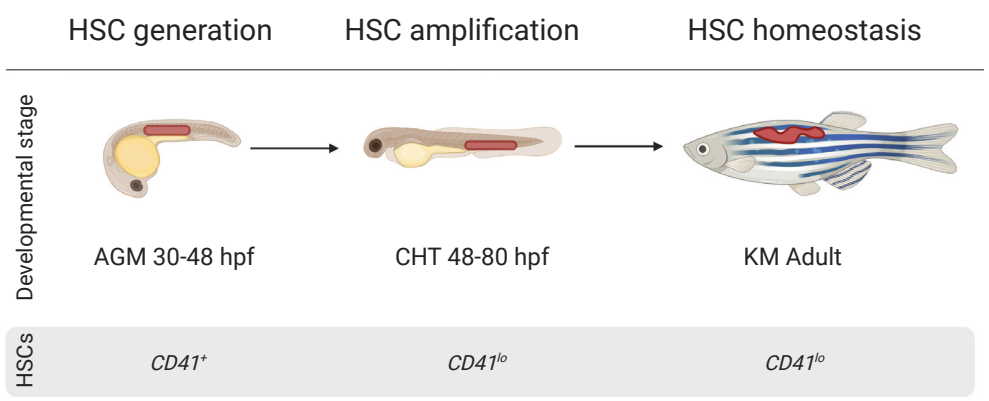


Figure 5. Hematopoietic organs and identification of HSCs throughout zebrafish development. HSCs occupy distinct stem-cell niches during their generation (aorta-gonad-mesonephros or AGM), amplification (caudal hematopoietic tissue or CHT) and homeostasis (kidney marrow or KM) and are characterized by the corresponding cell surface markers throughout the zebrafish development.

1.3. Hematopoiesis

Hematopoiesis refers to the process of blood cell formation. In vertebrates, hematopoiesis sequentially occurs through primitive and definitive waves. The primitive wave involves the formation of erythroid progenitors, which give rise to primitive erythrocytes and macrophages maintaining tissue formation and oxygenation during the early stages of embryonic development. However, the primitive wave is transitory and overtaken by the definitive wave following the generation of HSCs at the later stages of embryonic development (*Jagannathan-Bogdan and Zon, 2013; Orkin and Zon, 2008*).

HSCs are at the apex of definitive hematopoiesis. The differentiation of mature blood cells from HSCs is a continuous process involving multipotent, oligopotent and unipotent intermediate states of hematopoietic progenitor cells (HPCs) that have no or limited self-renewal capacity (Figure 6) (*Bryder et al., 2006; Orkin, 2000*). Unlike HSCs, which remain versatile in their lineage differentiation abilities, HPCs undergo serial fate decisions and are therefore committed to specific differentiation trajectories (*Brown and Ceredig, 2019*). The three main cell-fate trajectories in the blood cell differentiation continuum are the erythroid, myeloid, and lymphoid lineages. Mature blood cell (end-cell) types formed through these trajectories are generally short-lived and highly specialized in their functions; for instance, erythrocytes (half-life about 17 weeks) are most specialized in carrying oxygen to the tissues, while short-lived plasma cells (half-life of 3-5 days) are responsible for the immediate antigen-specific immunoglobulin production against pathogens (Figure 6) (*Mock et al., 2011; Fulcher et al., 1997*).

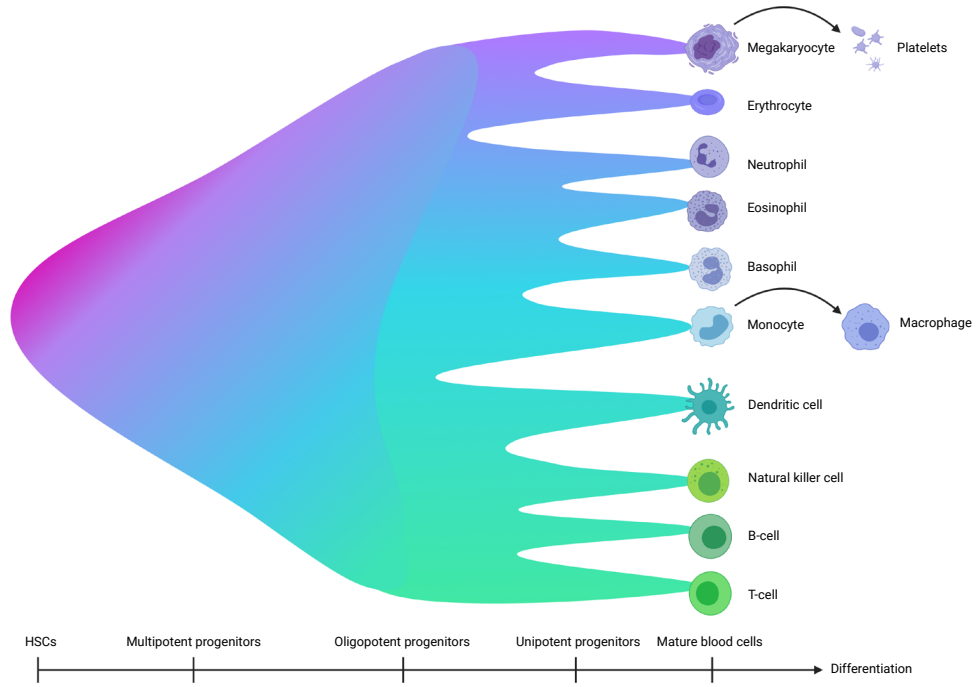


Figure 6. Definitive hematopoiesis. HSCs can differentiate into all mature blood cell types through the production of intermediate multipotent, oligopotent and unipotent lineage-committed hematopoietic progenitors. The lineage differentiation trajectories producing the mature blood cell (end-cell) types are the megakaryocyte lineage (megakaryocytes and platelets), erythroid lineage (erythrocytes), myeloid lineage (neutrophils, eosinophils, basophils, monocytes, macrophages and dendritic cells) and lymphoid lineage (dendritic cells, natural killer cells, B-cells and T-cells).

The cell-fate choice throughout the hematopoietic differentiation is dynamically regulated by the expression level and interaction of TFs. For example, the activity of *Spi1* and *Gata1* respectively contributes to myeloid/lymphoid and myeloid/erythroid lineage specification in HSCs (*Rhodes et al., 2005*). Furthermore, external signals, such as various cytokines and adhesion molecules, are also influencing factors during self-renewal, cell-fate choice, mobilization and survival of HSPCs (*Klamer and Voermans, 2014; Zhang and Lodish, 2008*)

1.3.1. Aging and clonal hematopoiesis

The HSC pool is heterogeneous and, besides the lineage-balanced HSCs, there are also myeloid-biased or lymphoid-biased HSCs (Figure 7A). Despite balanced-HSCs possessing the ability to differentiate into all mature blood cell types, they gradually lose their self-renewal and regeneration potential during aging (*López-Otín et al., 2013*). This results in age-associated functional decline and increase in myeloid-biased HSCs (Figure 7B) (*Rossi et al., 2008; 2005*). This age-related shift in the lineage potential of HSCs is associated with the longer life span of myeloid-biased HSCs compared to lymphoid-biased or lineage-balanced HSCs (*Cho et al., 2008*).

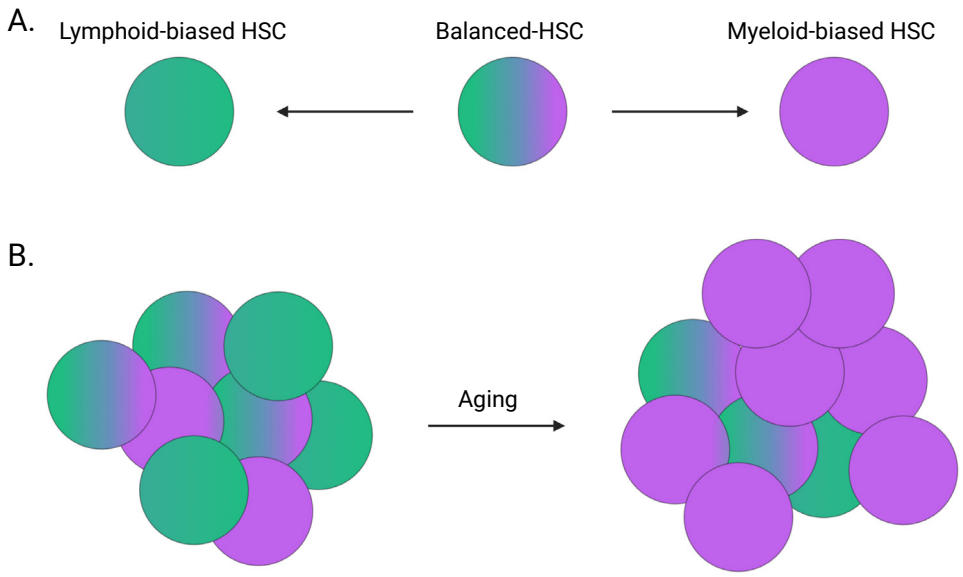


Figure 7. Aging increases the proportion of myeloid-biased HSCs. A) The heterogeneous HSC pool consists of lineage-balanced, lymphoid-biased and myeloid-biased HSCs. B) aging results in an increased proportion of myeloid-biased HSCs.

HSCs acquire somatic mutations during aging and this may result in functional heterogeneity in the HSC pool. Some HSCs that acquire somatic mutations gain proliferative advantage over others and form their distinct populations in the BM and sequentially in peripheral blood (PB), termed as clonal hematopoiesis (Figure 8) (*Beerman et al., 2010*). Although clonal hematopoiesis is a natural outcome of aging, mutations in some genes, e.g., *Dnmt3a*, *Tet2* and *Asxl1*, that result in the outgrowth of the corresponding HSC clones are associated with an elevated risk for the development of hematological malignancies such as myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) (Figure 8) (*Genovese et al., 2014; Xie et al., 2014; Busque et al., 2012; Boultwood et al., 2010*).

Although understanding the link between clonal evolution and functional heterogeneity of HSCs is challenging, recent advances coupling cellular barcoding and sequencing-based detection systems highlight the distinct transcriptional profile of functionally heterogeneous HSCs both in healthy and malignant hematopoiesis (*Avagyan et al., 2021; Velten et al., 2021; Pei et al., 2020; Nam et al., 2019; Rodriguez-Fraticelli et al., 2018*).

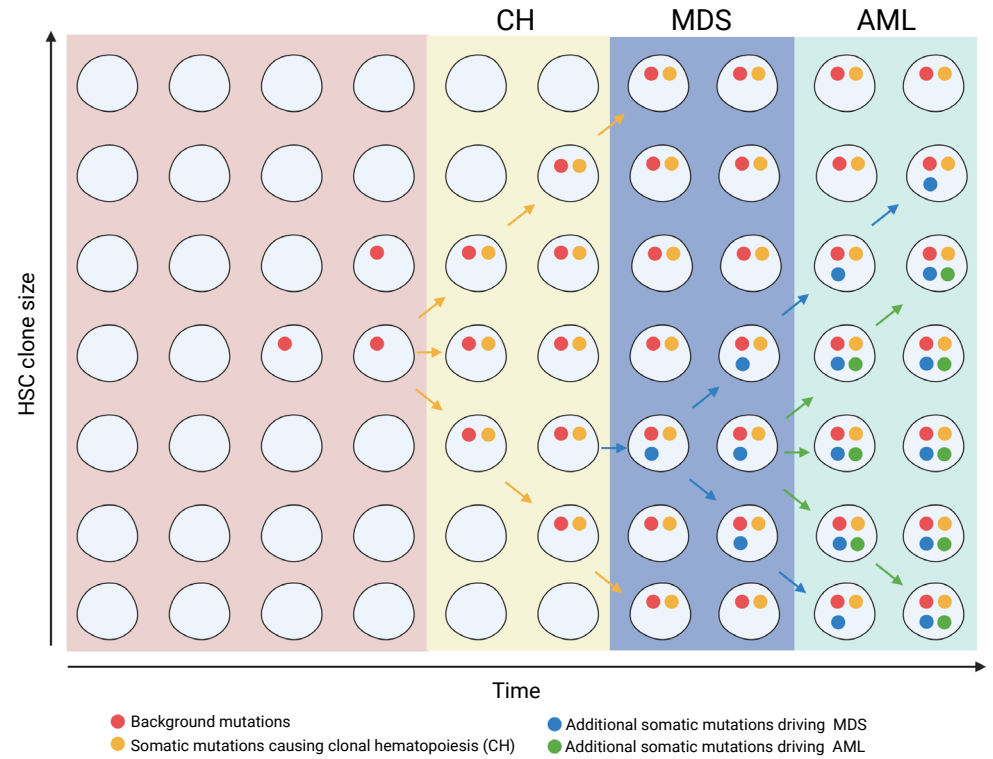


Figure 8. Acquisition of somatic mutations in HSCs can lead to clonal hematopoiesis and to the development of MDS and AML. Adapted from Kurosawa and Iwama (*Kurosawa and Iwama, 2020*). Acquired somatic mutations in HSCs can lead to the expansion of the corresponding HSC clones and result in clonal hematopoiesis (CH). Acquisition of additional somatic mutations in CH might lead to MDS and later to AML. CH might also directly lead to AML without an intervening MDS stage. MDS, myelodysplastic syndrome; AML, acute myeloid leukemia.

1.3.2. Bone marrow failure

Bone marrow failure (BMF) refers to an inefficient hematopoiesis resulting in reduced or absent hematopoietic precursors and associated PB cytopenias in one or more lineages. BMF syndromes comprise inherited or acquired forms; inherited forms that are associated with autosomal dominant mutations, such as the mutations in *GATA2* or *ELANE*, have variable penetrance with adolescence or adult onset of the disease (*Kallen et al., 2018; Boztug and Klein, 2009*). The acquisition of the BMF driving mutations in HSCs may impair hematopoiesis in a variety of ways; for instance, mutant HSCs may not differentiate into one or more lineages, undergo senescence, be eliminated by the immune system (*Challen and Goodell, 2020; Collado et al., 2005; Schreiber et al., 2011*). Additionally, aging may result in BMF as it reduces the functionality of HSCs due to, for instance, the acquired somatic mutations or prolonged exposure to chronic inflammation. (Zhao et al., 2021). Furthermore, BMF often predisposes to hematological malignancies such as MDS and AML (*Rio-Machin et al., 2020; Savage and Dufour, 2017*).

2. Transcription factor GATA2

The GATA family of zinc finger TFs consists of *GATA1-6* and has the ability to bind (T/A) GATA(A/G) DNA sequences. Three members of the GATA TF family, i.e., *GATA1*, *GATA2* and *GATA3*, have distinct and overlapping functions in the hematopoietic system: *GATA1* is expressed during the erythroid and megakaryocytic commitment, *GATA1* and *GATA2* have overlapping expression in the mast cell and eosinophil lineages, both *GATA2* and *GATA3* are expressed in HSCs and *GATA3* also plays an essential role during the T-cell development (*Gao et al., 2015*). Furthermore, during the formation of erythroid progenitors, reduced *GATA2* expression is followed by an increased *GATA1* expression through a process called 'GATA factor switch' indicating that cooperation between these two TFs is essential in erythroid differentiation (*Suzuki et al., 2013*).

In humans, germline heterozygous *GATA2* mutations are associated with various phenotypes since the initial studies in 2011: monocytopenia and mycobacterial infection (MonoMAC) syndrome (*Hsu et al., 2011*), monocyte, B cell, NK cell and dendritic cell deficiencies (DCML) (*Dickinson et al., 2011*), primary lymphedema with a predisposition to AML (Emberger syndrome) (*Ostergaard et al., 2011*) and familial MDS/AML (*Hahn et al., 2011*). Since 2011, hundreds of *GATA2* mutations have been identified in patients mostly predicted to be loss of function mutations. Therefore, germline GATA2 deficiency syndromes are classified as the loss of one allele of *GATA2* (GATA2 haploinsufficiency) which can manifest with immunodeficiencies, lymphedema, BMF and 80% risk of developing MDS/AML (*Donadieu et al., 2018; Hsu et al., 2015*).

In patients, *GATA2* mutations are found both in the coding and intronic regions of the gene. About 90% of *GATA2* mutations are either missense mutations within the zinc finger 2 (ZF2) or truncating mutations prior to the ZF2 domain of *GATA2* (*Wlodarski et al., 2016*). In addition, mutations in the intron 4 region of *GATA2* locus abrogate the function of a conserved +9.5 cis-element that regulates the expression level of *GATA2* and cause GATA2 haploinsufficiency (*Johnson et al., 2012*). However, the type or the location of *GATA2* mutations do not correlate with the phenotypic outcome of the patients. Although some *GATA2* mutation carriers remain asymptomatic, the risk of developing MDS/AML in *GATA2* mutation carriers increases from 6% at the age 10 to 81% at the age 40 (*Donadieu et al., 2018; Spinner et al., 2014*). This strongly suggests that GATA2 haploinsufficiency causes a vulnerable ground in the hematopoietic system for additional events to occur, such as secondary mutations or expansion of HSC clones harboring unfavorable molecular changes, that lead to leukemic transformation.

2.1. Gata2 in mouse hematopoiesis

In mice, germline homozygous deletion of *Gata2* (*Gata2*$^{-/-}$) is embryonically lethal due to hematopoietic failure and severe anemia at E10 (*Tsai et al., 1994*). On the other hand, heterozygous *Gata2* knockout mice (*Gata2*$^{+/-}$) have reduced numbers of HSPCs, during both embryogenesis (*Ling et al., 2004; Tsai et al., 1994*) and adulthood (*Guo et al., 2013; Rodrigues et al., 2005*). Furthermore, conditional deletion of *Gata2* in the HECs impairs the formation of HSCs during EHT, whereas conditional deletion of *Gata2* after HSC formation increases apoptosis in HSCs indicating that *Gata2* is required for the generation and survival of HSCs (*de Pater et al., 2013*). Besides, the disruption of the conserved +9.5 cis-element of *Gata2* locus impairs the formation of HSCs from HECs (*Lim et al., 2012; Gao et al., 2013*) indicating the regulatory function of this cis-acting element is conserved between human and mouse. Conversely, the overexpression of *Gata2* in HSCs results in the loss of HSC reconstitution potential (*Persons et al., 1999*) suggesting that fine-tuned *Gata2* expression is essential for HSC functionality.

Despite *Gata2*$^{+/-}$ mice are viable and have normal lineage differentiation throughout adulthood, conditional heterozygous deletion of *Gata2* in the hematopoietic cells increases proliferation and decreases lymphoid lineage differentiation ability of HSCs upon aging (*Abdelfattah et al., 2021*). This suggests that aging contributes to the progression of the Gata2 deficiency phenotype in *Gata2*$^{+/-}$ mice, as also observed in *GATA2*$^{+/-}$ patients. However, how aging deteriorates the phenotypic effects of GATA2 deficiency remains unknown and is investigated in **chapter 6**.

2.2. Gata2a and Gata2b in zebrafish hematopoiesis

As a result of an extra genome duplication event in teleost's, zebrafish have two orthologous of mammalian *Gata2*, *gata2a* and *gata2b,* showing 57% sequence identity (*Gillis et al., 2009*). The expression of *gata2a* and *gata2b* is detectable in the posterior lateral mesoderm around 10 hpf and 16 hpf respectively. At 25 hpf, both *gata2a* and *gata2b* are expressed in the dorsal aorta. However, *gata2a* expression is found throughout the posterior cardinal vein, whereas *gata2b* is mainly expressed in the ventral wall of the dorsal aorta, the site of HSC generation, indicating these two orthologues have distinct expression patterns in the dorsal aorta. By 36 hpf, *gata2a* expression is enriched in the vasculature, while *gata2b* expression is mainly found in nascent HSPCs (*Butko et al., 2015*).

The expression of *gata2a* is not affected by the loss of Notch signaling, which is an upstream regulator of mammalian *Gata2* in the hematopoietic system but dispensable for the arterial specification (*Clements et al., 2011*). Furthermore, the loss of *gata2a* causes lymphatic vascular defects and circulation (*Zhu et al., 2011*) indicating that *gata2a* is mainly required during lymphatic-vascular development. In the adult hematopoietic tissue KM, high levels of *gata2a* expression mark eosinophils (*Balla et al., 2010*). On the other hand, the presence of *gata2b* expression in the sites of hematopoiesis, e.g., CHT and KM, as well as in the majority of adult hematopoietic cells (*Butko et al., 2015*) shows that *gata2b* is essential in the hematopoietic system.

3. Outline and scope of the thesis

HSCs are the source of the hematopoietic system and *GATA2* is one of the master regulators of HSC generation and function. In humans, GATA2 deficiency syndromes present with a wide spectrum of phenotypes, and moreover, *GATA2* mutation carriers have more than 80% lifetime risk of developing MDS/AML. However, current *GATA2*-mutant models incompletely explain the underlying mechanisms driving the onset of hematological defects and leukemogenesis in GATA2 deficiency syndromes. In this thesis, we explore how *Gata2* dysregulation affects the formation and function of HSCs by investigating various *Gata2*-mutant mouse and zebrafish models.

Although *Gata2* haploinsufficiency severely reduces the formation of HSCs during EHT in *Gata2*$^{+/-}$ mice, the mechanism behind this reduction is unclear. In **chapter 2**, we investigate the effect of Gata2 haploinsufficiency on the individual subgroups of HSPC population undergoing EHT in germline *Gata2*$^{+/-}$ mice. We show that Gata2 haploinsufficiency does not abrogate the hematopoietic programming during EHT, but impairs HSC maturation through the downregulation of endothelial repressor *Gfi1b*. Furthermore, we explore whether the induction of *gfi1b* can restore the number of embryonic HSCs in *gata2b*-deficient zebrafish.

In zebrafish, the function of mammalian *Gata2* is shared between two orthologues; *gata2a* is mainly expressed in the vasculature and *gata2b* is expressed in the hematopoietic system allowing us to uniquely assess the role of *gata2b* in the hematopoietic system without disrupting the vascular morphogenesis or circulation. By taking advantage of this shared labor between the two orthologues, we investigate the role of *gata2b* by generating *gata2b*-null (**chapter 3**) and *gata2b*-heterozygous knockout (**chapter 4**) zebrafish. While complete deletion of *gata2b* (*gata2b*$^{-/-}$) attenuates embryonic HSC expansion and balanced lineage output in adult HSPCs (**chapter 3**), *gata2b* haploinsufficiency (*gata2b*$^{+/-}$) uniquely causes erythro-myeloid dysplasia in the adult KM (**chapter 4**). In the next chapter (**chapter 5**) we examine the function of a cis-regulatory element located in the 4th intron (i4) of *gata2a* locus corresponding to the +9.5 enhancer of the mouse *Gata2* locus. In **chapter 5,** we show that deleting *gata2a* i4 enhancer (*gata2a*$^{i4/i4}$) impairs EHT through the downregulation of *gata2b* by *gata2a* resulting in increased susceptibility to infections, oedema, neutropenia and hypocellular KM in adults. This variety in the phenotypic outcomes in *Gata2*-mutant zebrafish models (**chapter 3-5**) suggests that *Gata2* dosage is the determinant of the phenotypic consequences of Gata2 deficiency syndromes.

The risk of developing MDS/AML is substantially higher in the older group of GATA2 patients indicating aging elevates the effect of GATA2 haploinsufficiency. Although lymphoid lineage differentiation capacity of HSCs is impaired in aged-*Gata2*$^{+/-}$ mice, how aging contributes to this functional reduction is unexplained. In **chapter 6**, we address this by performing serial BM transplantation assays of aged-*Gata2*$^{+/-}$ and show that aging reduces the reconstitution ability and increases genomic instability of *Gata2*$^{+/-}$ HSCs resulting in B-cell cytopenia and monocytopenia, which recapitulates the phenotype of a subgroup of GATA2 patients. We also explore the transcriptional profile of *Gata2*$^{+/-}$ HSCs and show that *Gata2*$^{+/-}$ HSCs acquire a unique proliferative signature during embryonic development that they preserve throughout life.

Due to the incomplete penetrance in GATA2 deficiency syndromes and the lack of mechanistic insights on leukemogenesis in GATA2 patients, the only available treatment option for these patients is the allogenic HSC transplantation. In **chapter 7**, we discuss the mutational and phenotypic spectra of GATA2 patients, and the function of *Gata2* in mammalian hematopoiesis in detail to explore whether current advances in genome editing technologies can provide an alternative treatment for GATA2 patients.

Finally, in **chapter 8**, we summarize our findings and compare them to the current literature for their interpretation to the broader picture of the mechanisms behind GATA2 deficiency syndromes, and make recommendations for future directions.

REFERENCES

Abdelfattah, A., Hughes-Davies, A., Clayfield, L., Menendez-Gonzalez, J. B., Almotiri, A., Alotaibi, B., Tonks, A., & Rodrigues, N. P. (2021). Gata2 haploinsufficiency promotes proliferation and functional decline of hematopoietic stem cells with myeloid bias during aging. *Blood Adv*, *5*(20), 4285-4290.

Avagyan, S., Henninger, J. E., Mannherz, W. P., Mistry, M., Yoon, J., Yang, S., Weber, M. C., Moore, J. L., & Zon, L. I. (2021). Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis. *Science*, *374*(6568), 768-772.

Bakker, S. T., & Passegué, E. (2013). Resilient and resourceful: genome maintenance strategies in hematopoietic stem cells. *Exp Hematol*, *41*(11), 915-923.

Balla, K. M., Lugo-Villarino, G., Spitsbergen, J. M., Stachura, D. L., Hu, Y., Bañuelos, K., Romo-Fewell, O., Aroian, R. V., & Traver, D. (2010). Eosinophils in the zebrafish: prospective isolation, characterization, and eosinophilia induction by helminth determinants. *Blood*, *116*(19), 3944-3954.

Baron, C. S., Kester, L., Klaus, A., Boisset, J. C., Thambyrajah, R., Yvernogeau, L., Kouskoff, V., Lacaud, G., van Oudenaarden, A., & Robin, C. (2018). Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. *Nat Commun*, *9*(1), 2517.

Beerman, I., Bhattacharya, D., Zandi, S., Sigvardsson, M., Weissman, I. L., Bryder, D., & Rossi, D. J. (2010). Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc Natl Acad Sci U S A*, *107*(12), 5465-5470.

Bertrand, J. Y., Chi, N. C., Santoso, B., Teng, S., Stainier, D. Y., & Traver, D. (2010). Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature*, *464*(7285), 108-111.

Boisset, J. C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., & Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*, *464*(7285), 116-120.

Bonkhofer, F., Rispoli, R., Pinheiro, P., Krecsmarik, M., Schneider-Swales, J., Tsang, I. H. C., de Bruijn, M., Monteiro, R., Peterkin, T., & Patient, R. (2019). Blood stem cell-forming haemogenic endothelium in zebrafish derives from arterial endothelium. *Nat Commun*, *10*(1), 3577.

Boultwood, J., Perry, J., Pellagatti, A., Fernandez-Mercado, M., Fernandez-Santamaria, C., Calasanz, M. J., Larrayoz, M. J., Garcia-Delgado, M., Giagounidis, A., Malcovati, L., Della Porta, M. G., Jädersten, M., Killick, S., Hellström-Lindberg, E., Cazzola, M., & Wainscoat, J. S. (2010). Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia*, *24*(5), 1062-1065.

Boztug, K., & Klein, C. (2009). Novel genetic etiologies of severe congenital neutropenia. *Curr Opin Immunol*, *21*(5), 472-480.

Bradford, G. B., Williams, B., Rossi, R., & Bertoncello, I. (1997). Quiescence, cycling, and turnover in the primitive hematopoietic stem cell compartment. *Exp Hematol*, *25*(5), 445-453.

Brown, G., & Ceredig, R. (2019). Modeling the Hematopoietic Landscape. *Front Cell Dev Biol*, *7*, 104. stem cell. *Am J Pathol*, *169*(2), 338-346.

Busque, L., Patel, J. P., Figueroa, M. E., Vasanthakumar, A., Provost, S., Hamilou, Z., Mollica, L., Li, J., Viale, A., Heguy, A., Hassimi, M., Socci, N., Bhatt, P. K., Gonen, M., Mason, C. E., Melnick, A., Godley, L. A., Brennan, C. W., Abdel-Wahab, O., & Levine, R. L. (2012). Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet*, *44*(11), 1179-1181.

Butko, E., Distel, M., Pouget, C., Weijts, B., Kobayashi, I., Ng, K., Mosimann, C., Poulain, F. E., McPherson, A., Ni, C. W., Stachura, D. L., Del Cid, N., Espín-Palazón, R., Lawson, N. D., Dorsky, R., Clements, W. K., & Traver, D. (2015). Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. *Development*, *142*(6), 1050-1061.

Challen, G. A., & Goodell, M. A. (2020). Clonal hematopoiesis: mechanisms driving dominance of stem cell clones. *Blood*, *136*(14), 1590-1598.

Cheshier, S. H., Morrison, S. J., Liao, X., & Weissman, I. L. (1999). In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc Natl Acad Sci U S A*, *96*(6), 3120-3125.

Cho, R. H., Sieburg, H. B., & Muller-Sieburg, C. E. (2008). A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells. *Blood*, *111*(12), 5553-5561.

Ciau-Uitz, A., & Patient, R. (2019). Gene Regulatory Networks Governing the Generation and Regeneration of Blood. *J Comput Biol*, *26*(7), 719-725.

Clements, W. K., Kim, A. D., Ong, K. G., Moore, J. C., Lawson, N. D., & Traver, D. (2011). A somitic Wnt16/Notch pathway specifies haematopoietic stem cells. *Nature*, *474*(7350), 220-224.

Collado, M., Gil, J., Efeyan, A., Guerra, C., Schuhmacher, A. J., Barradas, M., Benguría, A., Zaballos, A., Flores, J. M., Barbacid, M., Beach, D., & Serrano, M. (2005). Tumour biology: senescence in premalignant tumours. *Nature*, *436*(7051), 642.

de Bruijn, M. F., Ma, X., Robin, C., Ottersbach, K., Sanchez, M. J., & Dzierzak, E. (2002). Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta. *Immunity*, *16*(5), 673-683.

de Pater, E., Kaimakis, P., Vink, C. S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P. S., Camper, S. A., Speck, N., & Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. *J Exp Med*, *210*(13), 2843-2850.

de Pater, E., & Trompouki, E. (2018). Bloody Zebrafish: Novel Methods in Normal and Malignant Hematopoiesis. *Front Cell Dev Biol*, *6*, 124.

Dickinson, R. E., Griffin, H., Bigley, V., Reynard, L. N., Hussain, R., Haniffa, M., Lakey, J. H., Rahman, T., Wang, X. N., McGovern, N., Pagan, S., Cookson, S., McDonald, D., Chua, I., Wallis, J., Cant, A., Wright, M., Keavney, B., Chinnery, P. F., Loughlin, J., Hambleton, S., Santibanez-Koref, M., & Collin, M. (2011). Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood*, *118*(10), 2656-2658.

Dieterlen-Lievre, F. (1975). On the origin of haemopoietic stem cells in the avian embryo: an experimental approach. *J Embryol Exp Morphol*, *33*(3), 607-619.

Donadieu, J., Lamant, M., Fieschi, C., de Fontbrune, F. S., Caye, A., Ouachee, M., Beaupain, B., Bustamante, J., Poirel, H. A., Isidor, B., Van Den Neste, E., Neel, A., Nimubona, S., Toutain, F., Barlogis, V., Schleinitz, N., Leblanc, T., Rohrlich, P., Suarez, F., Ranta, D., Chahla, W. A., Bruno, B., Terriou, L., Francois, S., Lioure, B., Ahle, G., Bachelerie, F., Preudhomme, C., Delabesse, E., Cave, H., Bellanné-Chantelot, C., Pasquet, M., & French, G. s. g. (2018). Natural history of GATA2 deficiency in a survey of 79 French and Belgian patients. *Haematologica*, *103*(8), 1278-1287.

Ema, H., & Nakauchi, H. (2000). Expansion of hematopoietic stem cells in the developing liver of a mouse embryo. *Blood*, *95*(7), 2284-2288.

Fleming, W. H., Alpern, E. J., Uchida, N., Ikuta, K., & Weissman, I. L. (1993). Steel factor influences the distribution and activity of murine hematopoietic stem cells in vivo. *Proc Natl Acad Sci U S A*, *90*(8), 3760-3764.

Fulcher, D. A., & Basten, A. (1994). Reduced life span of anergic self-reactive B cells in a double-transgenic model. *J Exp Med*, *179*(1), 125-134.

Gama-Norton, L., Ferrando, E., Ruiz-Herguido, C., Liu, Z., Guiu, J., Islam, A. B., Lee, S. U., Yan, M., Guidos, C. J., López-Bigas, N., Maeda, T., Espinosa, L., Kopan, R., & Bigas, A. (2015). Notch signal strength controls cell fate in the haemogenic endothelium. *Nat Commun*, *6*, 8510.

Ganuza, M., Hall, T., Finkelstein, D., Chabot, A., Kang, G., & McKinney-Freeman, S. (2017). Lifelong haematopoiesis is established by hundreds of precursors throughout mammalian ontogeny. *Nat Cell Biol*, *19*(10), 1153-1163.

Gao, J., Chen, Y. H., & Peterson, L. C. (2015). GATA family transcriptional factors: emerging suspects in hematologic disorders. *Exp Hematol Oncol*, *4*, 28.

Gao, X., Johnson, K. D., Chang, Y. I., Boyer, M. E., Dewey, C. N., Zhang, J., & Bresnick, E. H. (2013). Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *J Exp Med*, *210*(13), 2833-2842.

Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landén, M., Höglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., Sklar, P., Grönberg, H., Hultman, C. M., & McCarroll, S. A. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*, *371*(26), 2477-2487.

Gillis, W. Q., St John, J., Bowerman, B., & Schneider, S. Q. (2009). Whole genome duplications and expansion of the vertebrate GATA transcription factor gene family. *BMC Evol Biol*, *9*, 207.

Guo, G., Luc, S., Marco, E., Lin, T. W., Peng, C., Kerenyi, M. A., Beyaz, S., Kim, W., Xu, J., Das, P. P., Neff, T., Zou, K., Yuan, G. C., & Orkin, S. H. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell*, *13*(4), 492-505.

Hadland, B. K., Varnum-Finney, B., Poulos, M. G., Moon, R. T., Butler, J. M., Rafii, S., & Bernstein, I. D. (2015). Endothelium and NOTCH specify and amplify aorta-gonad-mesonephros-derived hematopoietic stem cells. *J Clin Invest*, *125*(5), 2032-2045.

Hahn, C. N., Chong, C. E., Carmichael, C. L., Wilkins, E. J., Brautigan, P. J., Li, X. C., Babic, M., Lin, M., Carmagnac, A., Lee, Y. K., Kok, C. H., Gagliardi, L., Friend, K. L., Ekert, P. G., Butcher, C. M., Brown, A. L., Lewis, I. D., To, L. B., Timms, A. E., Storek, J., Moore, S., Altree, M., Escher, R., Bardy, P. G., Suthers, G. K., D'Andrea, R. J., Horwitz, M. S., & Scott, H. S. (2011). Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*, *43*(10), 1012-1017.

Hsu, A. P., McReynolds, L. J., & Holland, S. M. (2015). GATA2 deficiency. *Curr Opin Allergy Clin Immunol*, *15*(1), 104-109.

Hsu, A. P., Sampaio, E. P., Khan, J., Calvo, K. R., Lemieux, J. E., Patel, S. Y., Frucht, D. M., Vinh, D. C., Auth, R. D., Freeman, A. F., Olivier, K. N., Uzel, G., Zerbe, C. S., Spalding, C., Pittaluga, S., Raffeld, M., Kuhns, D. B., Ding, L., Paulson, M. L., Marciano, B. E., Gea-Banacloche, J. C., Orange, J. S., Cuellar-Rodriguez, J., Hickstein, D. D., & Holland, S. M. (2011). Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*, *118*(10), 2653-2655.

Ivanovs, A., Rybtsov, S., Ng, E. S., Stanley, E. G., Elefanty, A. G., & Medvinsky, A. (2017). Human haematopoietic stem cell development: from the embryo to the dish. *Development*, *144*(13), 2323-2337.

Jagannathan-Bogdan, M., & Zon, L. I. (2013). Hematopoiesis. *Development*, *140*(12), 2463-2467.

Johnson, K. D., Hsu, A. P., Ryu, M. J., Wang, J., Gao, X., Boyer, M. E., Liu, Y., Lee, Y., Calvo, K. R., Keles, S., Zhang, J., Holland, S. M., & Bresnick, E. H. (2012). Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *J Clin Invest*, *122*(10), 3692-3704.

Kallen, M. E., Dulau-Florea, A., Wang, W., & Calvo, K. R. (2019). Acquired and germline predisposition to bone marrow failure: Diagnostic features and clinical implications. *Semin Hematol*, *56*(1), 69-82.

Kang, H., Mesquitta, W. T., Jung, H. S., Moskvin, O. V., Thomson, J. A., & Slukvin, II. (2018). GATA2 Is Dispensable for Specification of Hemogenic Endothelium but Promotes Endothelial-to-Hematopoietic Transition. *Stem Cell Reports*, *11*(1), 197-211.

Kim, A. D., Melick, C. H., Clements, W. K., Stachura, D. L., Distel, M., Panáková, D., MacRae, C., Mork, L. A., Crump, J. G., & Traver, D. (2014). Discrete Notch signaling requirements in the specification of hematopoietic stem cells. *Embo J*, *33*(20), 2363-2373.

Kissa, K., & Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*, *464*(7285), 112-115.

Klamer, S., & Voermans, C. (2014). The role of novel and known extracellular matrix and adhesion molecules in the homeostatic and regenerative bone marrow microenvironment. *Cell Adh Migr*, *8*(6), 563-577.

Kumaravelu, P., Hook, L., Morrison, A. M., Ure, J., Zhao, S., Zuyev, S., Ansell, J., & Medvinsky, A. (2002). Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. *Development*, *129*(21), 4891-4899.

Kurosawa, S., & Iwama, A. (2020). Aging and leukemic evolution of hematopoietic stem cells under various stress conditions. *Inflamm Regen*, *40*(1), 29.

Liakhovitskaia, A., Rybtsov, S., Smith, T., Batsivari, A., Rybtsova, N., Rode, C., de Bruijn, M., Buchholz, F., Gordon-Keylock, S., Zhao, S., & Medvinsky, A. (2014). Runx1 is required for progression of CD41+ embryonic precursors into HSCs but not prior to this. *Development*, *141*(17), 3319-3323.

Lim, K. C., Hosoya, T., Brandt, W., Ku, C. J., Hosoya-Ohmura, S., Camper, S. A., Yamamoto, M., & Engel, J. D. (2012). Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning. *J Clin Invest*, *122*(10), 3705-3717.

Lin, H. F., Traver, D., Zhu, H., Dooley, K., Paw, B. H., Zon, L. I., & Handin, R. I. (2005). Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. *Blood*, *106*(12), 3803-3810.

Ling, K. W., Ottersbach, K., van Hamburg, J. P., Oziemlak, A., Tsai, F. Y., Orkin, S. H., Ploemacher, R., Hendriks, R. W., & Dzierzak, E. (2004). GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med*, *200*(7), 871-882.

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, *153*(6), 1194-1217.

Ma, D., Zhang, J., Lin, H. F., Italiano, J., & Handin, R. I. (2011). The identification and characterization of zebrafish hematopoietic stem cells. *Blood*, *118*(2), 289-297.

Medvinsky, A., & Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*, *86*(6), 897-906.

Mikkola, H. K., & Orkin, S. H. (2006). The journey of developing hematopoietic stem cells. *Development*, *133*(19), 3733-3744.

Mock, D. M., Widness, J. A., Veng-Pedersen, P., Strauss, R. G., Cancelas, J. A., Cohen, R. M., Lindsell, C. J., & Franco, R. S. (2014). Measurement of posttransfusion red cell survival with the biotin label. *Transfus Med Rev*, *28*(3), 114-125.

Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E., & Weissman, I. L. (1997). Identification of a lineage of multipotent hematopoietic progenitors. *Development*, *124*(10), 1929-1939.

Müller, A. M., Medvinsky, A., Strouboulis, J., Grosveld, F., & Dzierzak, E. (1994). Development of hematopoietic stem cell activity in the mouse embryo. *Immunity*, *1*(4), 291-301.

Murke, F., Castro, S. V. C., Giebel, B., & Görgens, A. (2015). Concise Review: Asymmetric Cell Divisions in Stem Cell Biology. *Symmetry*, *7*(4), 2025-2037.

Nam, A. S., Kim, K. T., Chaligne, R., Izzo, F., Ang, C., Taylor, J., Myers, R. M., Abu-Zeinah, G., Brand, R., Omans, N. D., Alonso, A., Sheridan, C., Mariani, M., Dai, X., Harrington, E., Pastore, A., Cubillos-Ruiz, J. R., Tam, W., Hoffman, R., Rabadan, R., Scandura, J. M., Abdel-Wahab, O., Smibert, P., & Landau, D. A. (2019). Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature*, *571*(7765), 355-360.

North, T. E., de Bruijn, M. F., Stacy, T., Talebian, L., Lind, E., Robin, C., Binder, M., Dzierzak, E., & Speck, N. A. (2002). Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity*, *16*(5), 661-672.

Oatley, M., Bölükbası Ö, V., Svensson, V., Shvartsman, M., Ganter, K., Zirngibl, K., Pavlovich, P. V., Milchevskaya, V., Foteva, V., Natarajan, K. N., Baying, B., Benes, V., Patil, K. R., Teichmann, S. A., & Lancrin, C. (2020). Single-cell transcriptomics identifies CD44 as a marker and regulator of endothelial to haematopoietic transition. *Nat Commun*, *11*(1), 586.

Orkin, S. H. (2000). Diversification of haematopoietic stem cells to specific lineages. *Nat Rev Genet*, *1*(1), 57-64.

Orkin, S. H., & Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*, *132*(4), 631-644.

Osawa, M., Hanada, K., Hamada, H., & Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science*, *273*(5272), 242-245.

Ostergaard, P., Simpson, M. A., Connell, F. C., Steward, C. G., Brice, G., Woollard, W. J., Dafou, D., Kilo, T., Smithson, S., Lunt, P., Murday, V. A., Hodgson, S., Keenan, R., Pilz, D. T., Martinez-Corral, I., Makinen, T., Mortimer, P. S., Jeffery, S., Trembath, R. C., & Mansour, S. (2011). Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet*, *43*(10), 929-931.

Ottersbach, K. (2019). Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochem Soc Trans*, *47*(2), 591-601.

Pei, W., Shang, F., Wang, X., Fanti, A. K., Greco, A., Busch, K., Klapproth, K., Zhang, Q., Quedenau, C., Sauer, S., Feyerabend, T. B., Höfer, T., & Rodewald, H. R. (2020). Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding. *Cell Stem Cell*, *27*(3), 383-395 e388.

Persons, D. A., Allay, J. A., Allay, E. R., Ashmun, R. A., Orlic, D., Jane, S. M., Cunningham, J. M., & Nienhuis, A. W. (1999). Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. *Blood*, *93*(2), 488-499.

Porcheri, C., Golan, O., Calero-Nieto, F. J., Thambyrajah, R., Ruiz-Herguido, C., Wang, X., Catto, F., Guillén, Y., Sinha, R., González, J., Kinston, S. J., Mariani, S. A., Maglitto, A., Vink, C. S., Dzierzak, E., Charbord, P., Göttgens, B.,

Espinosa, L., Sprinzak, D., & Bigas, A. (2020). Notch ligand Dll4 impairs cell recruitment to aortic clusters and limits blood stem cell generation. *Embo J*, *39*(8), e104270.

Rhodes, J., Hagen, A., Hsu, K., Deng, M., Liu, T. X., Look, A. T., & Kanki, J. P. (2005). Interplay of pu.1 and gata1 determines myelo-erythroid progenitor cell fate in zebrafish. *Dev Cell*, *8*(1), 97-108.

Rio-Machin, A., Vulliamy, T., Hug, N., Walne, A., Tawana, K., Cardoso, S., Ellison, A., Pontikos, N., Wang, J., Tummala, H., Al Seraihi, A. F. H., Alnajar, J., Bewicke-Copley, F., Armes, H., Barnett, M., Bloor, A., Bödör, C., Bowen, D., Fenaux, P., Green, A., Hallahan, A., Hjorth-Hansen, H., Hossain, U., Killick, S., Lawson, S., Layton, M., Male, A. M., Marsh, J., Mehta, P., Mous, R., Nomdedéu, J. F., Owen, C., Pavlu, J., Payne, E. M., Protheroe, R. E., Preudhomme, C., Pujol-Moix, N., Renneville, A., Russell, N., Saggar, A., Sciuccati, G., Taussig, D., Toze, C. L., Uyttebroeck, A., Vandenberghe, P., Schlegelberger, B., Ripperger, T., Steinemann, D., Wu, J., Mason, J., Page, P., Akiki, S., Reay, K., Cavenagh, J. D., Plagnol, V., Caceres, J. F., Fitzgibbon, J., & Dokal, I. (2020). The complex genetic landscape of familial MDS and AML reveals pathogenic germline variants. *Nat Commun*, *11*(1), 1044.

Robert-Moreno, A., Espinosa, L., de la Pompa, J. L., & Bigas, A. (2005). RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells. *Development*, *132*(5), 1117-1126.

Rodrigues, N. P., Janzen, V., Forkert, R., Dombkowski, D. M., Boyd, A. S., Orkin, S. H., Enver, T., Vyas, P., & Scadden, D. T. (2005). Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood*, *106*(2), 477-484.

Rodriguez-Fraticelli, A. E., Wolock, S. L., Weinreb, C. S., Panero, R., Patel, S. H., Jankovic, M., Sun, J., Calogero, R. A., Klein, A. M., & Camargo, F. D. (2018). Clonal analysis of lineage fate in native haematopoiesis. *Nature*, *553*(7687), 212-216.

Rossi, D. J., Bryder, D., Zahn, J. M., Ahlenius, H., Sonu, R., Wagers, A. J., & Weissman, I. L. (2005). Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proc Natl Acad Sci U S A*, *102*(26), 9194-9199.

Rossi, D. J., Jamieson, C. H., & Weissman, I. L. (2008). Stems cells and the pathways to aging and cancer. *Cell*, *132*(4), 681-696.

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O., & Medvinsky, A. (2014). Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43(-) embryonic precursor. *Stem Cell Reports*, *3*(3), 489-501.

Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., Frampton, J., Zhao, S., & Medvinsky, A. (2011). Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J Exp Med*, *208*(6), 1305-1315.

Sánchez, M. J., Holmes, A., Miles, C., & Dzierzak, E. (1996). Characterization of the first definitive hematopoietic stem cells in the AGM and liver of the mouse embryo. *Immunity*, *5*(6), 513-525.

Savage, S. A., & Dufour, C. (2017). Classical inherited bone marrow failure syndromes with high risk for myelodysplastic syndrome and acute myelogenous leukemia. *Semin Hematol*, *54*(2), 105-114.

Schreiber, R. D., Old, L. J., & Smyth, M. J. (2011). Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science*, *331*(6024), 1565-1570.

Souilhol, C., Lendinez, J. G., Rybtsov, S., Murphy, F., Wilson, H., Hills, D., Batsivari, A., Binagui-Casas, A., McGarvey, A. C., MacDonald, H. R., Kageyama, R., Siebel, C., Zhao, S., & Medvinsky, A. (2016). Developing HSCs become Notch independent by the end of maturation in the AGM region. *Blood*, *128*(12), 1567-1577.

Spinner, M. A., Sanchez, L. A., Hsu, A. P., Shaw, P. A., Zerbe, C. S., Calvo, K. R., Arthur, D. C., Gu, W., Gould, C. M., Brewer, C. C., Cowen, E. W., Freeman, A. F., Olivier, K. N., Uzel, G., Zelazny, A. M., Daub, J. R., Spalding, C. D., Claypool, R. J., Giri, N. K., Alter, B. P., Mace, E. M., Orange, J. S., Cuellar-Rodriguez, J., Hickstein, D. D., & Holland, S. M. (2014). GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood*, *123*(6), 809-821.

Suzuki, M., Kobayashi-Osaki, M., Tsutsumi, S., Pan, X., Ohmori, S., Takai, J., Moriguchi, T., Ohneda, O., Ohneda, K., Shimizu, R., Kanki, Y., Kodama, T., Aburatani, H., & Yamamoto, M. (2013). GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation. *Genes Cells*, *18*(11), 921-933.

Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K. D., Yoshimoto, M., Yoder, M. C., Frampton, J., Schroeder, T., Enver, T., Göttgens, B., & de Bruijn, M. (2013). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat Commun*, *4*, 2924.

Tamplin, O. J., Durand, E. M., Carr, L. A., Childs, S. J., Hagedorn, E. J., Li, P., Yzaguirre, A. D., Speck, N. A., & Zon, L. I. (2015). Hematopoietic stem cell arrival triggers dynamic remodeling of the perivascular niche. *Cell*, *160*(1-2), 241-252.

Taoudi, S., Gonneau, C., Moore, K., Sheridan, J. M., Blackburn, C. C., Taylor, E., & Medvinsky, A. (2008). Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-cadherin+CD45+ pre-definitive HSCs. *Cell Stem Cell*, *3*(1), 99-108.

Thambyrajah, R., Mazan, M., Patel, R., Moignard, V., Stefanska, M., Marinopoulou, E., Li, Y., Lancrin, C., Clapes, T., Möröy, T., Robin, C., Miller, C., Cowley, S., Göttgens, B., Kouskoff, V., & Lacaud, G. (2016a). GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat Cell Biol*, *18*(1), 21-32.

Thambyrajah, R., Ucanok, D., Jalali, M., Hough, Y., Wilkinson, R. N., McMahon, K., Moore, C., & Gering, M. (2016b). A gene trap transposon eliminates haematopoietic expression of zebrafish Gfi1aa, but does not interfere with haematopoiesis. *Dev Biol*, *417*(1), 25-39.

Traver, D., Paw, B. H., Poss, K. D., Penberthy, W. T., Lin, S., & Zon, L. I. (2003). Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol*, *4*(12), 1238-1246.

Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W., & Orkin, S. H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, *371*(6494), 221-226.

Velten, L., Story, B. A., Hernández-Malmierca, P., Raffel, S., Leonce, D. R., Milbank, J., Paulsen, M., Demir, A., Szu-Tu, C., Frömel, R., Lutz, C., Nowak, D., Jann, J. C., Pabst, C., Boch, T., Hofmann, W. K., Müller-Tidow, C., Trumpp, A., Haas, S., & Steinmetz, L. M. (2021). Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. *Nat Commun*, *12*(1), 1366.

Wilkinson, A. C., & Göttgens, B. (2013). Transcriptional regulation of haematopoietic stem cells. *Adv Exp Med Biol*, *786*, 187-212.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C. F., Eshkind, L., Bockamp, E., Lió, P., Macdonald, H. R., & Trumpp, A. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*, *135*(6), 1118-1129.

Wilson, A., Laurenti, E., & Trumpp, A. (2009). Balancing dormant and self-renewing hematopoietic stem cells. *Curr Opin Genet Dev*, *19*(5), 461-468.

Wilson, A., & Trumpp, A. (2006). Bone-marrow haematopoietic-stem-cell niches. *Nat Rev Immunol*, *6*(2), 93-106.

Wlodarski, M. W., Hirabayashi, S., Pastor, V., Starý, J., Hasle, H., Masetti, R., Dworzak, M., Schmugge, M., van den Heuvel-Eibrink, M., Ussowicz, M., De Moerloose, B., Catala, A., Smith, O. P., Sedlacek, P., Lankester, A. C., Zecca, M., Bordon, V., Matthes-Martin, S., Abrahamsson, J., Kühl, J. S., Sykora, K. W., Albert, M. H., Przychodzien, B., Maciejewski, J. P., Schwarz, S., Göhring, G., Schlegelberger, B., Cseh, A., Noellke, P., Yoshimi, A., Locatelli, F., Baumann, I., Strahm, B., Niemeyer, C. M., & Ewog, M. D. S. (2016). Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. *Blood*, *127*(11), 1387-1397; quiz 1518.

Xie, H., Xu, J., Hsu, J. H., Nguyen, M., Fujiwara, Y., Peng, C., & Orkin, S. H. (2014). Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell*, *14*(1), 68-80.

Yokomizo, T., & Dzierzak, E. (2010). Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos. *Development*, *137*(21), 3651-3661.

Yzaguirre, A. D., Howell, E. D., Li, Y., Liu, Z., & Speck, N. A. (2018). Runx1 is sufficient for blood cell formation from non-hemogenic endothelial cells in vivo only during early embryogenesis. *Development*, *145*(2).

Zhang, C. C., & Lodish, H. F. (2008). Cytokines regulating hematopoietic stem cell function. *Curr Opin Hematol*, *15*(4), 307-311.

Zhao, L. P., Boy, M., Azoulay, C., Clappier, E., Sébert, M., Amable, L., Klibi, J., Benlagha, K., Espéli, M., Balabanian, K., Preudhomme, C., Marceau-Renaut, A., Benajiba, L., Itzykson, R., Mekinian, A., Fain, O., Toubert, A., Fenaux, P., Dulphy, N., & Adès, L. (2021). Genomic landscape of MDS/CMML associated with systemic inflammatory and autoimmune disease. *Leukemia*, *35*(9), 2720-2724.

Zhao, Y., Lin, Y., Zhan, Y., Yang, G., Louie, J., Harrison, D. E., & Anderson, W. F. (2000). Murine hematopoietic stem cell characterization and its regulation in BM transplantation. *Blood*, *96*(9), 3016-3022.

Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., Li, Z., Ni, Y., Mu, X., Wen, L., Cheng, T., Lan, Y., Yuan, W., Tang, F., & Liu, B. (2016). Tracing haematopoietic stem cell formation at single-cell resolution. *Nature*, *533*(7604), 487-492.

Zhu, C., Smith, T., McNulty, J., Rayla, A. L., Lakshmanan, A., Siekmann, A. F., Buffardi, M., Meng, X., Shin, J., Padmanabhan, A., Cifuentes, D., Giraldez, A. J., Look, A. T., Epstein, J. A., Lawson, N. D., & Wolfe, S. A. (2011). Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. *Development*, *138*(20), 4555-4564.

Zovein, A. C., Hofmann, J. J., Lynch, M., French, W. J., Turlo, K. A., Yang, Y., Becker, M. S., Zanetta, L., Dejana, E., Gasson, J. C., Tallquist, M. D., & Iruela-Arispe, M. L. (2008). Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell*, *3*(6), 625-636.

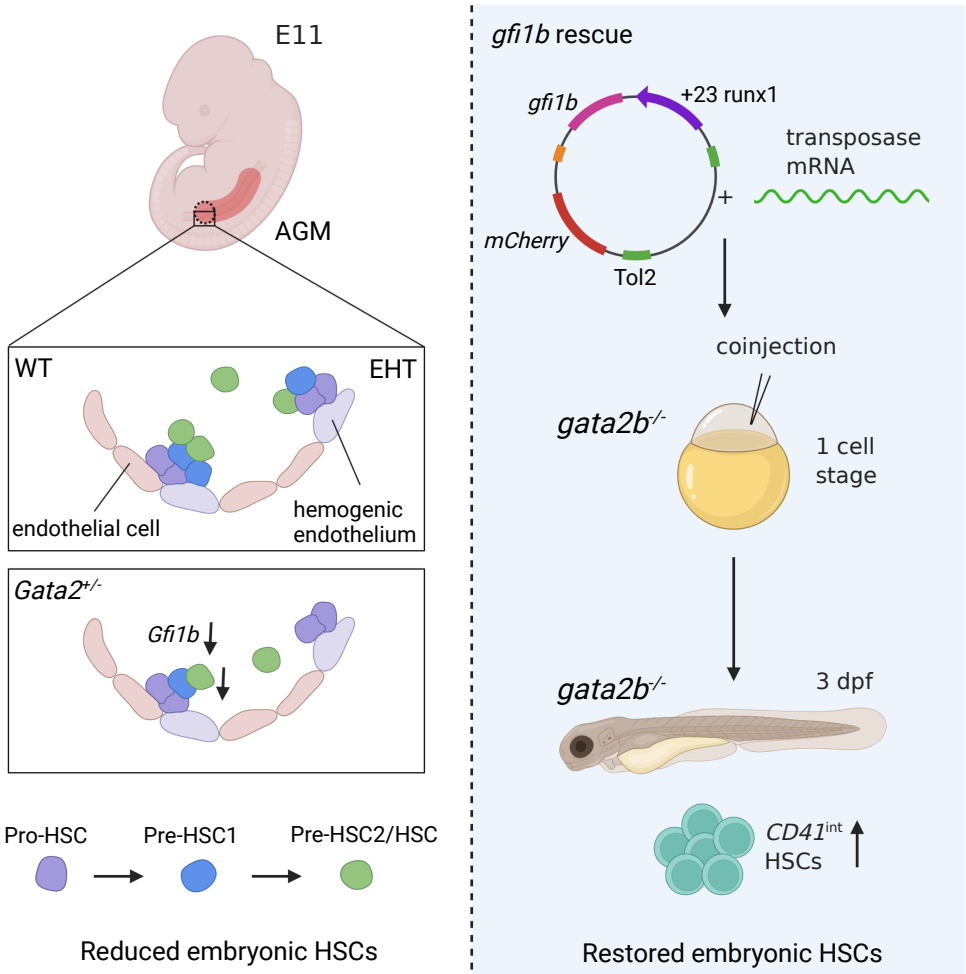
2

Gata2-regulated *Gfi1b* expression controls endothelial programming during endothelial-to-hematopoietic transition

Cansu Koyunlar[1], Emanuele Gioacchino[1], Disha Vadgama[1], Hans de Looper[1], Joke Zink[1], Remco Hoogenboezem[1], Marije Havermans[1], Eric Bindels[1], Elaine Dzierzak[2], Ruud, Delwel[1], Ivo Touw[1] and Emma de Pater[1].

[1]Department of Hematology, Erasmus Medical Center Cancer Institute, Rotterdam, the Netherlands
[2]The Queen's Medical Research Institute College of Medicine and Veterinary Medicine, Edinburgh, United Kingdom

Manuscript in submission



Abstract

The first hematopoietic stem cells (HSCs) are formed through endothelial-to-hematopoietic transition (EHT) events during embryonic development. The transcription factor *GATA2* is a crucial regulator of EHT and HSC function throughout life. Because *GATA2* haploinsufficiency patients have inborn mutations, prenatal defects are likely to have an influence on disease development. In mice, *Gata2* haploinsufficiency (*Gata2*$^{+/-}$) reduces the number and the functionality of embryonic hematopoietic stem and progenitor cells (HSPCs) generated through EHT. However, the embryonic HSPC pool is heterogeneous and the mechanisms underlying this defect in *Gata2*$^{+/-}$ embryos are unclear. Here, we investigated whether *Gata2* haploinsufficiency selectively affects a cellular subset undergoing EHT. We show that *Gata2*$^{+/-}$ HSPCs initiate but cannot fully activate hematopoietic programming during EHT. In addition, due to reduced activity of the endothelial repressor *Gfi1b*, *Gata2*$^{+/-}$ HSPCs cannot repress the endothelial identity to complete maturation. Finally, we show that hematopoietic-specific induction of *gfi1b* can restore HSC production in *gata2b*-null (*gata2b*$^{-/-}$) zebrafish embryos. This study illustrates pivotal roles of *Gata2* on the regulation of transcriptional network governing HSPC identity throughout EHT.

Highlights

- Embryonic *Gata2*$^{+/-}$ HSPCs are stuck during maturation due to aberrant endothelial gene expression and incomplete activation of hematopoietic transcriptional programming.
- *Gata2* activates *Gfi1b* to repress endothelial identity of embryonic HSPCs during maturation.
- Hematopoietic-specific induction of *gfi1b* restores the number of embryonic HSCs in *gata2b*$^{-/-}$ zebrafish.

INTRODUCTION

Hematopoiesis relies on multipotent, self-renewing HSCs. HSCs originate from the ventral wall of the embryonic dorsal aorta at the aorta-gonad-mesonephros (AGM) region. In the AGM region, definitive HSPCs are generated through a trans-differentiation process from a specialized endothelial cell (EC) compartment with hematopoietic potential (hemogenic endothelial cells or HECs). This process of endothelial-to-hematopoietic transition (EHT) is conserved between mammalian and non-mammalian vertebrates *(Ciau-Uitz and Patient, 2019; Ivanovs et al., 2017; Bertrand et al., 2010; Boisset et al., 2010; Kissa and Herbomel, 2010; Zovein et al., 2008; de Bruijn et al., 2002; North et al., 2002; Medvinsky and Dzierzak, 1996; Muller et al., 1994; Dieterlen-Lievre, 1975)*. In mice, EHT events occur between embryonic days (E)10.5-E12.5. Phenotypic HSPCs emerge from intra-aortic hematopoietic clusters (IAHCs) through EHT and co-express endothelial markers such as CD31 and hematopoietic markers like cKit *(Yokomizo and Dzierzak, 2010).* Previous studies showed that HSPCs develop through a multistep maturation process within IAHCs and that only a small fraction of IAHC cells become multipotent HSCs *(Rybtsov et al., 2014; 2011; Taoudi et al., 2008).* Throughout AGM maturation, HSPCs gradually repress the endothelial-specific gene expression and upregulate the expression of hematopoietic-specific genes *(Oatley et al., 2020; Baron et al., 2018; Zhou et al., 2016; Swiers et al., 2013).* Following their emergence from the AGM, HSCs migrate to the fetal liver and eventually colonize the BM around birth to maintain the hematopoietic system throughout life *(Zovein et al., 2008).*

The transcription factor *GATA2* is one of the key regulators of hematopoietic programming. In patients, germline heterozygous *GATA2* mutations result in *GATA2* haploinsufficiency syndromes. Typically, patients manifest with bone marrow failure and a high (80%) risk of developing myelodysplastic syndrome (MDS) or acute myeloid leukemia (AML) before the age of 40 *(Donadieu et al., 2018; Spinner et al., 2014; Dickinson et al., 2011; Hsu et al., 2011; Ostergaard et al., 2011; Hahn et al., 2011)*. Although *GATA2* expression is required in HSCs during both embryonic and adult stages, the consequences of embryonic *GATA2* haploinsufficiency for disease development are still unexplored.

In mice, *Gata2* is required for embryonic HSC generation and survival. While germline deletion of *Gata2* (*Gata2*$^{-/-}$) is lethal at E10, i.e., just before the appearance of the first HSCs, *Gata2* haploinsufficiency (*Gata2*$^{+/-}$) severely reduces the number of embryonic HSPCs, but these heterozygous mice survive to adulthood despite reduced numbers of HSCs *(Gao et al., 2013; de Pater et al., 2013; Rodrigues et al., 2005; Ling et al., 2004; Tsai et al., 1994).* Moreover, conditional deletion of *Gata2* in HECs does not fully abrogate the formation of IAHCs, but depletes the functional HSCs *(de Pater et al., 2013)*. In addition, HSCs do not survive from conditional deletion of *Gata2* after their emergence and become apoptotic *(de Pater et al., 2013).* Despite the requirement for *Gata2* during EHT is evident, mechanisms depleting functional HSCs within IAHCs in *Gata2*$^{+/-}$ embryos are incompletely understood.

Because phenotypic HSCs are generated during embryonic stages and *GATA2* patients have innate mutations, we aimed to understand as to why *GATA2* haploinsufficiency depletes phenotypic HSCs. In the present study, we explore how embryonic *Gata2* haploinsufficiency affects EHT and the development of first phenotypic HSCs in the AGM. We show that hematopoietic programming was not abrogated in *Gata2*$^{+/-}$ E11 HSPCs. However, *Gata2*$^{+/-}$ HSPCs were stuck during AGM maturation. Although some HSCs are matured, they are transcriptionally affected by Gata2 haploinsufficiency showing that *Gata2* is the key factor regulating transcriptional network in nascent HSCs. We further demonstrate that *Gata2* regulates *Gfi1b* to repress endothelial gene expression during HSPC maturation. Finally, we show that ectopic expression of *gfi1b* restores the number of phenotypic HSCs in *gata2b*-deficient zebrafish embryos. This study reveals previously unidentified roles of *Gata2* on modulating the transcriptional programming and maturation of HSPCs during EHT.

MATERIALS AND METHODS

Mouse and zebrafish models

Gata2$^{+/-}$ mice (*Tsai et al., 1994*), *gata2b*$^{-/-}$ zebrafish (*Gioacchino et al., 2021*) and *Tg(CD41:GFP)* zebrafish (*Ma et al., 2011*) were previously described. All animals were housed and bred in animal facilities at the Erasmus MC, Rotterdam, Netherlands. Animal studies were approved by the Animal Welfare and Ethics Committees of the EDC in accordance with legislation in the Netherlands.

Whole-mount immunofluorescence staining

E11 WT and *Gata2*$^{+/-}$ AGMs were dissected, prepared and mounted as previously described (*Yokomizo et al., 2012*). Primary antibody CD117 (cKit) rat anti-mouse (Invitrogen) was combined with secondary antibody Alexa Fluor-488 goat anti-rat (Invitrogen) for cKit visualization. CD31 was visualized by using biotinylated CD31 rat anti-mouse (BD Biosciences) and Cy5-conjugated Streptavidin (Jackson Immunoresearch) primary and secondary antibodies respectively. Whole AGM region for each sample was imaged using Leica SP5 confocal microscope. CD31 and cKit double positive cells were analyzed using Leica Application Suite X (version 4.3) software.

RNA isolation and sequencing

Cells were sorted in Trizol (Sigma) and total RNA isolation was performed according to the standard protocol using GenElute LPA (Sigma). RNA quality and quantity was assessed on 2100 Bioanalyzer (Agilent) using RNA 6000 Pico Kit (Agilent). cDNA was prepared using SMARTer procedure with SMARTer Ultra Low RNA kit (Clontech) and sequenced on Novaseq 6000 platform (Illumina).

Gene set enrichment and network analysis

Gene expression values were measured as FPKM (Fragments per kilobase of exon per million fragments mapped) and differential expression analysis was performed using the DESeq2 package in the R environment. Gene set enrichment analysis (GSEA) was performed on the FPKM values using the curated gene sets in the Molecular Signatures Database (MSigDB). GSEA results were used as an input for network analysis performed in Cytoscape software.

ATAC sequencing

Cells were processed for library preparation using the previously described protocol by Delwel group (*Ottema et al., 2021*). Libraries were quantified using Qubit and NEBNext Library Quant Kit for Illumina (NEB). Quality of the libraries were determined by the peak distribution visualization using 2100 Bioanalyzer (Agilent). Samples were sequenced on Novaseq 6000 platform (Illumina). Bigwig files were generated using the bamCoverage tool from deepTools and visualized using Integrative Genomics Viewer (IGV) software.

Flow cytometry and sorting (FACS)

AGM regions were dissected as described before (*Yokomizo et al., 2012*). Tissues were incubated with collagenase I (Sigma) in phosphate buffer solution (PBS) supplemented with 5 IU/mL penicillin, 5 µg/mL streptomycin, and 10% fetal calf serum (FCS) for 45 min at 37°C. Cells were stained using antibodies: PE-Cy7 anti-mouse CD31 (eBioscience), APC rat anti-mouse CD117 (cKit, BD Bioscience), FITC anti-mouse CD41 (Biolegend), PE rat anti-mouse CD43 (BD Bioscience) and Alexa Fluor-700 rat anti-mouse CD45 (BD Bioscience). All antibody incubations were performed in PBS + 10% FCS for 30 min on ice. After washing with PBS + 10% FCS at 1000 rpm for 10 min, cell pellets were resuspended with 1:1000 DAPI in PBS + 10% FCS for live/death cell discrimination. FACS events were recorded and cells were sorted using FACSAria III (BD Biosciences). Results were analyzed and visualized using FlowJo 7.6.5 software.

Colony-forming unit assay

Cells were incubated in MethoCult GF M3434 (Stem Cell Technologies) supplemented with 5 IU/mL penicillin and 5 µg/mL streptomycin at 37°C. Colony-forming units (CFU) were scored after 11 days of culture. Growth of primitive erythroid progenitor cells (BFU-E) and granulocyte-macrophage progenitor cells (CFU-GM, CFU-G and CFU-M) were scored using an inverted microscope.

Generation of a *gfi1b* construct

Wild type sequences of *gfi1b*, *mCherry* and *runx1* +23 enhancer were separately cloned into pJET1.2 vectors using CloneJet PCR Cloning Kit (Thermo Fisher). Following transformation, outgrown colonies were picked for DNA isolation. Presence of the insert was confirmed by restriction enzyme digestion using BglII (NEB) followed by agarose gel electrophoresis (1,5 %). DNA fragments then used as a PCR template for the Gibson cloning reaction. Fragments were amplified using overhang primers (Supplementary table 1) and purified using DNA Clean & Concentrator kit (Zymo Research). The pUC19-iTol2 backbone was digested with BamHI-HF (NEB) overnight at 37ºC and NEBuilder HiFi Assembly MasterMix (NEB) was used for the assembly of the fragments. Correct assembly was determined by HindIII restriction enzyme digestion and PCR amplification for the fragments. All transformations were done using E. coli and by performing heat shock for 30 seconds at 42ºC followed by recovery in SOC outgrowth medium (NEB) for 1 hour. All colonies were grown in LB medium plates supplemented with carbenicillin (50 mg/ml; 1000:1 v/v) and DNA from individual colonies were isolated using QIAprep Spin Miniprep Kit (Qiagen).

Generation of mRNA transposase

Plasmid with iTol2 sequence linearized using Not1 restriction enzyme (NEB). Linearized DNA was used as a template and RNA synthesis was performed using the HiScribe SP6 RNA Synthesis Kit (NEB) according to manufacturer's instructions. mRNA was precipitated using 3M sodium acetate (1:100) and 100% ethanol (3:1) and incubated overnight at -20 °C. Transposase mRNA was verified using 0.7% agarose gel electrophoresis.

Microinjection and embryo selection

Injection needles were prepared using the P-30 Magnetic Glass Microelectrode Vertical Needle Puller (Sutter Instrument). *gfi1b* construct and mRNA transposase were co-injected to the single cell of WT(*CD41:GFP*) or *gata2b*$^{-/-}$(*CD41:GFP*) zebrafish embryos at 1-cell stage using PV830 Pneumatic PicoPump (WPI). Embryos were anesthetized using 160 mg/L Tricaine (Sigma) for the selection of reporter expression. Reporter expression was assessed using the Leica DMLB fluorescence microscope.

***cmyb* in situ hybridization**

Following injections, embryos were treated with 0.003% 1-phenyl-2-thiourea (PTU, Sigma) at 24 hpf and fixed overnight with 4 % paraformaldehyde (PFA) in phosphate-buffered saline (PBS) containing 3% sucrose at 33 hpf. In situ hybridization (ISH) for *cmyb* was performed as previously described *(Gioacchino at al., 2021; Chocron et al., 2007).* The *cmyb* probe was a gift from Roger Patient. Results were imaged using an inverted microscope.

Statistics

Statistical analysis was carried out in GraphPad Prism 8.0.1 software. Normally distributed data were analyzed using unpaired t-test and otherwise Mann-Whitney test was used. Significance cut-off was set at $P < 0.05$.

RESULTS

E11 Gata2$^{+/-}$ HSPCs undergo incomplete EHT.

To explore the mechanisms leading to diminished HSPCs in *Gata2*$^{+/-}$ embryos, we sorted CD31^{+}cKit^{+} HSPCs from E11 WT and *Gata2*$^{+/-}$ AGMs and performed RNA-sequencing (RNA-seq) experiments (Figure 1A). We found significant differences in principal component analysis (PCA) between the transcriptomic signatures of WT and *Gata2*$^{+/-}$ HSPCs (Figure 1B). To understand the biological processes affected by the differentially expressed genes between the two genotypes, we performed gene-set enrichment analysis (GSEA) using curated gene-sets and compared our data to a previously published dataset showing upregulated and downregulated gene-sets found in EC, HEC and HSPC compartments during EHT (*Solaimani et al., 2015*). A hematopoietic-specific gene-set *Hematopoiesis stem cell*, which was upregulated in HECs and HSPCs compared to ECs, was overrepresented in *Gata2*$^{+/-}$ HSPCs compared to WT, indicating that there was no defect in the initiation of hematopoietic programming in these cells (Figure 1C). Surprisingly, an endothelial-specific gene-set *Vegfa Targets*, which was downregulated in HSPCs compared to ECs and HECs, was also enriched in *Gata2*$^{+/-}$ HSPCs (Figure 1D). Both hematopoietic and endothelial signatures were upregulated in *Gata2*$^{+/-}$ HSPCs suggesting these cells can initiate but are not able to complete EHT.

To further investigate the transcriptomic differences between WT and *Gata2*$^{+/-}$ HSPCs, we performed network analysis using Cytoscape software (Figure 1E). In this analysis, dots represent gene-sets and gene-sets sharing the same genes are connected by lines. Furthermore, gene-sets that are associated with the same biological processes form clusters indicated by the bigger circles. Earlier studies showed that HSPCs are highly proliferative during EHT *(Zape et al., 2017).* Strikingly, many gene-sets related to *Cell cycle*, *Proliferation*, *Transcription* and *Translation* networks were downregulated in *Gata2*$^{+/-}$ HSPCs, indicating that these processes are abrogated (Figure 1E). Conversely, gene-sets related to *Notch signaling*, *Integrin signaling*, *Cell junction organization* and *Extracellular matrix formation* were significantly upregulated in *Gata2*$^{+/-}$ HSPCs (Figure 1E).

These results suggest embryonic *Gata2*$^{+/-}$ HSPCs can initiate but cannot complete EHT due to an incomplete switch from endothelial programming to hematopoietic programming and are possibly stuck during their maturation into HSCs.

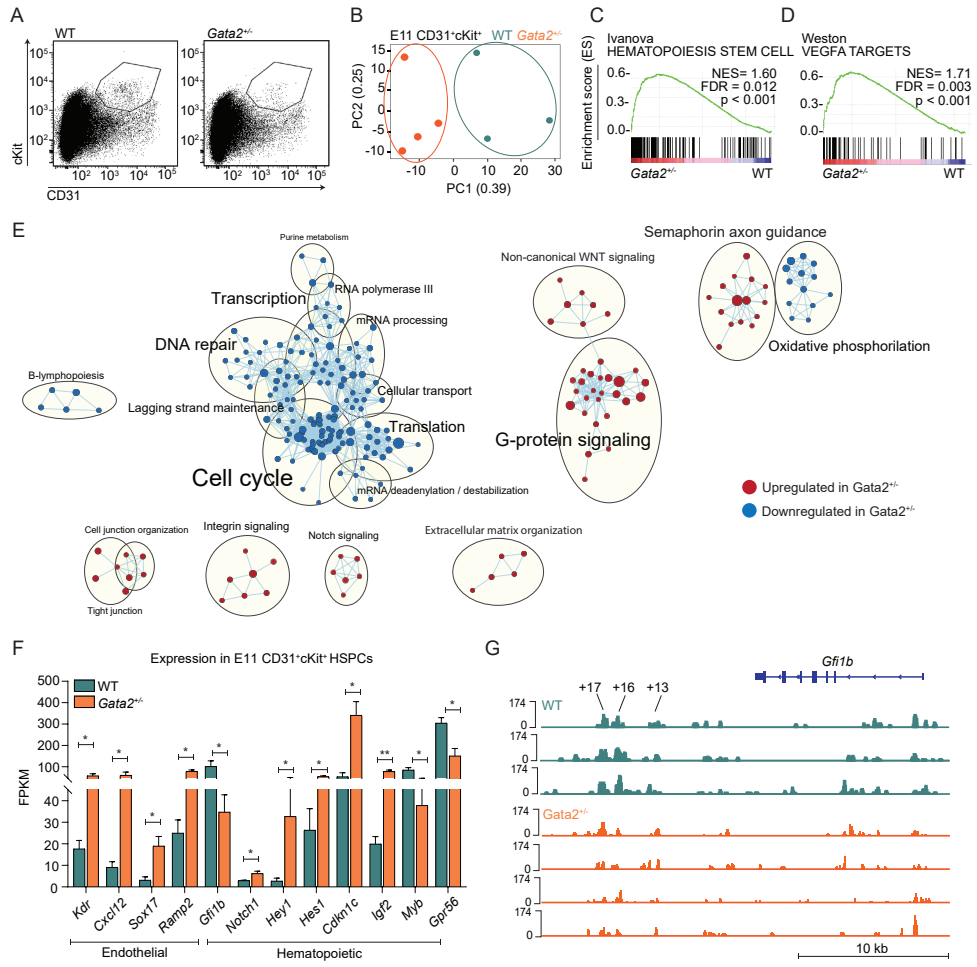


Figure 1. E11 *Gata2*$^{+/-}$ HSPCs have aberrant hematopoietic and endothelial transcriptome.
A) Sorting strategy for CD31^{+}cKit^{+} cells from E11 WT (left) or *Gata2*$^{+/-}$ (right) embryos. B) PCA of E11 WT (green) and *Gata2*$^{+/-}$ (orange) HSPCs. Dots represent the transcriptome of different samples. C-D) Gene sets upregulated in *Gata2*$^{+/-}$ HSPCs compared to WT HSPCs in GSEA; C) *Hematopoiesis stem cell* and D) *Vegfa targets*. E) Network analysis comparing E11 WT and *Gata2*$^{+/-}$ HSPCs. Red dots show upregulated and blue dots show downregulated gene sets in *Gata2*$^{+/-}$ HSPCs compared to WT. F) Comparison of the FPKM values of endothelial (*Kdr, Cxcl12, Sox17, Ramp2*) and hematopoietic (*Gfi1b, Notch1, Hey1, Hes1, Cdkn1c, Igf2, Myb, Gpr56*) specific genes between WT and *Gata2*$^{+/-}$ HSPCs. G) Comparison of open chromatin between E11 WT (N=3, green) or *Gata2*$^{+/-}$ (N=4, orange) HSPCs visualized by using IGV software. Accessible chromatin for *Gfi1b* and its + 13, + 16 and + 17 distal enhancer regions were analyzed. Peak range is set to min=0 and max= 174 for all samples. Tool bar represents 10 kb in length. Error bars represent standard error of the mean. *P < 0.05, **P < 0.01.

Endothelial-repressor transcription factor Gfi1b is downregulated in E11 Gata2$^{+/-}$ HSPCs.
Because GSEA and network analysis both indicated that endothelial and hematopoietic genes are enriched in *Gata2*$^{+/-}$ HSPCs, we looked into the expression of genes defining these transcriptomic signatures. Endothelial-specific genes such as *Kdr*, *Cxcl12*, *Sox17*

and *Ramp2* were significantly upregulated in *Gata2*$^{+/-}$ HSPCs (Figure 1F). In addition, we found an aberrant hematopoietic-specific transcriptome in *Gata2*$^{+/-}$ HSPCs. While some genes indicating hematopoietic programming such as *Notch1* and its targets *Hey1* and *Hes1*, *Cdkn1c* and *Igf2* were upregulated, other hematopoietic-specific genes such as *Myb* and *Gpr56* (*Adgrg1*) were significantly downregulated in *Gata2*$^{+/-}$ HSPCs (Figure 1F). These results support the hypothesis that *Gata2*$^{+/-}$ HSPCs can initiate hematopoietic programming but cannot fully gain hematopoietic characteristics due to impaired switching from the endothelial-specific to the hematopoietic-specific transcriptional program. Notably, *Gfi1b* expression was reduced in *Gata2*$^{+/-}$ HSPCs (Figure 1F). Previous work in mice showed that *Gfi1b* is responsible for the loss of endothelial identity during EHT and *Gfi1b* expression is essential for the formation of IAHCs from HECs *(Thambyrajah et al., 2016; Lancrin et al., 2012).* In addition, it was shown that *Gfi1b* is directly activated by *Gata2* through the hematopoietic specific +16 and +17 kb distal enhancer regions at E11.5 mouse embryos *(Moignard et al., 2013).* To further investigate whether *Gata2* haploinsufficiency causes *Gfi1b* downregulation due to the reduced activity in the enhancer regions of *Gfi1b*, we sorted CD31^{+}cKit^{+} HSPCs from E11 WT and *Gata2*$^{+/-}$ AGMs and performed ATAC-sequencing (ATAC-seq) to determine chromatin accessibility. Strikingly, both +16 and +17 kb distal enhancer regions of *Gfi1b* were less accessible in *Gata2*$^{+/-}$ HSPCs compared to WT HSPCs indicating *Gata2* regulates *Gfi1b* expression through these distal enhancer regions during EHT (Figure 1G).

Gata2 haploinsufficiency impairs HSPC maturation during EHT.

The above-mentioned results suggest that *Gata2*$^{+/-}$ HSPCs might get misprogrammed during EHT due to downregulation of *Gfi1b* and incomplete suppression of endothelial identity. Previous research showed that HSPCs develop within IAHCs through a multistep maturation process characterized as pro-HSC (*CD31*$^{+}$*cKit*$^{+}$*CD41*^{lo}*CD43*$^{-}$*CD45*$^{-}$) → pre-HSC Type I (pre-HSC1 or *CD31*$^{+}$*cKit*$^{+}$*CD41*^{lo}*CD43*$^{+}$*CD45*$^{-}$) → pre-HSC Type II (pre-HSC2) and HSCs (*CD31*$^{+}$*cKit*$^{+}$*CD41*^{lo}*CD43*$^{+}$*CD45*$^{+}$) *(Rybtsov et al., 2014; 2011; Taoudi et al., 2008)*. Furthermore, only the most mature compartment (pre-HSC2/HSCs) contains transplantable HSCs and can produce hematopoietic colonies in colony-forming unit culture (CFU-C) *(Taoudi et al., 2008).* Because HSC maturation requires activation of *Gfi1b* and consequently downregulation of endothelial genes *(Thambyrajah et al., 2016; Lancrin et al., 2012),* we asked whether *Gata2* haploinsufficiency causes a block in a specific stage of HSC maturation during EHT. To test this, we dissected WT and *Gata2*$^{+/-}$ AGMs from E11 embryos and performed flow cytometry experiments using antibodies against C*D31*, *cKit*, *CD41*, *CD43* and *CD45* (Figure S1A). Using this antibody combination, we analyzed the number of pro-HSC, pre-HSC1 and pre-HSC2/HSC populations in E11 WT and *Gata2*$^{+/-}$ AGMs (Figure 2A). We found that the number of E11 pro-HSCs were comparable between WT and *Gata2*$^{+/-}$ AGMs indicating the first step of HSPC formation was not hampered in *Gata2*$^{+/-}$ embryos (Figure 2B). In contrast, the number

of pre-HSC1 cells were moderately and pre-HSC2/HSC cells more prominently reduced in E11 *Gata2*$^{+/-}$ embryos, suggesting that *Gata2*$^{+/-}$ HSPCs cannot complete HSC maturation (Figure 2B).

To investigate if the maturation of pro-HSCs into pre-HSCs is impaired or delayed in *Gata2*$^{+/-}$ AGMs, we extended the flow cytometry analysis by including E12 and E13 AGMs dissected from WT and *Gata2*$^{+/-}$ embryos. At E12, the number of pre-HSC2/HSCs were higher compared to E11 in both WT and *Gata2*$^{+/-}$ indicating HSPCs were still actively undergoing maturation at this stage (Figure 2B, Figure S1B). In addition, both pre-HSC1 and pre-HSC2 cells were significantly reduced in *Gata2*$^{+/-}$ compared to WT AGMs at E12 (Figure S1B). As expected, the number of pro-HSC, pre-HSC1 and pre-HSC2/HSC populations were markedly reduced in both WT and *Gata2*$^{+/-}$ AGMs at E13 compared to E11 or E12 (Figure 2B, Figure S1B and C). *Gata2*$^{+/-}$ AGMs, however, still contained a decreased number of HSPCs at E13 compared to WT AGMs confirming that HSPC maturation is not delayed but blocked in *Gata2*$^{+/-}$ embryos (Figure S1C).

Finally, we asked whether the functionality of *Gata2*$^{+/-}$ HSCs is altered at E11. Earlier studies testing the functionality of E11 *Gata2*$^{+/-}$ HSPCs in CFU-C assays showed that *Gata2*$^{+/-}$ HSPCs produce fewer hematopoietic colonies compared to WT *(de Pater et al., 2013).* But it remained unclear whether this was due to a reduction in number of pre-HSC2/HSCs or due to a reduction in their potential to generate CFUs. To address this, we sorted pro-HSC, pre-HSC1 and pre-HSC2/HSC populations from E11 WT and *Gata2*$^{+/-}$ AGMs and performed CFU-C assay to assess their hematopoietic potential separately. After 11 days in CFU-C, we quantified and normalized the number of counted colonies to the number of cells plated per dish. The functionality of pre-HSC2/HSCs was preserved between WT and *Gata2*$^{+/-}$ AGMs, as this population from both genotypes produced similar type and number of hematopoietic colonies (Figure S1D). In addition, pro-HSC and pre-HSC1 populations of both WT and *Gata2*$^{+/-}$ AGMs did not form any colonies after 11 days in culture, confirming these cells have not yet acquired HSC potential (*data not shown*).

These results indicate that *Gata2*$^{+/-}$ HSPCs are arrested in pro-HSC to pre-HSC maturation during EHT. Furthermore, *Gata2*$^{+/-}$ AGMs are still able to produce fewer pre-HSC2/HSCs indicating HSC generation is not fully blocked but severely reduced by *Gata2* haploinsufficiency. Moreover, these cells do maintain similar CFU potential as WT Pre-HSC2/HSCs.

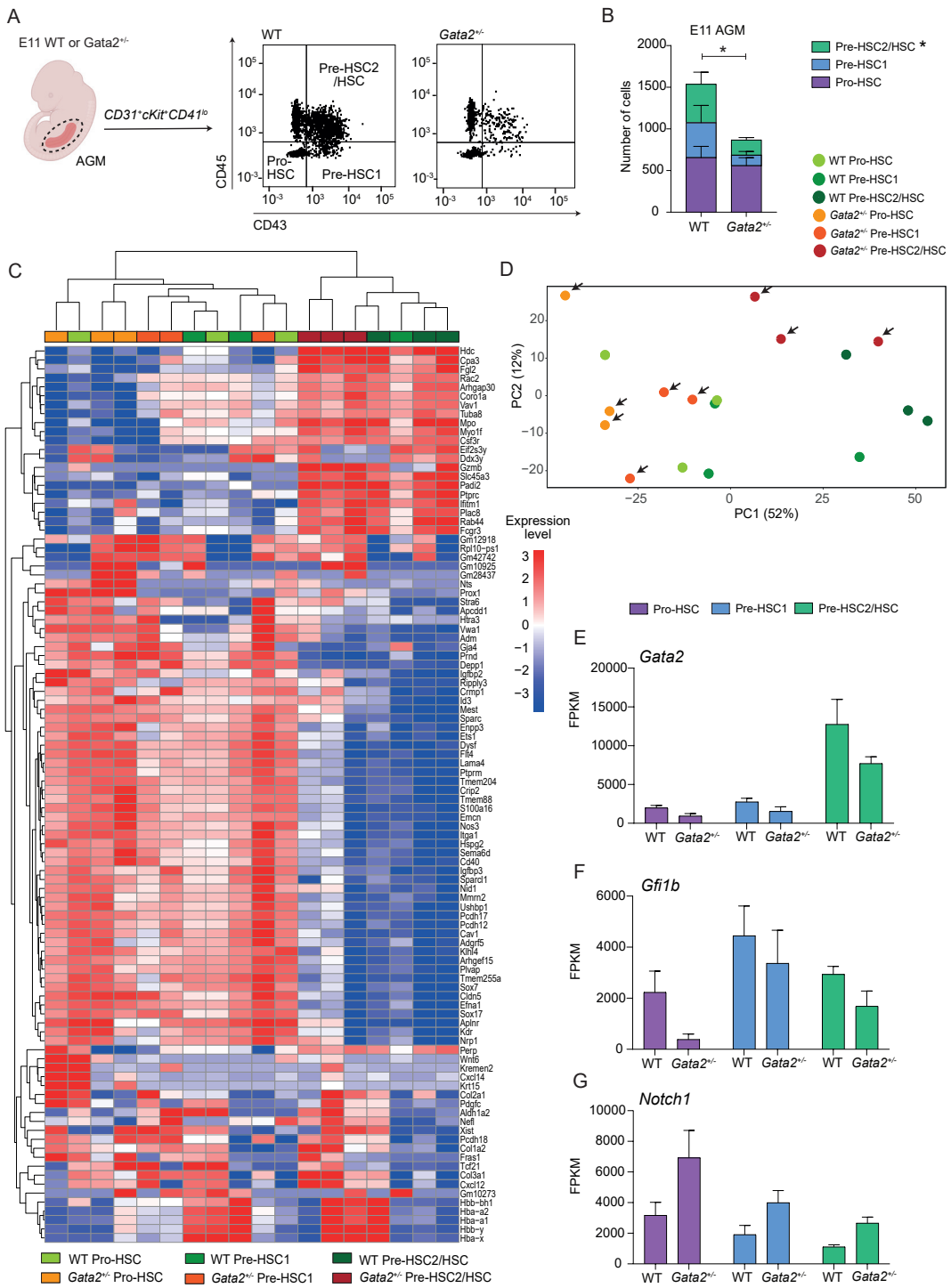


Figure 2. HSPC maturation within IAHCs is impaired in *Gata2*$^{+/-}$ during EHT.
A) Gating strategy to determine HSPC maturation at E11 WT or *Gata2*$^{+/-}$ AGMs. Representative image of pro-HSC, pre-HSC1 and pre-HSC2 gating obtained from WT (left) or *Gata2*$^{+/-}$ (right) AGMs. B) Quantification of the number of pro-HSC, pre-HSC1 and pre-HSC2 populations in E11 WT or *Gata2*$^{+/-}$ AGMs. C) Unbiased heatmap of the transcriptomic signatures of pro-HSC, pre-HSC1 and pre-HSC2 populations from E11 WT or *Gata2*$^{+/-}$ AGMs. D) PCA showing the transcriptome of each sample obtained from RNA sequencing. Black arrows indicate *Gata2*$^{+/-}$ samples. Green arrow indicates the maturation trajectory based on the transcriptome of WT HSPCs. FPKM values of E) *Gata2* F) *Gfi1b* and G) *Notch1* depicted for each stage of the maturation and compared between WT and *Gata2*$^{+/-}$. Error bars represent standard error of the mean. Color code for samples according to the genotype: WT samples in the shades of green and *Gata2*$^{+/-}$ samples in the shades of orange. Color code for maturation steps: pro-HSCs, purple; pre-HSC1, blue; pre-HSC2, green. *P < 0.05

Pre-HSC2/HSCs are marked by unique transcriptomic signatures.

To investigate why *Gata2*$^{+/-}$ HSPCs are affected in the pro-HSC stage of maturation, we sorted pro-HSC, pre-HSC1 and pre-HSC2/HSC compartments from E11 WT and *Gata2*$^{+/-}$ AGMs and performed RNA-seq experiments. The heatmap comparing the transcriptomic signatures of each cell population from WT and *Gata2*$^{+/-}$ AGMs shows that overall pro-HSC and pre-HSC1 compartments of both WT and *Gata2*$^{+/-}$ have comparable transcriptomes which dramatically change during pre-HSC2/HSC maturation (Figure 2C). PCA also confirmed that, in both WT and *Gata2*$^{+/-}$, the transcriptome of pre-HSC2/HSCs cluster separately from the pro-HSC and pre-HSC1 compartments (Figure 2D). As expected, pre-HSC2/HSCs from both genotypes uniquely express *Ptprc* (*CD45*) and are marked by the specific expression of transcripts such as *Hdc*, *Fgl2*, *Slc45a*, *Padi2*, *Rab44* and *Fcgr3* (Figure 2C). In addition, expression of many genes like *Flt4*, *Ptprm*, *Itga1* and *Sox17* were turned-off in pre-HSC2/HSCs of both WT and *Gata2*$^{+/-}$ indicating mature HSCs acquire a unique transcriptomic signature (Figure 2C).

Gata2$^{+/-}$ *HSPCs fail to repress endothelial programming throughout HSC maturation.*

Flow cytometry analysis showed that HSPC maturation in *Gata2*$^{+/-}$ was predominantly impaired during Pro-HSC to Pre-HSC1 transition (Figure 2B, Figure S1B). Comparison of the transcriptomes of WT and *Gata2*$^{+/-}$ by heatmap analysis and PCA, showed that the transcriptome profiles of WT pre-HSC1 cells were distributed between pro-HSC and pre-HSC2/HSCs states, whereas *Gata2*$^{+/-}$ pre-HSC1 cells were transcriptionally closer to a pro-HSC-like states (Figure 2C and D). Although some endothelial-specific genes like *Sox7* and *Flt4* were expressed in both WT and *Gata2*$^{+/-}$ pre-HSC1 cells (Figure 2C), other endothelial-specific genes such as *Igfbp2*, *Vwa1* and *Gja4* were upregulated in *Gata2*$^{+/-}$ compared to WT pre-HSC1 cells (Figure S2A-C and supplementary material 2).

Furthermore, WT and *Gata2*$^{+/-}$ pre-HSC2/HSCs did not show striking differences in transcriptome (Figure 2C and supplementary material 3) probably explaining the preserved functionality in CFU-C assays. Expression of some endothelial-specific genes such as *Sox17*, *Kdr* and *Flt1* were silenced during pre-HSC1 to pre-HSC2/HSC maturation in both WT and *Gata2*$^{+/-}$ (Figure 2C). However, some endothelial-specific genes such as *Cxcl12*, *Col3a1* and *Aplnr* were upregulated in *Gata2*$^{+/-}$ pre-HSC2/HSCs compared to WT indicating incomplete repression of endothelial programming throughout the maturation of *Gata2*$^{+/-}$ HSPCs (Figure 2C, Figure S2D-F).

These differences indicate that the mature Pre-HSC1 and Pre-HSC2/HSC compartments of *Gata2*$^{+/-}$ are transcriptionally distinguishable from WT and suggest that *Gata2*$^{+/-}$ HSPCs incompletely repress endothelial programming through maturation from pro-HSC to pre-HSC1 and pre-HSC2/HSC. However, despite this defect, some *Gata2*$^{+/-}$ HSPCs are still capable to undergo complete AGM maturation despite carrying endothelial transcriptomic signatures.

Gata2$^{+/-}$ *HSPCs incompletely activate hematopoietic programming throughout HSC maturation.*

Because some hematopoietic transcriptomic signatures were upregulated whereas others were downregulated in *Gata2*$^{+/-}$ HSPCs (Figure 1F), we asked whether this observation was due to the reduced number of mature pre-HSC2/HSCs within HSPCs. To explore this, we investigated the expression levels of hematopoietic-specific *cKit*, *Myb* and *CD44* at the individual steps of maturation. For both WT and *Gata2*$^{+/-}$, expression of these genes increased throughout the maturation (Figure S2G-I). However, expression levels of these genes were reduced in *Gata2*$^{+/-}$ HSPCs compared to WT throughout the maturation, confirming that hematopoietic transcriptional programming is hampered in *Gata2*$^{+/-}$ HSPCs (Figure S2G-I). Strikingly, these genes were most significantly reduced in the pro-HSCs in *Gata2*$^{+/-}$ (Figure S2G-I and supplementary material 1).

In addition, WT HSPCs were expressing hematopoietic specific genes such as *Vav1*, *Rac2* and *Mpo* in all stages and the expression level of these genes gradually increased throughout maturation (Figure S2J-L). On the other hand, these genes were downregulated in pro-HSCs and only activated later during the pre-HSC1 and pre-HSC2/HSCs maturation in *Gata2*$^{+/-}$ embryos (Figure S2J-L), indicating a direct effect of Gata2 haploinsufficiency in the onset of the hematopoietic programming.

Together these results propose a crucial role for *Gata2* in HSPCs undergoing EHT; downregulation of endothelial identity and upregulation of hematopoietic transcriptional programming to promote a complete transition to the HSC-like state.

Gfi1 and Gfi1b are downregulated in Gata2$^{+/-}$ *HSPCs during pro-HSC to pre-HSC maturation.*

Previous studies showed that *Gata2* is upregulated in HSPCs compared to ECs and HECs during EHT *(Eich et al., 2018).* By comparing the expression level of *Gata2* among pro-HSC, pre-HSC1 and pre-HSC2/HSC compartments we found that *Gata2* is predominantly expressed in the most mature compartment (pre-HSC2/HSC) of HSPCs (Figure 2E). In addition, *Gata2*$^{+/-}$ HSPCs in all maturation stages have reduced *Gata2* expression compared to WT confirming the *Gata2* haploinsufficiency due to the heterozygous deletion of *Gata2* (Figure 2E).

Because we hypothesized that *Gata2* regulates *Gfi1b* to repress endothelial gene expression during EHT and transcriptome analysis confirmed an incomplete repression of

endothelial programming during HSPC maturation in *Gata2*$^{+/-}$ embryos, we analyzed the expression of *Gfi1b* in each maturation stage. We found that *Gfi1b* expression was mainly increased during pro-HSC to pre-HSC1 maturation in both WT and *Gata2*$^{+/-}$ (Figure 2F). Furthermore, the expression level of *Gfi1b* was reduced in *Gata2*$^{+/-}$ Pro-HSCs but not as dramatically in pre-HSC1 and pre-HSC2/HSC compartments compared to WT (Figure 2F). These results suggest that *Gata2* haploinsufficiency impairs *Gfi1b* activation predominantly in the pro-HSC stage of the HSPC maturation.

Previous studies showed that *Gfi1*, a highly homologous interaction partner of *Gfi1b*, is also required for IAHC formation during EHT *(Thambyrajah et al., 2016; Lancrin et al., 2012).* In addition, *CD45*$^{+}$ pre-HSC2/HSCs co-express *Gfi1b* and *Gfi1* indicating that their co-activation is required during HSPC maturation within IAHCs *(Thambyrajah et al., 2016).* Therefore, we examined whether *Gfi1* expression was altered in *Gata2*$^{+/-}$ HSPCs throughout maturation. We found that, similar to *Gfi1b*, the expression of *Gfi1* was increased during pro-HSC to pre-HSC1 maturation (Figure S2M). Strikingly, *Gfi1* expression was completely abolished in *Gata2*$^{+/-}$ pre-HSC1 cells (Figure S2M). However, the expression level of *Gfi1* was normalized in *Gata2*$^{+/-}$ pre-HSC2/HSCs concomitant with the rescued expression level of *Gata2* (Figure 2E and S2M). This suggests *Gata2* is required for the expression of both *Gfi1* genes that are crucial for the repression of endothelial programming during HSPC maturation.

Notch signaling is upregulated in Gata2$^{+/-}$ *HSPCs.*

Notch signaling is essential for EHT and HSPC maturation and earlier studies indicated that activation of Notch1 and its targets such as Hes1 is also required for the initiation of EHT in the AGM (Souilhol et al., 2016; Robert-Moreno et al., 2005). However, HSCs become Notch-independent throughout AGM maturation and continuation of high Notch1 activity eventually blocks HSC maturation (Souilhol et al., 2016). Because transcriptome analysis of E11 HSPCs showed that genes related to Notch signaling were upregulated in Gata2$^{+/-}$ HSPCs, we explored the expression levels of key Notch signaling mediators. Both Notch1 and Hes1 were enriched in Gata2$^{+/-}$ HSPCs compared to WT (Figure 2G and S2N). However, the expression level of Notch2 was comparable between WT and Gata2$^{+/-}$ HSPCs throughout maturation (Figure S2O). Because Gata2 is a known downstream target of Notch1, these results suggest a possible feedback regulation of Notch signaling by Gata2 during EHT and HSPC maturation.

Both IAHCs and single bulging cells are diminished in Gata2$^{+/-}$ *AGMs.*

Our results indicated that HSPC maturation within IAHCs is inhibited in *Gata2*$^{+/-}$ AGMs. In addition, a recent study showed that acquisition of HSC fate is not exclusive to IAHCs and *Gata2*-expressing CD27^{+} single bulging cells (1- to 2-cell IAHCs or SBCs) also have HSC potential *(Vink et al., 2020)*. To clarify whether *Gata2* haploinsufficiency exclusively

diminishes HSPC maturation within IAHCs or has an effect on other intra-aortic HSPC subtypes, we dissected AGMs from E11 WT and *Gata2*$^{+/-}$ embryos and performed whole-mount immunofluorescence staining using antibodies against *CD31* and *cKit* (Figure 3A-C). In whole WT and *Gata2*$^{+/-}$ AGMs, we quantified the number of CD31^{+}cKit^{+} SBCs and the CD31^{+}cKit^{+} cells located in IAHCs (Figure 3D, E and G). We also assessed the number of single CD31^{+}cKit^{+} cells that are not attached to endothelium or IAHCs and are found in the aortic lumen (AL) (Figure 3F and G). Our results showed that the number of CD31^{+}cKit^{+} cells within all AGM compartments are decreased in E11 *Gata2*$^{+/-}$ embryos with significant reductions in both IAHCs and SBCs (Figure 3G). Furthermore, because *CD27* is expressed in both SBCs and IAHCs and is a marker for multipotent HSPCs *(Vink et al., 2020),* we examined the expression level of CD27 throughout the maturation stages. We found that the expression level of CD27 was greatly reduced in *Gata2*$^{+/-}$ pre-HSC1 compartment (Figure 3H). These results suggest that *Gata2* haploinsufficiency does not exclusively affect the HSC maturation within IAHCs, but reduces the HSC potential within AGM HSPC pool indicating a broader function for *Gata2* during EHT.

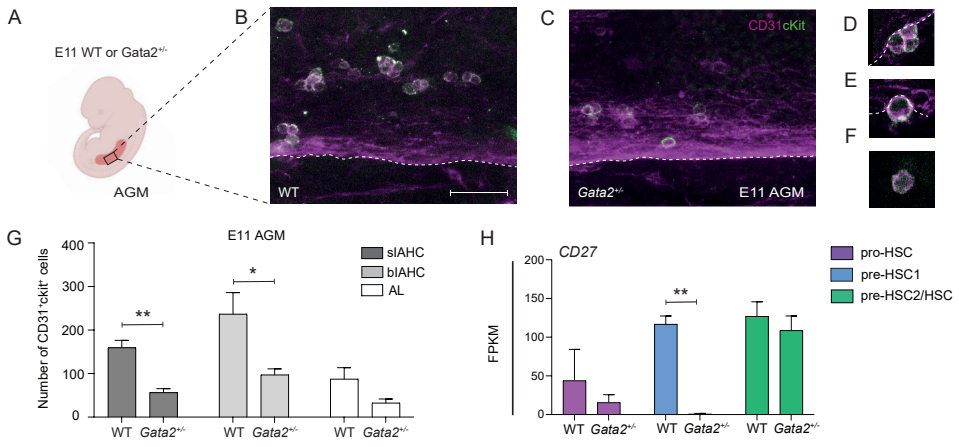


Figure 3. Both IAHCs and single bulging cells are depleted in *Gata2*$^{+/-}$ AGMs.
A) Illustration of AGM region dissected for analysis from E11 WT or *Gata2*$^{+/-}$ embryos. B-F) Representative images of CD31^{+}cKit^{+} cells obtained by confocal imaging at E11 B) WT AGM, C) *Gata2*$^{+/-}$ AGM, D) IAHCs, E) SBC and F) AL cell. White bar indicates 50 um in length. G) Quantification of CD31^{+}cKit^{+} cells located in IAHCs, as SBCs and as AL cell within E11 WT (N=3) and *Gata2*$^{+/-}$ (N=4) AGMs. H) FPKM values of CD27 depicted for each stage of the maturation and compared between WT and *Gata2*$^{+/-}$. Error bars represent standard error of the mean. *P < 0.05, **P < 0.01. AGM, aorta-gonad-mesonephros; IAHC, intra-aortic hematopoietic cluster; SBC, single bulging cell; AL, aortic lumen cell.

Hematopoietic-specific gfi1b induction restores embryonic HSCs in gata2b$^{-/-}$ *zebrafish.*
To test whether the effect of *Gata2* haploinsufficiency on EHT can be overcome by inducing *Gfi1b* expression, we took advantage of the previously described *gata2b*$^{-/-}$ zebrafish model. In *gata2b*$^{-/-}$ zebrafish, definitive HSPCs are marked by reduced *cmyb* signal from 33 hours

post fertilization (hpf) and onwards. Furthermore, the number of HSCs (*CD41*^{int}) are severely depleted in *gata2b*$^{-/-}$ embryos at 3 days post-fertilization (dpf) *(Gioacchino et al., 2021).* To test the effect of ectopic *gfi1b* expression on *gata2b*$^{-/-}$ HSPCs, we injected the *gfi1b* construct to WT and *gata2b*$^{-/-}$ *CD41:GFP* embryos at 1-cell stage (Figure 4A). We found that, upon *gfi1b* induction, *cmyb* signals were normalized in *gata2b*$^{-/-}$ HSPCs at 33 hpf (Figure 4B). In addition, the number of *CD41*^{int} HSCs in *gata2b*$^{-/-}$ zebrafish was restored to WT levels at 3 dpf (Figure 4C-E).

These results show that ectopic *gfi1b* expression can rescue the embryonic phenotype of *gata2b*$^{-/-}$ zebrafish and indicate that *Gata2* regulates Gfi1b during EHT.

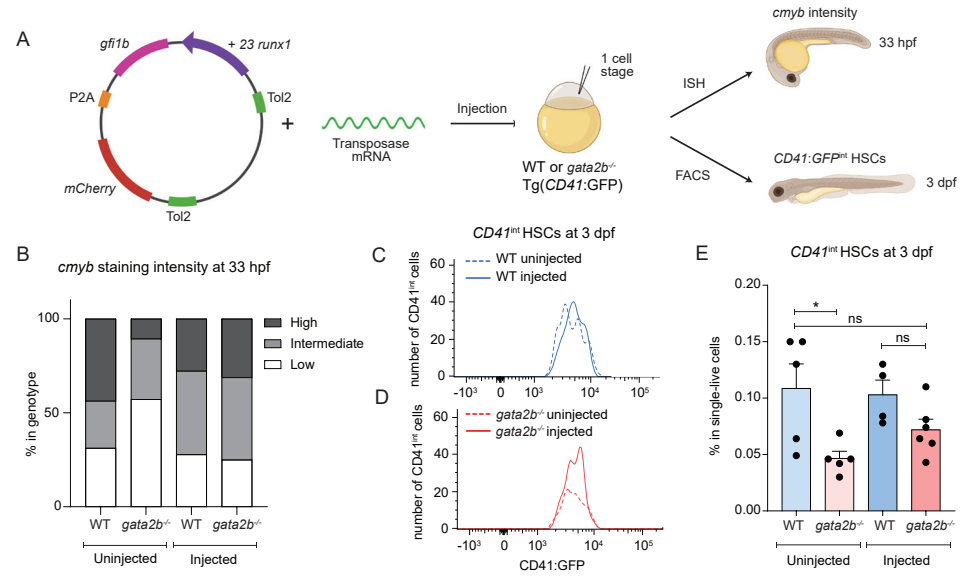


Figure 4. *gfi1b* induction restores the number of embryonic HSCs in *gata2b*$^{-/-}$ zebrafish.
A) Illustration of the experiment and analysis strategy for *gfi1b* induction in zebrafish embryos. Rescue construct containing +23 *runx1* enhancer, *gfi1b* and *mCherry* is injected to the single cells of Tg(CD41:GFP) WT or *gata2b*$^{-/-}$ zebrafish embryos in combination with transposase mRNA. Embryos were analyzed at 33 hpf for *cmyb* signal intensity and at 3 dpf for CD41^{+} HSCs. B) cmyb signal intensity analyzed at 33 hpf by ISH and compared between uninjected (WT N=2, *gata2b*$^{-/-}$ N=16) and injected (WT N=18, *gata2b*$^{-/-}$ N=16) WT and *gata2b*$^{-/-}$ embryos. C) Representative image of the number of CD41^{int} HSCs compared between uninjected and injected groups of WT and D) *gata2b*$^{-/-}$ embryos. E) Proportion of CD41^{int} HSCs compared between uninjected and injected groups of WT and *gata2b*$^{-/-}$ embryos. Dots represent individual samples and each sample contains n=4 pooled embryos. ISH, in situ hybridization; FACS, fluorescent activated cell sorting. ns, not significant.

DISCUSSION

In this study, we aimed to assess the role of *Gata2* on the generation of phenotypic HSCs through EHT. Although HSPCs do not completely switch-off their endothelial gene expression and remain expressing endothelial markers like *CD31* throughout EHT and maturation, moderated repression of endothelial programming is essential for the maturation and thus for establishing the HSC fate *(Oatley et al., 2020; Baron et al., 2018; Zhou et al., 2016; Swiers et al., 2013)*. We showed that *Gata2* haploinsufficiency does not completely abrogate the hematopoietic programming during EHT, but reduces the ability of HSPCs to complete HSC maturation in the AGM.

Gata2$^{+/-}$ HSPCs were predominantly blocked during pro-HSC to pre-HSC maturation and functional HSCs (pre-HSC2/HSCs) were significantly reduced at both E11 and E12 AGMs. We showed that phenotypic *Gata2*$^{+/-}$ HSCs (pre-HSC2/HSCs) not only fail to repress endothelial genes like *Cxcl12* and *Col3a1*, they also incompletely activate hematopoietic genes like *cKit* and *Myb* during AGM maturation. Together this implies that *Gata2* acts as a mediator between endothelial and hematopoietic transcriptional programs during EHT. Previous studies showed that *Gata2*$^{-/-}$ embryos do not form IAHCs in the AGM implying that *Gata2* expression is required for EHT *(de Pater et al., 2013; Tsai et al., 1994).* Our results extend these observations by showing that *Gata2* has multiple roles during EHT and is also a critical regulator of HSPC maturation.

Despite that the number of HSCs were reduced in *Gata2*$^{+/-}$ AGMs and their transcriptome was distinct from WT HSCs, CFU assays showed that their functionality was preserved. These results revise the observations from previous studies where it was shown that colony forming ability of *Gata2*$^{+/-}$ AGM HSPCs is reduced *(de Pater et al., 2013)* and clarify this was due to impaired HSC maturation in *Gata2*$^{+/-}$ AGMs and not due to a reduction in HSC functionality. However, CFU assays are limited when testing repopulating ability and lymphoid potential of HSCs and therefore transplantation studies are needed to elucidate the true potential of embryonic *Gata2*$^{+/-}$ HSCs.

Gfi1b was downregulated and hematopoietic distal enhancers of *Gfi1b* were less active in *Gata2*$^{+/-}$ HSPCs. Because *Gfi1b* expression is required to repress endothelial programming in HSPCs and for the formation of IAHCs during EHT, we propose that *Gata2* activates *Gfi1b* through its +16 and +17 distal enhancer regions to repress the endothelial identity of HSPCs. Restored number of HSCs in *gata2b*$^{-/-}$ zebrafish embryos upon ectopic *gfi1b* expression validated that *Gata2* activates *Gfi1b* and this regulatory mechanism is essential for the generation of embryonic HSCs. Although previous studies provided in silico and experimental evidence suggesting *Gata2* and *Gfi1b* regulate each other *(Moignard et al., 2013),* to our knowledge this study is the first experimental proof pointing out the phenotypic consequences of the disruption of this regulatory mechanism during EHT.

Importantly, we found an increased activity of *Notch signaling* in *Gata2*$^{+/-}$ HSPCs. Previous studies in mice and zebrafish showed that *Notch signaling* is an upstream activator of *Gata2 (Dobrzycki et al., 2020; Guiu et al., 2012).* Because the expression of *Gata2* is reduced in *Gata2*$^{+/-}$ HSPCs, increased *Notch signaling* could be activated as a compensatory mechanism in these cells. Therefore, our results suggest a possible feedback regulation of *Notch signaling* through *Gata2* expression. Because it was shown that downregulation of *Notch signaling* is required during HSPC maturation (*Souilhol et al., 2016*), we cannot exclude the possibility of impaired HSPC maturation in *Gata2*$^{+/-}$ AGMs is influenced by the high *Notch* activity. Although inducing *gfi1b* can sufficiently rescue the embryonic phenotype in *gata2b*$^{-/-}$ embryos, the contribution of *Notch signaling* herein requires further investigation.

Finally, we showed that many gene-sets related to *Cell cycle*, *Proliferation* and *DNA repair* were downregulated in *Gata2*$^{+/-}$ HSPCs throughout all maturation stages (Figure S3A-C). Although decreased proliferative signatures are hallmarks of impaired HSPC maturation, whether this has an influence on genome stability of *Gata2*$^{+/-}$ HSPCs remains to be explored. Because HSCs generated through EHT during embryonic stages are the source of the adult HSC pool, further studies are needed to understand the influence of prenatal *GATA2* haploinsufficiency to HSC fitness after birth and throughout adulthood.

REFERENCES

Baron, C. S., Kester, L., Klaus, A., Boisset, J. C., Thambyrajah, R., Yvernogeau, L., Kouskoff, V., Lacaud, G., van Oudenaarden, A., & Robin, C. (2018). Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. *Nat Commun*, *9*(1), 2517.

Bertrand, J. Y., Chi, N. C., Santoso, B., Teng, S., Stainier, D. Y., & Traver, D. (2010). Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature*, *464*(7285), 108-111.

Boisset, J. C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., & Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*, *464*(7285), 116-120.

Chocron, S., Verhoeven, M. C., Rentzsch, F., Hammerschmidt, M., & Bakkers, J. (2007). Zebrafish Bmp4 regulates left-right asymmetry at two distinct developmental time points. *Dev Biol*, *305*(2), 577-588.

Ciau-Uitz, A., & Patient, R. (2019). Gene Regulatory Networks Governing the Generation and Regeneration of Blood. *J Comput Biol*, *26*(7), 719-725.

de Bruijn, M. F., Ma, X., Robin, C., Ottersbach, K., Sanchez, M. J., & Dzierzak, E. (2002). Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta. *Immunity*, *16*(5), 673-683.

de Pater, E., Kaimakis, P., Vink, C. S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P. S., Camper, S. A., Speck, N., & Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. *J Exp Med*, *210*(13), 2843-2850.

Dickinson, R. E., Griffin, H., Bigley, V., Reynard, L. N., Hussain, R., Haniffa, M., Lakey, J. H., Rahman, T., Wang, X. N., McGovern, N., Pagan, S., Cookson, S., McDonald, D., Chua, I., Wallis, J., Cant, A., Wright, M., Keavney, B., Chinnery, P. F., Loughlin, J., Hambleton, S., Santibanez-Koref, M., & Collin, M. (2011). Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood*, *118*(10), 2656-2658.

Dieterlen-Lievre, F. (1975). On the origin of haemopoietic stem cells in the avian embryo: an experimental approach. *J Embryol Exp Morphol*, *33*(3), 607-619.

Dobrzycki, T., Mahony, C. B., Krecsmarik, M., Koyunlar, C., Rispoli, R., Peulen-Zink, J., Gussinklo, K., Fedlaoui, B., de Pater, E., Patient, R., & Monteiro, R. (2020). Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis. *Commun Biol*, *3*(1), 71.

Donadieu, J., Lamant, M., Fieschi, C., de Fontbrune, F. S., Caye, A., Ouachee, M., Beaupain, B., Bustamante, J., Poirel, H. A., Isidor, B., Van Den Neste, E., Neel, A., Nimubona, S., Toutain, F., Barlogis, V., Schleinitz, N., Leblanc, T., Rohrlich, P., Suarez, F., Ranta, D., Chahla, W. A., Bruno, B., Terriou, L., Francois, S., Lioure, B., Ahle, G., Bachelerie, F., Preudhomme, C., Delabesse, E., Cave, H., Bellanné-Chantelot, C., Pasquet, M., & French, G. s. g. (2018). Natural history of GATA2 deficiency in a survey of 79 French and Belgian patients. *Haematologica*, *103*(8), 1278-1287.

Eich, C., Arlt, J., Vink, C. S., Solaimani Kartalaei, P., Kaimakis, P., Mariani, S. A., van der Linden, R., van Cappellen, W. A., & Dzierzak, E. (2018). In vivo single cell analysis reveals Gata2 dynamics in cells transitioning to hematopoietic fate. *J Exp Med*, *215*(1), 233-248.

Gao, X., Johnson, K. D., Chang, Y. I., Boyer, M. E., Dewey, C. N., Zhang, J., & Bresnick, E. H. (2013). Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *J Exp Med*, *210*(13), 2833-2842.

Gioacchino, E., Koyunlar, C., Zink, J., de Looper, H., de Jong, M., Dobrzycki, T., Mahony, C. B., Hoogenboezem, R., Bosch, D., van Strien, P. M. H., van Royen, M. E., French, P. J., Bindels, E., Gussinklo, K. J., Monteiro, R., Touw, I. P., & de Pater, E. (2021). Essential role for Gata2 in modulating lineage output from hematopoietic stem cells in zebrafish. *Blood Adv*, *5*(13), 2687-2700.

Guiu, J., Shimizu, R., D'Altri, T., Fraser, S. T., Hatakeyama, J., Bresnick, E. H., Kageyama, R., Dzierzak, E., Yamamoto, M., Espinosa, L., & Bigas, A. (2013). Hes repressors are essential regulators of hematopoietic stem cell development downstream of Notch signaling. *J Exp Med*, *210*(1), 71-84.

Hahn, C. N., Chong, C. E., Carmichael, C. L., Wilkins, E. J., Brautigan, P. J., Li, X. C., Babic, M., Lin, M., Carmagnac, A., Lee, Y. K., Kok, C. H., Gagliardi, L., Friend, K. L., Ekert, P. G., Butcher, C. M., Brown, A. L., Lewis, I. D., To, L. B., Timms, A. E., Storek, J., Moore, S., Altree, M., Escher, R., Bardy, P. G., Suthers, G. K., D'Andrea, R. J., Horwitz, M. S., & Scott, H. S. (2011). Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*, *43*(10), 1012-1017.

Hsu, A. P., Sampaio, E. P., Khan, J., Calvo, K. R., Lemieux, J. E., Patel, S. Y., Frucht, D. M., Vinh, D. C., Auth, R. D., Freeman, A. F., Olivier, K. N., Uzel, G., Zerbe, C. S., Spalding, C., Pittaluga, S., Raffeld, M., Kuhns, D. B., Ding, L., Paulson, M. L., Marciano, B. E., Gea-Banacloche, J. C., Orange, J. S., Cuellar-Rodriguez, J., Hickstein, D. D., & Holland, S. M. (2011). Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*, *118*(10), 2653-2655.

Ivanovs, A., Rybtsov, S., Ng, E. S., Stanley, E. G., Elefanty, A. G., & Medvinsky, A. (2017). Human haematopoietic stem cell development: from the embryo to the dish. *Development*, *144*(13), 2323-2337.

Kissa, K., & Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*, *464*(7285), 112-115.

Lancrin, C., Mazan, M., Stefanska, M., Patel, R., Lichtinger, M., Costa, G., Vargel, O., Wilson, N. K., Möröy, T., Bonifer, C., Göttgens, B., Kouskoff, V., & Lacaud, G. (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood*, *120*(2), 314-322.

Ling, K. W., Ottersbach, K., van Hamburg, J. P., Oziemlak, A., Tsai, F. Y., Orkin, S. H., Ploemacher, R., Hendriks, R. W., & Dzierzak, E. (2004). GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med*, *200*(7), 871-882.

Ma, D., Zhang, J., Lin, H. F., Italiano, J., & Handin, R. I. (2011). The identification and characterization of zebrafish hematopoietic stem cells. *Blood*, *118*(2), 289-297.

Medvinsky, A., & Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*, *86*(6), 897-906.

Moignard, V., Macaulay, I. C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F. J., Kinston, S., Joshi, A., Hannah, R., Theis, F. J., Jacobsen, S. E., de Bruijn, M. F., & Göttgens, B. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat Cell Biol*, *15*(4), 363-372.

Müller, A. M., Medvinsky, A., Strouboulis, J., Grosveld, F., & Dzierzak, E. (1994). Development of hematopoietic stem cell activity in the mouse embryo. *Immunity*, *1*(4), 291-301.

North, T. E., de Bruijn, M. F., Stacy, T., Talebian, L., Lind, E., Robin, C., Binder, M., Dzierzak, E., & Speck, N. A. (2002). Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity*, *16*(5), 661-672.

Oatley, M., Bölükbası Ö, V., Svensson, V., Shvartsman, M., Ganter, K., Zirngibl, K., Pavlovich, P. V., Milchevskaya, V., Foteva, V., Natarajan, K. N., Baying, B., Benes, V., Patil, K. R., Teichmann, S. A., & Lancrin, C. (2020). Single-cell transcriptomics identifies CD44 as a marker and regulator of endothelial to haematopoietic transition. *Nat Commun*, *11*(1), 586.

Ostergaard, P., Simpson, M. A., Connell, F. C., Steward, C. G., Brice, G., Woollard, W. J., Dafou, D., Kilo, T., Smithson, S., Lunt, P., Murday, V. A., Hodgson, S., Keenan, R., Pilz, D. T., Martinez-Corral, I., Makinen, T., Mortimer, P. S., Jeffery, S., Trembath, R. C., & Mansour, S. (2011). Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet*, *43*(10), 929-931.

Ottema, S., Mulet-Lazaro, R., Erpelinck-Verschueren, C., van Herk, S., Havermans, M., Arricibita Varea, A., Vermeulen, M., Beverloo, H. B., Gröschel, S., Haferlach, T., Haferlach, C., B, J. W., Bindels, E., Smeenk, L., & Delwel, R. (2021). The leukemic oncogene EVI1 hijacks a MYC super-enhancer by CTCF-facilitated loops. *Nat Commun*, *12*(1), 5679.

Robert-Moreno, A., Espinosa, L., de la Pompa, J. L., & Bigas, A. (2005). RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells. *Development*, *132*(5), 1117-1126.

Rodrigues, N. P., Janzen, V., Forkert, R., Dombkowski, D. M., Boyd, A. S., Orkin, S. H., Enver, T., Vyas, P., & Scadden, D. T. (2005). Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood*, *106*(2), 477-484.

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O., & Medvinsky, A. (2014). Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43(-) embryonic precursor. *Stem Cell Reports*, *3*(3), 489-501.

Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., Frampton, J., Zhao, S., & Medvinsky, A. (2011). Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J Exp Med*, *208*(6), 1305-1315.

Solaimani Kartalaei, P., Yamada-Inagawa, T., Vink, C. S., de Pater, E., van der Linden, R., Marks-Bluth, J., van der Sloot, A., van den Hout, M., Yokomizo, T., van Schaick-Solernó, M. L., Delwel, R., Pimanda, J. E., van, I. W. F., & Dzierzak, E. (2015). Whole-transcriptome analysis of endothelial to hematopoietic stem cell transition reveals a requirement for Gpr56 in HSC generation. *J Exp Med*, *212*(1), 93-106.

Souilhol, C., Lendinez, J. G., Rybtsov, S., Murphy, F., Wilson, H., Hills, D., Batsivari, A., Binagui-Casas, A., McGarvey, A. C., MacDonald, H. R., Kageyama, R., Siebel, C., Zhao, S., & Medvinsky, A. (2016). Developing HSCs become Notch independent by the end of maturation in the AGM region. *Blood*, *128*(12), 1567-1577.

Spinner, M. A., Sanchez, L. A., Hsu, A. P., Shaw, P. A., Zerbe, C. S., Calvo, K. R., Arthur, D. C., Gu, W., Gould, C. M., Brewer, C. C., Cowen, E. W., Freeman, A. F., Olivier, K. N., Uzel, G., Zelazny, A. M., Daub, J. R., Spalding, C. D., Claypool, R. J., Giri, N. K., Alter, B. P., Mace, E. M., Orange, J. S., Cuellar-Rodriguez, J., Hickstein, D. D., & Holland, S. M. (2014). GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood*, *123*(6), 809-821.

Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K. D., Yoshimoto, M., Yoder, M. C., Frampton, J., Schroeder, T., Enver, T., Göttgens, B., & de Bruijn, M. (2013). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat Commun*, *4*, 2924.

Taoudi, S., Gonneau, C., Moore, K., Sheridan, J. M., Blackburn, C. C., Taylor, E., & Medvinsky, A. (2008). Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-cadherin+CD45+ pre-definitive HSCs. *Cell Stem Cell*, *3*(1), 99-108.

Thambyrajah, R., Mazan, M., Patel, R., Moignard, V., Stefanska, M., Marinopoulou, E., Li, Y., Lancrin, C., Clapes, T., Möröy, T., Robin, C., Miller, C., Cowley, S., Göttgens, B., Kouskoff, V., & Lacaud, G. (2016). GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat Cell Biol*, *18*(1), 21-32.

Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W., & Orkin, S. H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, *371*(6494), 221-226.

Vink, C. S., Calero-Nieto, F. J., Wang, X., Maglitto, A., Mariani, S. A., Jawaid, W., Göttgens, B., & Dzierzak, E. (2020). Iterative Single-Cell Analyses Define the Transcriptome of the First Functional Hematopoietic Stem Cells. *Cell Rep*, *31*(6), 107627.

Yokomizo, T., & Dzierzak, E. (2010). Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos. *Development*, *137*(21), 3651-3661.

Yokomizo, T., Yamada-Inagawa, T., Yzaguirre, A. D., Chen, M. J., Speck, N. A., & Dzierzak, E. (2012). Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos. *Nat Protoc*, *7*(3), 421-431.

Zape, J. P., Lizama, C. O., Cautivo, K. M., & Zovein, A. C. (2017). Cell cycle dynamics and complement expression distinguishes mature haematopoietic subsets arising from hemogenic endothelium. *Cell Cycle*, *16*(19), 1835-1847.

Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., Li, Z., Ni, Y., Mu, X., Wen, L., Cheng, T., Lan, Y., Yuan, W., Tang, F., & Liu, B. (2016). Tracing haematopoietic stem cell formation at single-cell resolution. *Nature*, *533*(7604), 487-492.

Zovein, A. C., Hofmann, J. J., Lynch, M., French, W. J., Turlo, K. A., Yang, Y., Becker, M. S., Zanetta, L., Dejana, E., Gasson, J. C., Tallquist, M. D., & Iruela-Arispe, M. L. (2008). Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell*, *3*(6), 625-636.

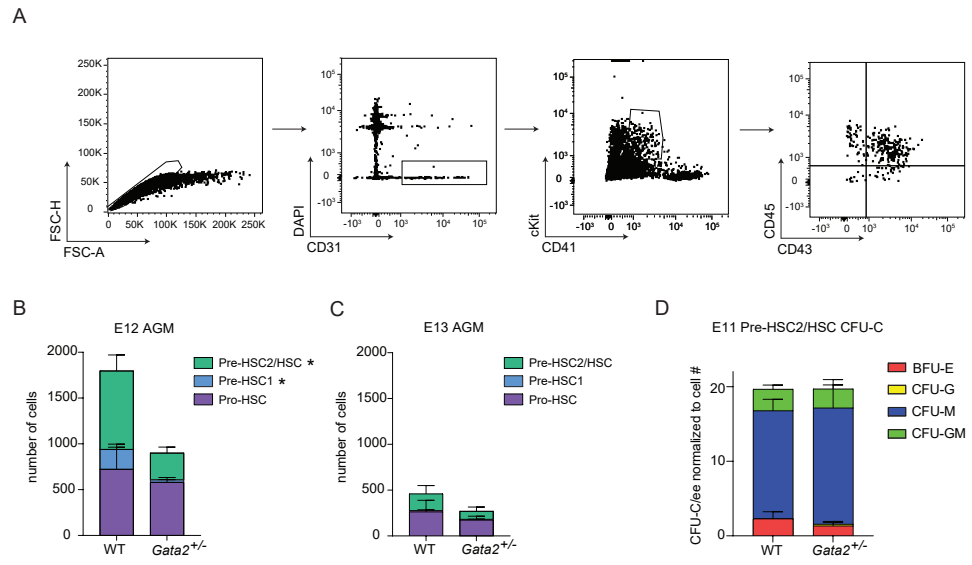
SUPPLEMENTARY INFORMATION

Table S1. Primers used for Gibson assembly reaction.

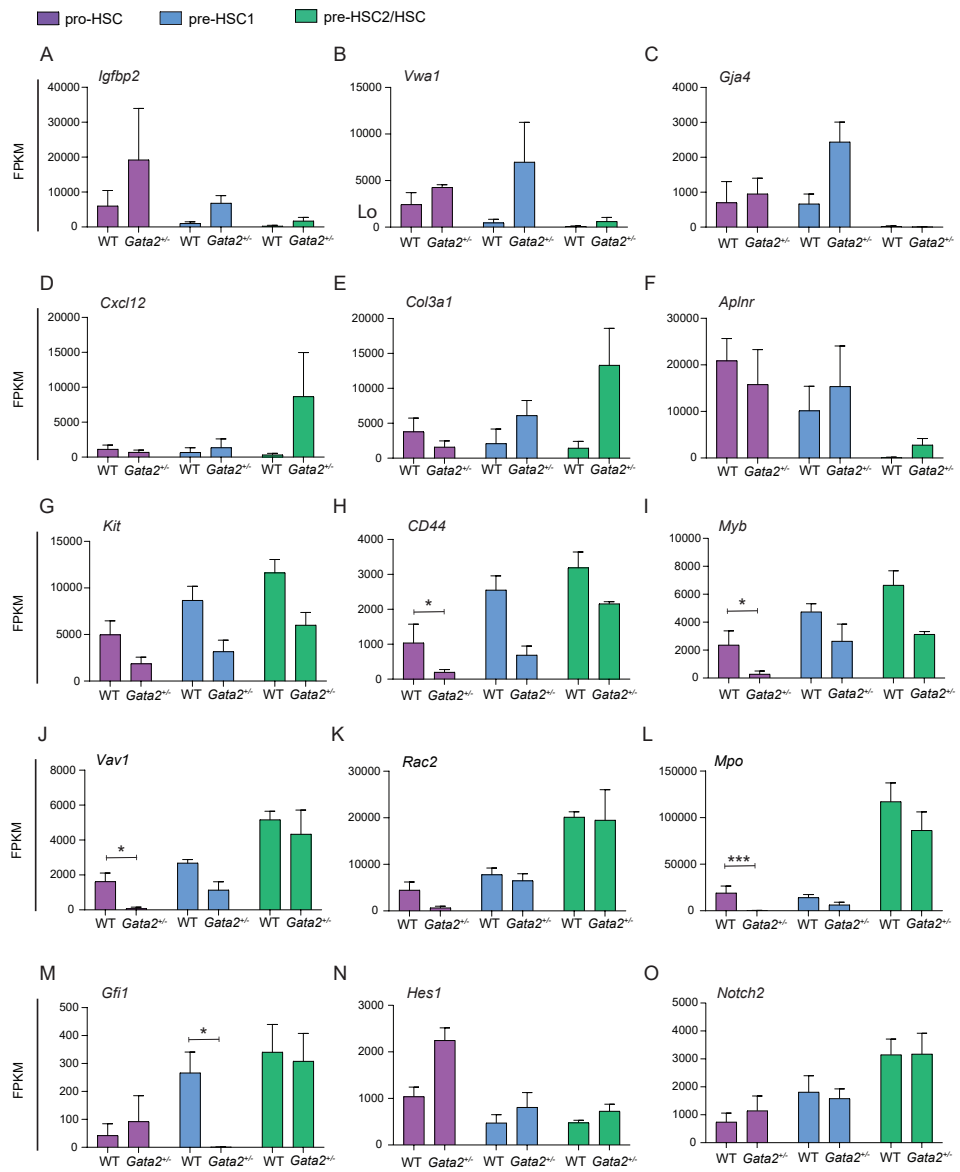
Fragment name/primer	Sequence (5'-3')
runx1 +23 enhancer Fw	agatctgatctagaggatcataatgGGGGTGGGAGGTGTAAGTTC
runx1 +23 enhancer Rv	acgaccgtggcatGGTGGTCTAGGGGATGTC
gfi1b Fw	cccctagaccaccATGCCACGGTCGTTTCTG
gfi1b Rv	tagtagctccggaaccTTTAAGGCTGTGCTGGCTC
mCherry Fw	gcacagccttaaaggttccggagctactaacttcagcctgctgaagcaggct ggagacgtggaggagaaccctggacctGTGAGCAAGGGCGAGGAG
mCherry Rv	cacttgggcccggctcgagcaggggCAGGGGCCCCCTGAACCT

Table S2. CD31^{+}cKit^{+} cell numbers in E11 WT and *Gata2*$^{+/-}$ AGMs.

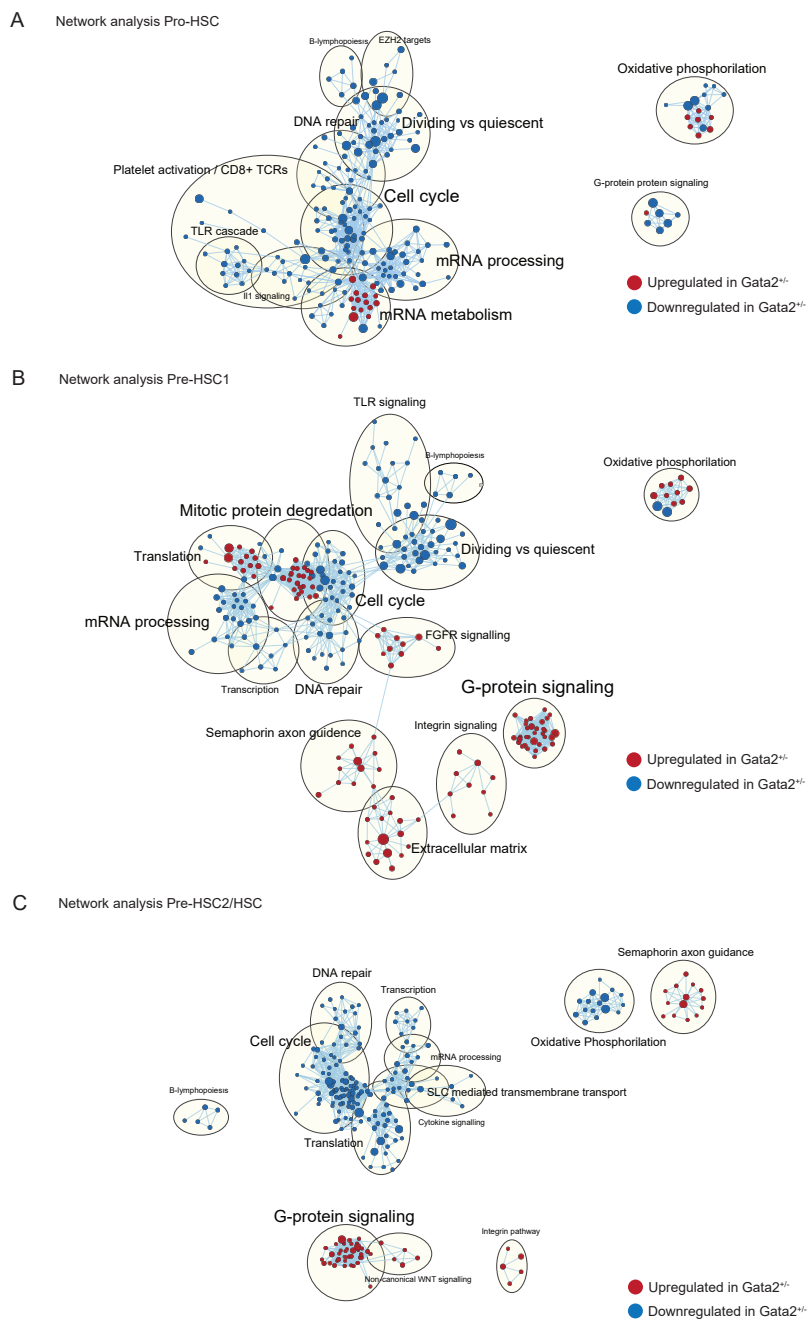
	sIAHC	bIAHC	AL
WT-1	187	301	129
WT-2	162	269	94
WT-3	130	140	39
Gata2$^{+/-}$ -1	31	85	13
Gata2$^{+/-}$ -2	67	88	33
Gata2$^{+/-}$ -3	69	77	57
Gata2$^{+/-}$ -4	59	138	27



Supplementary figure 1. HSPC maturation is not delayed but impaired in *Gata2*$^{+/-}$ AGMs.
A) Gating strategy to detect HSPC maturation in AGM. B) Quantification of the number of pro-HSC, pre-HSC1 and pre-HSC2 populations in E12 WT or *Gata2*$^{+/-}$ AGMs. C) Quantification of the number of pro-HSC, pre-HSC1 and pre-HSC2 populations in E13 WT or *Gata2*$^{+/-}$ AGMs. D) Quantification of BFU-E, CFU-G, CFU-M and CFU-GM colonies after 11 days of CFU culture of E11 WT and *Gata2*$^{+/-}$ pre-HSC2/HSCs. Colony numbers were normalized to the number of cells plated per dish.



Supplementary figure 2. *Gata2* haploinsufficiency impairs both endothelial and hematopoietic transcriptional programming throughout HSPC maturation.
A) Comparison of FPKM values of *Vav1*, B) *Rac2*, C) *Mpo,* D) *Igfbp2,* E) *Vwa1,* F) *Gja4,* G) *Cxcl12*, H) *Col3a1*, I) *Aplnr,* J) *Gfi1*, K) *Hes1,* L) *Notch2,* M) *Kit,* N) *Cd44* and O) *Myb* between WT or *Gata2*$^{+/-}$ pro-HSC, pre-HSC1 and pre-HSC2/HSCs. Color code for HSPC maturation steps: pro-HSCs, purple; pre-HSC1, blue; pre-HSC2, green. *P < 0.05, **P < 0.01, ***P < 0.001



Supplementary figure 3. Proliferation is abrogated in *Gata2*$^{+/-}$ HSPCs throughout maturation.
A) GSEA network analysis of pro-HSC, B) pre-HSC1 and C) pre-HSC2 populations. WT and *Gata2*$^{+/-}$ HSPCs were compared. Red dots are showing the upregulated and blue dots are showing the downregulated gene sets in *Gata2*$^{+/-}$ HSPCs compared to WT.

3

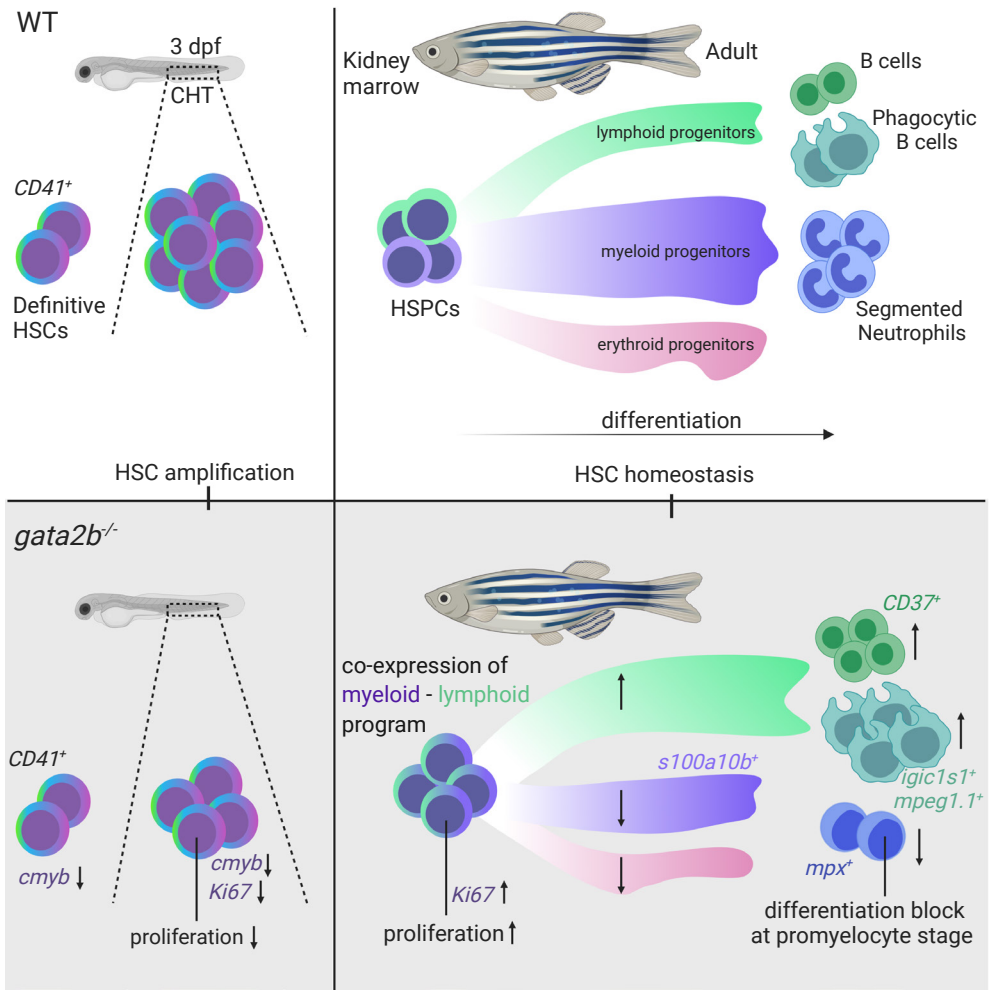
Essential role for Gata2 in modulating lineage output from hematopoietic stem cells in zebrafish

Cansu Koyunlar[1,7], Emanuele Gioacchino[1,7], Joke Zink[1], Hans de Looper[1,2], Madelon de Jong[1], Tomasz Dobrzycki[3], Christopher B Mahony[4], Remco Hoogenboezem[1], Dennis Bosch[1], Paulina M H van Strien[1], Martin E van Royen[5], Pim J French[6], Eric Bindels[1], Kirsten J Gussinklo[1], Rui Monteiro[4], Ivo P Touw[1] and Emma de Pater[1,2]*.

[1]Department of Hematology, Erasmus MC, Rotterdam, The Netherlands
[2]Cancer Genome Editing Canter of Erasmus MC, Rotterdam, The Netherlands.
[3]Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK.
[4]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham B152TT, UK.
[5]Department of Pathology, Cancer Treatment Screening Facility, Erasmus MC Optical Imaging Centre, Erasmus MC, Rotterdam, The Netherlands
[6]Department of Neurology, Cancer Treatment Screening Facility, Erasmus MC, Rotterdam, The Netherlands.
[7]Authors contributed equally

* Corresponding author

Blood Adv (2021) 5 (13): 2687–2700.



ABSTRACT

The differentiation of hematopoietic stem cells is tightly controlled to ensure a proper balance between myeloid and lymphoid cell output. GATA2 is a pivotal hematopoietic transcription factor required for generation and maintenance of hematopoietic stem cells. GATA2 is expressed throughout development but due to early embryonic lethality in mouse, its role during adult hematopoiesis is incompletely understood. Zebrafish contains two orthologues of GATA2; Gata2a and Gata2b that are expressed in different cell types. We show that the mammalian functions of GATA2 are split between these orthologues. Gata2b deficient zebrafish have a reduction in embryonic definitive hematopoietic stem and progenitor cell (HSPC) numbers, but are viable. This allows us to uniquely study the role of GATA2 in adult hematopoiesis. *gata2b* mutants have impaired myeloid lineage differentiation. Interestingly, this defect arises not in granulocyte-monocyte progenitors, but already in HSPCs. Gata2b deficient HSPCs showed impaired progression of the myeloid transcriptional program, concomitant with increased co-expression of lymphoid genes. This results in a decrease in myeloid programmed progenitors and a relative increase in lymphoid programmed progenitors. This shift in the lineage output could function as an escape mechanism to avoid a block in lineage differentiation. These studies help to deconstruct the functions of GATA2 during hematopoiesis and show that lineage differentiation flows towards a lymphoid lineage in the absence of Gata2b.

Highlights

1) Gata2b is required for embryonic HSPC expansion, but not HSPC generation in zebrafish
2) Gata2b plays an instructive role in the lineage output of HSPCs in zebrafish

INTRODUCTION

Hematopoietic stem cells (HSCs) have the capacity to self-renew and to generate all lineages of the hematopoietic system[1]. The HSC pool is a heterogeneous population of cells that are tightly controlled by cell-intrinsic and -extrinsic cues to maintain a balance between myeloid and lymphoid cell commitment[2-5]. It is currently under debate whether HSCs can flow between myeloid and lymphoid lineage commitment or whether the HSC pool consists of separate lymphoid biased and myeloid biased HSCs[6].

The transcription factor GATA2 has a key role in blood cell formation during mammalian embryonic development. GATA2 expression is tightly regulated during distinct stages of hematopoietic development and plays crucial roles in the specification of hemogenic endothelium (HE) and the generation and maintenance of HSCs[7-11]. A role for this transcription factor in myeloid/lymphoid commitment is supported by findings of reduced and impaired granulocyte-macrophage progenitors in *Gata2*$^{+/-}$ mice[12-14]. Conversely, retroviral mediated overexpression of *Gata2* results in enhanced self-renewal of the myeloid progenitors and a block in lymphoid differentiation[15]. Homozygous germline deletion of *Gata2* in mice results in embryonic lethality at E10, just before the generation of the first HSCs[16].

Zebrafish is an ideal *in vivo* model to study the function of GATA2 in hematopoiesis. Embryonic hematopoietic development in zebrafish is conserved with that of other vertebrates, including mammals. Like in mice, the first HSCs are generated in the dorsal aorta from hemogenic endothelial (HE) cells and are subsequently amplified in the fetal liver equivalent, the caudal hematopoietic tissue (CHT)[17-22]. The HSCs then populate the kidney marrow which is the site of adult hematopoiesis in zebrafish. In this organ all hematopoietic lineages are present[23] and hematopoietic cells morphologically resemble the corresponding human cells.

Zebrafish have two orthologues of GATA2; i.e., Gata2a and Gata2b. Previous studies have shown that *gata2b* is prominently expressed in HSPCs, whereas *gata2a* is mainly expressed in the vasculature, including the HE regulated by the conserved +9.5 enhancer previously identified in mice[24,25]. Knockdown of *gata2b* severely reduces definitive hematopoiesis during embryonic stages. Lineage tracing revealed that all definitive hematopoietic cells are derived from *gata2b* expressing cells[24], indicating that Gata2b is the predominant GATA2 orthologue required for the maintenance of hematopoietic stem cells.

In the present study, we show that Gata2b is not required for HE specification but regulates embryonic definitive HSPC expansion in the CHT. This allowed us to investigate the function of Gata2b in adult hematopoiesis and here, we demonstrate that Gata2b is necessary for balanced myeloid and lymphoid output during adulthood. Single cell transcriptome analysis revealed that Gata2b deficient HSPCs initiate an impaired myeloid gene expression program. As a result differentiation is not halted, but diverges into a lymphoid program, indicated by co-expression of lymphoid and myeloid genes within single HSPCs.

MATERIALS AND METHODS

Generation of *gata2b* mutant zebrafish

gata2b mutant zebrafish were generated using CRISPR/Cas9 targeting of exon 3. sgRNAs were designed using CHOPCHOP software and prepared according to Gagnon et al.[26] with minor adjustments.

qRT-PCR analysis

Total RNA was isolated from 6 pooled zebrafish embryos per genotype (n=6) using TRIzol Reagent (Life Technologies) and cDNA was synthesized using SuperScript III Reverse Transcriptase kit (Invitrogen). *gata2a* (FWD primer: 5'-CAAACTCCACAACGTCAACAG-3', REV primer: 5'-CCCTCACCAGATCGTTTACTC-3') and *gata2b* (FWD primer: 5'-TACACAATGTGAATCGCCCA-3', REV primer: 5'-GAAGGAGGATGGTTTGTCGT-3') expression levels were normalized to *elfa* (FWD primer: 5'-CCGCTAGCATTACCCTCC-3', REV primer: 5'-CTTCTCAGGCTGACTGTG-3') expression .

In situ hybridization (ISH) and analysis

0.003% 1-phenyl-2-thiourea (PTU) treated embryos were fixed O/N with 4% PFA in PBS containing 3% sucrose at appropriate stages and subsequently transferred to MeOH. KM smears were fixed in MeOH. ISH on embryos has been performed as previously described[27]. The *cmyb* and *runx1* probes were a kind gift from Roger Patient and quantified as described previously[28]. ISH on KM smears was performed as follows; DIG-11-UTP labelled *s100a10b* probe was incubated o/n at 68°C. Slides were blocked at RT in MABT (NaCl, Maleic Acid, 1% Tween 20)2% BSA and Sheep Serum for minimum 3 h and αDIG antibody was incubated o/n at 4°C. Staining was developed in Tris pH 9.5, $MgCl_2$, NaCl, Tween 20 with 5% PVA, NBT/BCIP at RT for two days. Cells were counterstained with Nuclear Fast Red (Sigma Aldrich) and imaged using a Leica microscope (63x magnification).

s100a10b probe synthesis

s100a10b was amplified from cDNA of adult kidney marrow (FWD primer: 5'-GAG AGC AAT GGA GAC CCT GA-3', REV primer: 5'-ACT TCT TGG CTG CTG CTT TC-3') and cloned into pCRII-TOPO. Plasmid was linearized with HindIII and antisense probe transcribed with the DIG labelling kit (Sigma-Alderich). Sense probe was used as negative control.

Transgenic lines, confocal imaging and adult KM FACS analysis

Embryos were anesthetized using tricaine (3-amino benzoic acidethylester) 160mg/L and selected for reporter expression. *Tg(fli:eGFP)*[29] and *Tg(CD41:GFP)*[30]*;Tg(flt1:RFP)*[31] embryos were imaged in 0.25% agarose with tricaine and imaged using a Leica SP5 confocal microscope pre-warmed at 28°C. *Tg(mpeg1.1:GFP)*[32], *Tg(mpx:GFP)*[33], *Tg(lck:GFP)*[34] embryos were placed

in a 96 well plate (ZFplate, Hashimoto Electronic Industry Co. Ltd, Japan) and imaged using a spinning disk confocal high-throughput microscope system (Opera Phenix, Perkin Elmer) equipped with a dry 10x objective (NA 0.3). B-cell populations were analysed using *Tg(IgM:GFP)*[35] zebrafish. Adult zebrafish were euthanized, KM was isolated and dissociated by pipetting in PBS/10% FCS. 7-AAD (7-amino-actinomycin D) 0.5mg/L (BDbiosciences)or DAPI 1mg/L were used for live/dead discrimination. For embryonic proliferation assay; 25 embryos per genotype were pooled in pre-warmed PBS/10% FCS, single cell suspension was prepared by adding 1% from each collagenase (I, II and IV)(Sigma) and incubating for 45 minutes at 37°C. Proliferation was assessed after 4% PFA fixation and α-Ki67 staining for both embryonic and adult stages. The analysis was performed using FACSAriaIII (BD).

Single cell RNA sequencing

70.000 single viable cells were sorted from 2 pooled KM of female *Tg(CD41:GFP)* zebrafish and supplemented with 114-1607 CD41:GFP^{low} expressing cells. cDNA was prepared using the manufacturers protocol (10x Chromium V2) and sequenced on a Novaseq 6000 instrument (Illumina). Two WT replicates and two *gata2b*$^{-/-}$ replicates were sequenced with the following read depth; WT1: 52384 reads per cell, WT2: 43876 reads per cell, *gata2b* KO1: 53836 reads per cell and *gata2b* KO2: 46761 reads per cell. Data was analyzed using the Seurat R package[36] and detailed description is provided in the supplementary information.

Statistics

All statistical analysis was carried out in GraphPad Prism 5 (GraphPad Software). Normally distributed data were analyzed using One-way ANOVA with Tukey multiple comparison test when comparing three sample sets or a t-test when comparing two sample sets. Data with non-normal distribution were analyzed using a non-parametric Kruskal-Wallis with Dunn correction test.

RESULTS

Generation of a Gata2b deficient zebrafish line

To generate *gata2b* zebrafish mutants, we used CRISPR/Cas9 to target the third exon of the *gata2b* gene (Figure 1A). A 28 bp insertion was introduced, leading to a frameshift truncation from amino acid 185 (Figure 1B-D). qRT-PCR analysis of *gata2b* on pooled WT and *gata2b*$^{-/-}$ embryos at 30 hpf indicated that *gata2b* expression levels, a known transcriptional tartget of Gata2, was significantly reduced in mutant embryos (Figure S1A, B)[25,37]. Hereafter, we refer to this mutant as *gata2b*$^{-/-}$.

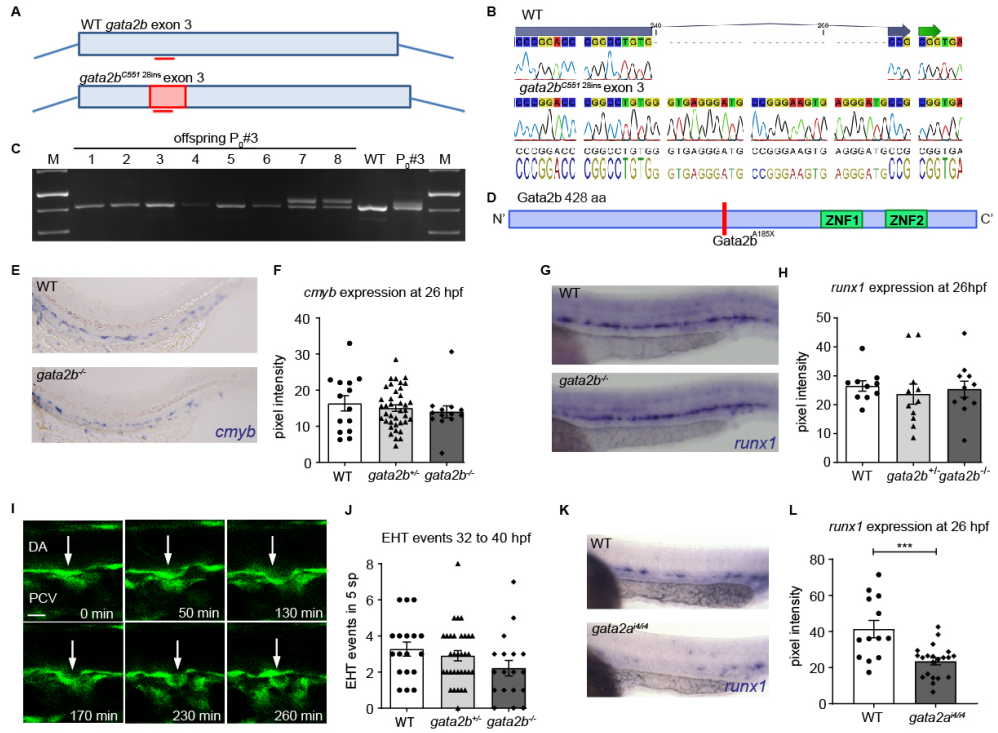


Figure 1. Newly generated Gata2b mutant does not show defects in HSPC generation.
A) Schematic representation of the CRISPR strategy targeting exon 3 of *gata2b* and the 28 nt integration in *gata2b* mutants. B) Alignment of sequencing data of WT *gata2b* exon 3, where the location of the guide is indicated in the blue arrow on top of the sequence and sequencing data from *gata2b*$^{-/-}$ DNA showing a 28 nucleotide integration. C) Gel picture showing genotyping PCR of founder 3 and the F1 with a 28 bp integration in embryo 7 and 8. D) Gata2b mutation leading to a STOP codon abrogating the protein before the DNA and protein binding znic fingers. E) Representative image of *cmyb* expression in WT and *gata2b*$^{-/-}$ embryos at 26 hpf. F) Quantitation of *cmyb* signal intensity relative to background in WT and *gata2b*$^{-/-}$ embryos at 26 hpf. G) Representative image of *runx1* expression in WT and *gata2b*$^{-/-}$ embryos at 26 hpf. H) Quantitation of *runx1* signal intensity relative to background in WT and *gata2b*$^{-/-}$ embryos at 26 hpf. I) Example of EHT event from WT *Tg(fli1a:eGFP)* transgenic zebrafish. Time indicated at the bottom right corner in minutes. Scale bar represents 10 µm. Arrow indicates endothelial cell undergoing hematopoietic transition. J) Quantitation of EHT events between 32-40 hpf in WT, *gata2b*$^{+/-}$ and *gata2b*$^{-/-}$ embryos. K) Representative example of *runx1* expression in WT and *gata2*$^{i4/i4}$ embryos at 26 hpf in the AGM region. L) Quantitation of signal intensity relative to background cells in WT and *gata2a*$^{i4/i4}$ embryos at 26 hpf, where each dot represents one embryo (41.4 ± 4.8 and 23.5 ± 2.0 , n = 13 WT and n = 21 *gata2a*$^{i4/i4}$). *** = $p < 0.001$, error bars represent SEM. Bp = basepair, EHT = endothelial to hematopoietic transition, DA = dorsal aorta, PVC = posterior cardinal vein, sp = somite pair, hpf = hours port fertilization. Error bars represent standard error of mean (SEM).

Gata2b is dispensable for the generation of hematopoietic stem cells from hemogenic endothelium

The first HSCs transdifferentiate from specialized hemogenic endothelial cells in the aorta-gonad-mesonephros (AGM) region, through a highly conserved process, known as endothelial-to-hematopoietic transition (EHT)[18,19,38-40]. In mice, *Gata2* is expressed in the endothelium, including the hemogenic endothelium (HE) of the dorsal aorta[8], and deletion

of *Gata2* results in a reduction in HSC generation[7,8]. *cmyb* and *runx1* are two *bona fide* marker genes for HE at 26 hours post fertilization (hpf) in zebrafish[18,41]. We quantified *cmyb* and *runx1* expression by measuring pixel intensity of the *in situ* hybridization staining compared to background[28]. Expression of *cmyb* (Figure 1E, F) and *runx1* (Figure 1G, H) was indistinguishable between WT and Gata2b deficient embryos at 26 hpf, indicating that specification of hemogenic endothelium occurs normally in the absence of Gata2b.

Next, we examined the ability of HE to undergo EHT using *Tg(fli1a:GFP)* reporter embryos, in which GFP marks all endothelial cells, including HE[29]. Consistent with our initial results, EHT events were not significantly reduced in *gata2b*$^{-/-}$ embryos compared to WT (p = 0.077, n= 18 WT and 18 *gata2b*$^{-/-}$ embryos, Figure 1I, J and Table S1). We conclude that neither HE specification, nor HSPC generation through EHT are impaired in *gata2b*$^{-/-}$ embryos.

Gata2a is required for HE specification

GATA2 is required for the generation of HSCs in mouse[7,16], but Gata2b deficient zebrafish have intact HE and EHT (Figure 1E-J). High maternal expression of *gata2b* has been reported previously[24], and therefore residual Gata2b protein levels could possibly rescue EHT. However, maternal zygotic *gata2b*$^{-/-}$ zebrafish, that do not contain functional maternally provided *gata2b* mRNA, are viable and survive to mendelian ratios (Figure S1C, D), indicating that maternal expression of *gata2b* does not contribute to embryonic hematopoiesis. By contrast, *gata2a* is expressed in hemogenic endothelium and regulates *runx1* expression in HE[25]. Thus, we analysed *runx1* expression in *gata2a* mutants (*gata2a*$^{i4/i4}$, lacking a conserved endothelial enhancer)[25]. *runx1* expression at 26 hpf was reduced in *gata2a*$^{i4/i4}$ embryos compared to WT embryos (Figure 1K, L). This confirmed that endothelial expression of *gata2a,* but not *gata2b,* is required for the specification of hemogenic endothelium and that the different functions of mammalian GATA2 are separated between Gata2a and Gata2b in zebrafish.

Gata2b is required for the expansion of definitive HSPCs during the CHT amplification phase

The CHT is temporally and spatially analogous to mouse fetal liver where HSPCs undergo amplification[22]. We investigated whether the loss of Gata2b affects the number of definitive HSPCs in the CHT between 2 and 3 days post fertilization (dpf). From 44 hpf onward, definitive HSPCs are marked by co-expression of the *Tg*(*CD41:GFP*)[30] marker and the arterial *Tg(flt:RFP*)[31] marker as definitive HSPCs are derived from arteries[42]. CD41:GFP^{+}Flt:RFP^{+} cell numbers were similar in WT and *gata2b*$^{-/-}$ embryos at 52-54 hpf (Figure 2A, B and Table S1) and 56-58 hpf (Figure 2C and Table S1). However, at 76 hpf, CD41:GFP^{+}Flt:RFP^{+} cells were significantly reduced in the CHT in *gata2b*$^{-/-}$ embryos compared to WT (Figure 2D and Table S1).

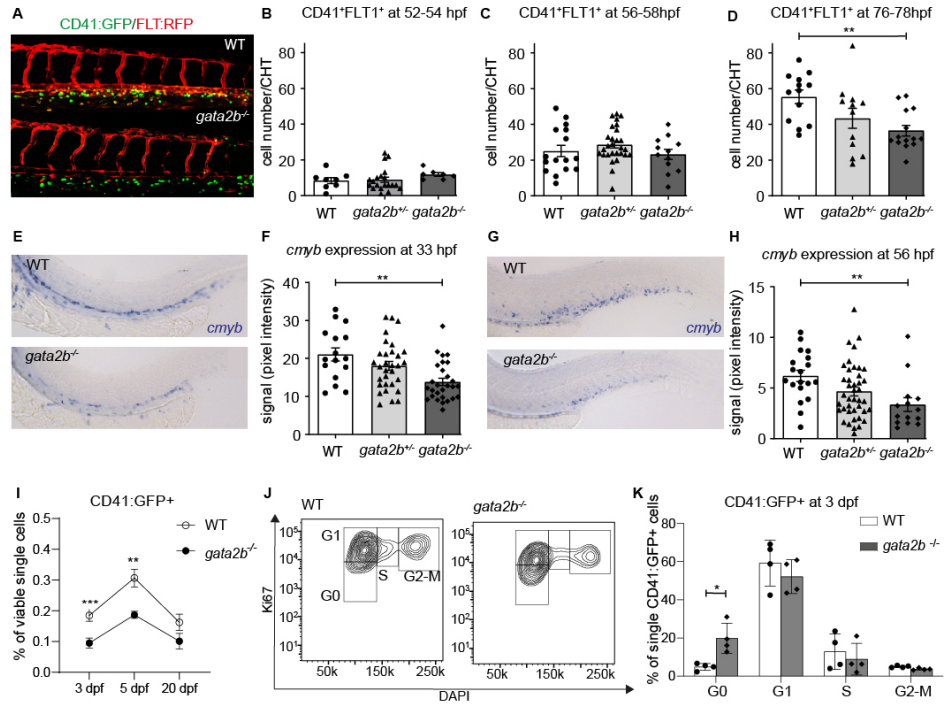


Figure 2. HSPC numbers are reduced in the CHT after 3 dpf in *gata2b*$^{-/-}$ embryos.
A) Example of *Tg(CD41:GFP); Tg(Flt:RFP)* expression in the CHT of WT and *gata2b*$^{-/-}$ embryos at 76 hpf. B-D) Quantitation of GFP^{+}RFP^{+} cells in WT, *gata2b*$^{+/-}$ and *gata2b*$^{-/-}$ embryos at different time points where each dot represents one embryo. B) 52-54 hpf (8.50 ± 1.7 vs 11.86 ± 1.0), C) 56-58 hpf (25.19 ± 3.2 vs 23.25 ± 2.9), and D) 76-78 hpf (55.46 ± 3.8 vs 36.50 ± 3.0; P < 0.001). E) Representative example of *cmyb* expression in WT and *gata2b*$^{-/-}$ embryos at 33 hpf and F) Quantitation of signal intensity relative to background at 33 hpf where each dot represents one embryo. G) Representative example of *cmyb* expression in WT and *gata2b*$^{-/-}$ embryos at 56 hpf and H) Quantitation of signal intensity relative to background at 56 hpf where each dot represents one embryo. I) Quantitation of CD41:GFP^{+} cell percentages by flow cytometry at 3 dpf, 5 dpf and 20 dpf. J) Cell cycle analysis by flow cytometry of Ki67 and DAPI staining of CD41:GFP^{+} cells in WT and *gata2b*$^{-/-}$ embryos at 75 hpf. K) Bar graph representing the quantitation of cell cycle of CD41:GFP^{+} cells in WT and *gata2b*$^{-/-}$ embryos at 75 hpf. Bars represent mean ± SEM with each dot indicating one pooled sample. hpf = hours post fertilization, dpf = days post fertilization, CHT = caudal hematopoietic tissue. * = p < 0.05, ** = p < 0.01. Error bars represent SEM.

The number of definitive HSPCs expands rapidly from 52 hpf to 76 hpf in WT embryos (6.5 fold); in *gata2b*$^{-/-}$ embryos that expansion was reduced (3.1 fold). To support our findings we investigated the expression of *cmyb,* which is a marker for proliferating HSPCs from 30 hpf[43,44]. A significant reduction in *cmyb* expression was detectable from 33 hpf onward in the AGM and CHT regions of *gata2b*$^{-/-}$ embryos compared to WT (Figure 2E-H and Table S1). This analysis detected a reduction in *cmyb* expression levels rather then quantifying HSPC numbers. However, because the number of CD41:GFP^{+}Flt:RFP^{+} cells was not affected at 33 hpf, but *cmyb* expression was already reduced at 33 hpf, this suggests that proliferation of definitive HSPCs is affected, resulting in a reduction of definitive HSPCs at 76 hpf (Figure 2D).

To test this, proliferation was assessed by flow cytometry of CD41:GFP^{+} cells at 75 hpf in WT and *gata2b*$^{-/-}$ embryos. This analysis shows that Gata2b deficient CD41:GFP^{+} cells have an increased proportion of cells in the G_0 phase of cell cycle explaining the reduction of HSPCs at 3 dpf (Figure 2I-K). At 5 dpf the difference in proliferation is no longer detectable although HSPC numbers are still reduced (Figure 2I and data not shown).

Single cell RNAseq identifies a lymphoid bias at the expense of myeloid lineage output in gata2b$^{-/-}$ *kidney marrow*

Because the functions of GATA2 are separated between Gata2a and Gata2b in zebrafish and Gata2b deficient zebrafish are viable, we can uniquely assess the function of Gata2b in adult hematopoiesis. To investigate the hematopoietic lineages in an unbiased manner and to assess the impact of Gata2b deficiency on the transcriptional profile of hematopoietic progenitors and differentiated cells, the progenitor population including lymphocytes from kidney marrow (KM) were isolated and processed for single-cell RNA sequencing (scRNA-seq)(Figure S2A and B). To enrich the scarce HSC population we used pooled KM from two WT and *gata2b*$^{-/-}$ *Tg(CD41:GFP)* zebrafish per sample and included all CD41:GFP^{low} expressing cells present in the kidney marrow pool as these cells were shown to contain transplantable HSCs[30] (Figure S2A and C). This resulted in a mild enrichment of phenotypic HSCs from 0.21-0.5% to 0.46 – 2.73% of CD41:GFP^{low} cells within the total progenitor population. We identified 20 different cell clusters were identified using the nearest neighbor algorithm in the R Seurat package[36] (Figure 3A and S2G). Most progenitors that were sequenced expressed previously characterized differentiation markers[45-49] (Figure 3B-E and S2H). We identified 2 HSPC populations. These clusters are characterized by the robust expression of HSC genes, like *fli1a* and *meis1b*[48-50] (Figure 3F,G), *gata2b* (Figure 3H), concomitant with intermediate levels of GFP derived from the CD41:GFP transgene (Figure 3I), and low expression of differentiation markers (Figure 3B-E and S2H). Compared to HSPC2, HSPC1 exhibited a lower expression of metabolic and proliferation markers like *pcna* and *myca*, suggesting that HSPC1 is more quiescent than HSPC2 (Figure 3J). Therefore, lineage trajectory analysis was started from this cluster identifying separate lineage differentiation trajectories for the erythroid-, myeloid- and lymphoid lineage (Figure 3K, L, N).

Proportion analysis regarding the distribution of WT and Gata2b deficient cells between the lymphoid and myeloid lineages indicated a bias towards the lymphoid lineage in *gata2b*$^{-/-}$ cells at the expense of the myeloid lineage compared to WT (Figure 3A, K-M and S2I and Table S2). The largest differences were observed in 3 clusters expressing high levels of *granulin 1 (grn1)*(Figure 3B) indicating that these clusters contains myeloid progenitors and were overrepresented by WT cells (Figure 3A, K, M and S2I). We defined these 3 clusters expressing myeloid specific genes with slight differences in their expression pattern as myeloid progenitors-1, -2 and -3 (Figure 3A and S2H).The *grn1* expressing cluster also showed high expression of *s100a10b,* a potential new marker for these cells (Figure S3A).

Expression analysis on KM smears showed that *s100a10b* is expressed in the neutrophil lineage (Figure S3B). The B-cell clusters were overrepresented by *gata2b*$^{-/-}$ cells and showed very high expression of *immunoglobulin heavy variable 1-4* (*ighv1-4)*(Figure S2H)*, CD37* (Figure 3C) and *pax5* (not shown), indicating that these were *bona fide* B-cell populations. Interestingly, we found a population of phagocytic B-cells previously identified in theleosts which express both *mpeg1.1* and B cell markers[51]. Lineage trajectory analysis indicates that these cells decend from lymphoid progenitors (Figure 3C, K, L). When pseudotime analyses was performed for WT and *gata2b*$^{-/-}$cells separately, the phagocytic B-cells did not only show a lineage differentiation trajectory from lymphoid progenitors, but also from the HSPC1 population, indicating a skewing in *gata2b*$^{-/-}$ HSPCs directly towards the lymphoid lineage (Figure S2J, K).

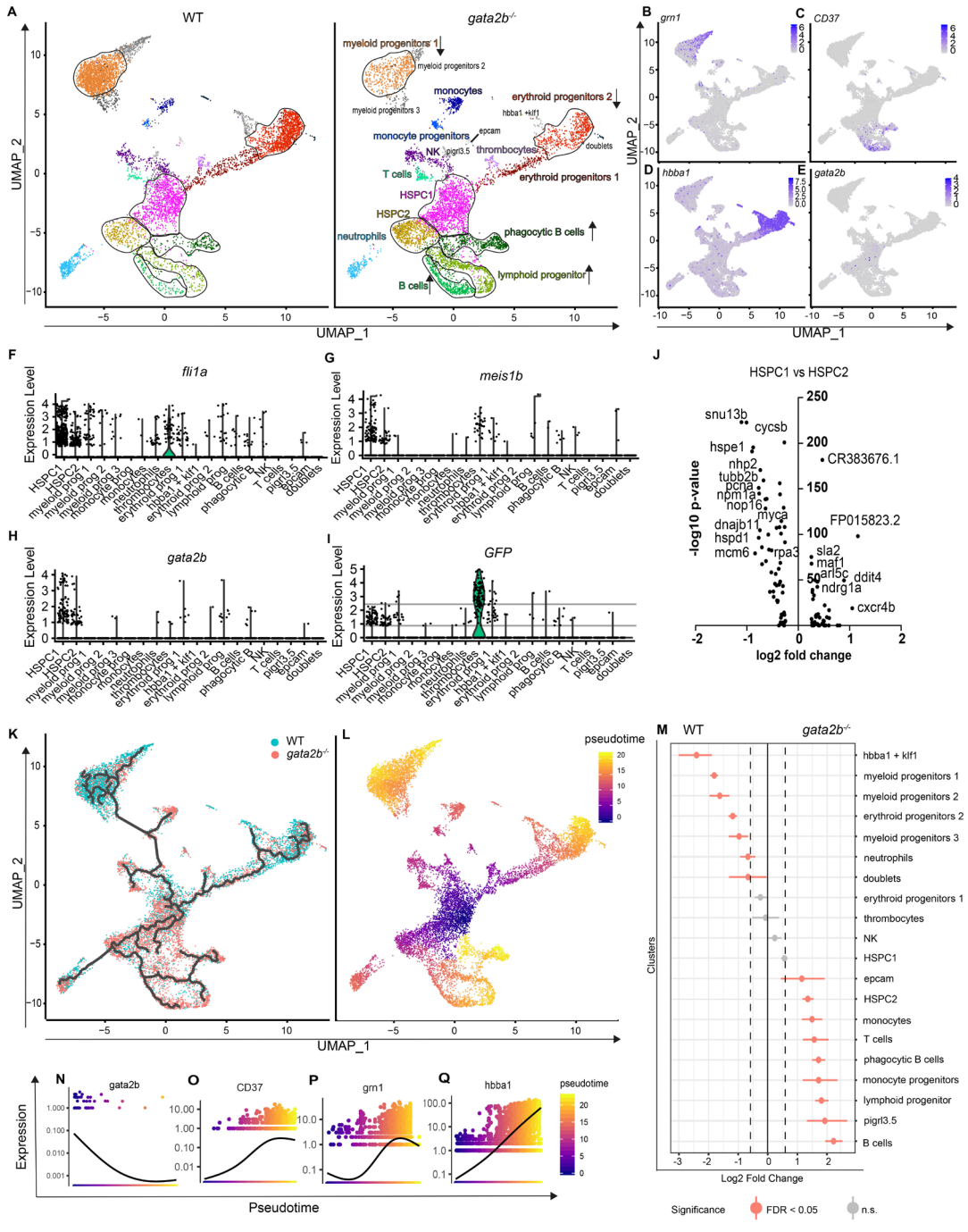


Figure 3. Single cell analysis reveals that *gata2b*$^{-/-}$ kidney marrow cells are overrepresented in lymphoid lineage clusters and reduced in erythroid and myeloid lineage clusters compared to WT.
A) Split UMAP of WT and *gata2b*$^{-/-}$ cells with cluster indication of enriched (arrow up) or reduced (arrow down) cell clusters in *gata2b*$^{-/-}$ cells. B-E) Pooled WT and *gata2b*$^{-/-}$ UMAP feature analysis with gradual gene expression in shades of blue. Expression pattern of B) *granulin1* (*grn1),* C) *cluster of differentiation 37* (*CD37),* D) *hemoglobin, beta adult (hbba1),* E) *GATA binding protein 2b* (*gata2b).* F-I) Violin plots representing the expression levels of genes within the different clusters and each dot represents expression in one cell. F) *fli-1 proto-oncogene* (*fli1a),* G) *meis homeobox 1b* (*meis1b),* H) *GATA binding protein 2b* (*gata2b),* I) *green fluorescent protein* (*GFP)*, indicating CD41:GFP^{low} cells. J) Volcano plot comparing HSPC1 vs HSPC2. At the left of the Y axis there are genes in HSPC1 with an average logarithmic fold change less than -0.25 and to the right are genes with a logarithmic fold change higher than 0.25 compared to HSPC2. K) Lineage differentiation trajectory depicted on UMAP with WT cells in blue and *gata2b*$^{-/-}$ cells in pink. L) Pseudotime analysis assuming HSPC1 as a starting point. M) quantitation of proportions of distribution between WT and *gata2b*$^{-/-}$ cells in the different clusters. Significant differences are indicated in pink. N-Q) pseudotime analysis of gene expression in lineage trajectory analysis of N) *gata2b,* O) *cluster of differentiation 37* (*CD37*), P) *granulin1* (*grn1)* ans Q) *hemoglobin, beta adult (hbba1)*. UMAP; uniform manifold approximation projection.

Lack of Gata2b leads to reduced neutrophil numbers and increased lymphoid progenitors in adult kidney marrow

Because scRNA-seq analysis showed a major switch in lineage differentiation, we asked whether hematopoietic differentiation was affected in the adult *gata2b*$^{-/-}$ KM using scatter profile-, transgenic marker- and morphological analysis[30,32,33,52]. While *gata2b*$^{-/-}$ embryos did not show signs of altered lineage differentiation up to 5 dpf (Figure S4A-F), scatter profiles of adult *gata2b*$^{-/-}$ zebrafish KM showed a significant reduction in the myeloid population (Figure 4A, B and Table 1) and a relative increase in the scatter population containing HSPCs and lymphoid cells at 4 months post fertilization (mpf) and onward (Figure 4A, C and Table 1). This skewing in the population frequencies persisted with age (Figure 4B, C). To further address how the myeloid lineage was affected by the loss of Gata2b, *Tg(mpx:GFP)* expression, specifically marking neutrophils[33,53] and *Tg(mpeg1.1:GFP),* marking monocytes and phagocytic B-cells[32,54] was assessed. No significant difference was observed in mpeg:GFP^{+} cells between WT and *gata2b*$^{-/-}$ KM (Figure S4O-Q). *gata2b*$^{-/-}$ zebrafish showed a severe reduction in mpx:GFP^{+} neutrophils in the kidney marrow at 4 mpf (Figure 4D-F and Table 1). Sorted mpx:GFP^{+} cells from these zebrafish showed that the remaining *gata2b*$^{-/-}$ mpx:GFP^{+} cells did not reach WT levels of GFP, had a more immature neutrophil morphology and a block at the promyelocyte stage (Figure 4G, H and S3E), indicating that Gata2b is required for terminal neutrophil differentiation. This could be a result of the reduction in myeloid progenitors in the single cell data (Figure 3A and M).

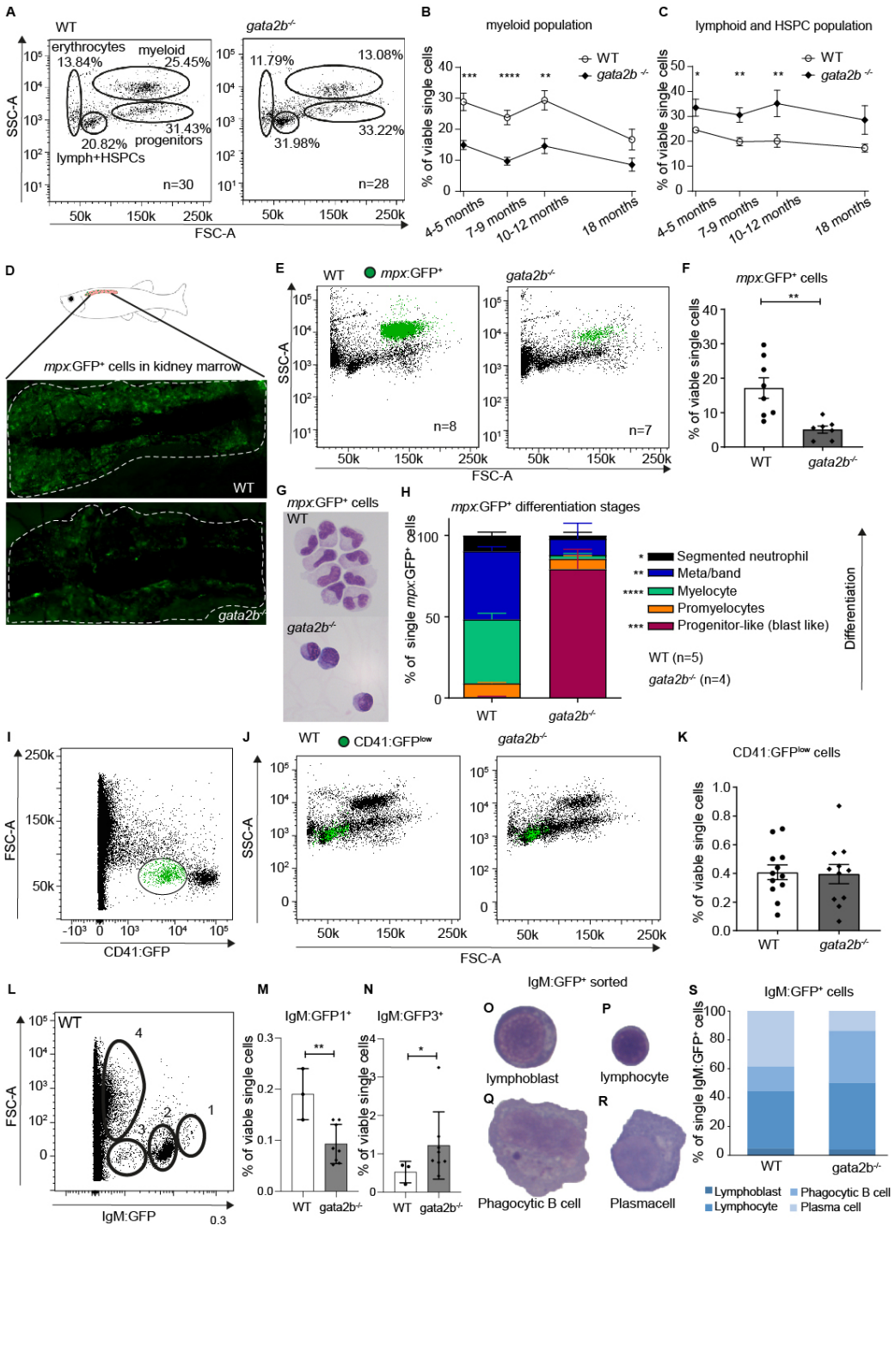


Figure 4. Gata2b deficiency results in decreased myeloid differentiation in adult zebrafish kidney marrow.
A) Gating strategy of FACS analysis of whole kidney marrow of WT and *gata2b*$^{-/-}$ zebrafish. Percentages represent the average of all zebrafish analyzed per genotype. B-C) Quantitation as percentages of single viable cells over time of B) myeloid and C) lymphoid and HSPC populations. D) Representative example of *Tg(mpx:GFP)* expression in WT and *gata2b*$^{-/-}$ zebrafish kidney marrow by fluorescence microscopy. E) Forward and side scatter profile of *Tg(mpx:GFP)* expression in WT and *gata2b*$^{-/-}$ zebrafish kidney marrow in green. F) Quantitation of *Tg(mpx:GFP)*$^{+}$ cells expressed as percentage in single viable cells. Each dot represents kidney marrow analysis of one zebrafish. G) Representative figure of sorted *Tg(mpx:GFP)*$^{+}$ cells from WT and *gata2b*$^{-/-}$ zebrafish kidney marrow after MGG staining. H) Quantification of *Tg(mpx:GFP)*$^{+}$ cells from WT and *gata2b*$^{-/-}$ zebrafish kidney marrow based on the differentiation phenotype using MGG staining. I) Gating strategy for CD41:GFP^{low} expressing cells in total kidney marrow in green. J) Forward- and side-scatter plot of WT and *gata2b*$^{-/-}$ kidney marrow cells and CD41:GFP^{low} expressing cells in green. K) Quantification of the frequency of CD41:GFP^{low} cells in single live cells of total kidney marrow. Each dot represents kidney marrow analysis of one zebrafish. L) FSC/GFP scatter profile of *Tg(IgM:GFP)* WT KM. M) Quantitation of Gating 1 of *Tg(IgM:GFP)* WT and *gata2b*$^{-/-}$ KM as percentage of single viable cells. Each dot represents kidney marrow analysis of one zebrafish. N) Quantitation of Gating 3 of *Tg(IgM:GFP)* WT and *gata2b*$^{-/-}$ KM as percentage of single viable cells. Each dot represents kidney marrow analysis of one zebrafish. O-R) representative image of sorted IgM:GFP^{+} cells indicating O) lymphoplastic cell, P) lymphocyte, Q) phagocytic B cell and R) plasmacell. S) quantitation of sorted IgM:GFP^{+} cells per genotype. SSC = side scatter, FSC = forward scatter, KM = kidney marrow. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. Error bars represent SEM.

Because GATA2 is also required for HSC maintenance in mice[7,12,13], we asked whether Gata2b deficiency resulted in a block in HSPC differentiation and thus an accumulation of HSPCs. In zebrafish, CD41:GFP^{low} expression marks the HSPC population most stringently[30]. Although CD41:GFP^{+} cell numbers and percentages were reduced during embryonic development (Figure 2D, I), at 20 dpf these percentages normalize resulting in comparable numbers of CD41:GFP^{low} cells during adulthood (Figure 2I and 4I-K) indicating that the accumulation of the population containing lymphoid cells and HSPCs in *gata2b*$^{-/-}$ KM is not due to a differentiation block in HSPCs, but due to an increase in lymphoid cells. *Tg(IgM:GFP)*, marking B-cells[35] and *Tg(lck:GFP)*, marking T-cells[34] were used to asses lymphoid differentiation. We did not find an increase in the lck:GFP^{+} population (Figure S4M, N and Table 1). However, in *Tg(IgM:GFP)* zebrafish we identified several populations of IgM:GFP^{+} cells with a significant increase in immature IgM:GFP^{+} cells (IgM:GFP3 fraction)(Figure 4L-N and S4G-M). We could classify the different IgM:GFP^{+} populations as lymphoblastic cells, lymphocytes, plasmacells and phagocytic B-cells[51,55](Figure 4O-S). In particular phagocytic B-cells were increased in *gata2b*$^{-/-}$ KM compared to WT, but mature plasmacells were significantly reduced (Figure 4S, P<0.01 and P<0.001). Although the majority of the cells in the lymphoid and HSPC population was not marked by known lymphoid lineage markers IgM:GFP or lck:GFP, we could still detect a significant increase in immature B-cells, confirming the increase in lymphoid output in KM in *gata2b*$^{-/-}$ zebrafish compared to WT.

Gene expression analysis reveals different HSPC populations in zebrafish

Next, we explored the molecular origin of the increase in lymphoid lineage output observed in Gata2b deficient zebrafish. Previous studies show that blocked neutrophil differentiation results in a shift towards monocytic lineage differentiation[56]. Our data shows indeed an increase in monocyte progenitors and monocytes (Figure 3M). However, we also detect a shift towards the lymphoid lineage, indicating that Gata2b is required for lineage

programming in more immature progenitors. First, we tested if the lymphoid lineage bias was detectable in the HSPC clusters we identified as HSPC1 and HSPC2 (Figure 3A).

HSPC1 makes up 18 percent of the total analyzed kidney marrow population. Because HSCs are a rare population of cells based on transplantation studies, we hypothesized that HSPC1 also contains other progenitor cells. We subclustered HSPC1 to subdivide these progenitors (Figure 5A,B). In this way, clustering is not based on gene expression differences found in comparison to more committed cells, but only based on gene expression differences within the HSPC1 population. Eight subclusters were identified based on differential gene expression analysis (Figure S5A). These subclusters were classified as a quiescent subcluster with very low gene expression, an HSC subcluster with expression of *meis1b* and *fli1a*, three myeloid subclusters, one lymphoid subcluster, a proliferative subcluster and an undefined subcluster (Figure 5B, E, F and S5A). Interestingly, proportion analysis of WT and *gata2b*$^{-/-}$ cells showed that Gata2b deficient cells almost entirely lost the quiescent subcluster and gained a proliferative subcluster and myeloid subcluster 3 (Figure 5C, D). A differential expression analysis of the whole HSPC1 cluster revealed a downregulation of myeloid genes like *s100a10b*, *grn1*, *csf3b* and *cepba* in *gata2b*$^{-/-}$ HSPC1 (Figure 5G) but not a clear upregulation of lymphoid genes. When the same comparison was done in HSPC2 cells (Figure S6A, B), we detected a larger reduction in the myeloid gene expression program in the entire HSPC2 cluster (Figure S6C, D) and found that *gata2b*$^{-/-}$ cells expressed higher levels of lymphoid genes like *ikzf1, fcer1gl, ighv1-4, ccr9a* and *xbp1* (Figure S6E-J). Psuedotime analysis showed the differentiation trajectories within the HSPC1 cluster when started from the quiescent subcluster, containing most CD41:GFP expressing cells (Figure 3E, F and H, I). When inferring gene expression in psuedotime analysis of HSPC1, we found that *gata2b* and *meis1* expression are highest in the quiescent population and decrease during differentiation (Figure 5J-L). Interestingly, *gata2a* expression does not overlap with *gata2b* indicating that *gata2a* does not compensate for the loss of *gata2b* in Gata2b deficient HSPCs (Figure 5K, M). As HSPCs become more mature, they first upregulate proliferation markers like *pcna* and *mki67* both highly expressed in the *gata2b*$^{-/-}$ unique subcluster (Figure 5N, O). Assessing proliferation by flow cytometry of CD41:GFP^{low} cells in WT and *gata2b*$^{-/-}$ KM, we found that *gata2b*$^{-/-}$ CD41:GFP^{low} cells have increased numbers of cells in S phase, indicating that *gata2b*$^{-/-}$ zebrafish HSPCs are more proliferative (Figure 5P-R). Together these data show that the absence of Gata2b leads to transcriptional changes in the HSPC compartment concomitant with a shift in lineage output from the myeloid lineage towards the lymphoid lineage.

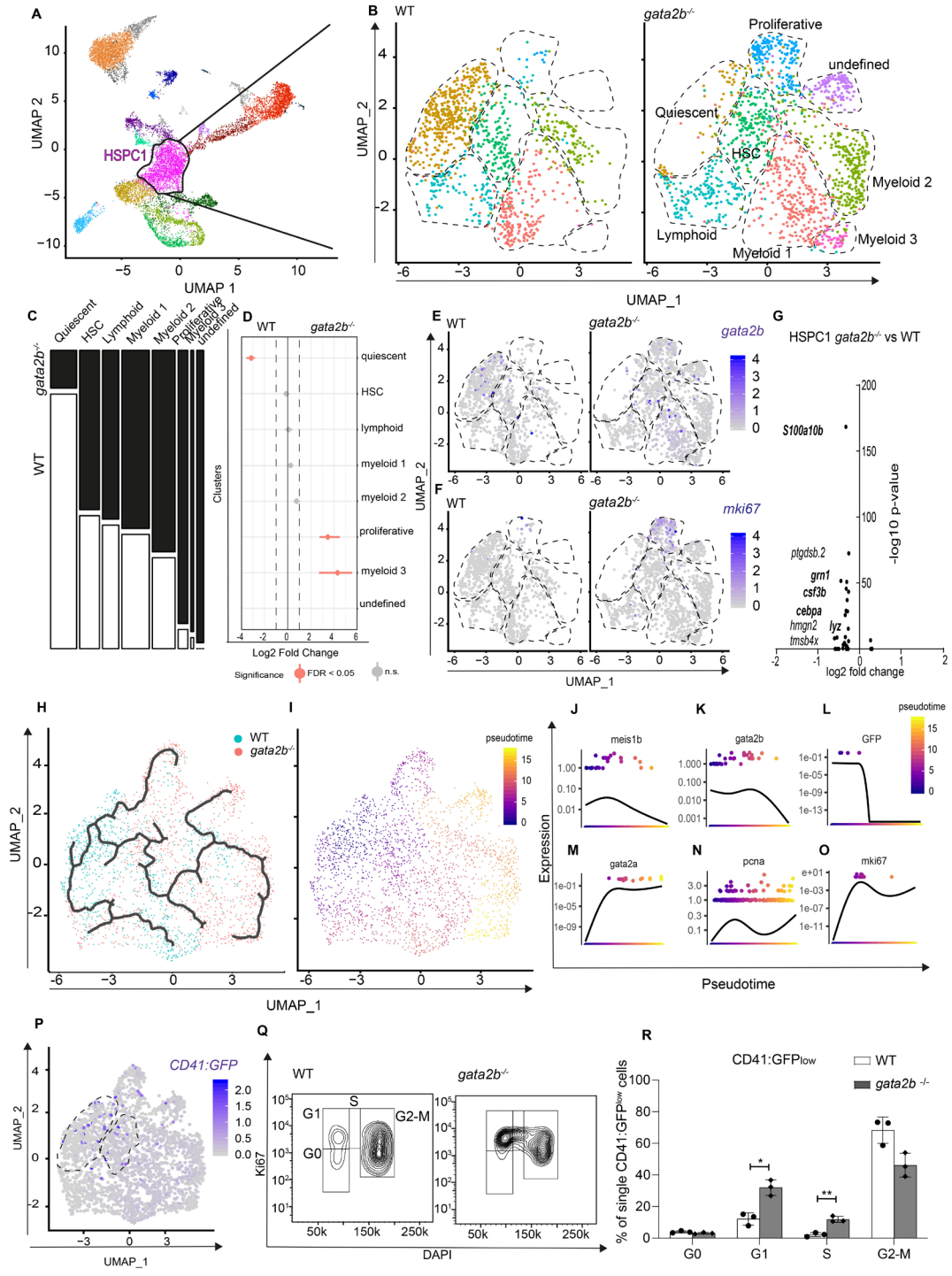


Figure 5. Subclustering of HSPC1 demonstrates the loss of a *gata2b* expressing quiescent subcluster and the appearance of a proliferative subcluster.
A) Cluster selection for subclustering. B) Reclustering of the HSPC1 population split between WT and *gata2b*$^{-/-}$cells. C) Genotype distribution of each of the clusters with WT cells in white and *gata2b*$^{-/-}$cells in black. D) quantitation of proportions of distribution between WT and *gata2b*$^{-/-}$ cells in the different clusters. Significant differences are indicated in pink. E-F) WT and *gata2b*$^{-/-}$ feature analysis with gradual gene expression in shades of blue within HSPC1 cells of D) *gata2b* E) *mki67*. G) Volcano plot comparing HSPC1 *gata2b*$^{-/-}$ vs WT. At the left of the Y axis gene expression in *gata2b*$^{-/-}$ HSPC1s with an average logarithmic fold change less than -0.25 and to the right gene expression with a logarithmic fold change higher than 0.25 compared to WT HSPC1s. Each dot represents a gene. H) Lineage differentiation trajectory depicted on UMAP with WT cells in blue and *gata2b*$^{-/-}$ cells in pink. I) Pseudotime analysis assuming the quiescent population as starting point. J-O) gene expression analysis on pseudotime analysis with J) *meis1b*, K) *gata2b* L) *GFP,* M) *gata2a,* N) *pcna*, O) *ki67*. P) WT and *gata2b*$^{-/-}$ feature analysis with gradual gene expression of *GFP* in shades of blue within HSPC1 cells. Dotted circles indicate the quiescent and HSC subcluster. Q) Cell cycle analysis by flow cytometry of Ki67 and DAPI staining of CD41:GFP^{low} cells in adult WT and *gata2b*$^{-/-}$ KM cells. R) Bar graph representing the quantitation of cell cycle of CD41:GFP^{low} cells in adult WT and *gata2b*$^{-/-}$ KM cells. Bars represent mean ± SEM, each dot indicates analysis from one zebrafish. UMAP; uniform manifold approximation projection.

Differential gene expression analysis reveals decreased myeloid marker expression in gata2b$^{-/-}$ *HSPCs and aberrant co-expression of myeloid and lymphoid genes.*

Overall, the expression of myeloid genes in *gata2b*$^{-/-}$ HSPC1 is reduced, but the percentage of *gata2b*$^{-/-}$ HSPC1s with detectable expression of myeloid genes such as *grn1* was increased (Figure 5G and 6A). This apparent contradiction was clarified by an overall transcript upregulation (Figure 6A, *actinb1* expression), indicative of a loss of quiescence. It is known that HSPCs can co-express myeloid and lymphoid genes before lineage decision[57,58]. While WT cells had a clear dichotomy in expression of myeloid and lymphoid genes, *gata2b*$^{-/-}$ HSPCs had a higher fraction of cells co-expressing lymphoid and myeloid genes (Figure 6B-E). For example, increased co-expression of a phagocytic B-cell marker, *igic1s1*, could be detected in *gata2b*$^{-/-}$ HSPC1 cells together with the myeloid marker *cebpb* (Figure 6B). This result suggests that the loss of Gata2b does not halt HSPC differentiation but re-directs this towards another lineage. Interestingly, when we infer psuedotime analysis of only WT and only *gata2b*$^{-/-}$ cells, this is exactly as we find. In *gata2b*$^{-/-}$ cells, phagocytic B cells can be formed from both lymphoid progenitors, as well as HSPC1 cells as opposed to WT phagocytic B cells (Figure S2J, K). Based on this data we conclude that the lymphoid bias in *gata2b*$^{-/-}$ zebrafish kidney marrow initiated in the most immature HSPC population. This is due to a failure to elicit proper expression of the myeloid differentiation program and concomitant upregulation of the lymphoid program, that redirects HSPCs towards a lymphoid fate.

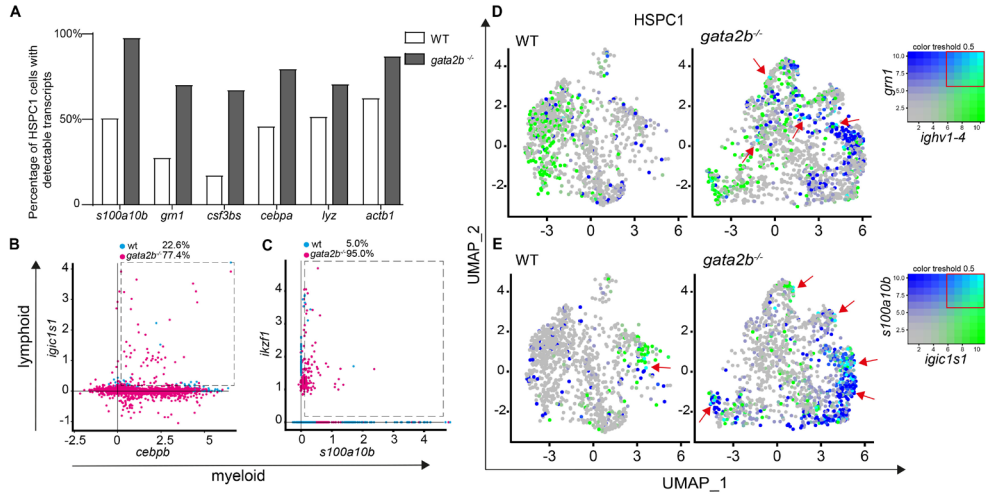


Figure 6. Gata2b deficient HSPC1 show co-expression of myeloid and lymphoid gene expression programs
A) Bar graph representing the percentage of WT HSPC1 cells in white and *gata2b*$^{-/-}$ HSPC1 cells in grey with at least one read of *s100a10b*, *grn1*, *colony stimulating factor 3b (*csf3bs), *CCAAT enhancer binding protein alpha (cebpa), lysozyme (lyz)* or *actin beta1 (actb1),* indicating more cells with detectable myeloid gene expression in Gata2b deficient HSPC1s B) Co-expression analysis of the lymphoid gene *igic1s1* with the myeloid gene *cebpb* and C) the lymphoid gene *IKAROS Family Zinc Finger 1* (*ikzf1*) with the myeloid gene *s100a10b*. Values represent percentages of WT and *gata2b*$^{-/-}$ HSPC1 cells co-expressing myeloid and lymphoid genes (within the dashed box). D) WT and *gata2b*$^{-/-}$ feature analysis representing co-expression analysis of the lymphoid gene *Immunoglobulin heavy variable 1-4* (*ighv1-4*) with the myeloid gene *grn1* and E) the phagocytic B cell marker immunoglobulin light iota constant 1-s1 (*igic1s1*) with the myeloid gene *s100a10b* with myeloid genes in blue and lymphoid genes in green. Co-expression of the myeloid and lymphoid genes is represented in turquoise indicated by red arrows in the WT and *gata2b*$^{-/-}$ feature analysis. Coloring threshold set in quantiles, min.cutoff= q10, max.cutoff= q90.

DISCUSSION

In this study, we showed that the function of mammalian GATA2 in zebrafish is split between Gata2a and Gata2b. Gata2a is required for HE specification upstream of Gata2b. Gata2b is not vital for embryonic generation of HSPCs, but supports their expansion in the caudal hematopoietic tissue. However, during adulthood, Gata2b is required for the quiescent HSPC population and in its absence HSPCs are more proliferative. In addition, Gata2b deficient kidney marrow from adult zebrafish showed a lymphoid bias at the expense of the myeloid lineage based on scatter profiles and transgenic marker analyses. Single cell transcriptome analysis showed that the stem and progenitor cells were the origin of the increased lymphoid lineage output in *gata2b*$^{-/-}$ kidney marrow cells, due to a failure to increase myeloid gene expression to sufficient levels and a subsequent co-expression of both myeloid and lymphoid genes in *gata2b*$^{-/-}$ HSPCs. These data establish that Gata2b is vital for maintaining the myeloid differentiation program while restricting lymphoid differentiation.

The molecular mechanism controlling lineage commitment has long been thought to be regulated by stochastic variations in the levels of transcription factors, and progenitors are committed to a lineage choice[59]. However, later reports suggested that some transcription factors have a reinforcing activity for terminal differentiation and propose that microenvironmental or upstream regulators are decisive for lineage commitment[60]. This would suggest that when these reinforcing factors are removed, cells can redirect their lineage. Our results are consistent with Gata2b being required for stemness of HSCs. Single cell transcriptome analysis showed a unique cluster of Gata2b deficient cells with upregulation of genes related to proliferation, suggestive of a role for Gata2b in cell cycle adaptation. The quiescent subcluster was almost entirely lost and the the CD41:GFP^{low} population showed increased proliferation. Loss of quiescence in HSPCs then increases the expression of commitment genes resulting in cells co-expressing lymphoid and myeloid lineage markers as detected in *gata2b*$^{-/-}$ HSPCs and Gata2b is therefore an essential cell-intrinsic regulator of lineage output in HSPCs.

In mouse, GATA2 is also required for the maintenance of HSCs after they are generated[7]. During embryonic hematopoiesis the number and percentage of HSPCs is reduced due to reduced proliferation, but during adult stages Gata2b deficient HSCs as marked by CD41:GFP^{low} expression are not reduced and proliferation is increased, probably responsible for the normalization in HSC numbers (Figure 4I-K and 5Q,R). We do not find upregulation of *gata2a* in these cells as a rescue mechanism (Figure 5M). Single cell transcriptome analysis identifies several HSPC populations with unique transcriptional signatures. Interestingly, the CD41:GFP^{low} expressing cells were scattered among different HSPC populations. Transplantation data suggest that only a minority of these cells are *bona fide* HSCs[30,61]. Because zebrafish are outbred, limiting dilution transplantation studies result in a gross underestimation of actual HSC numbers. This indicates that further research could provide us with a more stringent marker for HSCs in zebrafish. In Gata2b deficient HSPC1s, the quiescent HSPC population is absent (Figure 5B-D). This could represent the true quiescent HSC population. Interestingly, this does not affect survival of the zebrafish.

Not all myeloid lineage differentiation was abrogated in Gata2b deficient zebrafish and few intact neutrophils remained present. Also, the monocyte progenitor and -cluster, marked by *mpeg1.1* were present in Gata2b deficient KM. Previous studies found that if neutrophil development is blocked, myeloid differentiation progresses towards to monocytic lineage[56]. Besides an increase in monocytic progenitors, we also detect a redirection of lineage differentiation at a much earlier state leading to increases in B-cell populations (Figure 3M, S2J, K). This indicates that in Gata2b deficient HSPCs, a reprogramming occurs both in immature cells to delineate lineage differentiation towards the lymphoid lineage, but also in the myeloid lineage to redirect the lineage to monocytes, again indicating separate functions for Gata2 in lineage differentiation. Interestingly, the number and percentage of plasmacells was reduced (Figure 4S). Together with the severe neutropenia, this is very similar to patients sufferening from MonoMAC syndrome, which is characterized by neutropenia, monocytopenia, DC- and B-cell lymphopenia[62-64]. These syndromes are caused by haploinsufficiency of the GATA2 transcription factor. Despite the severe neutropenia, no infections were observed in Gata2b deficient zebrafish probably due to the SPF conditions of the animal facility.

In conclusion, we find that Gata2b is required for proliferation of the HSPC pool in the CHT and is vital for myeloid lineage differentiation in the adult, both in the HSPC compartment and for terminal differentiation. Loss of Gata2b consequently induces a differentiation diversion towards the lymphoid lineage.

Acknowledgements

We thank members of the de Pater, Touw, Raaijmakers and Schneider-Kramann labs for helpful discussions. We thank Dr Cupedo and Prof. Dr. Dzierzak (University of Edinburgh) for careful reading of the manuscript. We thank the Experimental Animal Facility of Erasmus MC for animal husbandry and the Erasmus Optical Imaging Center for confocal microscopy services. This research is supported by the European Hematology Association (junior non clinical research fellowship)(EdP), the Dutch Cancer Foundation KWF/Alpe d'HuZes (SK10321)(EdP), the British Heart Foundation (BHF IBSR Fellowship FS/13/50/30436 to RM, CBM), the Wellcome Trust (PhD Scholarship #WT102345/Z/13/Z. to TD), the Daniel den Hoed Foundation for support of the Cancer Genome Editing Center (IT) and the Josephine Nefkens Foundation for purchase of the Chromium 10x (IT). Graphical abstract is created with BioRender.com.

Authorship contributions

EdP, EG and CK conceived the study; EG, CK, HdL, JZ, DB, TD, CBM, PvS, MvR, and EB performed experiments; EG, CK, MdJ, RH, CBM, RM, KG and EdP analysed results; RM, PF and IT provided resources and EG, CK and EdP wrote the manuscript and IT revised the manuscript.

Disclosures

The authors declare no conflicts of interests

REFERENCES

1. Sawai CM, Babovic S, Upadhaya S, et al. Hematopoietic Stem Cells Are the Major Source of Multilineage Hematopoiesis in Adult Animals. *Immunity*. 2016;45(3):597-609.
2. Yamamoto R, Morita Y, Ooehara J, et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell*. 2013;154(5):1112-1126.
3. Sanjuan-Pla A, Macaulay IC, Jensen CT, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*. 2013;502(7470):232-236.
4. Muller-Sieburg CE, Cho RH, Karlsson L, Huang JF, Sieburg HB. Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness. *Blood*. 2004;103(11):4111-4118.
5. Dykstra B, Kent D, Bowie M, et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell*. 2007;1(2):218-229.
6. Watcham S, Kucinski I, Gottgens B. New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. *Blood*. 2019;133(13):1415-1426.
7. de Pater E, Kaimakis P, Vink CS, et al. Gata2 is required for HSC generation and survival. *J Exp Med*. 2013;210(13):2843-2850.
8. Gao X, Johnson KD, Chang YI, et al. Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *J Exp Med*. 2013;210(13):2833-2842.
9. Snow JW, Trowbridge JJ, Fujiwara T, et al. A single cis element maintains repression of the key developmental regulator Gata2. *PLoS Genet*. 2010;6(9):e1001103.
10. Mehta C, Johnson KD, Gao X, et al. Integrating Enhancer Mechanisms to Establish a Hierarchical Blood Development Program. *Cell Rep*. 2017;20(12):2966-2979.
11. Johnson KD, Kong G, Gao X, et al. Cis-regulatory mechanisms governing stem and progenitor cell transitions. *Sci Adv*. 2015;1(8):e1500503.
12. Ling KW, Ottersbach K, van Hamburg JP, et al. GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med*. 2004;200(7):871-882.
13. Rodrigues NP, Janzen V, Forkert R, et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood*. 2005;106(2):477-484.
14. Rodrigues NP, Boyd AS, Fugazza C, et al. GATA-2 regulates granulocyte-macrophage progenitor cell function. *Blood*. 2008;112(13):4862-4873.
15. Nandakumar SK, Johnson K, Throm SL, Pestina TI, Neale G, Persons DA. Low-level GATA2 overexpression promotes myeloid progenitor self-renewal and blocks lymphoid differentiation in mice. *Exp Hematol*. 2015;43(7):565-577 e561-510.
16. Tsai FY, Keller G, Kuo FC, et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*. 1994;371(6494):221-226.
17. Bertrand JY, Cisson JL, Stachura DL, Traver D. Notch signaling distinguishes 2 waves of definitive hematopoiesis in the zebrafish embryo. *Blood*. 2010;115(14):2777-2783.
18. Bertrand JY, Chi NC, Santoso B, Teng S, Stainier DY, Traver D. Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature*. 2010;464(7285):108-111.
19. Kissa K, Herbomel P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*. 2010;464(7285):112-115.
20. Warga RM, Kane DA, Ho RK. Fate mapping embryonic blood in zebrafish: multi- and unipotential lineages are segregated at gastrulation. *Dev Cell*. 2009;16(5):744-755.

21. Ciau-Uitz A, Monteiro R, Kirmizitas A, Patient R. Developmental hematopoiesis: ontogeny, genetic programming and conservation. *Exp Hematol*. 2014;42(8):669-683.
22. Tamplin OJ, Durand EM, Carr LA, et al. Hematopoietic stem cell arrival triggers dynamic remodeling of the perivascular niche. *Cell*. 2015;160(1-2):241-252.
23. Davidson AJ, Zon LI. The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis. *Oncogene*. 2004;23(43):7233-7246.
24. Butko E, Distel M, Pouget C, et al. Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. *Development*. 2015;142(6):1050-1061.
25. Dobrzycki T, Mahony CB, Krecsmarik M, et al. Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis. *Commun Biol*. 2020;3(1):71.
26. Gagnon JA, Valen E, Thyme SB, et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One*. 2014;9(5):e98186.
27. Chocron S, Verhoeven MC, Rentzsch F, Hammerschmidt M, Bakkers J. Zebrafish Bmp4 regulates left-right asymmetry at two distinct developmental time points. *Dev Biol*. 2007;305(2):577-588.
28. Dobrzycki T, Krecsmarik M, Bonkhofer F, Patient R, Monteiro R. An optimised pipeline for parallel image-based quantification of gene expression and genotyping after in situ hybridisation. *Biol Open*. 2018;7(4).
29. Lawson ND, Weinstein BM. In vivo imaging of embryonic vascular development using transgenic zebrafish. *Dev Biol*. 2002;248(2):307-318.
30. Ma D, Zhang J, Lin HF, Italiano J, Handin RI. The identification and characterization of zebrafish hematopoietic stem cells. *Blood*. 2011;118(2):289-297.
31. Bussmann J, Bos FL, Urasaki A, Kawakami K, Duckers HJ, Schulte-Merker S. Arteries provide essential guidance cues for lymphatic endothelial cells in the zebrafish trunk. *Development*. 2010;137(16):2653-2657.
32. Ellett F, Pase L, Hayman JW, Andrianopoulos A, Lieschke GJ. mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish. *Blood*. 2011;117(4):e49-56.
33. Renshaw SA, Loynes CA, Trushell DM, Elworthy S, Ingham PW, Whyte MK. A transgenic zebrafish model of neutrophilic inflammation. *Blood*. 2006;108(13):3976-3978.
34. Langenau DM, Ferrando AA, Traver D, et al. In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. *Proc Natl Acad Sci U S A*. 2004;101(19):7369-7374.
35. Page DM, Wittamer V, Bertrand JY, et al. An evolutionarily conserved program of B-cell development and activation in zebrafish. *Blood*. 2013;122(8):e1-11.
36. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36(5):411-420.
37. Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A*. 2003;100(15):8811-8816.
38. Medvinsky A, Dzierzak E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*. 1996;86(6):897-906.
39. Jaffredo T, Gautier R, Eichmann A, Dieterlen-Lièvre F. Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development*. 1998;125(22):4575-4583.
40. Boisset JC, van Cappellen W, Andrieu-Soler C, Galjart N, Dzierzak E, Robin C. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*. 2010;464(7285):116-120.
41. Bertrand JY, Kim AD, Teng S, Traver D. CD41+ cmyb+ precursors colonize the zebrafish pronephros by a novel migration route to initiate adult hematopoiesis. *Development*. 2008;135(10):1853-1862.
42. Bonkhofer F, Rispoli R, Pinheiro P, et al. Blood stem cell-forming haemogenic endothelium in zebrafish derives from arterial endothelium. *Nat Commun*. 2019;10(1):3577.

43. Bein K, Husain M, Ware JA, Mucenski ML, Rosenberg RD, Simons M. c-Myb function in fibroblasts. *J Cell Physiol*. 1997;173(3):319-326.
44. North TE, Goessling W, Walkley CR, et al. Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature*. 2007;447(7147):1007-1011.
45. Carmona SJ, Teichmann SA, Ferreira L, et al. Single-cell transcriptome analysis of fish immune cells provides insight into the evolution of vertebrate immune cell types. *Genome Res*. 2017;27(3):451-461.
46. Danilova N, Bussmann J, Jekosch K, Steiner LA. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat Immunol*. 2005;6(3):295-302.
47. Kortum AN, Rodriguez-Nunez I, Yang J, et al. Differential expression and ligand binding indicate alternative functions for zebrafish polymeric immunoglobulin receptor (pIgR) and a family of pIgR-like (PIGRL) proteins. *Immunogenetics*. 2014;66(4):267-279.
48. Macaulay IC, Svensson V, Labalette C, et al. Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep*. 2016;14(4):966-977.
49. Tang Q, Iyer S, Lobbardi R, et al. Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J Exp Med*. 2017;214(10):2875-2887.
50. Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat Commun*. 2017;8(1):2045.
51. Li J, Barreda DR, Zhang YA, et al. B lymphocytes from early vertebrates have potent phagocytic and microbicidal abilities. *Nat Immunol*. 2006;7(10):1116-1124.
52. Traver D, Paw BH, Poss KD, Penberthy WT, Lin S, Zon LI. Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol*. 2003;4(12):1238-1246.
53. Bennett CM, Kanki JP, Rhodes J, et al. Myelopoiesis in the zebrafish, Danio rerio. *Blood*. 2001;98(3):643-651.
54. Ferrero G, Gomez E, Lyer S, et al. The macrophage-expressed gene (mpeg) 1 identifies a subpopulation of B cells in the adult zebrafish. *J Leukoc Biol*. 2020;107(3):431-443.
55. Moss LD, Monette MM, Jaso-Friedmann L, et al. Identification of phagocytic cells, NK-like cytotoxic cell activity and the production of cellular exudates in the coelomic cavity of adult zebrafish. *Dev Comp Immunol*. 2009;33(10):1077-1087.
56. Olsson A, Venkatasubramanian M, Chaudhri VK, et al. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature*. 2016;537(7622):698-702.
57. Drissen R, Buza-Vidas N, Woll P, et al. Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nat Immunol*. 2016;17(6):666-676.
58. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. 2017;19(4):271-281.
59. Graf T, Enver T. Forcing cells to change lineages. *Nature*. 2009;462(7273):587-594.
60. Hoppe PS, Schwarzfischer M, Loeffler D, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*. 2016;535(7611):299-302.
61. Lin HF, Traver D, Zhu H, et al. Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. *Blood*. 2005;106(12):3803-3810.
62. Vinh DC, Patel SY, Uzel G, et al. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. *Blood*. 2010;115(8):1519-1529.
63. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*. 2011;118(10):2653-2655.
64. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*. 2011;43(10):1012-1017.

	WT	*gata2b*$^{-/-}$
Scatter analysis in freq. of single viable cells	**N = 26**	**N = 23**
Erythrocytes in KM (3 to 12 months)	25.4 ± 2.8	25.9 ± 3.0
Progenitors in KM (3 to 12 months)	26.8 ± 2.0	24.4 ± 1.9
Lymphocytes and HSCs in KM (3 to 12 months)	18.1 ± 1.1	29.0 ± 1.9****
Myeloid cells in KM (3 to 12 months)	20.3 ± 1.4	10.3 ± 0.9****
Reporter expression in freq. of single viable cells		
Tg(*mpx*:GFP) (GFP+ in single live cells)	17.2 ± 3.0 (n=8)	5.1 ± 1.0 (n=7)**
Tg(*mpeg*:GFP)^{gl22Tg} (GFP+ in single live cells)	3.5 ± 1.9 (n=4)	2.1 ± 0.7(n=6)
Tg(*LCK*:GFP) (GFP+ in single live cells)	3.1 ± 0.8 (n =4)	3.4 ± 0.6 (n=4)
Tg(CD41 :GFP) (GFP low in single live cells)	0.4 ± 0.1 (n=9)	0.4 ± 0.1 (n=8)

Table 1. Adult hematopoietic cell quantitations.
The data are mean± SEM. KM; kidney marrow, n= number of zebrafish used in analysis. *P< 0.05, **P< 0.01, ***P<0.001, ****P<0.0001. If data are normally distributed we used One-way ANOVA with Tukey post-test. If data are not normally distributed we used Kruskal-Wallis with Dunn's post-test.

SUPPLEMENTARY METHODS

Single cell transcriptome analysis of the progenitor compartment of WT and gata2b$^{-/-}$ zebrafish allows for unbiased lineage investigation

Kidney marrow from two 5 mpf female *Tg(CD41:GFP)* WT or *gata2b*$^{-/-}$ zebrafish were pooled and sorted and experiment was performed in replicate (Figure S2A-C). In total 70.000 cells from the gate in panel B were sorted and between 214 and 1607 CD41:GFP^{low} cells were added to this pool in PBS/10%FSC/2%BSA/2% carp serum. This resulted in final CD41:GFP^{low} percentage of 0.45 – 2.73%. The sorting strategy also included the CD41:GFP^{high} expressing cells which were previously identified as thrombocytes [1] (Figure S2C) so for every CD41:GFP^{low} cell, 1.9 CD41:GFP^{high} cells were present in our final population. 7630 cells in WT1, 4033 in WT2, 3675 in *gata2b*$^{-/-}$ 1 and 5229 in *gata2b*$^{-/-}$ 2 were obtained after quality control (Figure S2D-F), with a read depth of approximately 50.000 reads per cell. Gross differences in cell numbers between WT and *gata2b*$^{-/-}$ cells may influence cluster identification when a nearest-neighbor algorithm is used. Therefore, all replicates were randomly down-sampled to 3675 to match each other. First the replicates were aligned using anchor based integration and then the WT and *gata2b*$^{-/-}$ samples were aligned using the same method to correctly identify the clusters, avoiding batch specific differences. We could identify 20 different cell clusters using the R Seurat package[2](Figure S2G).

Cluster identification

Using the FindMarkers function in Seurat, differentially expressed genes were identified compared to the other clusters. Subsequently the functions FeaturePlot and VlnPlot were used to analyse gene expression patterns between clusters to test validity and exclusivity of individual clusters. Finally, marker analysis was visualized using the DoHeatmap function. Importantly, known Gata2 target genes like *cebpb* (p = 0,0002) and *alas2* (p = 4.77E-53) were found significantly downregulated in gata2b mutants.

Erythroid cells, thrombocytes, neutrophils, monocytes and their progenitors were identified by comparing our data to known single cell expression analysis and known lineage markers[3-7]. Canonical lineage differentiation markers are generally low expressing transcription factors and are poorly amplified by droplet based single cell sequencing methods, therefore we have presented the lineage differentiation using a combination of canonical lineage differentiation markers and high differentially expressed genes between clusters as identified by the FindMarkers function of Seurat and presented in a heatmap (Figure S2H). In short, cells of the erythroid lineage are known to express hemoglobins like *hbba1. itga2b* was used as a marker for thrombocytes in the form of *Tg(CD41:GFP)*, *mpeg1.1* for monocytes and *lysozyme (lyz)* for neutrophils (Figure S2H). Furthermore, the differentiated populations were devoid of expression of proliferation genes like *myca* and *pcna,* indeed suggesting that these are differentiated cells and not a progenitor compartment

(Figure S2H). Interestingly, one population expressed markers like *CCAAT enhancer binding protein alpha* (*cebpa)* and *granulin1 (grn1)* (Figure 3B) which are expressed in the monocyte lineage. This population also showed very high expression levels of *s100a10b* (Figure S3A). *In situ* hybridization confirmed that macrophages expressed high levels of *s100a10b* (Figure S3A, B).

The sorting strategy for the kidney progenitor population also contains cells of the lymphoid lineage. A lymphoid progenitor population expressing *rag1* was found. Furthermore, this population also expressed the *rag1* homologue *topoisomerase 2a (top2a)* (Figure S2H), probably as an alternative mechanism in V(D)J recombination[8], and high expression of proliferation markers *myca* and *pcna* suggestive of the lymphoid progenitors. T cells were marked by the expression of *tox, il2rb* and *dusp2*, NK cells were marked by *nkl.3, nkl.4* and *ccl33.3* and two distinct populations of B-cells were marked by expression of *CD37, CD79a* and *pax5* (Figure S2H, Figure 3C). Interestingly, the second B-cell population showed high levels of the immunoglobulins *ighv1-4* and *ighz*[7,9] but did not express the proliferative marker *pcna*, indicating this is a more mature or activated population of B-cells (Figure S2H). Analysis of IgM:GFP transgenic zebrafish [10] showed indeed the presence of several B-cell populations (Figure 4L-S, Figure S4G-L).

One cluster was marked by high expression of *pigrl3.5* which encodes a polymeric Ig receptor[11]. The *pigrl3.5* expressing cells are in close association in the UMAP with the NK population suggesting this is a lymphoid population (Figure S2G, H).

A small cluster with expression of *epithelial cell adhesion molecule (epcam)* was detected (Figure S2H). We hypothesize that this may be niche cells, but since this population is very small and our sorting strategy was not meant to obtain the niche cells, we have not further investigated this population.

Proportional difference between WT and *gata2b*$^{-/-}$ cells in clusters is calculated using scProportionTest package (Figure 3M, 5D and S6D)[12]

HSPC identification

The most immature population would be an hematopoietic stem or progenitor cell (HSPC). These cell types are marked by little expression of lineage markers and expression of proliferation markers. In zebrafish *fli1a* and *meis1b* are typical HSC markers[3-5]. The HSPC1 and HSPC2 populations meet these criteria by expressing low levels of lineage markers (Figure 3B-D). These populations showed high expression levels of proliferation markers like *myca, pcna* and *mki67* (Figure 3J, Figure S2H) and some cells with high levels of the stem cell markers *fli1a* and *meis1b* (Figure 3F, G). Putative HSC, represented by CD41:GFP^{low} sorted cells, are present in both the HSPC1 and HSPC2 clusters (Figure 3I). As expected, the GFP high expressing cells were thrombocytes. *gata2b* gene expression was enriched in the HSPC1 cluster compared to the other clusters (Figure 3H), indicating that this population would be most affected by the loss of Gata2b. Also, highest expression of *fli1a* and *meis1b* were found in HSPC1 (Figure 3F, G) indicating that this cluster contains some HSCs.

Lineage trajectory analysis

After defining the clusters we performed trajectory and pseudotime analysis using Monocle 3 on the integrated data set and HSPC1 cluster, both carrying Seurat embeddings[13]. Monocle 3 uses an algorithm to learn the differentiation trajectory according to the gene expression of each cell. Once the trajectory graph was learned (Figure 3K and 5H), we used get_earliest_principal_node function and chose the “root” to produce the pseudotime graph. For whole data set HSPC1 cluster was chosen as a starting point and for HSPC1 cluster quiescent subcluster was chosen to generate the pseudotime graphs (Figure 3L, 5I). We used plot_genes_in_pseudotime function to identify the gene expression in pseudotime and found that *gata2b* expression is the highest in the most immature HSPCs and decreases with differentiation (Figure 3N). We also showed that when HSPC1 cluster is chosen as a starting point, the expression of lineage specific genes such as *CD37*, *grn1* and *hbba1* were increased in pseudotime (Figure 3O-Q) which shows that HSPC1 cluster is indeed the most immature cluster within the whole dataset. Similarly, in HSPC1 cluster, when quiescent subcluster was chosen as a starting point, we found that the expression of stem cell marker *meis1b*, also *gata2b* and *CD41:GFP* were decreased in pseudotime, confirming the differentiation trajectory for these cells (Figure 5J-L).

Table S1.

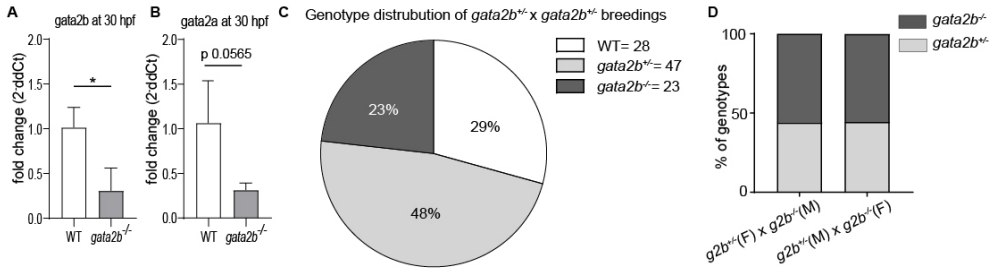
	WT	*gata2b*$^{+/-}$	*gata2b*$^{-/-}$
EHT events			
AGM (32 to 40hpf)	3.3 ± 0.4 (n=18)	2.9 ± 0.3 (n=33)	2.2 ± 0.4 (n=18)
CD41:GFP^{+}Flt1:RFP^{+} cells			
CHT (52-54hpf)	8.5 ± 1.7 (n=8)	9.0 ± 1.5 (n=20)	11.9 ± 1.0 (n=7)
CHT (58-60hpf)	25.2 ± 3.2 (n=16)	28.7 ± 1.9 (n=28)	23.3 ± 2.9 (n=12)
CHT (76-58hpf)	55.5 ± 3.8 (n=13)	43.4 ±5.6 (n=12)	36.5 ± 3.0 (n=14)**
***cmyb* expression intensity**			
AGM (24 hpf)	33.0 ± 6.3 (n=14)	28.5 ± 4.5 (n=39)	30.6 ± 2.6 (n=13)
AGM (33 hpf)	32.9 ± 10.9 (n=16)	30.9 ± 7.9 (n=31)	28.5 ± 6.5(n=28)**
CHT (56 hpf)	10.5 ± 2.5 (n=19)	12.8 ± 0.5 (n=41)	4.2 ± 1.1 (n=14)**
CHT (76 hpf)	21.3 ± 7.4 (n=19)	20.8 ± 1.125 (n=43)**	16.142 ± 1.950 (n=15)*
***Runx1* expression intensity**			
AGM (26 hpf)	24.1 ± 9.6 (n= 10)	23.7 ± 11.6 (n= 11)	25.4 ± 9.3 (n=11)
AGM (36 hpf)	30.0 ± 7.1 (n=18)	23.1 ± 8.1 (n=20)	25.3 ± 6.8 (n=11)
mpeg:GFP^{+} cells			
CHT (54hpf)	513.9±23.67 (n=18)	564.9±23.51 (n=26)	545.9±33.27 (n=14)
mpx:GFP^{+} cells			
CHT (75hpf)	227.5±14.40 (n=11)	221.1±8.63 (n=26)	235.0±15.49 (n=6)
LCK:GFP^{+} pixel number			
Thymus (5dpf)	2093±186.2 (n=9)	2449±165.5 (n=18)	2354±206.8 (n=17)

The data are mean ± SEM. n= number of zebrafish embryos used in analysis. AGM; aorta-gonad-mesonephros region, CHT; caudal hematopoietic tissue. *P< 0.05, **P< 0.01. If data were normally distributed we used One-way ANOVA with Tukey post-test. If data were not normally distributed we used Kruskal-Wallis with Dunn's post-test.

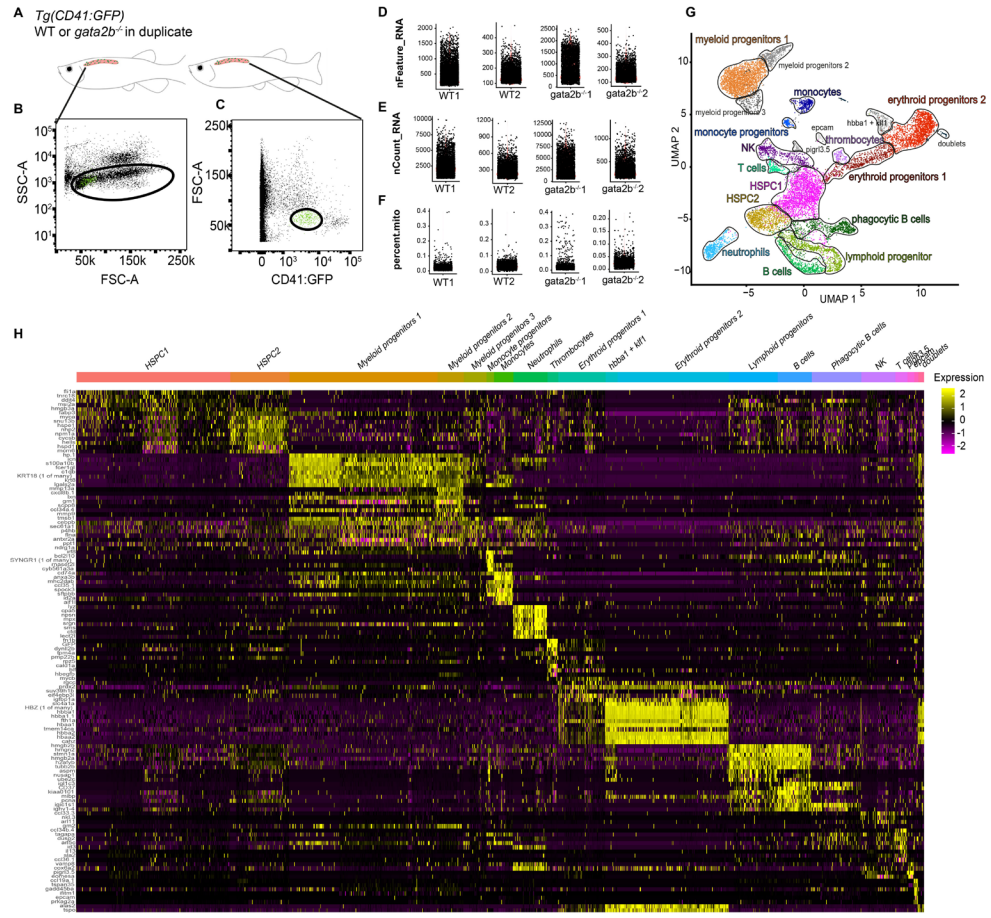
Table S2. Cell numbers within the different populations in single cell analysis

	HSPC1	HSPC2	Myeloid progenitors 1	Myeloid progenitors 2	Myeloid progenitors 3	Monocyte progenitor	Monocytes	Neutrophils	Trombocytes	Ery progenitors	hbba1 + klf1	Erythrocytes	Lymphoid progenitors	B-cells	Phagocytic B-cells	Nk cells	T cells	pigrl3.5	epcam	doublets
WT 2	588	188	1110	300	146	33	1	216	39	331	6	301	105	44	58	164	21	5	3	16
WT 1	498	110	822	31	109	59	30	142	55	112	168	1008	88	65	149	109	34	20	21	45
gata2b$^{-/-}$2	846	102	410	61	71	119	41	55	19	202	9	148	267	201	238	99	60	20	9	14
gata2b$^{-/-}$1	572	570	77	34	44	110	49	144	60	127	20	361	332	248	362	185	84	64	38	20
Total	2504	970	2419	426	370	121	321	557	173	772	203	1818	792	558	807	557	199	109	71	95

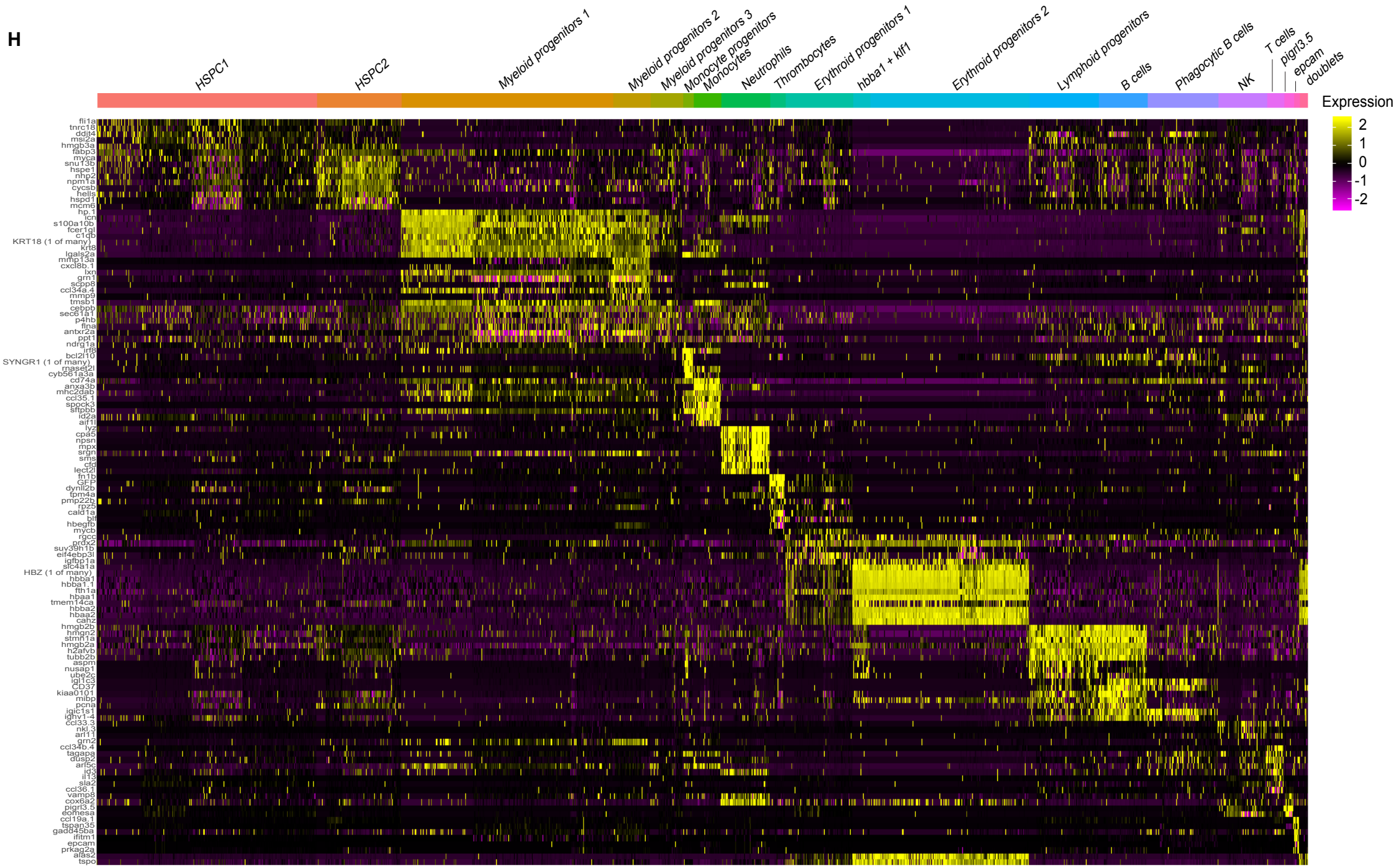
SUPPLEMENTARY FIGURES

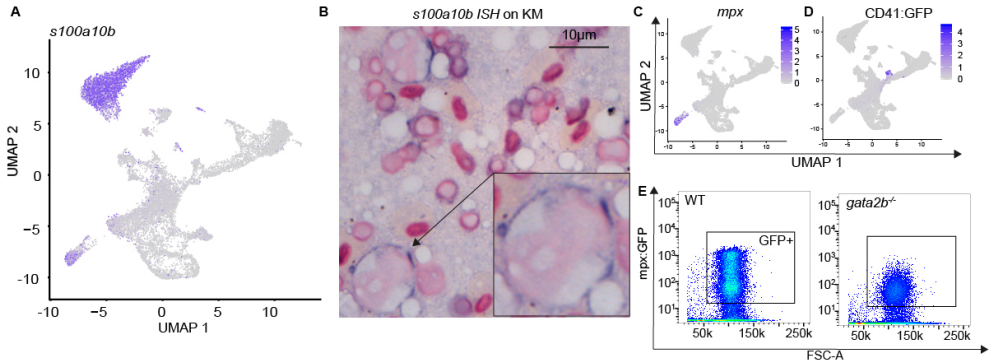


Supplementary Figure 1. Maternal contribution of *gata2b* does not affect *gata2b*$^{-/-}$ survival. A) qRT-PCR of *gata2b* references against *elfa* on pooled WT and *gata2b*$^{-/-}$ embryos at 30 hpf (n = 3) indicating that *gata2b* expression levels are significantly reduced in mutant embryos (p = 0.0218). B) qRT-PCR of *gata2a* references against *elfa* on pooled WT and *gata2b*$^{-/-}$ embryos at 30 hpf (n = 3) shows reduced *gata2a* expression levels are in mutant embryos (p = 0.0565). C) Genotype distribution of matings between *gata2b*$^{+/-}$ and *gata2b*$^{+/-}$ zebrafish. D) Genotype distribution of matings between *gata2b*$^{+/-}$ female and *gata2b*$^{-/-}$ zebrafish males or *gata2b*$^{+/-}$ male and *gata2b*$^{-/-}$ zebrafish females. *; P<0.05.



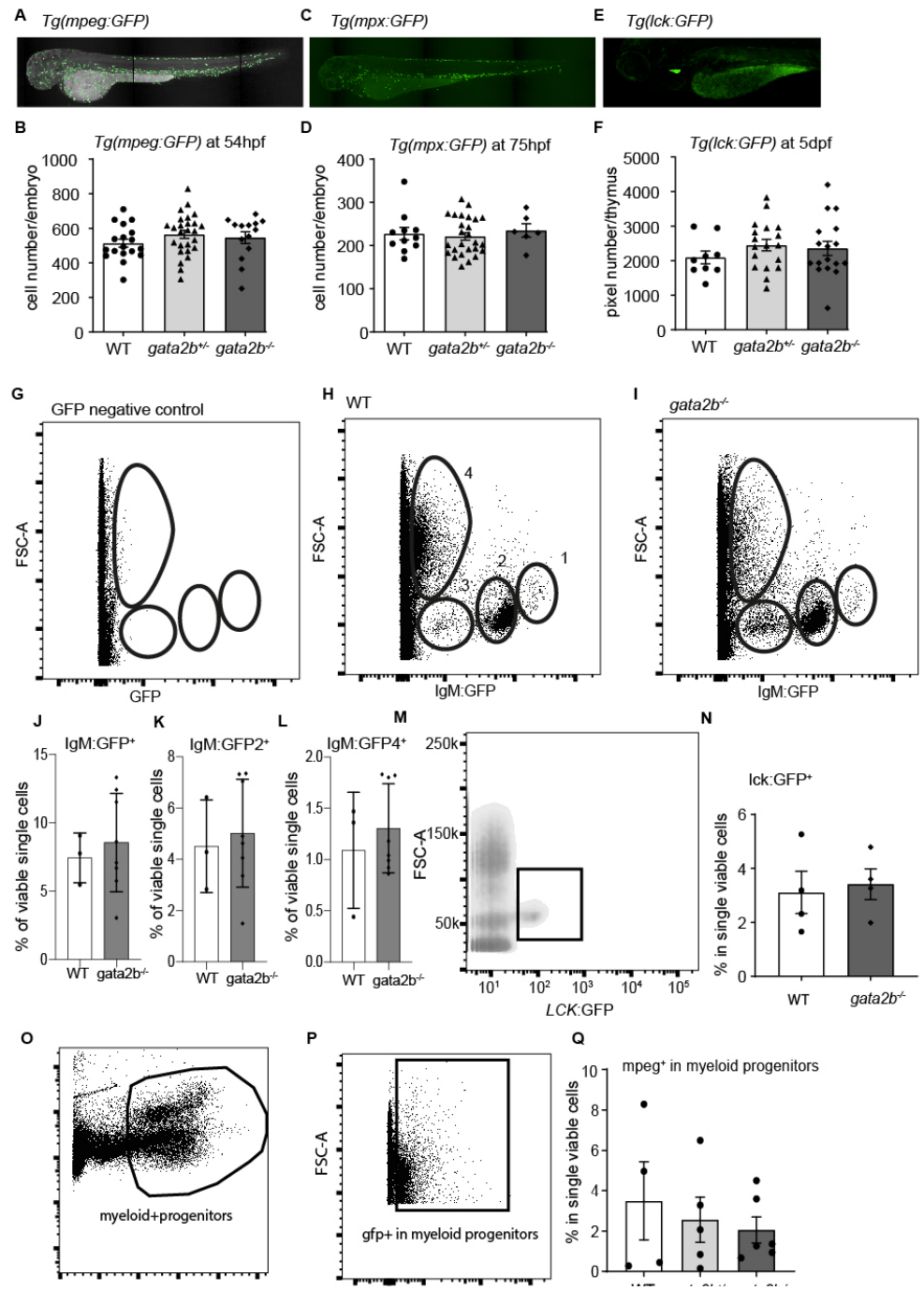
Supplementary Figure 2. Single cell RNA sequencing reveals several progenitor populations. A-C) Experimental strategy to obtain single cells for RNA sequencing. B) FACS plot indicating the progenitor population which was sorted and supplemented with the remaining C) CD41:GFP^{low} expressing cells from the kidney marrow pool of cells. D-F) Quality control parameters for each sample G) UMAP showing cluster analysis on aggregated data set of both WT and *gata2b*$^{-/-}$ cells indicating 20 different clusters with different colors. H)Heatmap showing top10 marker genes for each cluster calculated in an unbiased way. I) Genotype distribution of each of the clusters, area of the bars indicate the cell numbers in each cluster, white = WT, black = *gata2b*$^{-/-}$. Each replicate is depicted in a separate bar. J) Differentiation trajectory and pseudotime calculated only for WT cells. K) Differentiation trajectory and pseudotime calculated only for *gata2b*$^{-/-}$ cells.

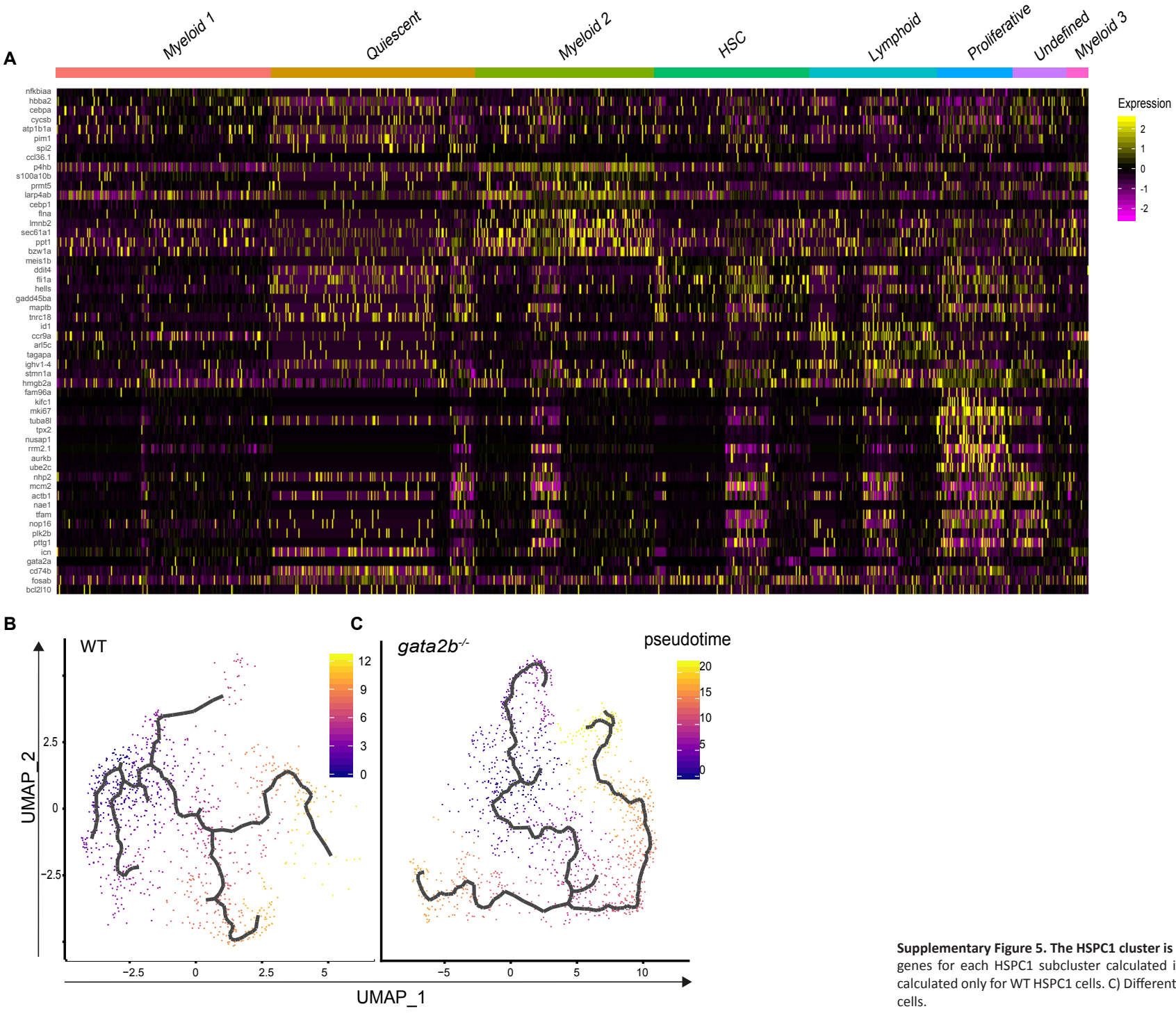

H
HSPC1
HSPC2
Myeloid progenitors 1
Myeloid progenitors 2
Myeloid progenitors 3
Monocyte progenitors
Monocytes
Neutrophils
Thrombocytes
Erythroid progenitors 1
hbba1 + klf1
Erythroid progenitors 2
Lymphoid progenitors
B cells
Phagocytic B cells
NK
T cells
pigri3.5
epcam
doublets
Expression
2
1
0
-1
-2
fli1a
tnrc18
ddit4
msi2a
hmgb3a
fabp3
myca
snu13b
hspe1
nhp2
npm1a
cycsb
hells
hspd1
mcm6
hp.1
lcn
s100a10b
fcer1gl
c1qb
KRT18 (1 of many)
krt8
lgals2a
mmp13a
cxcl8b.1
lxn
grn1
scpp8
ccl34a.4
mmp9
tmsb1
cebpb
sec61a1
p4hb
flna
antxr2a
ppt1
ndrg1a
irf8
bcl2l10
SYNGR1 (1 of many)
rnaset2l
cyb561a3a
cd74a
anxa3b
mhc2dab
ccl35.1
spock3
sftpbb
id2a
aif1l
lyz
cpa5
npsn
mpx
srgn
sms
cfd
lect2l
fn1b
GFP
dynll2b
tpm4a
pmp22b
rpz5
caid1a
blf
hbegfb
mycb
rgcc
prdx2
suv39h1b
eif4ebp3l
igfbp1a
slc4a1a
HBZ (1 of many)
hbba1
hbba1.1
fth1a
hbaa1
tmem14ca
hbba2
hbaa2
cahz
hmgb2b
hmgn2
stmn1a
hmgb2a
h2afvb
tubb2b
aspm
nusap1
ube2c
iglf1c3
CD37
kiaa0101
mibp
pcna
igic1s1
ighv1-4
ccl33.3
nkl.3
arl11
grn2
ccl34b.4
tagapa
dusp2
arl5c
id3
il13
sla2
ccl36.1
vamp8
cox6a2
pigri3.5
eomesa
ccl19a.1
tspan35
gadd45ba
ifitm1
epcam
prkag2a
alas2
tspo



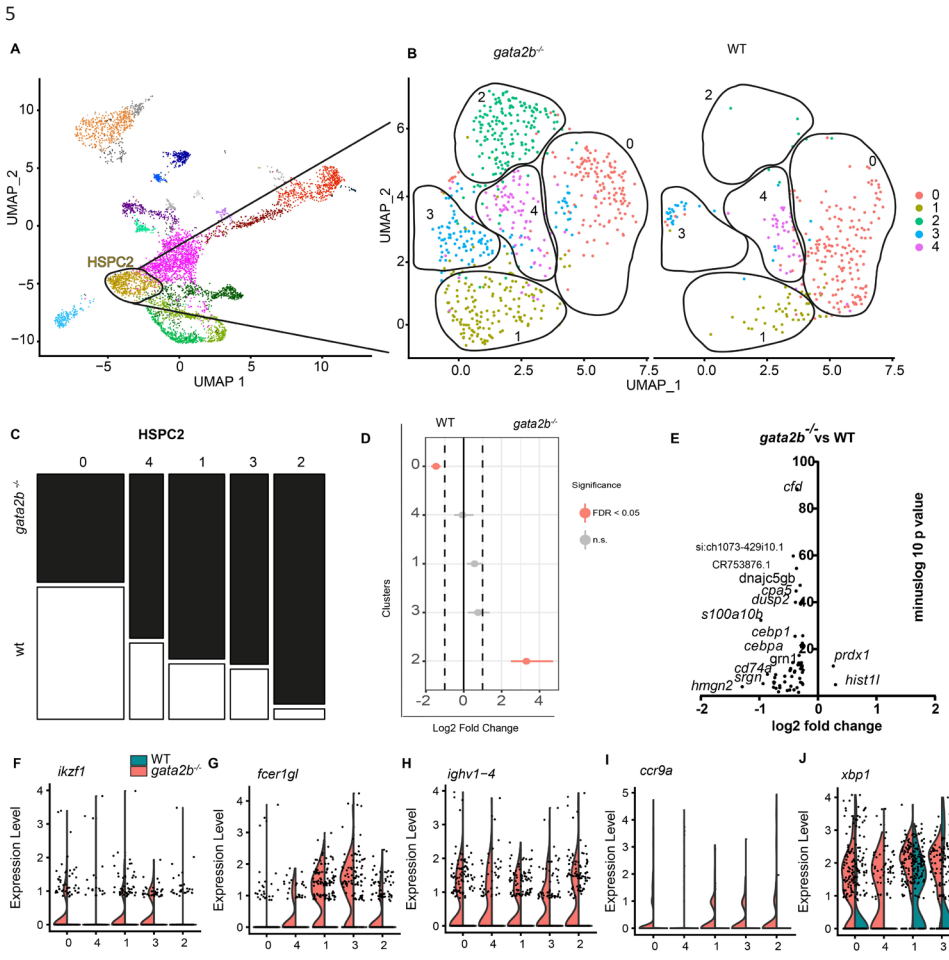
Supplementary Figure 3. *s100a10b* is expressed in the neutrophil lineage. A) Feature analysis with gradual gene expression in shades of blue of *s100a10b*. B) WT representative images of *s100a10b in situ* hybridization on zebrafish kidney marrow smears. A banded neutrophil is indicated by the arrow and shown enlarged in the corner. Scale bar indicates 10 µm. UMAP = Uniform manifold approximation and Projection. C) Feature analysis showing *mpx* expression in the neutrophil cluster. D) Feature analysis showing high *CD41:*GFP expression in the thrombocytes cluster and low *CD41:*GFP expression in the HSPC1 and HSPC2 clusters. E) Gating strategy for *Tg(mpx:GFP)* in WT and *Gata2b*$^{-/-}$ KM.

Supplementary Figure 4. Differentiation markers are not altered in *gata2b*$^{-/-}$ embryos.
A) Representative picture of *Tg(mpeg:GFP)* embryo at 54 hpf. B) Quantitation of *mpeg*:GFP^{+} cells in WT, *gata2b*$^{+/-}$ and *gata2b*$^{-/-}$ embryos at 54 hpf. C) Representative picture of *Tg(mpx:GFP)* embryos at 75 hpf. D) Quantitation of *mpx*:GFP^{+} cells in WT, *gata2b*$^{+/-}$ and *gata2b*$^{-/-}$ embryos. E) Representative picture of *Tg(lck:GFP)* embryos at 5 dpf F) Lck:GFP^{+} area represented as pixel number in WT, *gata2b*$^{+/-}$ and *gata2b*$^{-/-}$ embryos. Each dot represent one embryo, see Table S1 for exact cell numbers and numbers of embryos analyzed. G) Negative GFP control. H) WT *Tg(IgM:GFP)* zebrafish KM with 4 GFP populations gated with clear differences in size (FSC) or GFP positivity, I) Similar gating strategy for *Gata2b*$^{-/-}$ *Tg(IgM:GFP)* KM. J-L) Quantitation of total IgM:GFP^{+}, IgM:GFP2^{+} and IgM:GFP4^{+} populations for WT and *Gata2b*$^{-/-}$ KM cells as percentage of single viable cells. M) Gating strategy for *Tg(lck:GFP)* WT and *Gata2b*$^{-/-}$ KM. N) Quantitation of GFP^{+} cells in *Tg(lck:GFP)* WT and *Gata2b*$^{-/-}$ KM as percentage of single viable cells. O-P) Gating strategy of *Tg(mpeg1.1:GFP)* WT and *Gata2b*$^{-/-}$ KM. Q) Quantitation of GFP^{+} cells in *Tg(mpeg1.1:GFP)* WT, *gata2b*$^{+/-}$ and *Gata2b*$^{-/-}$ KM as percentage of single viable cells. Error bars represent SEM.





Supplementary Figure 5. The HSPC1 cluster is composed of multiple HSPC subtypes. A) Heatmap showing marker genes for each HSPC1 subcluster calculated in an unbiased way. B) Differentiation trajectory and pseudotime calculated only for WT HSPC1 cells. C) Differentiation trajectory and pseudotime calculated only for *gata2b*$^{-/-}$ HSPC1 cells.



Supplementary Figure 6. HSPC2 shows reduced myeloid differentiation and increased in lymphoid differentiation. A) Selection of HSPC2 for further analysis in B) subclusters (0-4) of *gata2b*$^{-/-}$ cells on the left and WT cells on the right. C) Proportion analysis of the different subclusters, indicating unequal distribution of WT and *gata2b*$^{-/-}$ cells in the individual subclusters. D) Pointrange plot showing the difference between proportion of WT and *gata2b*$^{-/-}$ cells for each HSPC2 subcluster calculated by permutation test. If FDR < 0.05, point is colored in pink and if not in grey. E) Volcanoplot showing differential gene expression analysis between WT and *gata2b*$^{-/-}$ cells showing robust downregulation of myeloid genes in *gata2b*$^{-/-}$ cells. F-J) violin plots of individual lymphoid gene expression within subclusters of HSPCs with WT cells in green and *gata2b*$^{-/-}$ cells in pink of F) *ikzf2*, G) *fcer1gl*, H) *ighv1-4*, I) *ccr9a* and J) *xpb1*.

SUPPLEMENTARY REFERENCES

1. Ma D, Zhang J, Lin HF, Italiano J, Handin RI. The identification and characterization of zebrafish hematopoietic stem cells. *Blood*. 2011;118(2):289-297.
2. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36(5):411-420.
3. Macaulay IC, Svensson V, Labalette C, et al. Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep*. 2016;14(4):966-977.
4. Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat Commun*. 2017;8(1):2045.
5. Tang Q, Iyer S, Lobbardi R, et al. Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J Exp Med*. 2017;214(10):2875-2887.
6. Carmona SJ, Teichmann SA, Ferreira L, et al. Single-cell transcriptome analysis of fish immune cells provides insight into the evolution of vertebrate immune cell types. *Genome Res*. 2017;27(3):451-461.
7. Moore JC, Tang Q, Yordan NT, et al. Single-cell imaging of normal and malignant cell engraftment into optically clear prkdc-null SCID zebrafish. *J Exp Med*. 2016;213(12):2575-2589.
8. Riou JF, Grondard L, Petitgenet O, Abitbol M, Lavelle F. Altered topoisomerase I activity and recombination activating gene expression in a human leukemia cell line resistant to doxorubicin. *Biochem Pharmacol*. 1993;46(5):851-861.
9. Danilova N, Bussmann J, Jekosch K, Steiner LA. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat Immunol*. 2005;6(3):295-302.
10. Page DM, Wittamer V, Bertrand JY, et al. An evolutionarily conserved program of B-cell development and activation in zebrafish. *Blood*. 2013;122(8):e1-11.
11. Kortum AN, Rodriguez-Nunez I, Yang J, et al. Differential expression and ligand binding indicate alternative functions for zebrafish polymeric immunoglobulin receptor (pIgR) and a family of pIgR-like (PIGRL) proteins. *Immunogenetics*. 2014;66(4):267-279.
12. Miller SA, Policastro RA, Sriramkumar S, et al. LSD1 promotes secretory cell specification to drive BRAF mutant colorectal cancer. *bioRxiv*. 2020:2020.2009.2025.313536.
13. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019;566(7745):496-502.

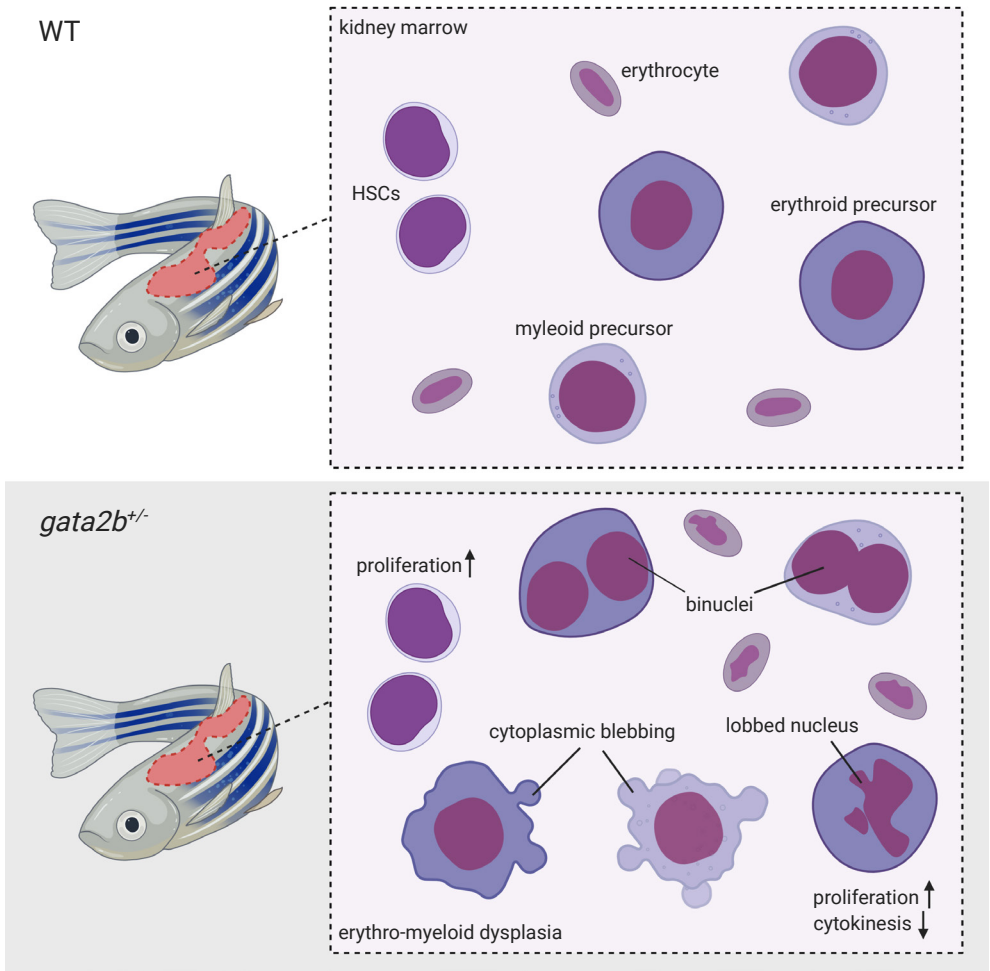
4

Gata2b haploinsufficiency causes aberrant transcriptional signatures in HSPCs resulting in myeloid and erythroid dysplasia in zebrafish

Emanuele Gioacchino[1], Cansu Koyunlar[1], Joke Zink[1], Hans de Looper[1,2], Kirsten J. Gussinklo[1], Remco Hoogenboezem[1], Dennis Bosch[1], Eric Bindels[1], Ivo P. Touw[1] and Emma de Pater[1,2,*].

[1] Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
[2] Cancer Genome Editing Center, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
* Corresponding author

Manuscript in submission



ABSTRACT

The transcription factor *GATA2* has pivotal roles in hematopoiesis. Germline *GATA2* mutations result in GATA2 haploinsufficiency characterized by immunodeficiency, bone marrow failure and predispositions to myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). Clinical symptoms in GATA2 patients are diverse and mechanisms driving *GATA2* related phenotypes are largely unknown. To explore the impact of GATA2 haploinsufficiency on hematopoiesis, we generated a zebrafish model carrying a heterozygous mutation in *gata2b*, an orthologue of *GATA2*. Morphological analysis revealed progression of myeloid and erythroid dysplasia in *gata2b*$^{+/-}$ kidney marrow (KM). Single cell RNA sequencing on KM cells showed that the erythroid dysplasia in *gata2b*$^{+/-}$ zebrafish was preceded by a differentiation block in erythroid progenitors, hallmarked by downregulation of cytoskeletal transcripts, aberrant proliferative signatures and ribosome biogenesis. Additionally, transcriptional and functional analysis of Gata2b haploinsufficient hematopoietic stem cells (HSCs) indicated that proliferative stress within the HSC compartment possibly contributes to the development of myeloid and erythroid dysplasia in *gata2b*$^{+/-}$ zebrafish.

INTRODUCTION

The transcription factor GATA2 plays a major role in the generation and maintenance of the hematopoietic system[1-3]. In humans, heterozygous germline mutations in *GATA2* often lead to loss of function of one allele, causing GATA2 haploinsufficiency. The clinical features of GATA2 haploinsufficiency are broad and include immunodeficiency, pulmonary-, vascular- and/or lymphatic dysfunctions and a strong propensity to develop MDS or AML[4,5]. GATA2 haploinsufficiency has been linked to the pathogenesis of Emberger syndrome[6], Monocytopenia and Mycobacterium Avium Complex (MonoMAC) syndrome[7], dendritic cell, monocyte, B- and natural killer cell deficiency (DCML)[8], and familial forms of AML[9]. In addition to the various disease phenotypes, the risk of developing MDS/AML in GATA2 patients is approximately 80%[4,5]. There is no clear correlation between the occurrence of *GATA2* mutations and the severity of hematopoietic deficiencies, even among family members who share the same mutation[10,11]. Therefore, it is essential to gain insight into the mechanism of underlying GATA2 deficiencies in well-defined experimental models.

In mice, Gata2 has an essential regulatory activity in HSC generation and maintenance. *Gata2*-null mice are lethal at embryonic day (E) 10.5[3], whereas *Gata2* heterozygous (*Gata2*$^{+/-}$) mice survive to adulthood with normal blood values. However, *Gata2*$^{+/-}$ mice have a diminished HSC compartment in the bone marrow (BM) and *Gata2*$^{+/-}$ HSCs have a reduced repopulation capacity in competitive transplantation studies[12,13]. Whereas mouse models thus emerged as a precious source to identify the function of *GATA2* in HSC generation and fitness, they leave the mechanisms causing the different aspects of GATA2 deficiency syndromes largely undiscovered. To better understand the biology of GATA2 haploinsufficiency syndromes, zebrafish serves as an attractive alternative model. Zebrafish have the advantage of having 2 *GATA2* orthologues; *gata2a* and *gata2b*. Gata2a is expressed predominantly in the vasculature[14] and is required for programming of the hemogenic endothelium[15,16]. Gata2b is expressed in hematopoietic stem/progenitor cells (HSPCs)[14] and homozygous deletion (*gata2b*$^{-/-}$) redirects HSPC differentiation to the lymphoid lineage in expense of myeloid differentiation causing a lymphoid bias with an incomplete B-cell differentiation in the KM, thus mimicking a portion of the GATA2 haploinsufficiency phenotypes found in patients. Additionally, the most primitive HSC compartment was lost in *gata2b*$^{-/-}$ KM, but none of the animals displayed signs of dysplasia[16,17]. Because patients carry heterozygous rather than homozygous *GATA2* mutations, we hypothesize that Gata2b haploinsufficiency could be a driver for leukemia onset. To investigate this, we assessed hematopoietic cell differentiation in *gata2b* heterozygous zebrafish (*gata2b*$^{+/-}$) KM. Upon morphological assessment of KM smears, we observed erythroid dysplasia in all *gata2b*$^{+/-}$ but not in WT zebrafish and myeloid dysplasia in 25% of the *gata2b*$^{+/-}$ zebrafish. Single cell transcriptome analysis revealed downregulation of cytoskeletal transcripts, aberrant proliferative signatures and ribosome biogenesis in erythroid progenitor cells. Further characterization of the HSC transcriptome indicated increased proliferative stress in the hematopoietic system originating in the HSC compartment, possibly underlying the erythroid dysplasia in *gata2b*$^{+/-}$ zebrafish. These results highlight a concentration-dependent function of Gata2b in zebrafish hematopoiesis.

MATERIAL AND METHODS

Generation and genotyping of Gata2b heterozygous zebrafish

Gata2b$^{+/-}$ and wild type (WT) clutch mates were used for all analyses[16] and animals were maintained under standard conditions. A knockout allele was generated by introducing a 28bp out-of-frame insertion in exon 3 as previously described[16].

Zebrafish embryos were kept at 28,5°C on a 14h/10h light-dark cycle in HEPES-buffered E3 medium. Zebrafish were anesthetized using tricaine and euthanized by ice-water. Animal studies were approved by the animal Welfare/Ethics Committee in accordance to Dutch legislation.

Kidney marrow isolation and analysis

Kidney marrow was dispersed mechanically using tweezers and dissociated by pipetting in phosphate buffered saline (PBS)/10% fetal calf serum (FCS) to obtain single-cell suspensions as previously described[16]. The KM cells were sorted in non-stick cooled micro tubes (Ambion) containing 10% FCS in PBS. Proliferation was assessed by anti-Ki67 staining in fixed (4% paraformaldehyde) KM cells. 7-AAD (7-amino-actinomycin D) (Stem-Kit Reagents) 0.5mg/L or DAPI 1mg/L was used for live/dead discrimination. FACS sorting and analysis were performed using FACS AriaIII (BD Biosciences).

May–Grünwald-Giemsa stain of KM smears

Kidney marrow smears were fixed in 100% MeOH before staining in May Grünwald solution (diluted 1:1 in phosphate buffer) and Giemsa solution (diluted 1:20 in phosphate buffer) followed by a last rinsing step in tap water.

Morphological analysis was performed by counting 200-500 hematopoietic cells of each kidney marrow smear; excluding mature erythrocytes and thrombocytes. Cells were categorized as: blast, myelocyte, neutrophil, eosinophil, lymphocyte or erythroblast. Furthermore, if dysplasia was observed within a specific lineage, the percentage of dysplastic cells within that lineage was determined by additional counting of at least 50 cells within that specific lineage.

Single cell RNA sequencing

Kidney marrow cells were isolated and 7x10^4 viable cells were sorted from 2 pooled

Tg(CD41:GFP[18]; runx1:DsRed[19]) wild type (WT) or *gata2b*$^{+/-}$ male zebrafish at 1 year of age. For additional replicates, 7x10^4 single viable cells were sorted from kidney marrows of one WT or *gata2b*$^{+/-}$ female zebrafish between 18-20 months of age. cDNA was prepared using the manufacturers protocol (10x Chromium V3) and sequenced on a Novaseq 6000 instrument (Illumina). After sample processing and quality control analysis, 18147 cells for WT and 10849 cells for *gata2b*$^{+/-}$ were processed for further analysis. The read depth was over 20K reads for all replicates. Data was analyzed using the Seurat and Monocle3 R packages[20,21].

RNA extraction and RNA quality control

Total sample RNA isolation was performed according to the standard protocol of RNA isolation with Trizol and GenElute LPA (Sigma). Quality and quantity of the total RNA was determined on a 2100 Bioanalyzer (Agilent) using the Agilent RNA 6000 Pico Kit.

RNA Sequencing and gene set enrichment analysis (GSEA)

cDNA was prepared using the SMARTer Ultra Low RNA kit (Clontech) for Illumina Sequencing. The gene expression values were measured as FPKM (Fragments per kilobase of exon per million fragments mapped). Fragment counts were determined per gene with HTSeq-count, utilizing the strict intersection option, and subsequently used for differential expression analysis using the DESeq2 package, with standard parameters, in the R environment. GSEA (Gene Set Enrichment Analysis) was performed on the FPKM values using collection of curated gene sets from Molecular Signatures Database (MSigDB). Network analysis was performed using the output from GSEA analysis in Cytoscape Software.

Statistics

Statistical analysis was carried out in GraphPad Prism 8 (GraphPad Software). Unless otherwise specified, data were analyzed using unpaired, 2-tailed Student's t-test. Statistical significance was defined as $p<0.05$. Graphs are means ± standard error of mean (SEM) and the number of replicates is indicated in the figure legend.

RESULTS

Gata2b$^{+/-}$ *kidney marrow shows erythroid and myeloid dysplasia*

We first assessed hematopoietic cell morphology in KM smears of WT and *gata2b*$^{+/-}$ zebrafish with ages ranging from 9 months post fertilization (mpf) to 18 mpf. Morphological analysis showed that, while WT zebrafish had normal KM cell morphology, all *gata2b*$^{+/-}$ KM samples had a considerable fraction of dysplastic cells in the erythroid lineage (Figure 1A, B, panel 2, 4, 6, 7 and 8). On average 0.5% of WT erythroid cells showed dysplastic features, compared to 9.9% of *gata2b*$^{+/-}$ erythroid cells (Figure 1B, C), the latter representing 4.5% of the total kidney marrow population of *gata2b*$^{+/-}$ zebrafish. Furthermore, we found evidence of myeloid lineage dysplasia in the *gata2b*$^{+/-}$ KM in 25 % of the fish (Figure 1B, panel 1, 3 and 5 and C). In these samples, 30% of myeloid cells were dysplastic compared to 0.3% in WT. While the myeloid dysplasia was mostly represented by multi-lobulated nuclei, the erythroid abnormalities were ranging from nuclear deformities and double nuclei to irregular cytoplasm or an almost complete lack of cytoplasm (Figure 1B). The remaining cell types were not affected morphologically by Gata2b haploinsufficiency. These results indicate that *gata2b* heterozygosity induces dysplasia, predominantly in erythroid progenitors.

Gata2b$^{+/-}$ *kidney marrow differentiation remains grossly intact over time*

Next, we assessed hematopoietic lineage differentiation in *gata2b*$^{+/-}$ KM by scatter analysis (Figure 1D). *Gata2b*$^{+/-}$ zebrafish KM showed no significant difference in the distribution of either mature myeloid-, erythroid-, lymphoid- or progenitor- and HSPC populations compared to WT (Figure 1E-H). Since GATA2 haploinsufficiency manifestations might require longer periods of time to become evident[4,5], we tested the effect of *gata2b* heterozygosity after aging zebrafish up to 18 months. No significant differences between the proportions of the different lineages were detected over time in *gata2b*$^{+/-}$ kidney marrow compared to WT (Figure 1E-H). However, the erythroid population increased, indiscriminately of genotype, after 12 months of age ($P<0.001$) (Figure 1G), while the remaining myeloid, lymphoid and HSPC and progenitor populations did not vary as dramatically (Figure 1E, F and H). Lineage analysis in transgenic lines specifically marking neutrophils by *Tg(mpx:GFP)* (Suppl Figure 1A-B), T-cells by *Tg(lck:GFP)* (Suppl. Figure 1C-D), B-cells by *Tg(lgm:GFP)* and *Tg(mpeg:GFP*) (Suppl. Figure 1E-F and I-J) or macrophages by *Tg(mpeg:GFP)* (Suppl. Figure 1G-H)[22,23] showed no significant alterations in lineage distribution[24-27]. However, KM smear quantification by MGG staining revealed that *gata2b*$^{+/-}$ KM had a significant decrease in eosinophils compared to WT (Figure 1I), representing less than 5% of the total KM cells. In summary, based on scatter analysis and transgenic reporters we conclude that the differentiation in major hematopoietic lineages is not altered in *gata2b*$^{+/-}$ KM.

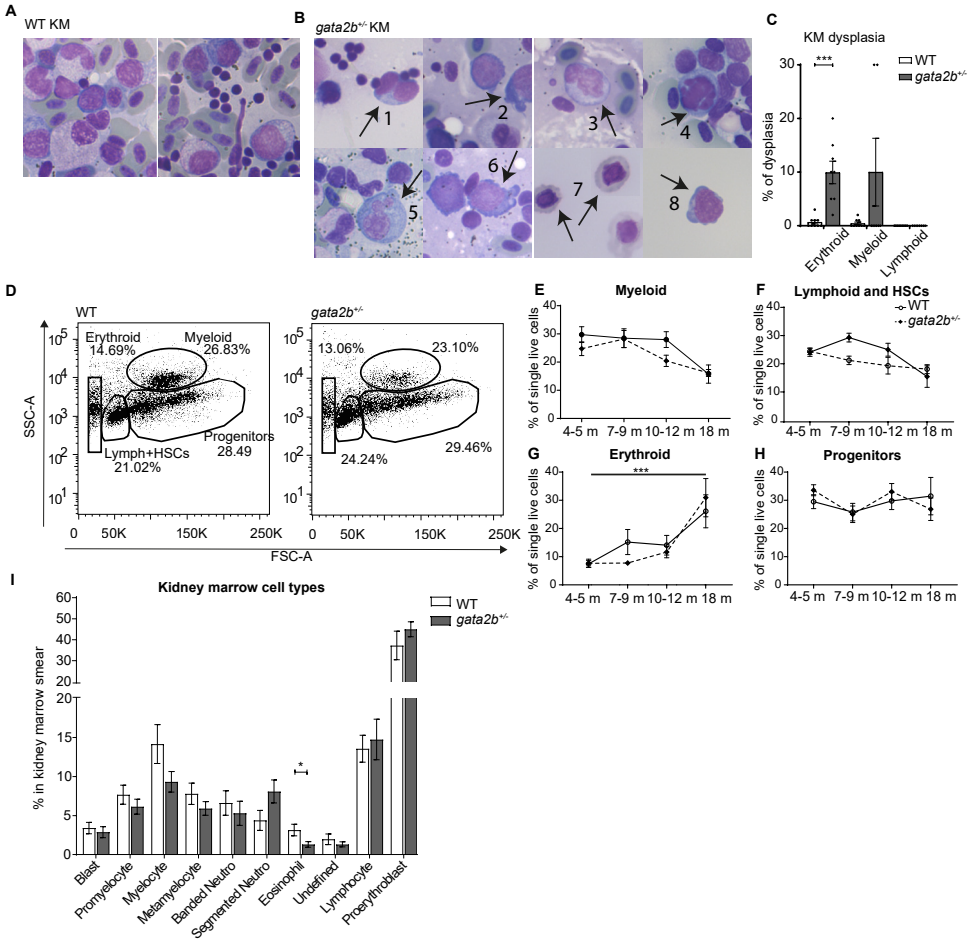


Figure. 1: *Gata2b$^{+/-}$ kidney marrow shows erythroid dysplasia and normal lineage distribution*
Representative pictures of kidney marrow smears after May Grünwald-Giemsa staining of A) WT KM smears and B) *gata2b$^{+/-}$* KM smears. 1) Binucleated erythroblast; 2) Blebbing in cytoplasm of proerythroblast; 3) Binucleated promyelocyte; 4) Lobed nucleus and micronucleus in erythroblast; 5) Multinucleated promyelocyte; 6) Irregular cytoplasm in erythroid precursor; 7) Lobed nucleus in erythrocytes of sorted cell after cytospin; 8) Blebbing in cytoplasm of blast of sorted cell after cytospin. C) Frequency of dysplastic cells of the erythroid, myeloid and lymphoid lineage in KM smears of WT (n=9) and *gata2b$^{+/-}$* (n=9) zebrafish. D) Gating strategy of FACS analysis of whole kidney marrow of WT and *gata2b$^{+/-}$* zebrafish. Percentages represent the average of all zebrafish analyzed per genotype. Quantitation as percentages of the different cell populations in single viable cells of E) myeloid F) lymphoid and HSCs G) erythrocytes H) progenitors in WT and *gata2b$^{+/-}$* zebrafish kidney marrow over time. Mean ± SEM. I) Frequency of differentiated cell types in KM cells in smears of WT (n=9) and *gata2b$^{+/-}$* (n=9) zebrafish. Differentiated erythrocytes were excluded from quantification. * P value<0.5, ***P value<0.001. Data represents mean ± Standard error of the mean.

Single cell transcriptome analysis reveals unique subpopulations in WT and gata2b$^{+/-}$ KM indicative of a cell maturation arrest.

To investigate molecular mechanisms underlying the dysplasia in *gata2b$^{+/-}$* KM, we sought to identify the transcriptional differences in the dysplastic cell population. Through flow cytometry, we sorted 4 cell populations based on light scatter and observed dysplastic cells in the progenitor and lymphoid + HSPCs population of *gata2b$^{+/-}$* KM, indicating that dysplastic cells could be viably sorted (Figure 1B, panel 7 and 8). However, flow cytometry did not identify a uniquely separated population of dysplastic cells, possibly caused by a heterogeneity in their shape. Therefore, we sorted progenitor and lymphoid + HSPC populations from WT and *gata2b$^{+/-}$* KMs and performed single cell RNA sequencing (scRNA-seq) to identify transcriptome signatures defining dysplastic cells (Figure 2A). We identified 19 clusters based on nearest neighbor algorithm in the R Seurat package[20] (Figure 2B). Each cluster was classified based on differentially expressed genes (Suppl. Figure 2A-C) and known differentiation markers (Figure 2C-F)[28-33]. In short, cells of the erythroid lineage are known to express hemoglobin genes like *hbba1. itga2b (CD41)* was used as a marker for thrombocytes and *mpeg1.1* for monocytes and *lysozyme (lyz)* for neutrophils. Furthermore, the differentiated populations were devoid of expression of proliferation genes like *mki67* and *pcna* indeed suggesting that these are differentiated cells and not a progenitor compartment. The populations expressing markers like *CCAAT enhancer binding protein alpha* (*cebpa)* and *granulin1 (grn1)* were identified as myeloid progenitors 1 and 2. T cells were marked by the expression of *tox, il2rb* and *dusp2*, NK cells were marked by *nkl.3, nkl.4* and *ccl33.3* and B cells were marked by the expression of *CD37, CD79a*, *pax5* and high levels of the immunoglobulin *ighv1-4*. Corroborating a conserved lymphoid and myeloid differentiation program, the cluster proportion analysis indicated an overall similar distribution of cells within clusters between *gata2b$^{+/-}$* and WT (Figure 2G, H).

Considering that dysplastic cells were readily recognized as myeloid or erythroid origin in KM smears, we hypothesized that dysplastic cells do not form one major cluster but instead, they occupy a proportion of erythroid or myeloid clusters. For a more precise examination of small populations, we subclustered erythroid and myeloid clusters (including their progenitors) and identified 5 significantly overrepresented sub-clusters in *gata2b$^{+/-}$* KM (Figure 2I, J and K and Supplementary Figure 3A-H). These sub-clusters represented 5.6% of the total sequenced *gata2b$^{+/-}$* cells, comparable to the proportion of dysplastic cells observed in *gata2b$^{+/-}$* KM. Differential gene expression analysis of overrepresented sub-clusters in *gata2b$^{+/-}$* KM compared to the rest of that cluster showed downregulation of tubulin transcripts (not shown), suggesting a loss of cytoskeletal structure, a characteristic of dysplasia. Further investigation of clusters revealed that ''HSPCs'', ''proliferative progenitors'', "progenitors", "myeloid progenitors 1" and "erythroid progenitors 2" clusters contain underrepresented sub-clusters in *gata2b$^{+/-}$* KM (Figure 2I, J and K and Supplementary Figure 3A-D). The loss of these cells suggests that *gata2b$^{+/-}$* cell maturation

could be impaired resulting in a differentiation block, as this is often the case in dysplasia[34]. In line with this hypothesis, *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster cells had upregulated signatures of DNA replication together with downregulation of transcripts necessary for mitosis indicating an impaired cell cycle progression of these cells (Figure 2L). A block in the cell cycle progression can explain the origin of multi-lobulated nuclei and other nuclear abnormalities observed in *gata2b*$^{+/-}$ dysplastic cells[35,36]. To investigate if the loss of sub-clusters was caused by an erythroid differentiation block in *gata2b*$^{+/-}$ KM, we performed lineage trajectory analysis separately on WT and *gata2b*$^{+/-}$ "erythroid progenitor 2" clusters. Because immature erythroid markers were highest in sub-cluster 1, we chose this cluster as the starting point for the trajectory analysis. Compared to WT, differentiation was clearly diminished in *gata2b*$^{+/-}$ "erythroid progenitor 2" cells resulting in an incomplete trajectory and overall reduced pseudo-time value of 12.5 in comparison to 20 in WT "erythroid progenitor 2" cells. Next, we examined the pseudo-time expression of erythroid lineage specific genes in the lineage trajectory. As expected, all cells in both WT and *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster showed high expression of hemoglobin gene *hbba1*. During erythroid differentiation, expression of *GATA1* is known to be highest in immature erythroid cells and gradually decreases towards mature erythroid cells[37]. WT "erythroid progenitor 2" cells indeed show high expression for the zebrafish *GATA1* orthologue *gata1a* at the start of the differentiation trajectory and decrease its expression during differentiation, *gata2b*$^{+/-}$ "erythroid progenitor 2" cells showed a complete opposite expression pattern for *gata1a*. Normal erythroid differentiation is dependent on the "GATA factor switch", which is characterized by a shift from *GATA2* to *GATA1* occupation during erythroid lineage differentiation[38-40]. Our results suggest that these mechanisms might be altered in *gata2b*$^{+/-}$ KM, causing deficiencies in erythroid lineage differentiation.

Because we identified transcriptome signatures indicative of increased proliferation in *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster, we analyzed the pseudo-time expression of *bona fide* proliferation marker *mki67* in these cells. As we described before for *gata2b*$^{-/-}$ HSCs[16], we observed upregulation of *mki67* in *gata2b*$^{+/-}$ "erythroid progenitor 2" cluster cells in contrast to WT cells that gradually decreased *mki67* expression throughout lineage trajectory. Similarly, we performed lineage trajectory and pseudo-time analysis for ''HSPCs'', ''proliferative progenitors'', "progenitors" and "myeloid progenitors 1" clusters (Supplementary Figure 4A-P). In these clusters, we found underrepresented sub-clusters indicative of an accumulation of the cell types in *gata2b*$^{+/-}$ compared to WT, but lineage differentiation was not severely affected from Gata2b haploinsufficiency based on pseudo-time values and the pseudo-time expression of cluster marker genes (Supplementary Figure 4A-P). Nevertheless, in "myeloid progenitors 1" cluster, the expression of proliferation marker *mki67* was upregulated in *gata2b*$^{+/-}$ cells in contrast to WT cells throughout lineage differentiation, indicating that altered proliferation is not solely specific to "erythroid progenitor 2" cluster but broadly present in other clusters of *gata2b*$^{+/-}$ cells (Supplementary Figure 4E and H). Overall, in single cell transcriptome analysis, dysplastic cells do not form a separate cluster but are likely scattered in the myeloid and erythroid progenitor clusters. Furthermore, our findings indicate that erythroid dysplasia in *gata2b*$^{+/-}$ KM possibly originates as a result of proliferative alterations and impaired differentiation.

Gata2b$^{+/-}$ *HSPCs have a high nucleic acid metabolism and downregulated cytoskeletal and protein ubiquitination transcripts*

In *gata2b*$^{+/-}$ KM, dysplasia was observed in both myeloid and erythroid cells, suggesting an aberrancy derived from a common progenitor. Additionally, Gata2b expression was present mostly in the HSPC cluster (Figure 2F) and complete deletion of *gata2b* was shown to predominantly affect HSPCs in zebrafish[16]. The single cell analysis of HSPCs showed great heterogeneity with scattered HSPCs expressing differentiation markers like *ighv1-4* indicative of B cell differentiation, *Nkl.2* indicative of NK-differentiation and *icn* indicative of myeloid differentiation (Supplementary Figure 2A). To investigate the effect of Gata2b haploinsufficiency in the most immature cells of the HSPC compartment, we further purified the HSC population by flow sorting of the CD41:GFP^{low} expressing cells from WT or *gata2b*$^{+/-}$ KMs (Figure 3A), as these were shown to contain transplantable HSCs[18]. Although numerically unaffected by *gata2b* heterozygosity (Figure 3B), gene expression was markedly different in *gata2b*$^{+/-}$ HSCs compared to WT. To elucidate the effect of Gata2b haploinsufficiency in specific biological pathways, we performed gene-set enrichment (GSEA) and KEGG analysis that revealed 71 upregulated and 80 downregulated gene-sets (P<0.05) in *gata2b*$^{+/-}$ HSCs. We found an enrichment of gene-sets related to DNA replication and ribosome biogenesis in *gata2b*$^{+/-}$ HSCs (Figure 3C-E). Simultaneously, gene-sets related to cytoskeleton and protein degradation were underrepresented in *gata2b*$^{+/-}$ HSCs (Figure 3F-H). Similar to the *gata2b*$^{+/-}$ transcriptome signatures we observed in scRNA-seq, we found upregulation of proliferation related genes such as *pcna* and *mcm2-8* in *gata2b*$^{+/-}$ CD41:GFP^{low} cells, indicating these signatures originate from the most immature compartment of *gata2b*$^{+/-}$ KM (Supplementary Figure 4Q). To understand the consequences of under- and over- represented gene-sets found in *gata2b*$^{+/-}$ HSCs, we performed network analysis using Cytoscape software. In this analysis, gene-sets (dots) are connected with lines if they share similar genes and form networks if they are assigned in similar biological processes. Network analysis further revealed the enrichment of gene-sets related to proliferation and transcription at the expense of gene-sets related to protein ubiquitination and cytoskeleton structure (Figure 3I). Because our results showed an aberrant proliferative signature, we assessed proliferation by flow cytometry in *gata2b*$^{+/-}$ CD41:GFP^{low} cells. We found a relative increase in cells in G1 phase of the cell cycle and a significant decrease in cells in G2-M phase of the cell cycle phase in *gata2b*$^{+/-}$ HSCs indicating that despite increased proliferative signatures cell cycle progression is impaired in *gata2b*$^{+/-}$ HSCs (Figure 3J and K). These data suggest that *gata2b*$^{+/-}$ HSPCs have an aberrant expression of the genes controlling cell structure and replication and fail to progress through cell cycle, which is possibly the source of dysplasia observed in the KM of *Gata2b*$^{+/-}$ zebrafish.

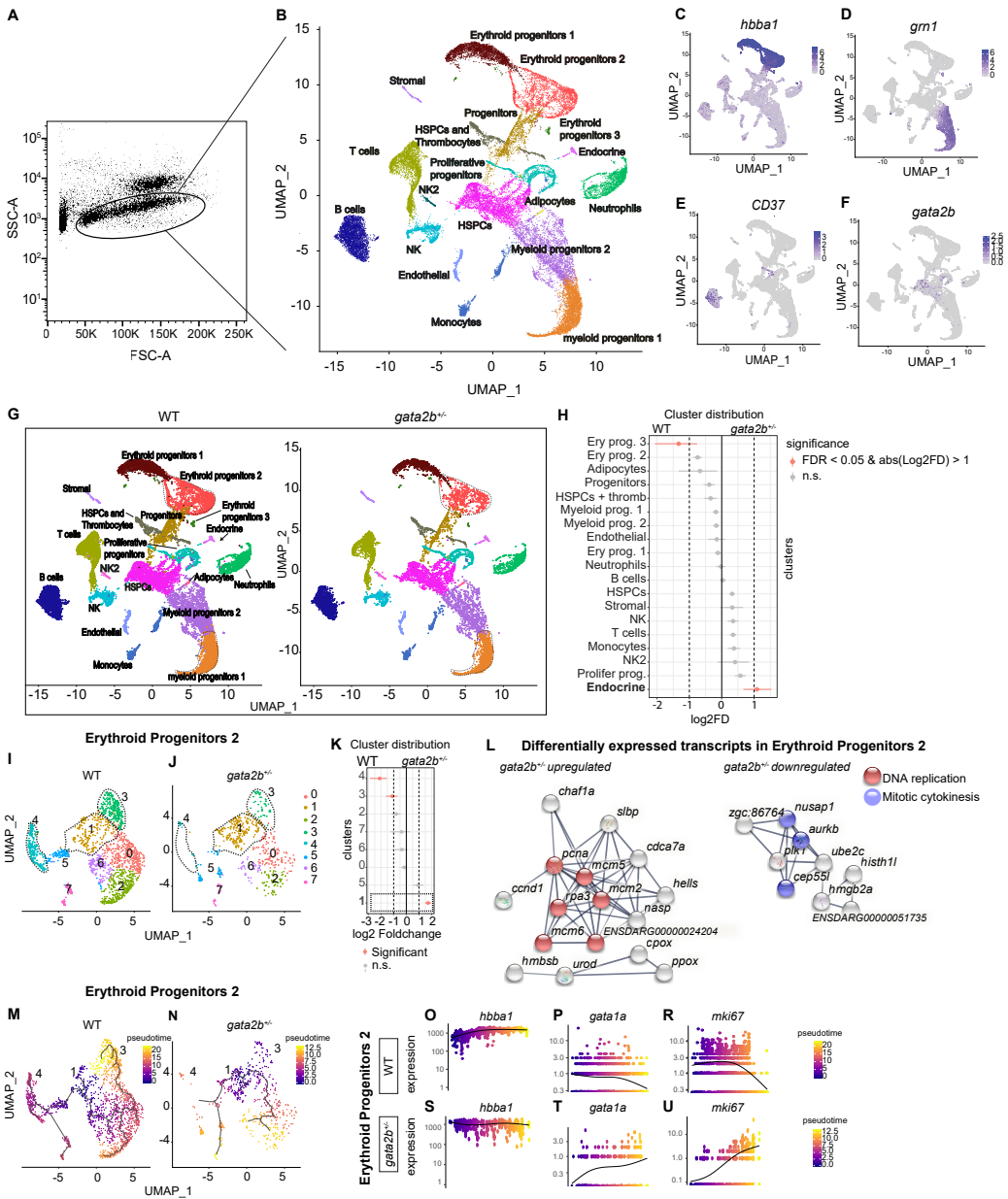


Figure. 2: Single cell analysis reveals unique *gata2b*$^{+/-}$ cells scattered among clusters and a differentiation block in erythroid progenitors

A) Kidney marrow light scatter sorting strategy for scRNA sequencing analysis and B) UMAP analysis of scRNA-seq data showing different cell types based on expression analysis. FeaturePlot gene expression analysis of C) *hemoglobin beta adult 1* (*hbba1*) expressed in the erythrocyte lineage D*) granulin1* (*grn1)* in the myeloid lineage E) *cluster of differentiation 37* (*CD37)* expressed in the B cell lineage and F) *GATA binding protein 2b* (*gata2b)* in HSPCs. G) Split UMAP of all cells separating WT and *gata2b*$^{+/-}$ cells. H) Quantification of distribution between WT and *gata2b*$^{+/-}$ cells in each cluster. Significantly differentially distributed clusters in pink. FDR<0.05 & Log2 fold change >1. I) Split UMAP of the subclustering of the erythroid progenitors 2 cluster. K) Quantification of distribution between WT and *gata2b*$^{+/-}$ erythroid progenitor 2 subcluster cells. Significantly differentially distributed clusters in pink. Dotted box indicates overrepresented subcluster in *gata2b*$^{+/-}$ cells L) STRING network of upregulated and downregulated transcripts in *gata2b*$^{+/-}$ "erythroid progenitors 2". Only networks with more than 2 interactions were represented. Highlighted in red DNA replication genes from KEGG pathways and highlighted in blue mitotic cytokinesis genes form biological processes (gene Ontology). M) Lineage trajectory analysis of WT and N) *gata2b*$^{+/-}$ erythroid progenitor 2 subclusters. Note the difference in pseudotime scale between the WT and *gata2b*$^{+/-}$ trajectories. O-R) WT and S-U) *gata2b*$^{+/-}$ pseudotime expression of individual genes (O+S) *hemoglobin beta adult 1* (*hbba1*), (P+T) *GATA binding protein 1 (gata1),* (R+U) *mki67.* Fold change >0.05 & adjusted P value <0.05. FDR=False discovery rate. FD=Fold difference.

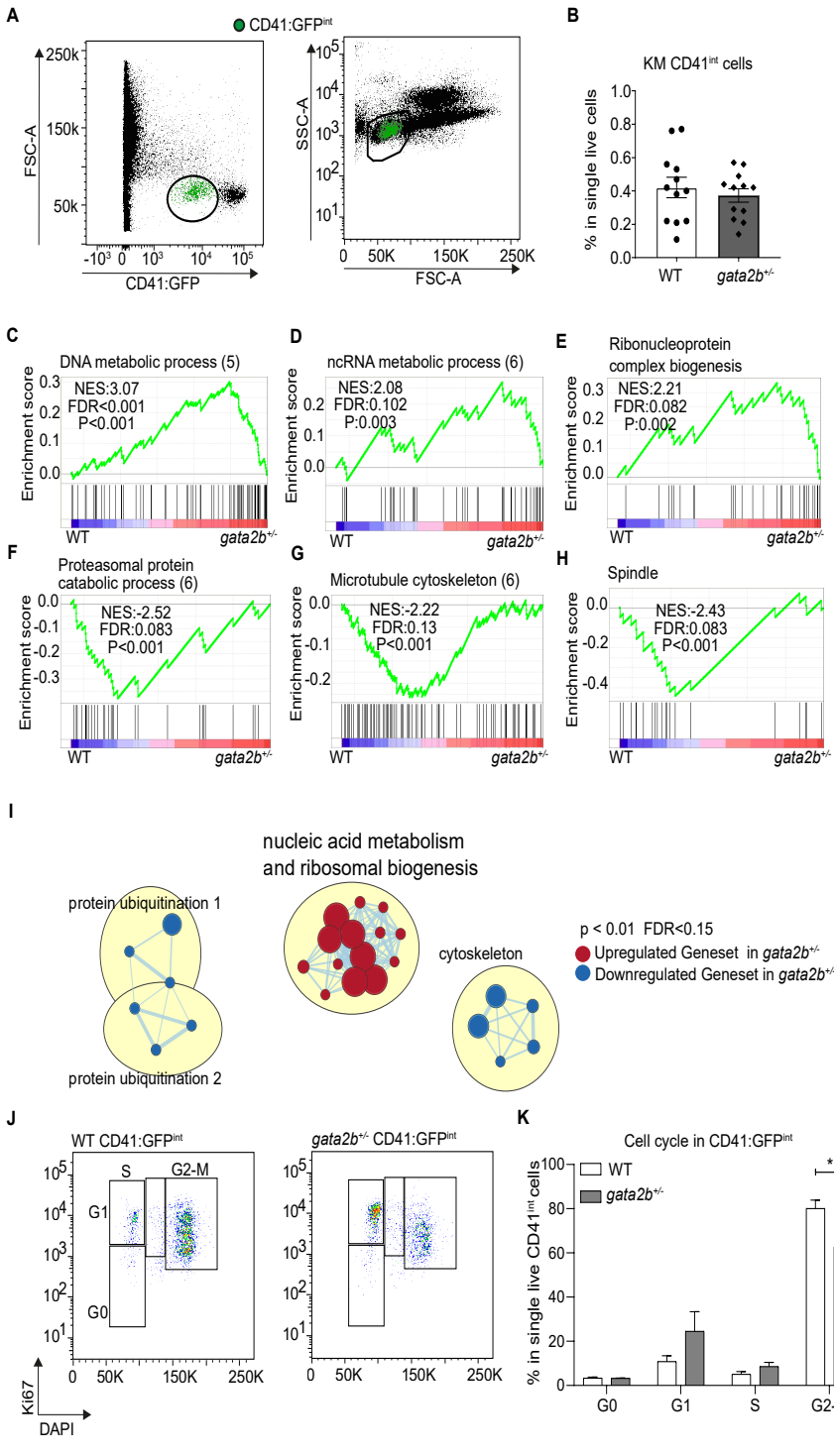


Figure. 3: *Gata2b*$^{+/-}$ HSPCs have a high nucleic acid metabolism and downregulated cytoskeletal and protein ubiquitination transcripts

A) Representative figure of identification of CD41:GFP^{int} population (in green) and distribution in the FSC-A SSC-A kidney marrow population. B) Quantification of the percentage of CD41:GFP^{int} cells in WT and *gata2b*$^{+/-}$ kidney marrow single live cells. C-H) Representative gene set enrichments plots showing the profile of the enrichment score (ES) and positions of gene set members on the rank ordered list. I) Cytoscape 3.8.2 enrichment map depicting significantly ($p < 0.01$ and FDR<0.15) upregulated (red) and downregulated (blue) gene sets. The network clusters in yellow are generated using the AutoAnnotation application of Cytoscape 3.8.2. J) Representative flow cytometry plots of cycle analysis by anti-Ki67 and DAPI staining in WT and *gata2b*$^{+/-}$ CD41:GFP^{int} KM. K) Quantification of the percentages of WT (n=3) and *gata2b*$^{+/-}$ (n=3) CD41:GFP^{int} cells in different cell cycle stages (* P value<0.5). Data represents mean ± Standard error of the mean.

DISCUSSION

In humans, a balanced *GATA2* expression is essential for proper hematopoiesis. Consequently, more than 80% of *GATA2* mutation carriers progress to hematological malignancy by 40 years of age[4]. While the clinical consequences of *GATA2* mutations became obvious over the last decades, the regulation of GATA2 activity and its contribution to human bone marrow failure syndromes and progression to hematological malignancy remain incompletely understood.

Here, we used transgenic reporters, morphological phenotyping and RNA sequencing to analyze *gata2b* heterozygous zebrafish. We showed that, while major differentiation lineages remain intact, *gata2b* heterozygosity causes dysplasia in erythroid and myeloid progenitors of zebrafish KM. ScRNA-seq analysis did not identify a single population of dysplastic cells indicating that these cells are likely scattered through erythroid and myeloid clusters. In fact, after sub-clustering we identified uneven distribution of erythroid and myeloid sub-clusters between WT and *gata2b*$^{+/-}$ KM. Further examination with lineage trajectory analysis captured a differentiation block in one cluster of erythroid progenitors, concomitant with aberrant expression of *gata1a*. This suggested that alterations in GATA switch mechanisms may also contribute to the deterioration of erythroid dysplasia in *gata2b*$^{+/-}$ KM. Additionally, in the same erythroid progenitor cluster, an increased proliferative signature together with decreased expression of genes related to mitosis indicated an impaired cell cycle progression. We propose that these alterations could play a role in the onset of dysplasia found in *gata2b*$^{+/-}$ zebrafish.

Because we found both myeloid and erythroid dysplasia in *gata2b*$^{+/-}$ zebrafish and the expression of *gata2b* was highest in immature HSPCs compare to differentiated cells, this suggested that Gata2b haploinsufficiency affects the HSPC compartment. However, even after sub-clustering, we did not observe dramatic differences between the distribution of WT and *gata2b*$^{+/-}$ cells in the HSPC cluster. Because the HSPC cluster also contained lineage primed cells, we purified the most primitive HSPCs (CD41 intermediate cells or HSCs) from WT and *gata2b*$^{+/-}$ KMs and performed additional RNA sequencing experiments to identify the effect of Gata2b haploinsufficiency in HSCs. After analyzing HSCs, an increased proportion of cells in G1 phase of the cell cycle were found in *gata2b*$^{+/-}$ HSCs compared to WT and these cells transcriptionally showed loss of quiescence as observed in Gata2b knockout (*gata2b*$^{-/-}$) HSCs[16]. Furthermore, the proliferation assay indicated that cell cycle progression is diminished in these cells in G2-M phase. Because erythroid progenitors showed upregulation of proliferation genes and downregulation of mitosis genes, we propose that Gata2 haploinsufficiency causes proliferative stress in HSPCs. As cells become multi-lobulated towards the beginning of G2-M phase of the cell cycle, a block in cell cycle progression in G2-M phase could possibly explain the presence of multi-lobulated dysplastic cells in *gata2b*$^{+/-}$ KM.

Interestingly, whereas *gata2b*$^{-/-}$ zebrafish display abrogated myeloid lineage differentiation and a bias toward lymphoid differentiation[16,17], *gata2b*$^{+/-}$ did not simply result in an intermediate phenotype between WT and *gata2b*$^{-/-}$. Instead, Gata2b haploinsufficiency uniquely cause dysplasia, not observed in *gata2b*$^{-/-}$. The differences in the homozygous and heterozygous *gata2b* knockout phenotype support a role for gene dosage underlying GATA2 deficiency phenotype, possibly explaining the phenotypic heterogeneity between patients. Since both erythroid and myeloid dysplasia can be observed in GATA2 patients[4], we propose that the presence of dysplastic cells in *gata2b*$^{+/-}$ resembles the clinical phenotypes associated with *GATA2* heterozygosity. In the future, the isolation of single dysplastic cells could help us to further explore the effect of Gata2 haploinsufficiency in malignant transformation. Nevertheless, it remains to be established how *gata2b*$^{+/-}$ HSPCs would respond to secondary insults such as infections or severe bleeding.

In conclusion, while the major lineage differentiation remains intact, *gata2b*$^{+/-}$ zebrafish possess a stressed proliferative HSPC compartment which leads to the generation of erythroid and myeloid dysplastic cells. Taken together, our model provides insights into the consequences of Gata2b dosage, and reveals transcriptional networks affected by Gata2b haploinsufficiency in zebrafish.

Acknowledgements

We thank Dr Monteiro (University of Birmingham) for careful reading of the manuscript. We thank the Experimental Animal Facility of Erasmus MC for animal husbandry and the Erasmus Optical Imaging Center for confocal microscopy services. This research is supported by the European Hematology Association (junior non clinical research fellowship)(EdP), the Dutch Cancer Foundation KWF/Alpe d'HuZes (SK10321)(EdP), the Daniel den Hoed Foundation for support of the Cancer Genome Editing Center (IT) and the Josephine Nefkens Foundation for purchase of the Chromium 10x (IT).

Authorship contributions

EdP and EG conceived the study; EG, CK, HdL, JZ, DB, and EB performed experiments; EG, CK, RH, KG and EdP analyzed results; IT provided resources and EG, CK and EdP wrote the manuscript and IT revised the manuscript.

Disclosures

The authors declare no conflicts of interests

REFERENCES

1 Gao, X. *et al.* Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *J Exp Med* **210**, 2833-2842, doi:jem.20130733 [pii], 10.1084/jem.20130733 (2013).

2 de Pater, E. *et al.* Gata2 is required for HSC generation and survival. *J Exp Med* **210**, 2843-2850, doi:jem.20130751 [pii] 10.1084/jem.20130751 (2013).

3 Tsai, F. Y. *et al.* An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371**, 221-226, doi:10.1038/371221a0 (1994).

4 Donadieu, J. *et al.* Natural history of GATA2 deficiency in a survey of 79 French and Belgian patients. *Haematologica* **103**, 1278-1287, doi:haematol.2017.181909 [pii]

10.3324/haematol.2017.181909 (2018).

5 Spinner, M. A. *et al.* GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood* **123**, 809-821, doi:S0006-4971(20)36040-7 [pii]

10.1182/blood-2013-07-515528 (2014).

6 Ostergaard, P. *et al.* Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet* **43**, 929-931, doi:ng.923 [pii], 10.1038/ng.923 (2011).

7 Hsu, A. P. *et al.* Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* **118**, 2653-2655, doi:blood-2011-05-356352 [pii], 10.1182/blood-2011-05-356352 (2011).

8 Dickinson, R. E. *et al.* Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* **118**, 2656-2658, doi:S0006-4971(20)40687-1 [pii], 10.1182/blood-2011-06-360313 (2011).

9 Hahn, C. N. *et al.* Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet* **43**, 1012-1017, doi:ng.913 [pii], 10.1038/ng.913 (2011).

10 Mutsaers, P. G. *et al.* Highly variable clinical manifestations in a large family with a novel GATA2 mutation. *Leukemia* **27**, 2247-2248, doi:leu2013105 [pii], 10.1038/leu.2013.105 (2013).

11 Wang, X. *et al.* GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies. *Haematologica* **100**, e398-401, doi:haematol.2015.127092 [pii], 10.3324/haematol.2015.127092 (2015).

12 Rodrigues, N. P. *et al.* Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood* **106**, 477-484, doi:2004-08-2989 [pii], 10.1182/blood-2004-08-2989 (2005).

13 Ling, K. W. *et al.* GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med* **200**, 871-882, doi:jem.20031556 [pii], 10.1084/jem.20031556 (2004).

14 Butko, E. *et al.* Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. *Development* **142**, 1050-1061, doi:10.1242/dev.119180 (2015).

15 Dobrzycki, T. *et al.* Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis. *Commun Biol* **3**, 71, doi:10.1038/s42003-020-0798-3, 10.1038/s42003-020-0798-3 [pii] (2020).

16 Gioacchino, E. *et al.* Essential role for Gata2 in modulating lineage output from hematopoietic stem cells in zebrafish. *Blood Adv* **5**, 2687-2700, doi:S2473-9529(21)00348-7 [pii], 10.1182/bloodadvances.2020002993 (2021).

17 Avagyan, S. *et al.* Single-cell ATAC-seq reveals GATA2-dependent priming defect in myeloid and a maturation bottleneck in lymphoid lineages. *Blood Adv* **5**, 2673-2686, doi:S2473-9529(21)00347-5 [pii], 10.1182/bloodadvances.2020002992 (2021).

18 Ma, D., Zhang, J., Lin, H. F., Italiano, J. & Handin, R. I. The identification and characterization of zebrafish hematopoietic stem cells. *Blood* **118**, 289-297, doi:blood-2010-12-327403 [pii], 10.1182/blood-2010-12-327403 (2011).

19 Tamplin, O. J. *et al.* Hematopoietic stem cell arrival triggers dynamic remodeling of the perivascular niche. *Cell* **160**, 241-252, doi:10.1016/j.cell.2014.12.032 (2015).

20 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821, doi:S0092-8674(19)30559-8 [pii], 10.1016/j.cell.2019.05.031 (2019).

21 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).

22 Boatman, S. *et al.* Assaying hematopoiesis using zebrafish. *Blood Cells Mol Dis* **51**, 271-276, doi:S1079-9796(13)00160-5 [pii], 10.1016/j.bcmd.2013.07.009 (2013).

23 Traver, D. *et al.* Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol* **4**, 1238-1246, doi:10.1038/ni1007 ni1007 [pii] (2003).

24 Langenau, D. M. *et al.* In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. *Proc Natl Acad Sci U S A* **101**, 7369-7374, doi:10.1073/pnas.0402248101, 0402248101 [pii] (2004).

25 Renshaw, S. A. *et al.* A transgenic zebrafish model of neutrophilic inflammation. *Blood* **108**, 3976-3978, doi:blood-2006-05-024075 [pii], 10.1182/blood-2006-05-024075 (2006).

26 Ellett, F., Pase, L., Hayman, J. W., Andrianopoulos, A. & Lieschke, G. J. mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish. *Blood* **117**, e49-56, doi:blood-2010-10-314120 [pii], 10.1182/blood-2010-10-314120 (2011).

27 Page, D. M. *et al.* An evolutionarily conserved program of B-cell development and activation in zebrafish. *Blood* **122**, e1-11, doi:blood-2012-12-471029 [pii], 10.1182/blood-2012-12-471029 (2013).

28 Carmona, S. J. *et al.* Single-cell transcriptome analysis of fish immune cells provides insight into the evolution of vertebrate immune cell types. *Genome Res* **27**, 451-461, doi:10.1101/gr.207704.116 (2017).

29 Danilova, N., Bussmann, J., Jekosch, K. & Steiner, L. A. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat Immunol* **6**, 295-302, doi:ni1166 [pii], 10.1038/ni1166 (2005).

30 Kortum, A. N. *et al.* Differential expression and ligand binding indicate alternative functions for zebrafish polymeric immunoglobulin receptor (pIgR) and a family of pIgR-like (PIGRL) proteins. *Immunogenetics* **66**, 267-279, doi:10.1007/s00251-014-0759-4 (2014).

31 Macaulay, I. C. *et al.* Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep* **14**, 966-977, doi:10.1016/j.celrep.2015.12.082 (2016).

32 Tang, Q. *et al.* Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J Exp Med* **214**, 2875-2887, doi:10.1084/jem.20170976 (2017).

33 Athanasiadis, E. I. *et al.* Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat Commun* **8**, 2045, doi:10.1038/s41467-017-02305-6 (2017).

34 Monika Belickova, M. *et al.* Up-regulation of ribosomal genes is associated with a poor response to azacitidine in myelodysplasia and related neoplasms. *Int J Hematol* **104**, 566-573, doi:10.1007/s12185-016-2058-3, 10.1007/s12185-016-2058-3 [pii] (2016).

35 Nakayama, Y. & Yamaguchi, N. Multi-lobulation of the nucleus in prolonged S phase by nuclear expression of Chk tyrosine kinase. *Exp Cell Res* **304**, 570-581, doi:S0014-4827(04)00703-7 [pii], 10.1016/j.yexcr.2004.11.027 (2005).

36 Ullah, Z., Lee, C. Y. & Depamphilis, M. L. Cip/Kip cyclin-dependent protein kinase inhibitors and the road to polyploidy. *Cell Div* **4**, 10, doi:1747-1028-4-10 [pii], 10.1186/1747-1028-4-10 (2009).

37 Moriguchi, T. & Yamamoto, M. A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation. *Int J Hematol* **100**, 417-424, doi:10.1007/s12185-014-1568-0 (2014).

38 Grass, J. A. *et al.* GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A* **100**, 8811-8816, doi:10.1073/pnas.1432147100 1432147100 [pii] (2003).

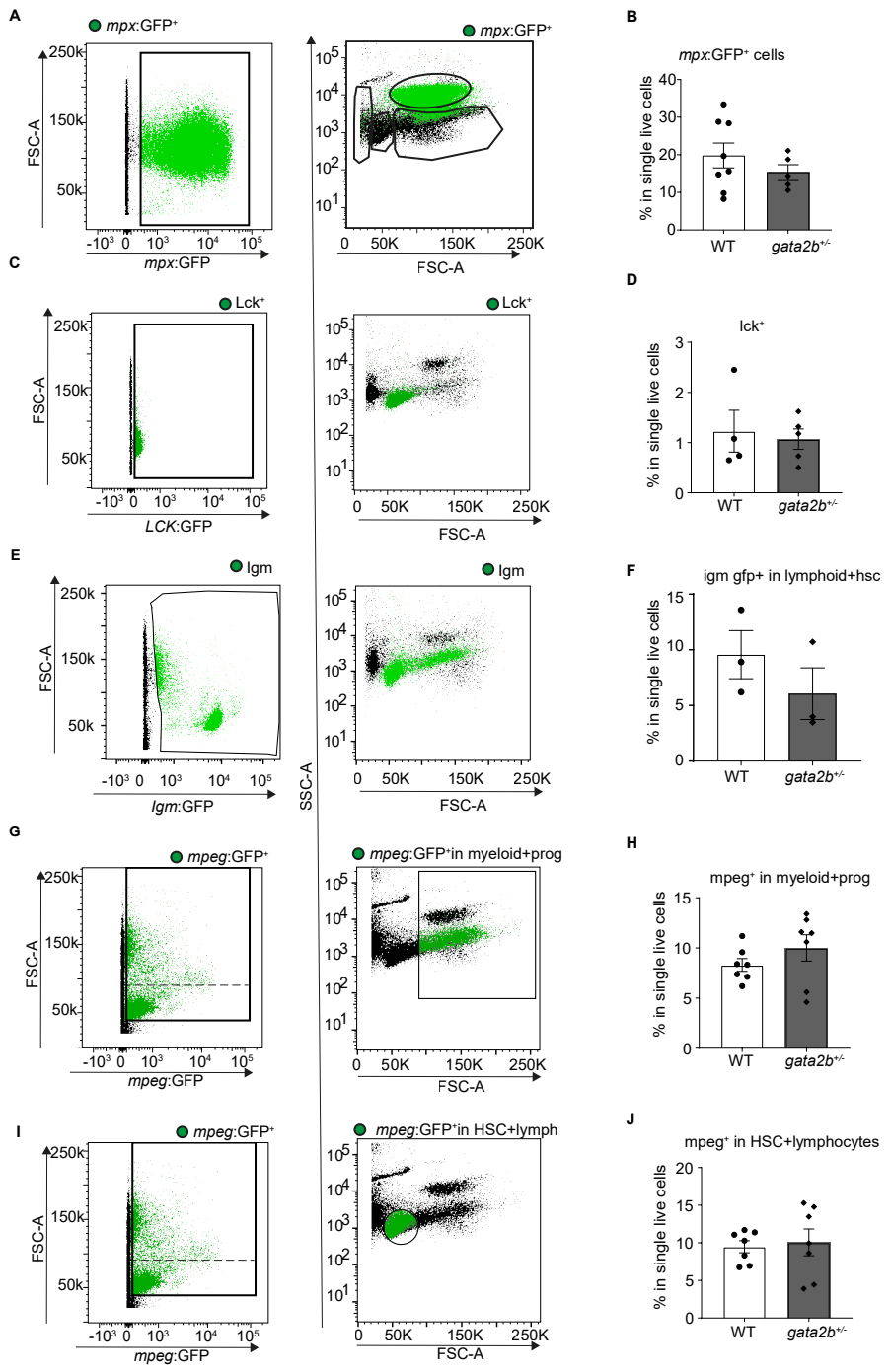
39 Pal, S. *et al.* Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A* **101**, 980-985, doi:10.1073/pnas.0307612100 0307612100 [pii] (2004).

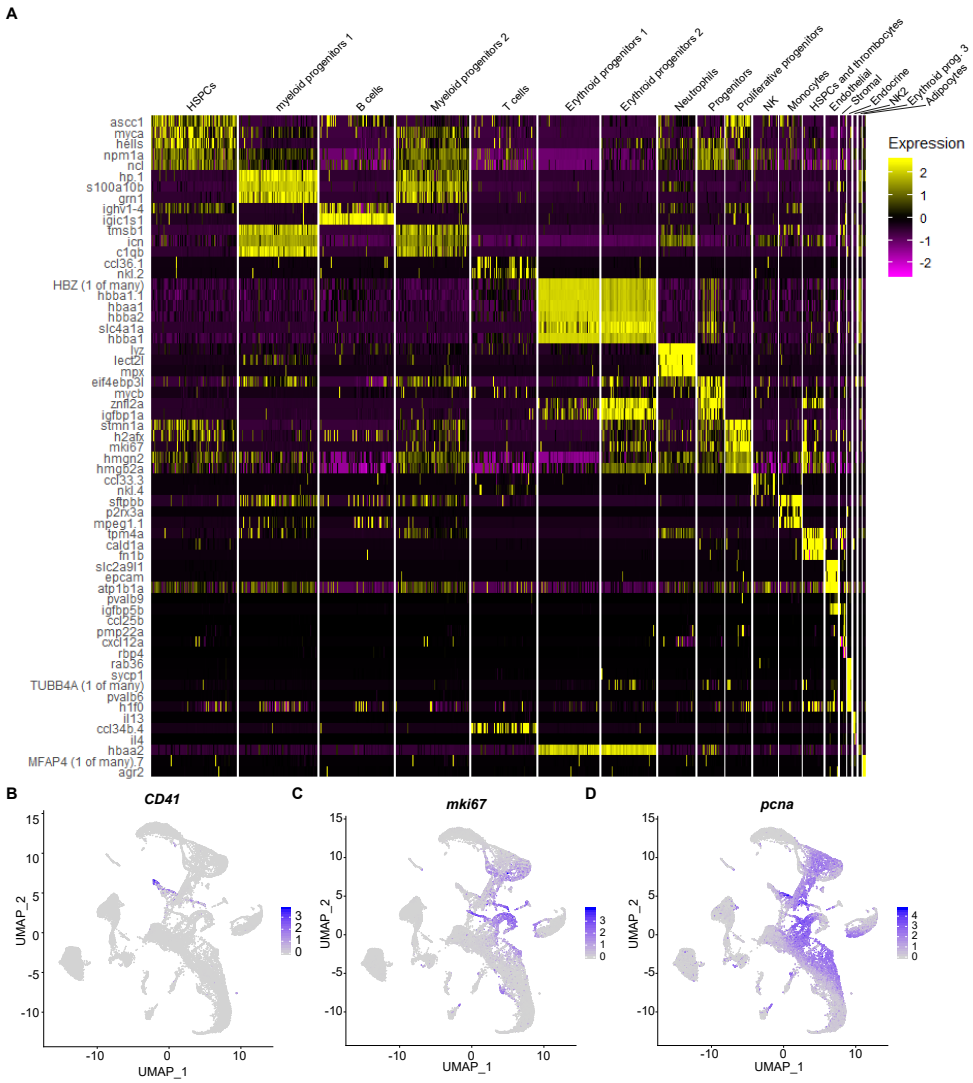
40 Bresnick, E. H., Lee, H. Y., Fujiwara, T., Johnson, K. D. & Keles, S. GATA switches as developmental drivers. *J Biol Chem* **285**, 31087-31093, doi:S0021-9258(19)88840-3 [pii], 10.1074/jbc.R110.159079 (2010).

SUPPLEMENTARY FIGURES

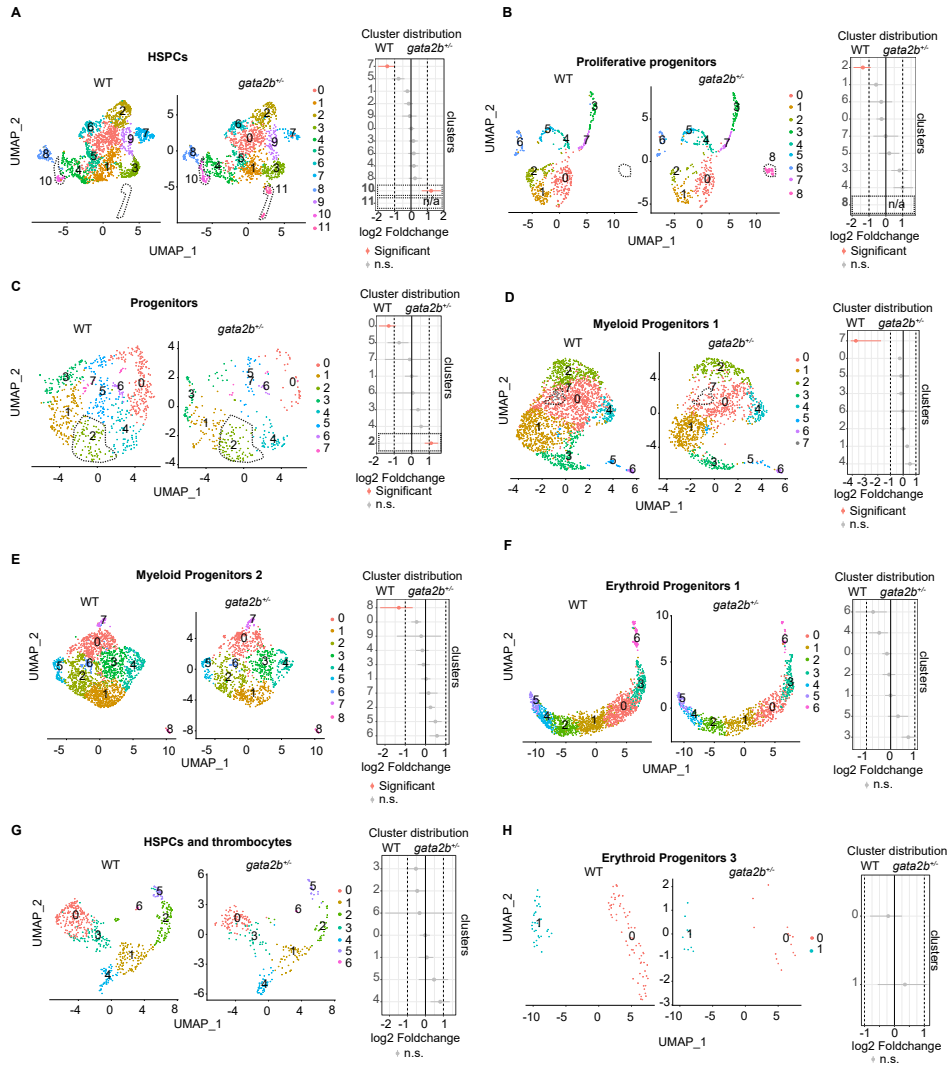
Supplementary Figure. 1: Transgenic zebrafish reporter lines show no overall differentiation difference in *gata2b*$^{+/-}$

A) *mpx* positive cells, depicted in green, with distribution in FSC-A SSC-A graph. Followed by B) Quantification of the percentage of mpx:GFP+ cells in WT and *gata2b*$^{+/-}$ kidney marrow single live cells. C) *Lck* positive cells, depicted in green, with distribution in FSC-A SSC-A graph. Followed by D) Quantification of the percentage of lck:GFP+ cells in WT and *gata2b*$^{+/-}$ kidney marrow single live cells. E) *Igm* positive cells, depicted in green, with distribution in FSC-A SSC-A graph. Followed by F) Quantification of the percentage of IgM:GFP+ cells in WT and *gata2b*$^{+/-}$ kidney marrow single live cells. G) *mpeg* positive cells, depicted in green, mark monocytes and I) phagocytic B-cells. Followed by H and J) Quantification of the percentage of mpeg:GFP+ cells in WT and *gata2b*$^{+/-}$ kidney marrow single live cells. Data represents mean ± Standard error of the mean.

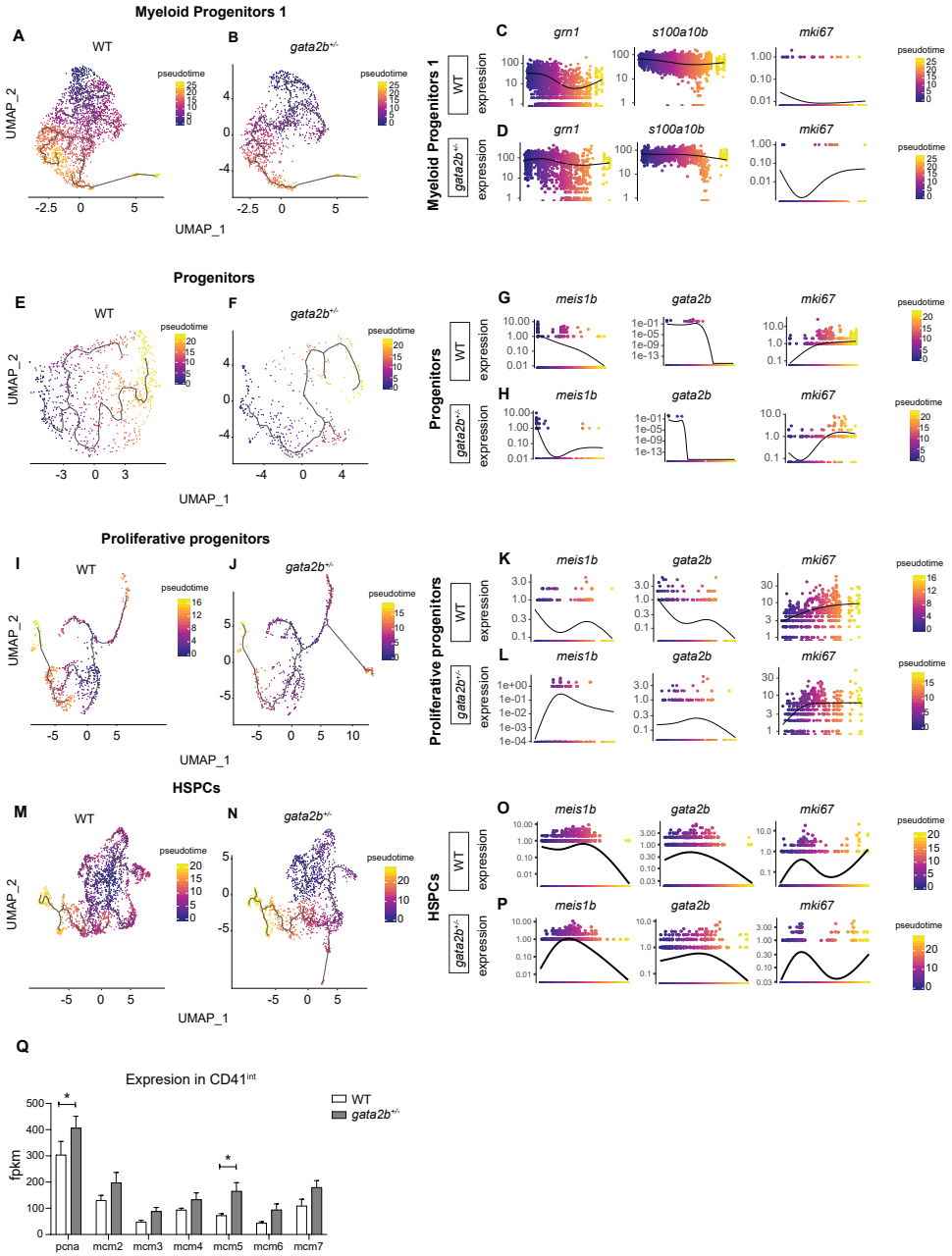




Supplementary Figure. 2: Heatmap and FeaturePlots
A) Unbiased heatmap representing the expression level of the top 5 expressed transcripts per cluster. Transcripts highly expressed in multiple clusters were not repeated in the list (like hemoglobin in erythroid clusters). Transcripts identified only by their chromosome location were not included. FeaturePlot gene expression analysis of B) *CD41* in HSPCs and thrombocytes, C) *mki67* in Proliferative progenitors and D) *pcna* predominantly in undifferentiated clusters.



Supplementary Figure. 3: *gata2b*$^{+/-}$ cells have a different distribution of subclusters compared to WT
Split UMAP representing the cell distribution in the various subclusters (distinguishable by different colors) in WT and *gata2b*$^{+/-}$ for A) HSPCs B) Proliferative progenitors C) Progenitors D) Myeloid progenitors 1, E) Myeloid progenitors 2, F) Erythroid progenitors 1, G) HSPCs and thrombocytes , H) Erythroid progenitors 3, with respective quantification of distribution between genotypes. Significantly differentially distributed subclusters in pink. FDR<0.05 & Log2 fold change >1.



Supplementary Figure. 4: Differentiation trajectory and gene expression analysis of *gata2b*$^{+/-}$ cells
Lineage trajectory analysis of WT and *gata2b*$^{+/-}$ A-B) Myeloid progenitors 1, E-F) Progenitors, I-J) Proliferative progenitors, M-N) HSPCs clusters. C) WT and D) *gata2b*$^{+/-}$ pseudotime expression of individual genes in Myeloid progenitors 1. G) WT and H) *gata2b*$^{+/-}$ pseudotime expression of individual genes in Progenitors. K) WT and L) *gata2b*$^{+/-}$ pseudotime expression of individual genes in Proliferative progenitors. O) WT and P) *gata2b*$^{+/-}$ pseudotime expression of individual genes in HSPCs. Q) FPKM values of proliferation related genes *pcna* and *mcm2-7* in WT and *gata2b*$^{+/-}$ CD41:GFP^{int} cells. * P value<0.5 and data represents mean ± Standard error of the mean.

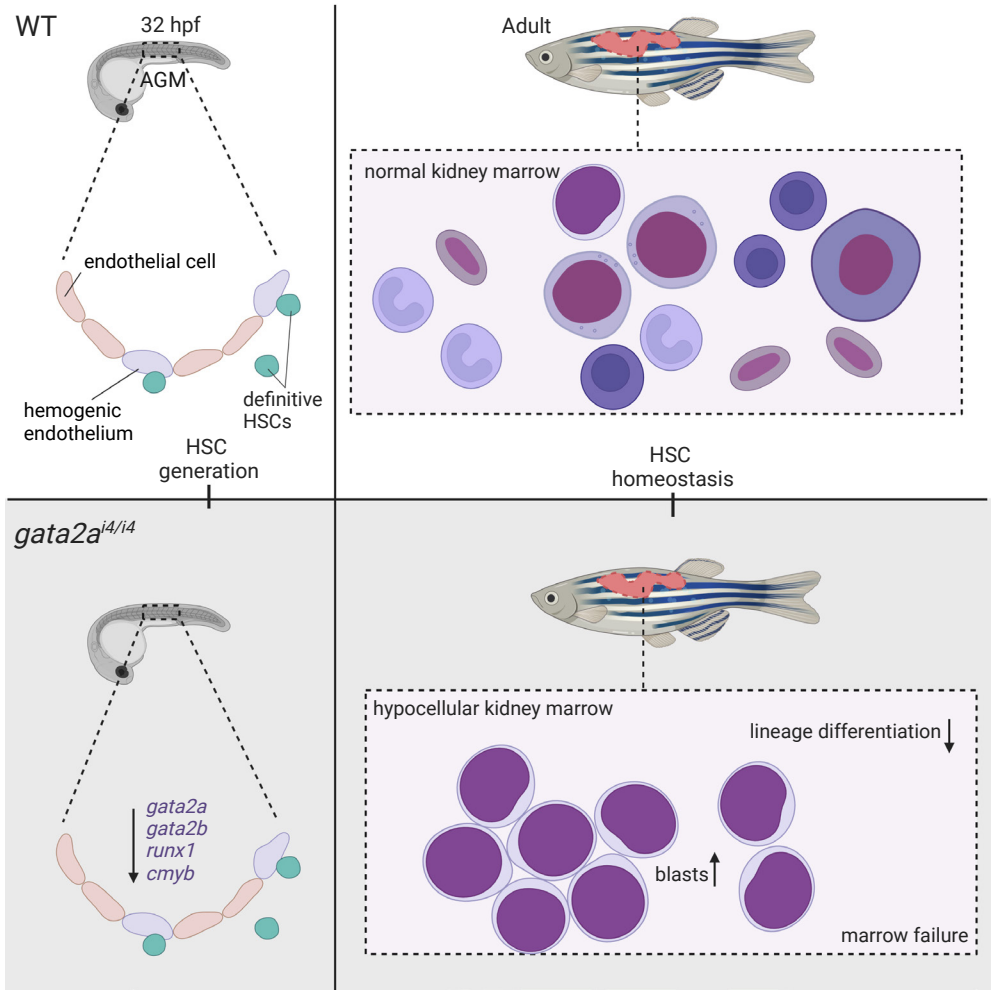
5

Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis

Tomasz Dobrzycki[1], Christopher B. Mahony[2], Monika Krecsmarik[1,3], Cansu Koyunlar[4], Rossella Rispoli[1,5], Joke Peulen-Zink[4], Kirsten[4], Bakhta[2], Emma de Pater[4], Roger Patient[1,3] & Rui Monteiro[1,2,3]
*

[1] MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK.
[2] Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK.
[3] BHF Centre of Research Excellence, Oxford, UK.
[4] Department of Hematology, Erasmus MC, Rotterdam, The Netherlands.
[5] Division of Genetics and Molecular Medicine, NIHR Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust and King's College London, London, UK.

Commun Biol 3, 71 (2020)



ABSTRACT

Gata2 is a key transcription factor required to generate Haematopoietic Stem and Progenitor Cells (HSPCs) from haemogenic endothelium (HE); misexpression of Gata2 leads to haematopoietic disorders. Here we deleted a conserved enhancer (i4 enhancer) driving pan-endothelial expression of the zebrafish *gata2a* and showed that Gata2a is required for HE programming by regulating expression of *runx1* and of the second Gata2 orthologue, *gata2b*. By 5 days, homozygous *gata2a*$^{\Delta i4/\Delta i4}$ larvae showed normal numbers of HSPCs, a recovery mediated by Notch signalling driving *gata2b* and *runx1* expression in HE. However, *gata2a*$^{\Delta i4/\Delta i4}$ adults showed oedema, susceptibility to infections and marrow hypocellularity, consistent with bone marrow failure found in GATA2 deficiency syndromes. Thus, *gata2a* expression driven by the i4 enhancer is required for correct HE programming in embryos and maintenance of steady-state haematopoietic stem cell output in the adult. These enhancer mutants will be useful in exploring further the pathophysiology of GATA2-related deficiencies in vivo.

INTRODUCTION

Haematopoietic stem cells (HSCs) are the source of all blood produced throughout the lifetime of an organism. They are capable of self-renewal and differentiation into progenitor cells that generate specialised blood cell types. DNA-binding transcription factors are fundamental players in the inception of the haematopoietic system as it develops in the embryo, but also play a crucial role in maintaining homeostasis of the haematopoietic system in the adult organism. They coordinate differentiation, proliferation and survival of haematopoietic cells and ensure their levels are appropriate at all times throughout life. Misexpression of key transcription factors may thus lead to a failure to produce HSCs or, alternatively, to haematopoietic disorders and eventually leukaemia. Therefore, understanding how transcription factors drive the haematopoietic process provides opportunities for intervention when haematopoiesis is dysregulated.

The development of blood occurs in distinct waves: primitive, pro-definitive and definitive, each of them characterised by the generation of blood progenitors in a specific location and restricted in time, where the definitive wave produces multi- lineage self-renewing HSCs[1]. The specification of HSCs initiates in cells with arterial characteristics[2] and proceeds through an endothelial intermediate, termed the haemogenic endothelium (HE)[3]. In zebrafish and other vertebrates, expression of *runx1* defines the *bona fide* HE population[4,5]. Haematopoietic stem and progenitor cells (HSPCs) emerge from the HE by endothelial- to-haematopoietic transition (EHT), both in zebrafish and in mice[6-8]. They arise between 28 and 48 h post fertilisation (hpf) from the HE in the ventral wall of the dorsal aorta (DA)[9], the analogue of the mammalian aorta-gonad-mesonephros (AGM)[10]. After EHT, the HSCs enter the bloodstream through the posterior cardinal vein (PCV)[9] to colonise the caudal haematopoietic tissue(CHT), the zebrafish equivalent of the mammalian foetal liver[11]. Afterwards the HSCs migrate again within the bloodstream to colonise the kidney marrow (WKM) and thymus[9], the final nichefor HSCs, equivalent to the bone marrow in mammals[1].

Gata2 is a key haematopoietic transcription factor (TF) in development. In humans, *GATA2* haploinsufficiency leads to blood disorders, including MonoMAC syndrome (Monocytopenia, Mycobacterium avium complex) and myelodysplasticsyndrome (MDS)[12,13]. While its presentation is variable, Mono-MAC syndrome patients always show cytopenias, ranging from mild to severe, and hypocellular bone marrow[13,14]. These patientsare susceptible to mycobacterial and viral infections, and have a propensity to develop MDS and Acute Myeloid Leukaemia (AML), with a 75% prevalence and relatively early onset at age 20[13].

Gata2 knockout mice are embryonic lethal and die by E10.5[15]. Conditional *Gata2* knockout under the control of the endothelial *VE-cad* promoter abolished the generation of intra-aortic clusters[16], suggesting that Gata2 is required for HSPC formation. Further studies in the mouse revealed a decrease in HSC numbersin *Gata2* heterozygous mutants, but also a dose-dependency of adult HSCs on Gata2[17].

Gata2 expression in the endothelium is regulated by an intronic enhancer element termed the +9.5 enhancer[18,19]. Deletion of this enhancer results in the loss of HSPC emergence fromHE, leading to lethality by E14[19]. The same element is also mutated in 10% of all the MonoMAC syndrome patients[12].

Because of a partial genome duplication during the evolution of teleost fish, numerous zebrafish genes exist in the form of two paralogues, including *gata2*[20]. This provides an opportunity to separately identify the temporally distinct contributions made by each Gata2 orthologue. *Gata2a* and *gata2b* are only 57% identical and are thought to have undergone evolutionary sub-functionalisation from the ancestral vertebrate *Gata2* gene[21,22]. *Gata2b* is expressed in HE from 18hpf and is thought to regulate *runx1* expression in HE[21]. Lineage tracing experiments showed that *gata2b*-expressing HE cells gave rise to HSCs in the adult[21]. Similar to the mouse Gata2, *gata2b* expression depends on Notch signalling and is a *bona fide* marker of HE, currently regarded asthe functional 'haematopoietic homologue' of Gata2 in zebrafish[21]. By contrast, *gata2a* is expressed in all endothelial cells and in the developing central nervous system[21,23]. Homozygous *gata2a*^{um27} mutants showed arteriovenous shunts in the dorsal aorta at 48hpf[24]. However, *gata2a* is expressed at 11hpf in the haemangioblast population in the posterior lateral mesoderm (PLM) that gives rise to the arterial endothelial cells in the trunk[25], well before *gata2b* is expressed in HE. This suggests that *gata2a* might play a role in endothelial and HE programming andthus help to elucidate an earlier role for Gata2 in HSC development.

Here we show that the *gata2a* locus contains a conserved enhancer in its 4th intron, corresponding to the described +9.5 enhancer in the mouse Gata2 locus[18,19]. Using CRISPR/Cas9 genome editing, we demonstrated that this region, termed the i4 enhancer, is required for endothelial-specific *gata2a* expression. Homozygous mutants (*gata2a*$^{\Delta i4/\Delta i4}$ mutants) showed decreased expression of the HE-specific genes *runx1* and *gata2b*. Thus, endothelial expression of *gata2a*, regulated by the i4 enhancer, is required for *gata2b* and *runx1* expression in the HE. Strikingly, their expression recovers and by 48hpf, the expression of haematopoietic markers in *gata2a*$^{\Delta i4/\Delta i4}$ mutants is indistinguishable from wild-type siblings. We have demonstrated that this recoveryis mediated by an independent input from Notch signalling, sufficient to recover *gata2b* and *runx1* expression in HE and thusHSPC emergence by 48hpf. We conclude that *runx1* and *gata2b* are regulated by two different inputs, one Notch-independent input from Gata2a and a second from the Notch pathway, acting as a fail-safe mechanism for the initial specification of HSPCs in the absence of the input by Gata2a. Despite the early rescue, *gata2a*$^{\Delta i4/\Delta i4}$ adults showed increased susceptibility to infections, oedema, a hypocellular WKM and neutropenia, a phenotype resembling key features of GATA2 deficiency syndromes in humans. We conclude that Gata2a is required for HE programming in the embryo and to maintain the steady-state haematopoietic output from adult HSPCs and that this function requires the activity of the i4 enhancer.

METHODS

Maintenance of zebrafish.

Zebrafish (Danio rerio) were maintained in flowing system water at 28.5 °C, conductance 450–550 µS and pH 7.0 ± 0.5 as described[47]. Fish suffering from infections or heart oedemas were culled according to Schedule 1 of the Animals (Scientific Procedures) Act 1986. Eggs were collected by natural mating. Embryos were grown at 24–32 °C in E3 medium with methylene blue and staged according to morphological features corresponding to respective age in hours or days post-fertilization (hpf or dpf, respectively). Published strains used in this work were wild-type (wt^{KCL}), Tg(−6.0*itga2b*:EGFP)^{la2} [38,48] and Tg(*kdrl*:GFP)^{s843} [28]; animals were used at embryonic and larval stages and as adults (male and female) as specified in the figures. All animal experiments were approved by the relevant University of Oxford, University of Birmingham and Erasmus University ethics committees.

ATAC-seq.

Tg(*kdrl*:GFP)^{s843} embryos were dissociated for FACS at 26-27hpf to collect *kdrl*$^{+}$ and *kdrl*$^{-}$ cell populations (40,000–50,000 cells each). They were processed for ATAC library preparation using optimised standard protocol[27]. Briefly, after sorting into Hanks' solution (1xHBSS, 0.25% BSA, 10mM HEPES pH8), the cells were spun down at 500 g at 4 °C, washed with ice-cold PBS and resuspended in 50 µl cold Lysis Buffer (10mM Tris-HCl, 10 mM NaCl, 3mM MgCl2, 0.1% IGEPAL, pH 7.4). The nuclei were pelleted for 10 min. at 500 × g at 4 °C and resuspended in the TD Buffer with Tn5 Transposase (Illumina), scaling the amounts of reagents accordingly to the number of sorted cells. The transposition reaction lasted 30 min. at 37 °C. The DNA was purified with PCR Purification MinElute Kit (QIAGEN). In parallel, transposase-untreated genomic DNA from *kdrl*$^{+}$ cells was purified with the DNeasy® Blood & Tissue Kit (QIAGEN). The samples were amplified with appropriate Customized Nextera primers[27] in NEBNext High-Fidelity 2x PCR Master Mix (NEB). The libraries were purified with PCR Purification MinElute Kit (QIAGEN) and Agencourt AMPure XP beads (Beckmann Coulter). The quality of each library was verified using D1000 ScreenTape System (Agilent). Four biological replicas of the libraries were quantified with the KAPA Library Quantification Kit for Illumina® platforms (KAPA Biosystems). The libraries were pooled (including the Tn5-untreated control), diluted to 1 ng/µl and sequenced using 75 bp paired-end reads on Illumina HiSeq 4000 (Wellcome Trust Centre for Human Genetics, Oxford). Raw sequenced reads were checked for base qualities, trimmed where 20% of the bases were below quality score 20, and filtered to exclude adapters using Trimmomatic (Version 0.32) and mapped to Zv9 reference genome (comprising 14,612 genes)[49] using BWA with default parameters. The results were visualised using UCSC Genome Browser (http://genome-euro.ucsc.edu/)[50]. The eight data sets were analysed with Principal Component Analysis (PCA) to identify outliers. Correlation among *kdrl*:GFP^{+} and *kdrl*:GFP^{-} samples was assessed with a tree map. The peaks were called for each sample using the Tn5-untreated control as input. We identified the common peaks between replicates and then used DiffBind (EdgeR method) to identify differential peaks between *kdrl*:GFP^{+} and *kdrl*:GFP^{-} samples (Supplementary Data 2). The threshold for differential peaks was $p < 0.05$.

Generation of transgenic and mutant zebrafish lines.

Genomic regions containing the identified 150bp-long gata2a-i4 enhancer flanked by ±500 bp (i4–1.1 kb) or ±150 bp (i4–450bp) were amplified from wild-type zebrafish genomic DNA with NEB Phusion® polymerase (see Supplementary Table 1 for primer sequences) and cloned upstream of E1b minimal promoter and GFP into a Tol2 recombination vector (Addgene plasmid #3784551) with Gateway® cloning technology (Life Technologies™) following the manufacturer's protocol. One-cell zebrafish embryos were injected with 1 nl of an injection mix, containing 50 pg *gata2a*-i4-E1b-GFPTol2 construct DNA + 30 pg tol2 transposase mRNA31. Transgenic founders (Tg (*gata2a*-i4-1.1 kb:GFP) and (*gata2a*-i4-450 bp:GFP)) were selected under a wide field fluorescent microscope and outbred to wild-type fish. Carriers of monoallelic insertions were detected by the Mendelian distribution of 50% fluorescent offspring coming from wild-type outcrosses. These transgenics were then inbred to homozygosity.

To generate the i4 deletion mutant, we identified potential sgRNA target sites flanking the 150 bp conserved region within intron 4 of the *gata2a* locus (see Fig. 1a, Supplementary Fig 3a). sgRNAs were designed with the CRISPR design tool (http:/crispr.mit.edu/, see Supplementary Table 1 for sequences) and prepared as described[32]. To reduce potential off-target effects of CRISPR/Cas9, we utilized the D10A 'nickase' version of Cas9 nuclease[52,53], together with two pairs of sgRNAs flanking the enhancer (Supplementary Table 1, Supplementary Fig. 3a, b). We isolated two mutant alleles with deletions of 215 bp (Δ78–292) and 231 bp (Δ73–303) (Supplementary Fig. 3b). Both deletions included the highly conserved E-box, Ets and GATA transcription factor binding sites (Supplementary Fig. 3b). The Δ73–303 allele was selected for further experiments and named Δi4. Adult zebrafish were viable and fertile as heterozygous (*gata2a*$^{\Delta i4/+}$) or homozygous (*gata2a*$^{\Delta i4/\Delta i4}$). To unambiguously genotype wild types, heterozygotes and homozygous mutants, we designed a strategy consisting of two PCR primer pairs (Supplementary Fig. 3a, c). One primer pair flanked the whole region, producing a 600 bp wild-type band and 369 bp mutant band. In the second primer pair, one of the primers was designed to bind within the deleted region, only giving a 367 bp band in the presence of the wild-type allele (Supplementary Fig. 3c).

To generate the gata2b mutant we designed a CRISPR/Cas9 strategy for a frameshift truncating mutant in exon 3 deleting both zinc fingers. sgRNAs were designed as described above and guides were prepared according to Gagnon et al.[54] with minor adjustments. Guide RNAs were generated using the Agilent SureGuide gRNA Synthesis Kit, Cat# 5190–7706. Cas9 protein (IDT) and guide were allowed to form ribonucleoprotein structures (RNPs) at

RT and injected in 1 cell stage oocytes. 8 embryos were selected at 24 hpf and lysed for DNA isolation. Heteroduplex PCR analysis was performed to test guide functionality and the other embryos from the injection were allowed to grow up. To aid future genotyping we selected mutants by screening F1 for a PCR detectable integration or deletion in, exon 3. Sequence verification showed that founder 3 had a 28 nt integration resulting in a frameshift truncating mutation leading to 3 new STOP codons in the third exon. To get rid of additional mutations caused by potential off-target effects, founder 3 was crossed to WT for at least 3 generations. All experiments were performed with offspring of founder 3.

Fluorescence-activated cell sorting (FACS).

Approximately 100 embryos at the required stage were collected in Low Binding® SafeSeal® Microcentrifuge Tubes (Sorenson) and pre-homogenized by pipetting up and down in 500 µl Deyolking Buffer (116mM NaCl, 2.9mM KCl, 5 mM HEPES, 1 mM EDTA). They were spun down for 1 min. at 500 g and incubated for 15 min. at 30 °C in Trypsin + Collagenase Solution (1xHBSS, 0.05% Gibco® Trypsin + EDTA (Life Technologies), 20 mg/ml collagenase (Sigma)). During that time, they were homogenized by pipetting up and down every 3 min. The lysis was stopped by adding 50 µl foetal bovine serum and 650 µl filter-sterilized Hanks' solution (1xHBSS, 0.25% BSA, 10 mM HEPES pH8). The cells were rinsed with 1 ml Hanks' solution and passed through a 40 µm cell strainer (Falcon®). They were resuspended in ~400 µl Hanks' solution with 1:10,000 Hoechst 33258 (Molecular Probes®) and transferred to a 5 ml polystyrene round bottom tube for FACS sorting. The cells were sorted on FACSAria Fusion sorter by Kevin Clark (MRC WIMM FACS Facility). The gates of GFP (488–530) and DsRed (561–582) channels were set with reference to samples derived from non-transgenic embryos. The fluorescence readouts were compensated when necessary. For ATAC-seq library preparation, the cells were sorted into Hank's solution. For RNA isolation, the cells were sorted directly into RLT Plus buffer (QIAGEN) + 1% β-mercaptoethanol and processed with the RNEasy® Micro Plus kit (QIAGEN), according to the accompanying protocol. The RNA was quantified and its quality assessed with the use of Agilent RNA 6000 Pico kit. All RNA samples were stored at −80 °C.

SYBR® Green qRT-PCR.

3 µl of the cDNA diluted in H2O were used for technical triplicate qRT-PCR reactions of 20 µl containing the Fast SYBR® Green Master Mix (Thermo Fisher Scientific) and appropriate primer pair (see Supplementary Table 1). The reactions were run on 7500 Fast Real-Time PCR System (Applied Biosystems) and the results were analysed with the accompanying software. Notemplate controls were run on each plate for each primer pair. Each reaction was validated with the melt curve analysis. The baseline values were calculated automatically for each reaction. The threshold values were manually set to be equal for all the reactions run on one plate, within the linear phase of exponential amplification. The relative mRNA

levels in each sample were calculated by subtracting the geometric mean of Ct values for housekeeping genes *eef1a1l1* and *ubc* from the average Ct values of the technical triplicates for each gene of interest. This value

(ΔCt) was then converted to a ratio relative to the housekeeping genes with the formula 2−ΔCt.

Fluidigm Biomark qRT-PCR.

To quantify the differences in gata2a expression between wild-type and mutant ECs, we crossed homozygous *gata2a*$^{\Delta i4/\Delta i4}$ mutants to Tg(*kdrl*:GFP) transgenics to generate Tg(*kdrl*:GFP); *gata2a*$^{\Delta i4/\Delta i4}$ embryos. These fish, along with non-mutant Tg(*kdrl*:GFP), were used for FACS-mediated isolation of *kdrl*:GFP^{+} and *kdrl*:GFP^{-} cells to quantitatively compare mRNA expression levels of *gata2a* in the endothelial and non-endothelial cells of wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ embryos, using the Fluidigm Biomark™ qRT-PCR platform. Briefly, 1 ng RNA from FACS-sorted cells was used for Specific Target Amplification in a 10 µl reaction with the following reagents: 5 µl 2xBuffer and 1.2 µl enzyme mix from SuperScript III One-Step Kit (Thermo Fisher Scientific), 0.1 µl SUPERase• In™ RNase Inhibitor (Ambion), 1.2 µl TE buffer (Invitrogen), 2.5 µl 0.2x TaqMan® assay mix (see Supplementary Table 2 for the details of TaqMan® assays). The reaction was incubated for 15 min. at 50 °C, for 2 min. at 95 °C and amplified for 20 cycles of 15 s at 95 °C/4 min. at 60 °C. The cDNA was diluted 1:5 in TE buffer and stored at −20 °C. Diluted cDNA was used for qRT-PCR according to the Fluidigm protocol for Gene Expression with the 48.48 IFC Using Standard TaqMan® Assays (Supplementary Table 2). Each sample was run in 3–4 biological replicates. The collected data were analysed with Fluidigm Real-Time PCR Analysis software (version 4.1.3). The baseline was automatically corrected using the built-in Linear Baseline Correction. The thresholds were manually adjusted for each gene to fall within the linear phase of exponential amplification, after which they were set to equal values for the housekeeping genes: *rplp0*, *rpl13a*, *cops2*[55], *lsm12b*[56] and *eef1a1l1*. The relative mRNA levels for each sample were calculated by subtracting the geometric mean of Ct values for the housekeeping genes from the Ct value for each gene of interest. This value (ΔCt) was then converted to a ratio relative to the housekeeping genes with the formula 2−ΔCt. The ΔCt values were analysed with 2-tailed paired-samples t-tests with 95% confidence levels.

Flow cytometry and isolation of WKM haematopoietic cells.

Single cell suspensions of WKM cells were prepared from adult zebrafish kidneys of the required genotypes as described57. Briefly, adult zebrafish were first euthanized in 0.5% tricaine in PBS and dissected to remove the kidney. WKM cells were recovered by vigorous pipetting in 0.5 ml PBS with 10% Foetal Calf Serum (PBS + 10%FCS), followed by filtration in a cell strainer (FALCON, ref 352235) pre-coated with PBS + 10%FCS. Strainers were rinsed with PBS + 10%FCS and the cells spun down (~300 g, 10 min at 4 °C) and ressuspended in 200–500

µl PBS + 10%FCS with Hoechst 33342 1:10000 (Hoechst 33342, H3570, ThermoScientific). Flow cytometry analysis was performed on a FACS Aria II (BD Biosciences) after exclusion of dead cells by uptake of Hoechst dye, as described[41]. WKM cell counts were performed on a PENTRA ES60 (Hariba Medical) following the manufacturer's instructions. Note that the cell counter does not recognize the zebrafish nucleated erythrocytes, so these were excluded from this analysis. Cell counts for each genotype were analysed with 2-tailed paired-samples t-tests with 95% confidence levels, using a Mann-Whitney test for non-parametric distribution. The scatter plots were generated using GraphPad Prism 8.0 and show medians ± SD.

May-Grunwald and Wright-Giemsa staining.

Cell staining with May-Grunwald (MG) stain (Sigma MG500) and Giemsa (GIEMSA STAIN, FLUKA 48900) was performed on haematopoietic cell samples. After cytospin, slides are allowed to airdry and were stained for 5 min at room temperature with a 1:1 mix of MG:distilled water. Next, slides were drained and stained with a 1:9 dilution of Giemsa:distilled water solution for 30 min at room temperature. Excess solution was drained and removed by further washes in distilled water. Finally, the slides were air-dried and mounted in DPX (06522, Sigma) for imaging.

Whole-mount in situ hybridization and immunohistochemistry.

Whole-mount in situ hybridization (ISH) was carried out as described previously[58], using probes for *kdrl*, *runx1*, *cmyb*, *gata2a*, *gata2b*, *rag1*[4,37,59,60] and *gfp* (Supplementary Table 1). For conventional ISH embryos were processed, imaged and the ISH signal quantified as described34. Briefly, the pixel intensity values were assessed for normal distribution with a Q-Q plot and transformed when necessary. Mean values (µ) of each experimental group were analysed with 2-tailed independent-samples t-tests or with ANOVA with 95% confidence levels, testing for the equality of variances with a Levene's or Brown-Forsythe test and applying the Welch correction when necessary. For ANOVA, differences between each two groups were assessed with either Tukey's post-hoc test (for equal variances) or with Games-Howell test (for unequal variances). For all these analyses, the IBM® SPSS® Statistics (version 22) or GraphPad Prism 8.0 package were used.

For the analysis of cmyb expression in the CHT at 4dpf, the embryos scored as 'high' or 'low' were tested for equal distribution between morphants and uninjected controls or among wild-type, heterozygous and mutant genotypes with contingency Chi-squared tests, applying Continuity Correction for 2 × 2 tables, using IBM® SPSS® Statistics (version 22).

For fluorescent ISH (FISH) combined with immunohistochemistry, ISH was performed first following the general whole-mount in situ hybridisation protocol. The signal was developed with SIGMAFAST Fast Red TR/Naphthol, the embryos rinsed in phosphate-buffered saline with tween20 (PBT) and directly processed for immunohistochemistry. Embryos were

blocked in blocking buffer (5% goat serum/0.3% Triton X-100 in PBT) for 1 h at RT before incubated with primary antibody against GFP (rabbit, 1:500, Molecular Probes), diluted in blocking buffer overnight at 4 °C. Secondary antibody raised in goat coupled to AlexaFluor488 (Invitrogen) was used in 1:500 dilutions for 3 h at RT. Hoechst 33342 was used as a nuclear counterstain.

Fluorescent images were taken on a Zeiss LSM880 confocal microscope using ×40 or ×63 oil immersion objectives. Images were processed using the ZEN software (Zeiss).

Fluorescence microscopy and cell counting.

For widefield fluorescence microscopy, live embryos were anaesthetised with 160 µg/ml MS222 and mounted in 3% methylcellulose and imaged on a AxioLumar V.12 stereomicroscope (Zeiss) equipped with a Zeiss AxioCam MrM. To count *itga2b*-GFP^{high} and *itga2b*-GFP^{low} cells in the CHT, Tg(*itga2b*:GFP;*kdrl*:mCherry); *gata2a*$^{\Delta i4/+}$ animals were incrossed and grown in E3 medium supplemented with PTU to prevent pigment formation. At 5dpf, the larvae were anaesthetised with MS222 and the tail was cut and fixed for 1 h at room temperature in 4% PFA. Next, the tails were mounted on 35 mm glass bottomed dishes (MAtTEK) in 1% low melt agarose and imaged using a ×40 oil objective on an LSM880 confocal microscope (Zeiss). Cells in the CHT region were counted manually on Z-stacks as 'itga2b:GFPlow' (HSPCs) or '*itga2b*:GFP^{high}' (thrombocytes). Genomic DNA from the heads was extracted and used for genotyping as described above. Cell counts for each genotype were analysed with 2-tailed paired-samples t-tests with 95% confidence levels, using a Mann-Whitney test for non-parametric distribution. The graphs were generated using GraphPad Prism 8.3.0 and show medians ± SD.

Statistics and reproducibility.

Data were analysed using either IBM® SPSS® Statistics (version 22) or GraphPad Prism software (v8.02). In situ quantification data was analysed with 2-tailed independent-samples t-tests or with ANOVA with 95% confidence levels, testing for the equality of variances with a Levene's or Brown-Forsythe test and applying the Welch correction when necessary. For ANOVA, differences between each two groups were assessed with either Tukey's post-hoc test (for equal variances) or with Games-Howell test (for unequal variances). Alternatively, data were analysed with an appropriate non-parametric test (Kruskall-Wallis) followed by Dunn's multiple comparisons test or uncorrected Dunn's test where appropriate. Cell count data were analysed with 2-tailed paired- samples t-tests with 95% confidence levels, using a Mann-Whitney test for nonparametric distribution. Gene expression data were analysed with 2-tailed paired samples t-tests with 95% confidence levels. All experiments were repeated at least three times with similar results obtained;
sample sizes are shown in the respective figure legends.

RESULTS

Analysis of open chromatin regions in the gata2a locus.

Because Gata2 genes are duplicated in zebrafish, we set out to unpick the different roles Gata2a and Gata2b play during HSC generation and homeostasis by identifying their regulatory regions. Analysis of sequence conservation revealed that one region within the fourth intron of the zebrafish *gata2a* locus was conserved in vertebrates, including mouse and human (Fig. 1a–c). This region, which we termed 'i4 enhancer', corresponds to the endothelial +9.5 Gata2 enhancer identified previously in the mouse[18,19] and human[26]. Notably, the *gata2b* locus did not show broad conservation in non-coding regions (Supplementary Fig. 1a).

To investigate whether the i4 element was a potentially active enhancer, we first performed ATAC-seq[27] to identify open chromatin regions in endothelial cells (ECs) in zebrafish. We used a Tg(*kdrl*:GFP) transgenic line that expresses GFP in all endothelium[28] and isolated the higher GFP-expressing ECs (*kdrl:GFP*^{high}, termed *kdrl:GFP*$^{+}$ for simplicity) as this fraction was enriched for endothelial markers compared to the *kdrl:GFP*^{low} fraction (Supplementary Fig. 1b, c). Principal Component Analysis on the ATAC-seq data from 26hpf *kdrl*:GFP^{+} cells (n = 2) and kdrl:GFP^{-} cells (n = 4) revealed strong differences between the open chromatin regions in the two cell populations, further supported by a correlation analysis (Supplementary Fig. 1d–f). 78,026 peaks were found in common between replicates of the ATACseq in kdrl:GFP^{+} cells (Supplementary Fig. 1g). 44,025 peaks were differentially expressed between the kdrl:GFP^{+} and kdrl:GFP^{-} fractions (Supplementary Fig. 1h). An analysis of known motifs present in the kdrl:GFP^{+} population revealed an enrichment for the ETS motif (Supplementary Fig. 1i). ETS factors are essential regulators of gene expression in endothelium[29]. In addition, we performed gene ontology (GO) term analysis on the peaks showing >3-fold enrichment or depletion in ECs (Supplementary Fig. 1j–l). As expected, non-ECs showed a broad range of GO terms whereas EC-enriched peaks were associated with terms like angiogenesis or blood vessel development (Supplementary Fig. 1k, l).

Differential peak analysis in the gata2a locus identified four differentially open sites within a 20 kb genomic region (Fig. 1b), including one peak in intron 4 corresponding to the predicted i4 enhancer. It contained a core 150 bp-long element that included several binding motifs for the GATA, E-box and Ets transcription factor families (Fig. 1b). Although the positioning of the E-box site relative to the adjacent GATA site differs in zebrafish and mammals (Fig. 1b, c), the necessary spacer distance of ~9 bp between the two sites[30] was conserved. Thus, this site may be a target for TF complexes containing an E-box-binding factor and a GATA family TF.

Thus, the intronic enhancer (i4) identified in the zebrafish gata2a locus is accessible to transposase in endothelial cells and contains highly conserved binding sites for key haematopoietic transcription factors, suggesting that genetic regulation of gata2a expression in zebrafish HE is a conserved feature of vertebrate gata2 genes.

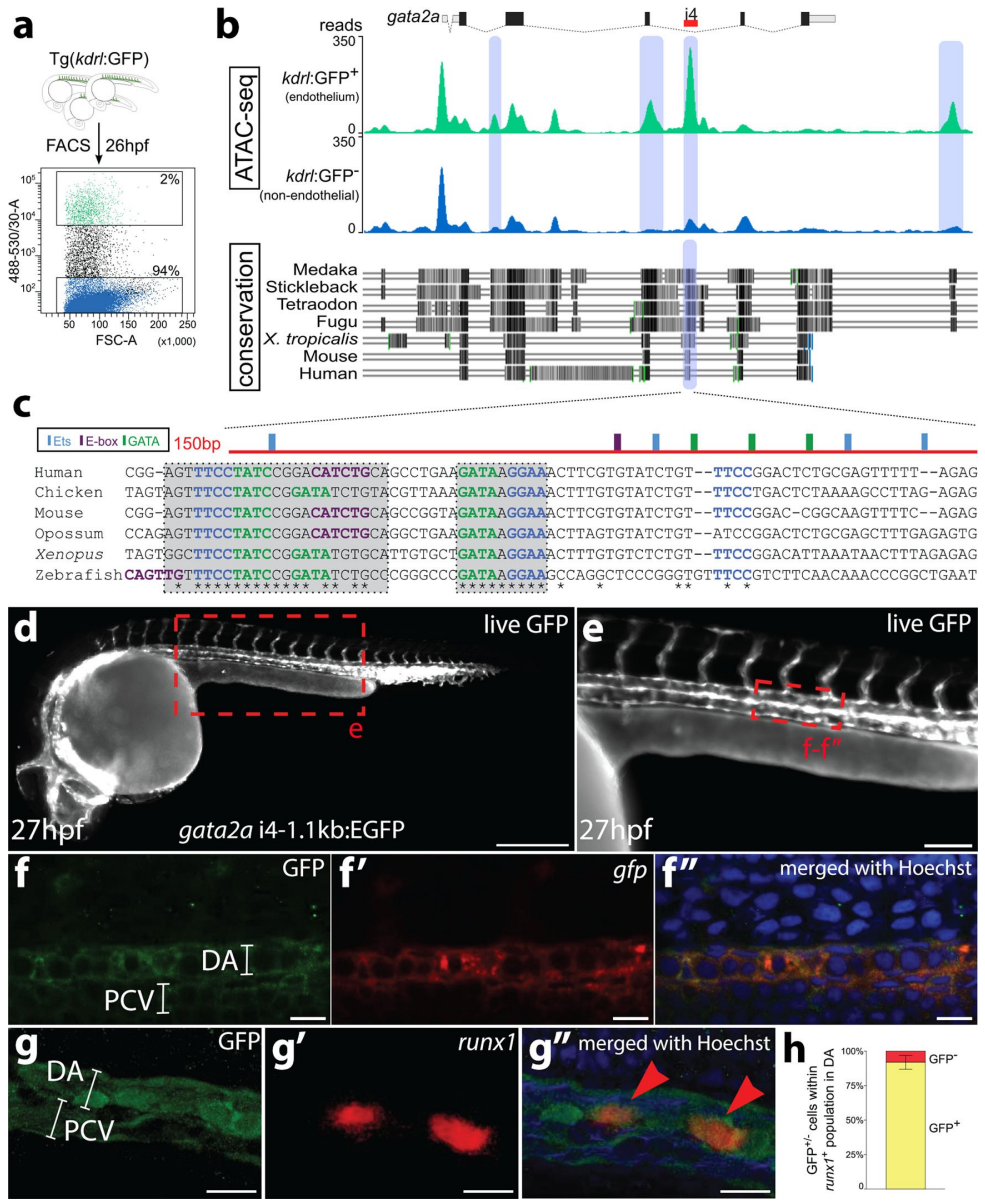


Fig. 1 The i4 enhancer in the gata2a locus is conserved and drives pan-endothelial expression of a GFP reporter in zebrafish.

a *Kdrl*:GFP^{+} (green) and *kdrl*:GFP^{-} (blue) cells were FACS-sorted from 26hpf embryos and used for preparation of ATAC-seq libraries. **b** The image of the mapped reads represents stacked means of two biological ATAC-seq replicates. Differential peak analysis identified four chromatin regions (blue shading) in the locus of *gata2a* that are significantly more open in the *kdrl*:GFP^{+} population ($p < 0.0001$). A region in the fourth intron (termed i4 enhancer) is conserved throughout vertebrates. Black and grey shading denotes regions of high conservation between the species analysed. **c** The highly conserved 150 bp region (red) contains putative transcription factor binding sites, mapped computationally. Light blue: Ets binding sites; purple: E-box binding sites; green: GATA binding sites;

asterisks: conserved residues. **d** Widefield fluorescent image of a live Tg(*gata2a*-i4-1.1 kb:GFP) zebrafish embryo at 27hpf showing GFP fluorescence in the endothelial cells and in the heart (endocardium). **e** Higher magnification image of the trunk of the embryo from panel **d. f–f″** Confocal images of a trunk fragment of a Tg(*gata2a*-i4-1.1 kb:GFP) embryo immunostained with anti-GFP antibody (**f**) and probed for gfp mRNA (**f′**) at 25hpf. **f″** Merged images from panels **f–f′** with Hoechst nuclear staining in blue, showing complete overlap of GFP protein and mRNA. **g–g″** Confocal images of the dorsal aorta (DA) and posterior cardinal vein (PCV) of a Tg(*gata2a*-i4-1.1 kb:GFP) embryo immunostained with anti-GFP antibody (**g**) and probed for runx1 mRNA (**g′**) at 25hpf. See panel e for approximate position within the embryo. **g″** Merged images from panels **g–g′**, also showing Hoechst nuclear staining in blue. **h** Counting of the *runx1*$^{+}$ cells represented in panels **g′–g″** in 25 embryos shows that >90% of *runx1*$^{+}$ cells are also GFP^{+}. N = 3. Error bars: ± SD. See also Supplementary Fig. 1.

The gata2a-i4 enhancer drives GFP expression in endothelium.

To investigate the activity of the gata2a-i4 enhancer in vivo, the conserved genomic 150 bp region (Fig. 1b, c), together with flanking ±500 bp (gata2a-i4-1.1 kb:GFP) or ±150 bp (gata2a-i4- 450 bp:GFP) was cloned into a Tol2-based reporter E1b:GFP construct[31] and used to generate stable transgenic lines (Supplementary Fig. 2). The earliest activity of the enhancer was observed at the 14-somite stage (14ss), when gfp mRNA was detected in the PLM (Supplementary Fig. 2a, b). After 22hpf, the reporter signal was pan-endothelial (Fig. 1d–e, Supplementary Fig. 2c–i). Around 27hpf, higher intensities of GFP fluorescence and correspondingly higher levels of gfp mRNA were visible in the floor of the DA (Fig. 1d–e, Supplementary Fig. 2e–h). While the GFP protein was still visible in the vasculature around 3dpf, it was likely carried over from earlier stages, since the gfp mRNA was not detectable any more (Supplementary Fig. 2i, j). We focused our subsequent analysis on the gata2a-i4-1.1 kb:GFP transgenics as they showed stronger expression of the transgene. At 25hpf, the expression of GFP protein and gfp mRNA overlapped completely in the endothelial cells of the DA (Fig. 1f–f″). Overall, these data confirm that the i4 enhancer is active in vivo in endothelial cells at the correct time to regulate definitive haematopoiesis. The endothelial activity of the corresponding +9.5 enhancer was also observed in mouse embryos[18], indicating functional conservation of the gata2a-i4 enhancer across vertebrates.

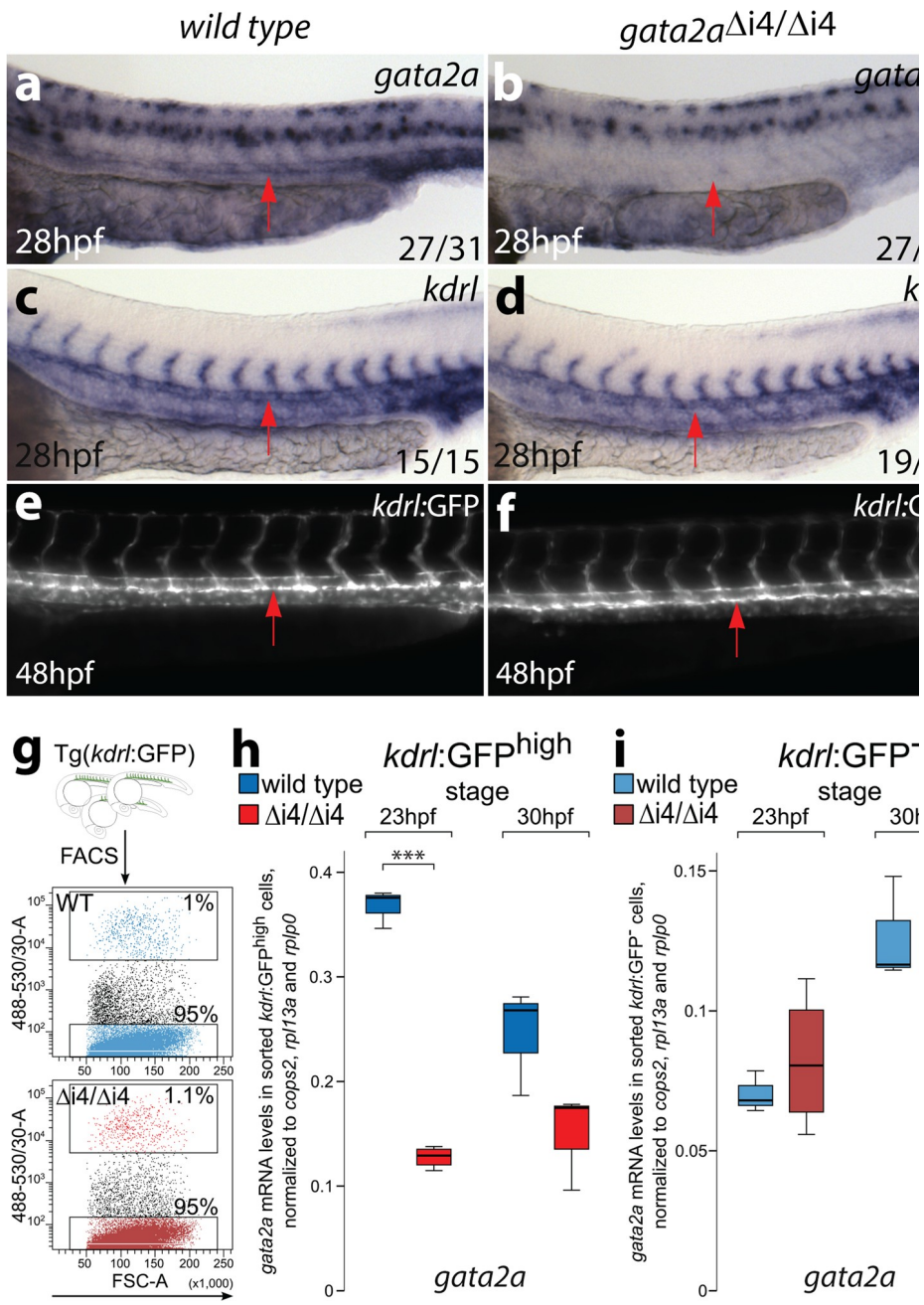
To further characterise the enhancer activity in vivo, Tg (gata2a-i4-1.1 kb:GFP) embryos were stained for gata2a mRNA and for GFP protein (Supplementary Fig. 2k–o). We found a large overlap between gata2a+ and GFP+ cells at 30hpf in the DA, with a small proportion of GFP+ cells that did not express gata2a mRNA (<5%, Supplementary Fig. 2o). This could suggest that some cells require activity of other endothelial enhancers to trigger transcription of *gata2a* or that *gfp* mRNA has a longer half-life than *gata2a* mRNA. Importantly, the GFP signal was absent in *gata2a*-expressing neural cells (Supplementary Fig. 2l–n), indicating that the i4 enhancer is specifically activein (haemogenic) endothelial cells.

Next we examined the expression of the HE marker runx14 in gata2a-i4-1.1 kb:*GFP* embryos at 25hpf. At this stage, over 90% of runx1^{+} cells were GFP^{+} (Fig. 1g–h). We conclude that the *GFP* expression under the gata2a-i4 enhancer marks the majority of the HE population.

Endothelial gata2a required gata2a-i4 enhancer activity.

To investigate whether endothelial-specific expression of *gata2a* is required for definitive haematopoiesis, we deleted the conserved *gata2a*-i4 enhancer using CRISPR/Cas9 genome editing[32]. We generated a deletion mutant lacking 231 bp of the i4 enhancer (Supplementary Fig. 3a–c) and named it *gata2a*$^{\Delta i4/\Delta i4}$. Homozygous *gata2a*$^{\Delta i4/\Delta i4}$ mutants showed decreased levels of *gata2a* expression in endothelial cells when compared to wild-type embryos (Fig. 2a, b). By contrast, *gata2a* expression in the neural tube appeared unaffected in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Fig. 2a, b). At 28hpf, expression of the pan-endothelial marker *kdrl* was indistinguishable between wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Fig. 2c, d). To verify these results, we crossed homozygous *gata2a*$^{\Delta i4/\Delta i4}$ mutants to Tg(*kdrl*:GFP) transgenics and analysed vascular morphology. *Gata2a*$^{\Delta i4/\Delta i4}$ embryos showed no gross vascular abnormalities at 48hpf as assessed by the expression of the Tg(*kdrl*:GFP) transgene (Fig. 2e, f).

Next, we isolated endothelial cells from Tg(*kdrl*:GFP) andTg(*kdrl*:GFP);*gata2a*$^{\Delta i4/\Delta i4}$ embryos by FACS (Fig. 2g) at 23hpf and 30hpf and confirmed by qRT-PCR that the endothelial markers *kdrl, dld* and *dll4* were unaffected in *gata2a*$^{\Delta i4/\Delta i4}$ embryos (Supplementary Fig. 3d–f). The arterial marker *efnb2a*[33] was decreased at 23hpf in *gata2a*$^{\Delta i4/\Delta i4}$ mutants but recovered by30hpf (Supplementary Fig. 3g). In addition, *gata2a* was significantly decreased in the *kdrl*:GFP^{+} ECs in 23hpf *gata2a*$^{\Delta i4/\Delta i4}$ embryos compared to wild types (Fig. 2h). At 30hpf this decrease was not statistically significant. This was likely due to a decrease in expression of *gata2a* that appears to occur inwild-type ECs during development, whereas *gata2a* expression inmutants remained low (Fig. 2h). Importantly, there was no difference in *gata2a* expression in the non-endothelial population (*kdrl*:GFP^{-} cells) between wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants at either 23hpf or 30hpf (Fig. 2i). Altogether, these data suggest that genomic deletion of the *gata2a*-i4 enhancer is sufficient to reduceexpression of *gata2a* specifically in endothelium.



Gata2a regulates runx1 and gata2b in haemogenic endothelium.

To investigate a potential role of *gata2a* in HSC development, we compared the expression of *runx1*, the key marker of HE in zebrafish[4], in wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ embryos. Quantitative in situ hybridization (ISH) analysis[34] showed that *runx1* expression was decreased in *gata2a*$^{\Delta i4/\Delta i4}$ embryos at 24hpf (Supplementary Fig. 4a–c) and 28hpf compared to wild-typesiblings (Fig. 3a–c). Further analysis in *kdrl*:GFP^{+} ECs showed that this decrease in *runx1* expression was already detectable at 23hpf in *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Fig. 3d), at the onset of its expression in HE[35]. Thus, deletion of the *gata2a*-i4 enhancer results in impaired *runx1* expression in the early stages of HE programming. This correlates well with decreased *runx1* expression levels in +9.5$^{-/-}$ mouse AGM explants[19], further supporting the critical evolutionary role of the intronic enhancer of *Gata2* in HSC specification.

Next, we tested whether Gata2a could act upstream of *gata2b* by measuring its expression in *gata2a*$^{\Delta i4/\Delta i4}$ embryos. Quantitation of the ISH signal showed that *gata2b* expression was decreased in *gata2a*$^{\Delta i4/\Delta i4}$ embryos compared to wild-type siblings at 26hpf (Supplementary Fig. 4d) and 28hpf (Fig. 3e–g), but recovered to wild-type levels by 30hpf (SupplementaryFig. 4e). Accordingly, *kdrl*:GFP^{+}; *gata2a*$^{\Delta i4/\Delta i4}$ cells expresssignificantly lower levels of *gata2b* mRNA than the wild-type *kdrl*:GFP^{+} endothelial population at 23hpf, but not at 30hpf(Fig. 3h). These data suggest that endothelial expression of *gata2a* is required upstream of *gata2b* and *runx1* for the proper specification of HE, uncovering a previously unrecognized role for Gata2a in definitive haematopoiesis.

Fig. 2 Deletion of the i4 enhancer in *gata2a*$^{\Delta i4/\Delta i4}$ mutants leads to reduced levels of *gata2a* mRNA in the endothelium.

a, b A significant majority of *gata2a*$^{\Delta i4/\Delta i4}$ mutants have reduced levels of *gata2a* mRNA in the dorsal aorta (arrows) at 28hpf, compared to wild-type siblings, as detected with in situ hybridization. (X2 = 10.720, d.f. = 1, $p < 0.01$). The expression in the neural tube appears unaffected. **c, d** In situ hybridization for the endothelial marker kdrl at 28hpf reveals no difference between *gata2a*$^{\Delta i4/\Delta i4}$ mutants and wild-type siblings. The dorsal aorta (arrows) appears unaffected. **e, f** Live images of the trunks of 48hpf Tg(*kdrl*:GFP) and Tg(*kdrl*:GFP); *gata2a*$^{\Delta i4/\Delta i4}$ embryos show normal vascular morphology in the mutants. The endothelium of the dorsal aorta (arrows) appears normal in the gata2aΔi4/Δi4 embryos. **g** *Kdrl*:GFP^{high} and *kdrl*:GFP^{-} cells were sorted from non-mutant (WT, blue) and *gata2a*$^{\Delta i4/\Delta i4}$ (red) embryos carrying the Tg(*kdrl*:GFP) transgene. **h, i** qRT-PCR on RNA isolated from the sorted *kdrl*:GFP^{high} or *kdrl*:GFP^{-} cells (panel g) shows decreased levels of gata2a mRNA in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 23hpf (t =20.026,d.f. = 5, $p < 0.001$) compared to wild-type. At 30hpf this difference is not statistically significant (t = 2.146, d.f. = 4, p = 0.098). There is no difference in gata2a mRNA levels in non-endothelial cells between wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (23hpf: t = 0.69, d.f. = 5, $p > 0.5$; 30hpf: t = 0.618, d.f. = 4, $p > 0.5$). N = 4 for *gata2a*$^{\Delta i4/\Delta i4}$ at 23hpf, N = 3 for other samples. Note different scales of expression levels. ***$p < 0.001$. See also Supplementary Fig. 2.

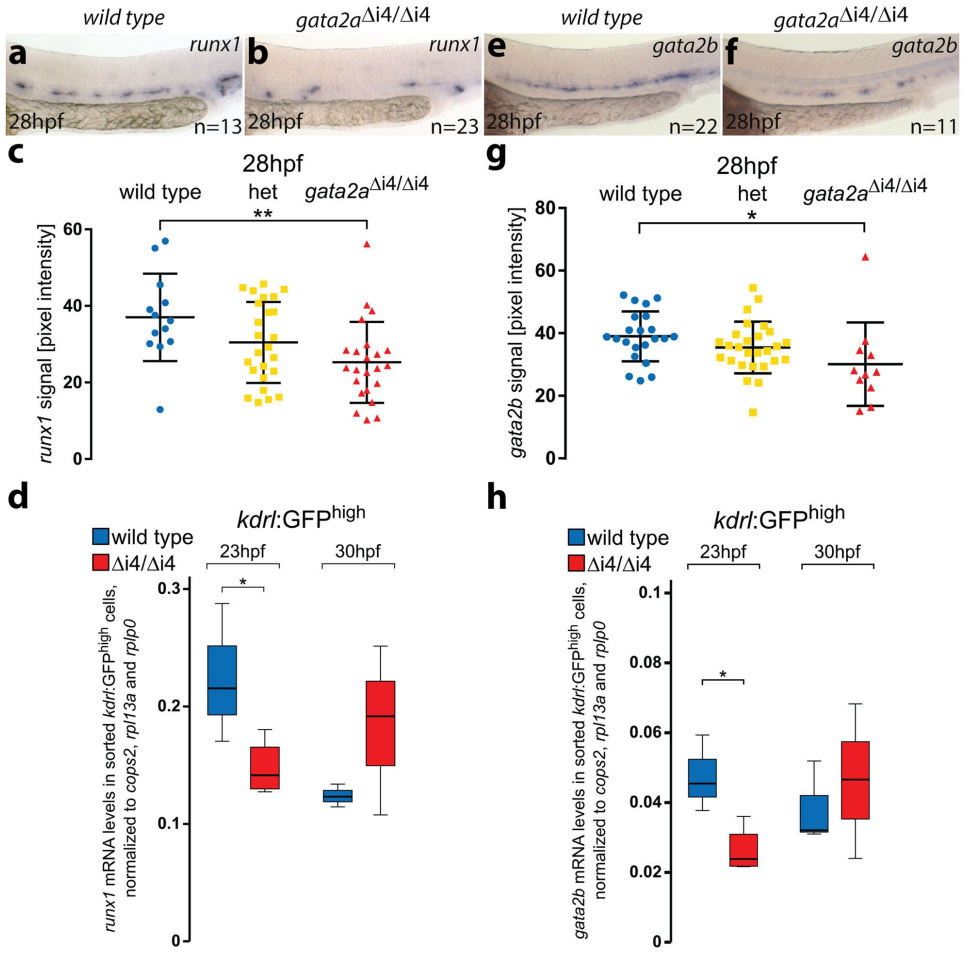


Fig. 3 Loss of *gata2a* expression in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants leads to decreased levels of *runx1* and *gata2b* in the HE.

a, b In situ hybridization for *runx1* expression in the HE of wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ embryos at 28hpf. **c** Quantification of the runx1 in situ hybridization signal from wild-type (blue), heterozygous *gata2a*$^{+/\Delta i4}$ (het, yellow) and *gata2a*$^{\Delta i4/\Delta i4}$ (red) siblings at 28hpf shows significant decrease in *runx1* pixel intensity in the DA in the homozygous mutants compared to wild-type (μ_{wt} = 34.8, μ_{mut} = 25.3; F =4.956, d.f. = 2, 58; ANOVA),**p < 0.01. n= 14, wild-type; n = 25, het; n = 23, *gata2a*$^{\Delta i4/\Delta i4}$. Error bars: mean ± SD. **d** qRT-PCR on RNA isolated from the sorted *kdrl*:GFP^{+} cells shows decreased levels of *runx1* mRNA in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 23hpf (t = 2.585, d.f. = 5, p < 0.05) but not at 30hpf (t = 1.326, d.f. =4, p > 0.2), compared to wild-type. N= 4 for *gata2a*$^{\Delta i4/\Delta i4}$ at 23hpf, N= 3 for other samples. Note different scales of expression levels. *p < 0.05. **e, f** Gata2b expression in the HE of wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ embryos at 28hpf. **g** Quantification of the *gata2b* mRNA signal, detected by in situ hybridization, from wild-type (blue), heterozygous *gata2a*$^{+/\Delta i4}$ (het; yellow) and *gata2a*$^{\Delta i4/\Delta i4}$ (red) siblings at 28hpf shows significant decrease in *gata2b* pixel intensity in the DA in the homozygous mutants compared to wild-type (μ_{wt} = 39, μ_{mut} = 30.1; F = 5.05, d.f. = 2, 54; ANOVA), *p < 0.05. n = 22, wild-type; n =24, het; n =11, gata2aΔi4/Δi4. Error bars: mean ± SD. **h** qRT-PCR in sorted *kdrl*:GFP^{+} cells showed decreased levels of gata2b mRNA in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 23hpf (t = 3.334, d.f. = 5, p < 0.05) but not at 30hpf (t = 0.373, d.f. = 4, p > 0.7), compared to wild-type. N =4 for gata2aΔi4/Δi4 at 23hpf, N =3 for other samples. *p < 0.05. See also Supplementary Figs. 3 and 4.

Recovery of embryonic HSC activity in gata2a$^{\Delta i4/\Delta i4}$ *mutants.*

The qPCR analysis in sorted *kdrl*:GFP^{+}; *gata2a*$^{\Delta i4/\Delta i4}$ and wild type *kdrl*:GFP^{+} ECs (Fig. 3d) already suggested a recovery of *runx1* expression from 30hpf. Of note, the *kdrl*:GFP^{+} population likely includes the *kdrl*$^{+}$, *runx1*-expressing erythro-myeloid progenitors (EMPs) located in the caudal region[36]. This region was not included in the quantification of ISH but cannot be separated by sorting for *kdrl*:GFP^{+} and could thus explain the discrepancy between image quantification and qRT-PCR. To further characterize the haematopoietic phenotype in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants, we tested whether expression of markers of haematopoietic activity in the embryo was affected from 48hpf onwards (Fig. 4).

At 48hpf, the expression of *runx1* in the DA showed no significant difference between *gata2a*$^{\Delta i4/\Delta i4}$ mutants and wild-type controls (Fig. 4a, b). These data suggest that the decrease of *runx1* expression at early stages of HE programming in *gata2a*$^{\Delta i4/\Delta i4}$ mutants is transient and recovers by 2dpf. Indeed, analysis of the HSPC marker *cmyb*[11] in the CHT at 4dpf showed no differences between *gata2a*$^{\Delta i4/\Delta i4}$ and wild-type larvae (Fig. 4c, d). Expression of the T-cell progenitor marker *rag1* in the thymus[37] showed that around half of the *gata2a*$^{\Delta i4/\Delta i4}$ larvae had reduced *rag1* expression at 4dpf compared to wild-type(Fig. 4e, f). This effect was absent at 5dpf (Fig. 4g, h), suggesting that HSPC activity was normal in *gata2a*$^{\Delta i4/\Delta i4}$ mutants from 4dpf onwards. Next, we crossed the *gata2a*$^{\Delta i4/\Delta i4}$ mutants to Tg(*itga2b*:GFP) transgenics, where *itga2b*-GFP^{high} and *itga2b*-GFP^{low} cells in the CHT mark thrombocytes and HSPCs, respectively[9,38]. Our analysis revealed no difference in *itga2b*-GFP^{low} HSPC or *itga2b*-GFP^{high} thrombocyte numbers in the CHT region at 5dpf between wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants(Fig.4i–l). Taken together, our data suggest that endothelial *gata2a* expression mediated by the i4 enhancer is required for the initial expression of *gata2b* and *runx1* in the HE but largely dispensable after 2dpf.

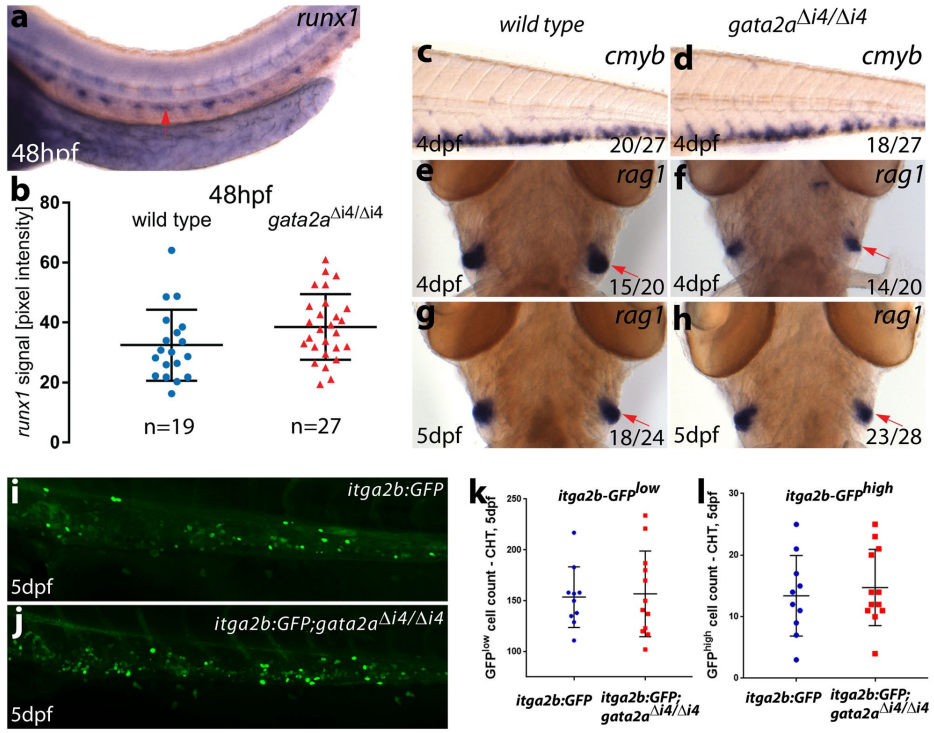


Fig. 4 *Gata2a*$^{\Delta i4/\Delta i4}$ mutants display a recovery of the initial haematopoietic defects from 48hpf.
a Representative image of *runx1* expression in the trunk of a wild-type embryo at 48hpf showing *runx1* mRNA in the dorsal aorta (arrow). **b** Quantification of the *runx1* in situ hybridization signal in wild-type (blue) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (red) siblings at 48hpf. There is no significant difference in *runx1* pixel intensity in the DA between the homozygous mutants and wild-type (μ_{wt} = 33.1, μ_{mut} = 37.5, t =1.410, d.f. =44, p= 0.17. n =19, wild-type; n =27, *gata2a*$^{\Delta i4/\Delta i4}$). Error bars: mean ± SD. **c, d** In situ hybridization for *cmyb* in the CHT. We detected no difference in expression between wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ siblings at 4dpf. (**e–h**) In situ hybridization (ventral view) for rag1 in the thymii, showing a slight decrease (relative to wild-type) in *rag1* (red arrows) in approximately half of the homozygous mutant embryos at 4dpf. This effect is absent at 5dpf. (**i, j**) Maximum projections of *itga2b*:GFP transgenic embryos in the CHT at 5dpf in **i** wild-type and **j** *gata2a*$^{\Delta i4/\Delta i4}$ siblings. **k** HSPC (*itga2b*:GFP^{low}) counts in the CHT of wild-type (n =10) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (n =12) at 5dpf. No difference was detected between genotypes (μ_{wt} = 153.5; μ_{mut} = 145.5; p = 0.98, Mann-Whitney test). **l** Thrombocyte (*itga2b*:GFP^{high}) counts in the CHT of wild-type (n =10) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (n = 12) at 5dpf. No difference was detected between genotypes (μ_{wt} = 13; μ_{mut} = 13; p = 0.71, Mann-Whitney test). Error bars: median ± SD.

Notch recovers haematopoiesis in gata2a$^{\Delta i4/\Delta i4}$ *mutants.*

The recovery of *gata2b* expression by 30hpf (Fig. 3h, Supplementary Fig. 4e) coincides temporally with the observed decrease in *gata2a* in wild-type endothelial cells (Fig. 2h). Thus, we reasoned that other regulators of *gata2b* might compensate for the lack of endothelial *gata2a* in *gata2a*$^{\Delta i4/\Delta i4}$ mutants and thus lead to a recovery of the initial haematopoietic phenotype. Therefore, we investigated whether the loss of *gata2b* in *gata2a*$^{\Delta i4/\Delta i4}$ background resulted in a more severe haematopoietic phenotype than observed in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants. For this, we injected *gata2a*$^{\Delta i4/\Delta i4}$ and wild-type controls with a suboptimal amount (7.5 ng) of a gata2b morpholino oligonucleotide (MO)[21]. Quantitative ISH analysis confirmed that this amount of *gata2b* MO had no effect on *runx1* expression at 32hpf (Fig. 5a, b). As expected, *runx1* expression in *gata2a*$^{\Delta i4/\Delta i4}$ embryos was significantly reduced compared to wild-type siblings (Fig. 5a, b). *Gata2b* knockdown in *gata2a*$^{\Delta i4/\Delta i4}$ embryos further reduced *runx1* expression (Fig. 5a, b). To test whether this stronger reduction of *runx1* at 32hpf affected later stages of embryonic haematopoiesis, we assessed *cmyb* expression in the CHT at 4dpf(Fig. 5c). We scored *cmyb* expression levels as 'wild-type' or 'reduced' and found that the 'reduced' embryos were substantially overrepresented in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants injected with the *gata2b* MO, compared to wild-type fish and non-injected *gata2a*$^{\Delta i4/\Delta i4}$ siblings (Fig. 5c).

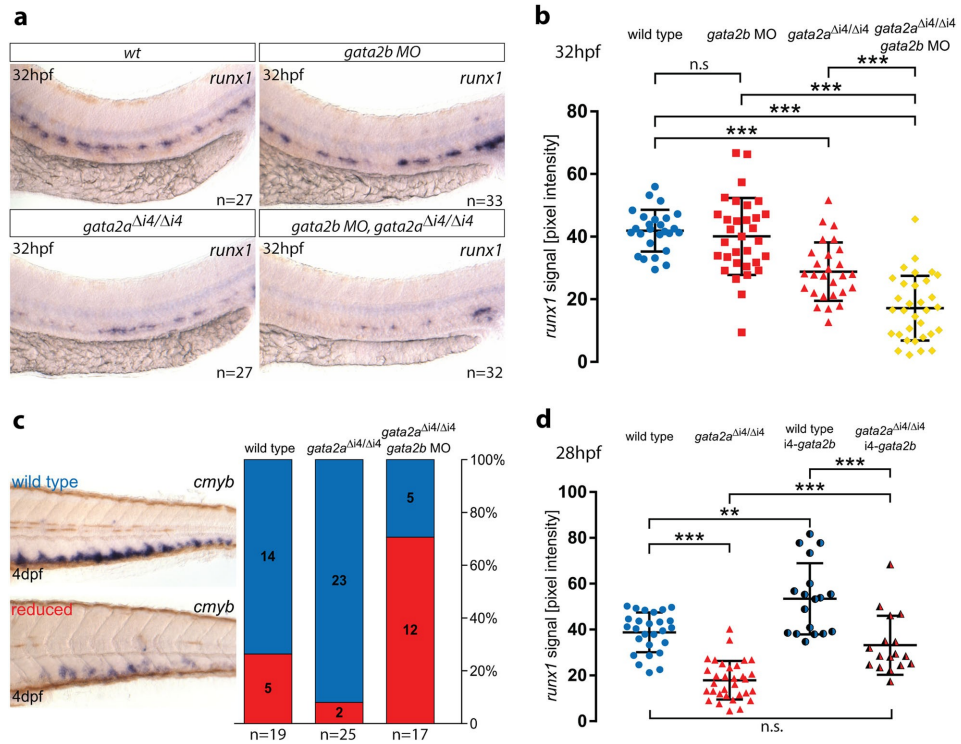
To verify whether Gata2b is required for definitive haematopoiesis downstream of Gata2a, we generated a frameshift truncating mutant for Gata2b and incrossed *gata2a*$^{\Delta i4/+}$*; gata2b*$^{+/-}$ adults to investigate *cmyb* expression at 33hpf in their progeny. *Gata2b*$^{-/-}$ mutants showed a more severe decrease in *cmyb* expression than *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Supplementary Fig. 4f–i). Double *gata2b*$^{-/-}$; *gata2a*$^{\Delta i4/\Delta i4}$ mutants showed no further reduction in *cmyb* expression compared to *gata2b*$^{-/-}$ mutants, suggesting that Gata2a was not sufficient to drive *cmyb* expression in HE in the absence of Gata2b (Supplementary Fig. 4f–i). Taken together, we conclude that Gata2b is regulated by Gata2a and is required for definitive haematopoiesis.

Next, we tested whether forced ectopic expression of *gata2b* was sufficient to speed up the haematopoietic recovery of *gata2a*$^{\Delta i4/\Delta i4}$ embryos. Thus, we overexpressed *gata2b* under the control of the *gata2a*-i4-450bp enhancer in wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ mutant embryos and measured *runx1* expression at 28hpf in the DA. *Gata2a*$^{\Delta i4/\Delta i4}$ embryos showed a significant decrease in *runx1* expression in comparison to wild-type (Fig. 5d). Ectopic expression of *gata2b* under the *gata2a*-i4 enhancer significantly increased *runx1* expression in wild-type and mutants(Fig. 5d). Importantly, it was sufficient to bring the *runx1* expression levels in the mutants up to the levels detected in uninjected wild-type embryos (Fig. 5d), demonstrating that *gata2b* alone was sufficient to drive *runx1* expression and drive the haematopoietic recovery in *gata2a*$^{\Delta i4/\Delta i4}$ mutants. Thus, *gata2b* can recover the definitive haematopoietic programme in the absence of endothelial *gata2a*.

Because the expression of *gata2b* is regulated by Notch signalling[21], we investigated whether inhibition of Notch would also prevent the haematopoietic recovery of *gata2a*$^{\Delta i4/\Delta i4}$ embryos. For this, we used the Notch inhibitor DAPM[39], and titrated it down to a suboptimal dose (25 µM) that did not significantly affect *runx1* expression at 30hpf in wild-type embryos (Supplementary Fig. 5a). This dose induced a small but measurable decrease in *gata2b* expression in DAPM-treated embryos while higher doses had a more robust effect (Supplementary Fig. 5b). Next, we treated wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ mutant embryos with DAPM and measured *runx1* expression in the DA at 36hpf (Fig. 5e, f). Suboptimal DAPM treatment did not affect *runx1* expression in wild-type embryos (Fig. 5e, f), but *gata2a*$^{\Delta i4/\Delta i4}$ mutants showed lower *runx1* levels and DAPM treatment further reduced

runx1 expression (Fig. 5e, f). Treatment with 25 µM DAPM did not significantly affect *gata2b* expression at 36hpf in either wild-type or *gata2a*$^{\Delta i4/\Delta i4}$ mutant embryos (Supplementary Fig. 5c). These experiments suggested that Notch signalling alone may be sufficient to rescue expressionof HE markers in *gata2a*$^{\Delta i4/\Delta i4}$ mutants. Indeed, ectopic activation of Notch signalling in endothelium using a *fli1a*-NICD:GFP construct[40] led to increased *runx1* and *gata2b* expression in wild-type embryos (Supplementary Fig. 5d, e). When overexpressed in *gata2a*$^{\Delta i4/\Delta i4}$ mutants, *fli1a*-NICD:GFP rescued *runx1* expression to near wild-type levels at 26hpf, whereas *gata2b* was increased beyond normal levels independently of the genotype (Supplementary Fig. 5f, g). Taken together,these experiments confirm that Notch activity regulates *runx1* and *gata2b* in HE and is sufficient to drive haematopoietic recovery in *gata2a*$^{\Delta i4/\Delta i4}$ mutants. Thus, we conclude that HE programming requires two independent inputs on *runx1* and *gata2b* expression; one from Gata2a, driven in ECs by the i4 enhancer, and the other from Notch signalling, necessary and sufficient to drive HE programming even in the absence of *gata2a*.

Fig. 5 *Gata2b* and Notch signalling are sufficient to recover haematopoietic markers in *gata2a*$^{\Delta i4/\Delta i4}$ mutants.
a Expression of *runx1* in HE at 32hpf in wild-type (wt), *gata2b* MO-injected (7.5 ng) wt embryos, *gata2a*$^{\Delta i4/\Delta i4}$ mutants and *gata2b* MO-injected (7.5 ng) *gata2a*$^{\Delta i4/\Delta i4}$ mutants. **b** Quantification of the *runx1* in situ hybridization (ISH) signal in wt, *gata2b* morphants, *gata2a*$^{\Delta i4/\Delta i4}$ mutants and *gata2a*$^{\Delta i4/\Delta i4}$ mutants injected with *gata2b* MO. *runx1* expression is decreased in *gata2a*$^{\Delta i4/\Delta i4}$ mutants (μ_{wt} = 41.9, μ_{mut} = 28.9; F =44.641, d.f. = 3, 62.3; p < 0.001). *Gata2b* MO knockdown significantly decreases *runx1* in the DA of *gata2a*$^{\Delta i4/\Delta i4}$ mutants (μ_{mut} =28.9, μ_{mut+MO} = 17.2), but not wt embryos at 32hpf (μ_{wt} = 41.9, μ_{MO} = 40.1; p = 0.89, p = 0.89, Games-Howell post-hoc test, Welch's ANOVA). n = 27, wt; n = 27, *gata2a*$^{\Delta i4/\Delta i4}$; n = 33, wt + *gata2b* MO; n = 32, *gata2a*$^{\Delta i4/\Delta i4}$ + *gata2b* MO. **c** Scoring *cmyb* expression at 4dpf in wt, *gata2a*$^{\Delta i4/\Delta i4}$ mutants and *gata2a*$^{\Delta i4/\Delta i4}$ mutant embryos injected with *gata2b* MO as wt (blue) or reduced (red). *Gata2b* MO knockdown (7.5 ng) inhibits the haematopoietic recovery of *gata2a*$^{\Delta i4/\Delta i4}$ mutants. (X2 = 18.784, d.f. = 2, p < 0.001). **d** Quantification of the *runx1* ISH signal, from 28hpf wt embryos (blue), *gata2a*$^{\Delta i4/\Delta i4}$ mutants (red) and their siblings injected with a *gata2a*-i4-450bp:*gata2b* construct (shaded blue and red). Ectopic expression of *gata2b* increases *runx1* expression in the HE of wt embryos (μ_{wt} = 38.8, $\mu_{wt+gata2b}$ = 53.4; p < 0.01) and rescues *runx1* expression in the DA of *gata2a*$^{\Delta i4/\Delta i4}$ mutants to wt levels (μ_{mut} = 17.9, $\mu_{mut+gata2b}$ = 33.2; p < 0.001; μ_{wt} =38.8, $\mu_{mut+gata2b}$ = 33.2; p = 0.31, Tukey HSD post-hoc test). n = 25, wt; n = 33, *gata2a*$^{\Delta i4/\Delta i4}$;n = 18, wt+ *gata2a*-i4-450bp:*gata2b* construct; n = 17, *gata2a*$^{\Delta i4/\Delta i4}$ + *gata2a*-i4-450bp:*gata2b* construct. **e** Quantification of the *runx1* ISH signal at 36hpf in embryos treated with a suboptimal dose (25 µM) of the Notch inhibitor DAPM. 25 µM DAPM showed no effect on *runx1* expression in wt compared to DMSO-treated embryos (μ_{DMSO} = 40.5, μ_{DAPM} = 38; p =0.735, Tukey HSD post-hoc test). DMSO-treated *gata2a*$^{\Delta i4/\Delta i4}$ mutants show a decrease in *runx1* expression (μ_{DMSO} = 40.5, $\mu_{mut+DMSO}$ = 31.5; F =25.774, d.f. = 3, 91; ANOVA). DAPM treatment significantly reduced *runx1* expression in the DA *gata2a*$^{\Delta i4/\Delta i4}$ mutants ($\mu_{mut+DMSO}$ = 31.5, $\mu_{mut+DAPM}$ = 19.4). n = 27, wt +DMSO; n = 20, *gata2a*$^{\Delta i4/\Delta i4}$ + DMSO; n = 30, wt + DAPM; n = 20, *gata2a*$^{\Delta i4/\Delta i4}$ + DAPM. **f** Representative images of the average *runx1* expression at 36hpf in wt and *gata2a*$^{\Delta i4/\Delta i4}$ mutants treated with 25 µM DAPM. Error bars: mean ± SD. **p < 0.01; ***p < 0.001. See also Supplementary Fig. 5.



Impaired haematopoiesis in adult gata2a$^{\Delta i4/\Delta i4}$ mutants.

To investigate whether Gata2a plays a role in adult haematopoiesis, we first asked whether the *gata2a*-i4-1.1 kb:GFP reporter was active in haematopoietic cells in the adult. Whole kidney marrow (WKM) cells isolated from the transgenic fish showed that the i4 enhancer is active in haematopoietic cells previously defined by flow cytometry[41] as progenitors, lymphoid + HSPC (containing the HSPCs) and myeloid cells (Supplementary Fig. 6a–c). Accordingly, single cell transcriptional profiling showed higher levels of *gata2a* in HSPCs, progenitors, neutrophils and thrombocytes (Supplementary Fig. 6d–f)[42,43]. Consistent with this notion, we observed a high incidence of infections and heart oedemas in *gata2a*$^{\Delta i4/\Delta i4}$ adult fish, with over 25% suffering from one of these defects by 6 months of age, compared to <1% of wild-type fish (Fig. 6a–c). The heart oedemas and the infections are suggestive of lymphatic defects and immune deficiency as observed in human patients bearing genetic GATA2 haploinsufficiency syndromes such as MonoMAC syndrome[13]. Notably, around 10% of MonoMAC syndrome patients show mutations in the homologous enhancer region of *GATA2*[12,14].

Next, we counted the total number of haematopoietic cells in wild-type and *gata2a*$^{\Delta i4/\Delta i4}$ mutant WKM (Fig. 6d–f). To avoid any confounding effects in our analysis, we compared wild-type to *gata2a*$^{\Delta i4/\Delta i4}$ mutants without overt signs of infection. The *gata2a*$^{\Delta i4/\Delta i4}$ mutants showed a ~2-fold decrease in the total number of haematopoietic cells in the WKM (Fig. 6d–f). In addition, neutrophils were similarly reduced (Fig. 6g), another characteristic in common with MonoMAC syndrome patients[14]. Lastly, kidney marrow smears of ten 9-month old *gata2a*$^{\Delta i4/\Delta i4}$ mutants were assessed. One of the ten mutants showed an excess of immature myeloid blast cells in the WKM (>98%) and only minor erythrocyte differentiation (Fig. 6h, i). The presence of excess blasts is usually an indication of AML in humans. Together these data strongly suggest that the i4 enhancer is a critical driver of *gata2a* expression in adult haematopoietic cells. The enhancer deletion in *gata2a*$^{\Delta i4/\Delta i4}$ mutants leads to a hypocellular WKM and neutropenia, strongly suggestive of marrow failure, a hallmark of disease progression in Gata2 deficiency syndromes.

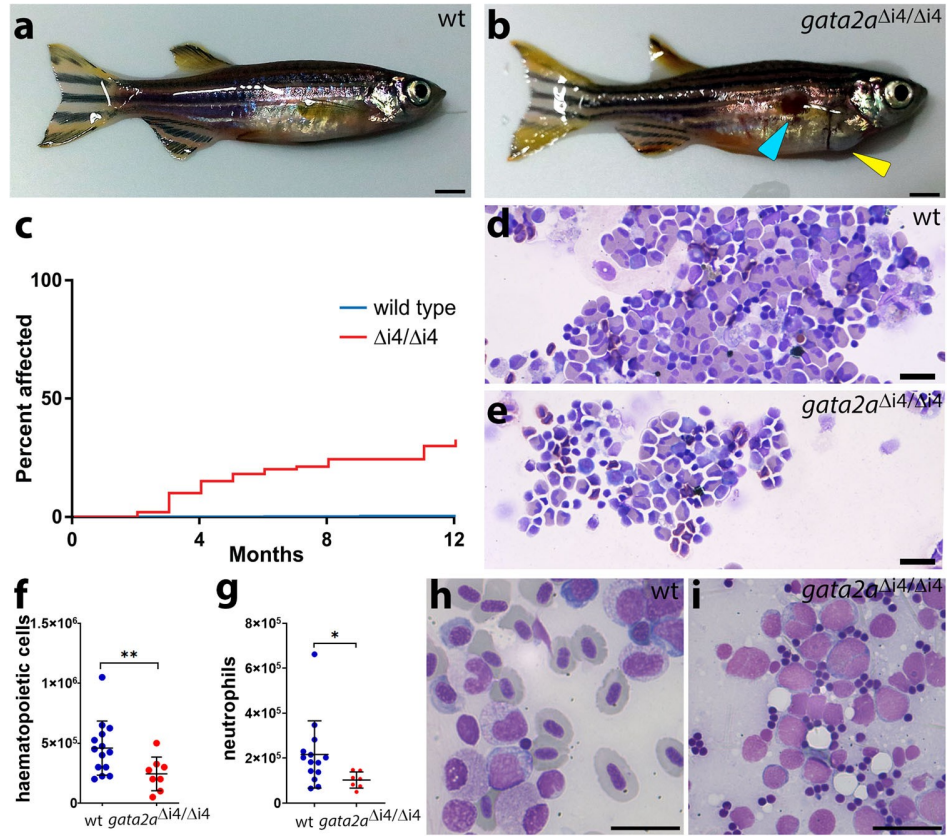


Fig. 6 *gata2a*$^{\Delta i4/\Delta i4}$ mutants show cardiac oedema, hypocellularity and marrow failure.
a, b General morphology of zebrafish adults: **a** wild-type; **b** *gata2a*$^{\Delta i4/\Delta i4}$ mutant showing skin infection (blue arrowhead) and pericardial oedema (yellow arrowhead). **c** Over 25% (n = 29/108) of *gata2a*$^{\Delta i4/\Delta i4}$ mutants (red) catch infections or suffer from heart oedemas by 6 months. Only around 65% (n =69/108) survive for more than 12 months without overt signs of infections. Fewer than 1% (n = 2/500) of wild-type fish (blue) exhibit such defects. The graph does not include deaths by other causes. **d, e** May-Grunwald/Wright-Giemsa staining in cytospins of haematopoietic cells isolated from the WKM of zebrafish adults: **d** wild-type; **e** *gata2a*$^{\Delta i4/\Delta i4}$ mutant. Note the decrease in cell numbers. **f** Cell counts of haematopoietic cells isolated from WKM of wild-type (n= 14) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (n = 8). The *gata2a*$^{\Delta i4/\Delta i4}$ mutants show a ~2-fold decrease in haematopoietic cell numbers in the WKM (μ_{wt} = 4.37 × 105; μ_{mut} =2.37 × 105, p = 0.0185, Mann-Whitney test). (**g**) Number of neutrophils isolated from WKM of wild-type (n = 14) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (n =7). The *gata2a*$^{\Delta i4/\Delta i4}$ mutants show a ~2-fold decrease in neutrophil numbers in the WKM (μ_{wt} = 2.17 × 105; μ_{mut} =1.03 × 105, p = 0.0269, Mann-Whitney test). Error bars: median cell number ± SD. **h, i** Kidney smears from 9 months post-fertilization adult animals were assessed. **h** Wild-type shows various stages of lineage differentiation. **i** WKM smear; 1 of 10 *gata2a*$^{\Delta i4/\Delta i4}$ mutants showed the presence of excess blasts with very little erythroid differentiation (98% blasts, >200 cells assessed). Scalebars: 2mm (**a, b**) and 10 µm (**d, e, h, i**). See also Supplementary Fig.6.

DISCUSSION

The sub-functionalisation of the Gata2 paralogues in zebrafish provided an opportunity to unpick the different roles of *Gata2* in the multi-step process of definitive haematopoiesis. Here we have investigated the conservation of the Gata2 +9.5 enhancer and identified a homologous region in intron 4 of the zebrafish *gata2a* locus (*gata2a*-i4) that is not present in the *gata2b* locus. The zebrafish *gata2a*-i4 enhancer, like the mouse enhancer[18], is sufficient to drive pan-endothelial expression of GFP and necessary for endothelial expression of *gata2a* (Figs. 1 and 2). We traced the activity of the i4 enhancer back to the PLM, the source of precursors of endothelium and HSCs[10]. This degree of sequence and functional conservation of the i4 enhancer led us to hypothesize that Gata2a might play a role in definitive haematopoiesis. Indeed, homozygous deletion of the i4 enhancer (*gata2a*$^{\Delta i4/\Delta i4}$) allowed us to uncover a previously unknown function of Gata2a in regulating the initial expression of *runx1* and *gata2b* in HE. Although *cmyb* expression in HE was decreased in *gata2a*$^{\Delta i4/\Delta i4}$ mutants, it was more severely reduced in *gata2b*$^{-/-}$ mutants, suggesting that Gata2b is more important for *cmyb* regulation than Gata2a. However, both Gata2 orthologues regulate gene expression in the HE before the first reported EHT events at 34hpf[6].

Gata2 binds to the +9.5 enhancer to maintain its own expression in endothelial and haematopoietic cells[26,44]. In zebrafish, it is likely that Gata2a binds the GATA motifs in the i4 enhancer and loss of *gata2a* in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants (Fig. 2) seems to support this view. Interestingly, we detected a small region in intron 4 of the *gata2b* locus that was not identified as a peak in our ATACseq experiment but is conserved in some fish species (Supplementary Fig. 1a) and thuscould potentially represent a divergent gata2b intronic enhancer. We speculate that the positive autoregulation of Gata2 was likelyretained by both *gata2* orthologues in zebrafish, but this possibility remains to be investigated.

The *gata2a*$^{\Delta i4/\Delta i4}$ mutants recovered from the early defects in HE programming and displayed normal expression levels of cmyb in the CHT at 4dpf and rag1 in the thymus at 5dpf, used as indicators of the definitive haematopoietic programme[11]. We hypothesized that this could be due to the presence of the two homologues of Gata2 in zebrafish[20], despite Gata2a and Gata2b proteins being only 50% identical[21]. Indeed, forced expression of *gata2b* under the gata2a-i4 enhancer rescued DA expression of *runx1* in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants to wild-type levels and suboptimal depletion of *gata2b* in the *gata2a*$^{\Delta i4/\Delta i4}$ mutants resulted in more severe reduction in cmyb expression in the CHT by 4dpf (Fig. 4). In addition, we demonstrated that Notch signalling, a known regulator of *gata2b* expression[21], is sufficient to rescue the initial HE programming defect induced by deletion of the gata2a i4 enhancer. We propose a model in which *gata2a* acts upstream of *runx1* and *gata2b* independently of Notch to initiate HE programming. The regulation of *gata2b* by Gata2a is transient, and the timing largely coincides with the natural decrease in endothelial expression of *gata2a* by 30hpf. After this stage, endothelial Notch signalling takes over the regulation of *runx1* and *gata2b* expression, acting as a fail-safe mechanism that buffers against fluctuations in the system caused by loss of one or more of the initial inputs (in this case, Gata2a). Despite the apparent haematopoietic recovery, we observed a high incidence of infections and oedema in *gata2a*$^{\Delta i4/\Delta i4}$ adults, and a striking decrease in the number of haematopoietic cells in the WKM. The decrease in haematopoietic cells in particular is reminiscent of the loss of proliferative potential of haematopoietic *Gata2*$^{+/-}$ heterozygous cells in the mouse[17,45]. This raises the possibility that in zebrafish the *gata2a* and *gata2b* paralogues may function as two *Gata2* 'alleles' that together regulate the haematopoietic output of the WKM. This will be addressed by comparing the adult phenotypes of *gata2a*$^{\Delta i4/\Delta i4}$ and *gata2b*$^{-/-}$ mutants. Taken together, our initial characterization of WKM shows that *gata2a*$^{\Delta i4/\Delta i4}$ mutants present a phenotype consistent with Gata2 deficiency syndromes in humans brought about by GATA2 haploinsufficiency[12,26]. Strikingly, about 10% of all MonoMAC patients show mutations in the conserved +9.5 enhancer[12,14], the corresponding regulatory element to the i4 enhancer. The i4 enhancer is active in the lymphoid + HSPC fraction that contains the HSC activity[46], in the progenitor cells and in the myeloid fraction that contains eosinophils, previously identified as expressing high levels of a gata2a-GFP BAC transgenic reporter[41]. Thus, it is likely that *gata2a*$^{\Delta i4/\Delta i4}$ adult fish show lineage specific differentiation defects. Further characterization of the *gata2a*$^{\Delta i4/\Delta i4}$ mutants will uncover which haematopoietic cells are most affected by the loss of i4 enhancer activity and how Gata2a regulates haematopoietic output, thus establishing a zebrafish animal model for human diseases linked to Gata2 haploinsufficiency.

Data availability

All data generated or analyzed during this study (images, quantitation data in the form of graphs and ATACseq data) are included in this published article and its supplementary information files. The source data are available in Supplementary Data 1 and the list of the called ATACseq peaks is available in Supplementary Data 2. The ATACseq data was deposited in GEO (Accession number GSE143763).

Acknowledgements

We thank the staff of the Biomedical Services Units (Oxford, Birmingham and Rotterdam) for fish husbandry. We thank Kevin Clark and Sally-Ann Clark from the WIMM flow cytometry facility for cell sorting. The flow cytometry facility is supported by the MRC HIU, MRC MHU (MC_UU_12009), NIHR Oxford BRC and John Fell Fund (131/030 and 101/517), the EPA fund (CF182 and CF170). We thank the Wolfson Imaging Centre Oxford for imaging. The Centre is supported by a Wolfson Foundation (grant 18272), and an MRC/BBSRC/EPSRC grant (MR/K015777X/1) to MICA – Nanoscopy Oxford. Both facilities were supported by

the WIMM Strategic Alliance awards G0902418 and MC_UU_12025. We thank Fatma Kok and Douglas Vernimmen for critical reading of the manuscript. We thank Feng Liu for the generous gift of the fli1a-NICD:GFP construct. This research was supported by the British Heart Foundation (BHF Oxford CoRE and BHF IBSR Fellowship FS/13/50/30436 to R.M. and M.K.), by a Wellcome Trust Chromosome and Developmental Biology PhD Scholarship (#WT102345/Z/13/Z. to T.D.) and by the MRC MHU programme (MC_UU_12009/8 to R.P.).

Author contributions

T.D., M.K., C.K., E.P. and R.M. designed the study. T.D., M.K., C.K., J.P-Z., K.G., C.B.M. and R.M. performed experiments and analyzed the data. J.P-Z., B.F. and K.G. performed experiments. R.R. performed the bioinformatics analyses, T.D. and R.M. wrote the paper and R.P., E.P. and R.M. edited the paper. R.P., E.P. and R.M. secured funding.

Competing interests

The authors declare no competing interests.

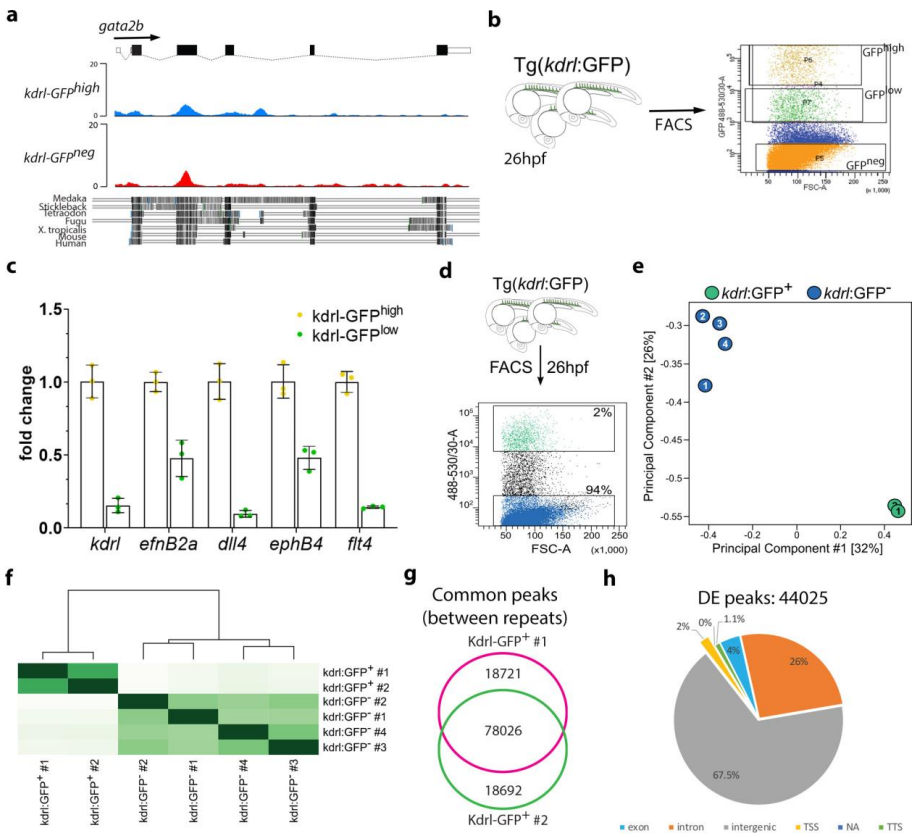
REFERENCES

1. Ciau-Uitz, A., Monteiro, R., Kirmizitas, A. & Patient, R. Developmental hematopoiesis: ontogeny, genetic programming and conservation. Exp. Hematol. 42, 669–683 (2014).
2. Bonkhofer, F. et al. Blood stem cell-forming haemogenic endothelium in zebrafish derives from arterial endothelium. Nat. Commun. 10, 3577 (2019).
3. Gritz, E. & Hirschi, K. K. Specification and function of hemogenic endothelium during embryogenesis. Cell Mol. Life Sci. 73, 1547–1567 (2016).
4. Kalev-Zylinska, M. L. et al. Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. Development 129, 2015–2030 (2002).
5. Swiers, G. et al. Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. Nat. Commun. 4, 2924 (2013).
6. Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. Nature 464, 112–115 (2010).
7. Bertrand, J. Y. et al. Haematopoietic stem cells derive directly from aortic endothelium during development. Nature 464, 108–111 (2010).
8. Boisset, J. C. et al. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. Nature 464, 116–120 (2010).
9. Kissa, K. et al. Live imaging of emerging hematopoietic stem cells and early thymus colonization. Blood 111, 1147–1156 (2008).
10. Davidson, A. J. & Zon, L. I. The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis. Oncogene 23, 7233–7246 (2004).
11. Murayama, E. et al. Tracing hematopoietic precursor migration to successive hematopoietic organs during zebrafish development. Immunity 25, 963–975 (2006).
12. Hsu, A. P. et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. Blood 121, 3830–3837 (2013). S3831-3837.
13. Wlodarski, M. W., Collin, M. & Horwitz, M. S. GATA2 deficiency and related myeloid neoplasms. Semin. Hematol. 54, 81–86 (2017).
14. Wlodarski, M. W. et al. Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. Blood 127, 1387–1397 (2016). quiz 1518.
15. Tsai, F. Y. et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature 371, 221–226 (1994).
16. de Pater, E. et al. Gata2 is required for HSC generation and survival. J. Exp. Med. 210, 2843–2850 (2013).
17. Ling, K. W. et al. GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. J. Exp. Med. 200, 871–882 (2004).
18. Khandekar, M. et al. A Gata2 intronic enhancer confers its pan-endothelia- specific regulation. Development 134, 1703–1712 (2007).
19. Gao, X. et al. Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. J. Exp. Med. 210, 2833–2842 (2013).
20. Gillis, W. Q., St John, J., Bowerman, B. & Schneider, S. Q. Whole genome duplications and expansion of the vertebrate GATA transcription factor gene family. BMC Evol. Biol. 9, 207 (2009).
21. Butko, E. et al. Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. Development 142, 1050–1061 (2015).

22. Liu, J., Jiang, J., Wang, Z., He, Y. & Zhang, Q. Origin and evolution of GATA2a and GATA2b in teleosts: insights from tongue sole, Cynoglossus semilaevis. PeerJ 4, e1790 (2016).
23. Yang, L., Rastegar, S. & Strahle, U. Regulatory interactions specifying Kolmer- Agduhr interneurons. Development 137, 2713–2722 (2010).
24. Zhu, C. et al. Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. Development 138, 4555–4564 (2011).
25. Patterson, L. J. et al. The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish. Blood 109, 2389–2398 (2007).
26. Johnson, K. D. et al. Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. J. Clin. Invest. 122, 3692–3704 (2012).
27. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218 (2013).
28. Jin, S. W., Beis, D., Mitchell, T., Chen, J. N. & Stainier, D. Y. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. Development 132, 5199–5209 (2005).
29. Meadows, S. M., Myers, C. T. & Krieg, P. A. Regulation of endothelial cell development by ETS transcription factors. Semin Cell Dev. Biol. 22, 976–984 (2011).
30. Wozniak, R. J. et al. Molecular hallmarks of endogenous chromatin complexes containing master regulators of hematopoiesis. Mol. Cell Biol. 28, 6681–6694 (2008).
31. Kawakami, K. et al. A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. Dev. Cell 7, 133–144 (2004).
32. Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J. L. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep. 4, 220–228 (2013).
33. Lawson, N. D. et al. Notch signaling is required for arterial-venous differentiation during embryonic vascular development. Development 128, 3675–3683 (2001).
34. Dobrzycki, T., Krecsmarik, M., Bonkhofer, F., Patient, R. & Monteiro, R. An optimised pipeline for parallel image-based quantification of gene expression and genotyping after in situ hybridisation. Biol. Open 7, bio031096 (2018).
35. Wilkinson, R. N. et al. Hedgehog and Bmp polarize hematopoietic stem cell emergence in the zebrafish dorsal aorta. Dev. Cell 16, 909–916 (2009).
36. Bertrand, J. Y. et al. Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. Development 134, 4147–4156 (2007).
37. Willett, C. E., Zapata, A. G., Hopkins, N. & Steiner, L. A. Expression of zebrafish rag genes during early development identifies the thymus. Dev. Biol. 182, 331–341 (1997).
38. Lin, H.-F. F. et al. Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. Blood 106, 3803–3810 (2005).
39. Walsh, D. M. et al. Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo. Nature 416, 535–539 (2002).
40. Liu, Z. et al. Primary cilia regulate hematopoietic stem and progenitor cell specification through Notch signaling in zebrafish. Nat. Commun. 10, 1839 (2019).
41. Traver, D. et al. Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. Nat. Immunol. 4, 1238–1246 (2003).
42. Athanasiadis, E. I. et al. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. Nat. Commun. 8, 2045 (2017).

43. Macaulay, I. C. et al. Single-cell RNA-sequencing reveals a continuous spectrum of differentiation in Hematopoietic Cells. Cell Rep. 14, 966–977 (2016).
44. Sanalkumar, R. et al. Mechanism governing a stem cell-generating cis- regulatory element. Proc. Natl Acad. Sci. USA 111, E1091–1100 (2014).
45. Rodrigues, N. P. et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. Blood 106, 477–484 (2005).
46. Ma, D., Zhang, J., Lin, H.-f, Italiano, J. & Handin, R. I. The identification and characterization of zebrafish hematopoietic stem cells. Blood 118, 289–297 (2011).
47. Westerfield, M. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 5th edn, (Univ of Oregon Press, 2007).
48. Chi, N. C. et al. Foxn4 directly regulates tbx2b expression and atrioventricular canal formation. Genes Dev. 22, 734–739 (2008).
49. Howe, K. et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496, 498–503 (2013).
50. Kent, W. J. et al. The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002).
51. Birnbaum, R. Y. et al. Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. 22, 1059–1068 (2012).
52. Ran, F. A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380–1389 (2013).
53. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–838 (2013).
54. Gagnon, J. A. et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PLoS ONE 9, e98186 (2014).
55. Xu, H. et al. Genome-wide identification of suitable zebrafish Danio rerio reference genes for normalization of gene expression data by RT-qPCR. J. Fish. Biol. 88, 2095–2110 (2016).
56. Hu, Y., Xie, S. & Yao, J. Identification of novel reference genes suitable for qRT-PCR normalization with respect to the Zebrafish developmental stage. PLoS ONE 11, e0149277 (2016).
57. Stachura, D. L. & Traver, D. Cellular dissection of zebrafish hematopoiesis. Methods Cell Biol. 133, 11–53 (2016).
58. Jowett, T. & Yan, Y. L. Double fluorescent in situ hybridization to zebrafish embryos. Trends Genet 12, 387–389 (1996).
59. Thompson, M. A. et al. The cloche and spadetail genes differentially affect hematopoiesis and vasculogenesis. Dev. Biol. 197, 248–269 (1998).
60. Monteiro, R. et al. Transforming growth factor beta drives hemogenic endothelium programming and the transition to hematopoietic stem cells. Dev. Cell 38, 358–370 (2016).

SUPPLEMENTARY FIGURES



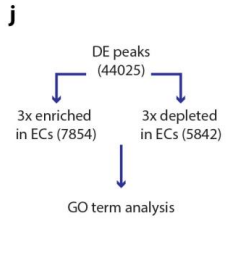
i

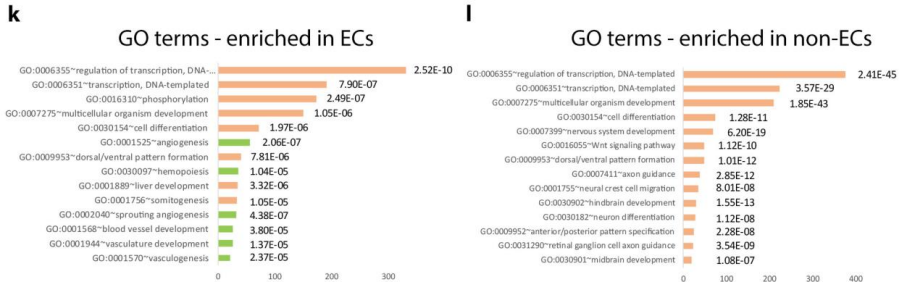
Unique EC peaks=5883

Rank	Motif name	P-value	% of Targets Sequences	Fold enrichment
1	ERG(ETS)	1e-1006	49.23%	3.89
2	ETS1(ETS)	1e-779	37.57%	4.21
3	ETV1(ETS)	1e-777	42.12%	3.69
4	EWS:ERG-fusion(ETS)	1e-676	30.17%	4.73
5	Fli1(ETS)	1e-673	38.60%	3.58
6	Ets1-distal(ETS)	1e-580	20.53%	6.46
7	GABPA(ETS)	1e-579	30.75%	3.98
8	EWS:FLI1-fusion(ETS)	1e-563	24.87%	4.91
9	EHF(ETS)	1e-536	35.49%	3.24
10	SPDEF(ETS)	1e-357	27.60%	3.00

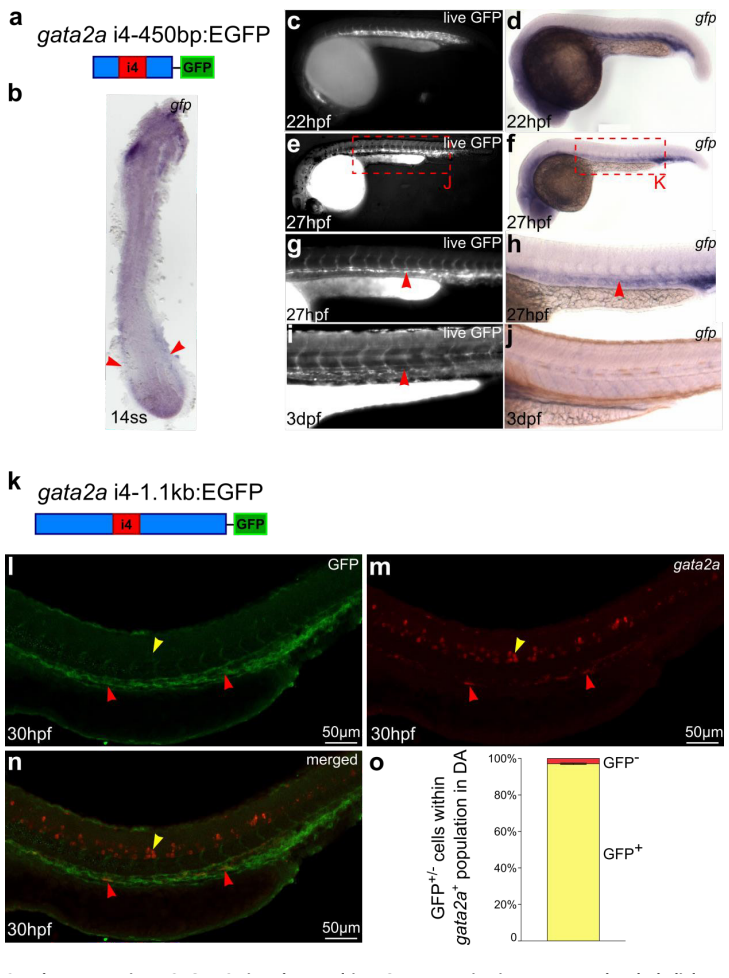
Common EC and non-ECs peaks =1320

Rank	Motif name	P-value	% of Target Sequences	Fold enrichment
1	TR4(NR),DR1	1.00E-14	4.09%	3.62
2	Sox10(HMG)	1.00E-11	21.44%	1.50
3	PPARE(NR),DR1	1.00E-10	11.06%	1.78
4	RXR(NR),DR1	1.00E-10	12.05%	1.72
5	GAGA-repeat	1.00E-10	52.42%	1.21
6	ERG(ETS)	1.00E-10	16.52%	1.55
7	GAGA-repeat	1.00E-10	27.95%	1.36
8	Sox3(HMG)	1.00E-09	21.89%	1.43
9	Nur77(NR)	1.00E-09	3.71%	2.75
10	Erra(NR)	1.00E-08	21.59%	1.40

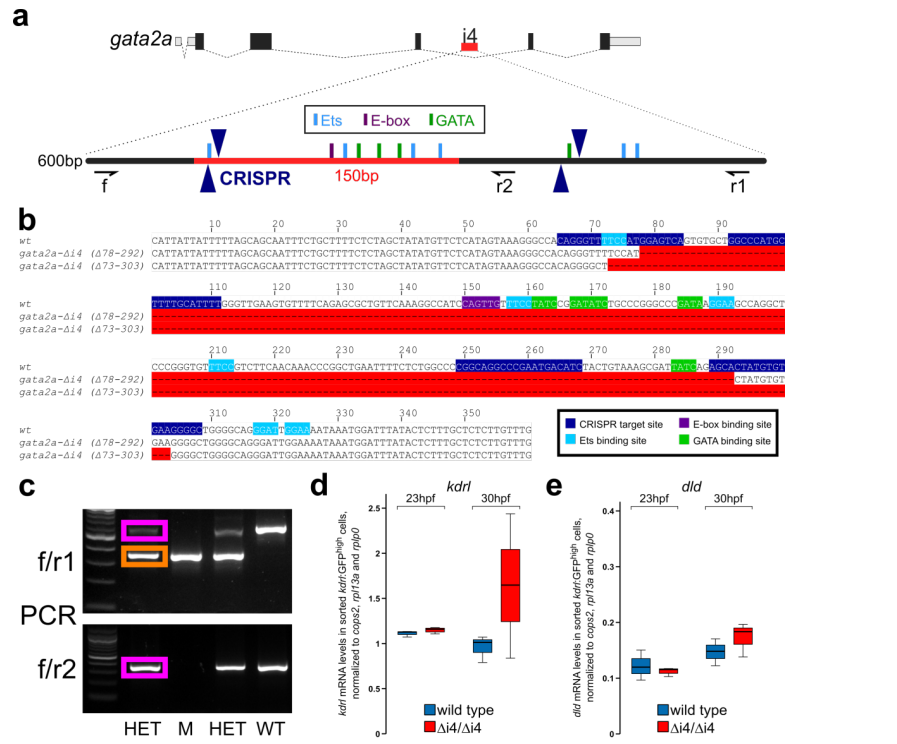


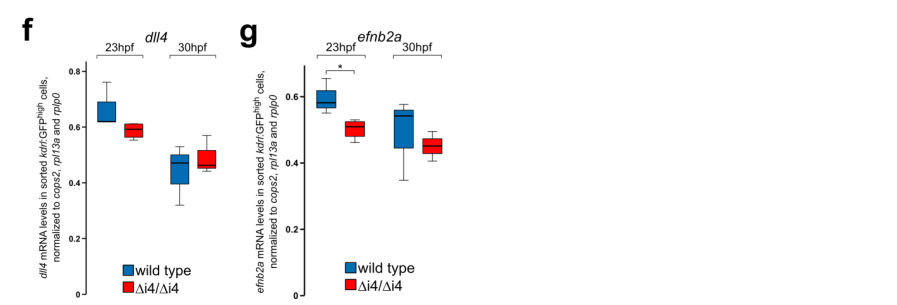


Supplementary Figure 1. Analysis of open chromatin in endothelial cells by ATAC-seq. (a) Overviewof the zebrafish *gata2b* locus and genome alignment with other species, including mouse and human. Note the conservation within exon sequences in all the species used for comparison. Some intronic sequences are conserved only within fish species. (b) Schematic representation of the gating strategy to test expression of endothelial genes in two distinct *kdrl*:GFP^{+} cell populations: *kdrl*:GFP^{high}(orange, top panel) and Kdrl:GFP^{low} (green). (c) qPCR in kdrl:GFP^{high} and Kdrl:GFP^{low} cell populations, showing significant enrichment for endothelial-specific markers in the kdrl:GFP^{high} population. This population was selected for the ATACseq experiment and will be further referred to as *kdrl*:GFP^{+} for simplicity. (d) *Kdrl*:GFP^{+} (green) and *kdrl*:GFP^{-} (blue) cells were FACS-sorted from 26hpf embryos and used for preparation of ATAC-seq libraries. (e) Principal Component Analysis of 2 *kdrl*:GFP^{+} replicas and 4 *kdrl*:GFP^{-} replicas, showing strong separation of the two cell populations. The same numbers inside the circles denote replicas coming from the same FACS. (f) Clustering of the ATAC-seq samplesshowing high correlation between *kdrl*:GFP^{+} replicates and among *kdrl*:GFP^{-} replicas. The same numbers represent replicas coming from the same FACS. (g) Venn diagram showing 78026 common peaks between the *kdrl*:GFP^{+} ATACseq profile replicates. (h) Genome-wide distribution of the ATAC peaks from the differential peaks analysis. Note that most peaks are intergenic and intragenic; TSS peaks are only 2% of the total (44026 total peaks). (i) Motif enrichment analysis in endothelial cells (EC-enriched) and common between ECs and non-ECs showing a clear enrichment for ETS binding sites in ECs. (j) Scheme depicting the strategy for the Gene Ontology (GO) term analysis using DAVID[1]. GO term analysis was performed for genes associated to ATACseq peaks that were (k) >3-foldenriched in *kdrl*:GFP^{+} endothelial cells (ECs) or (l) >3-fold depleted in *kdrl*:GFP^{+} cells.

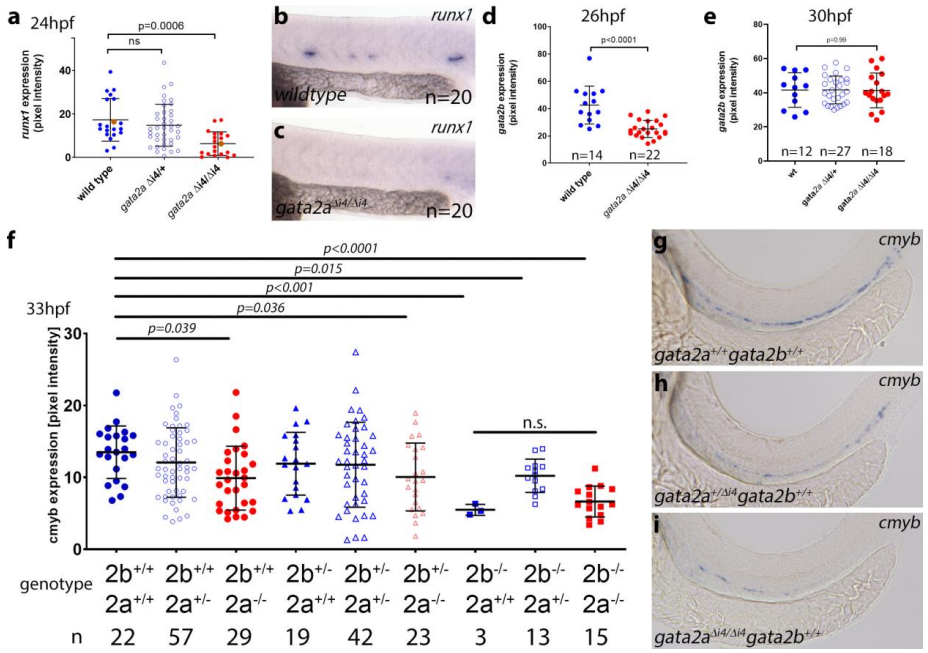


Supplementary Figure 2. *Gata2a*-i4 enhancer drives GFP expression in PLM, DA and endothelial cells. (a) Schematic representation of the construct used to generate the Tg(*gata2a*-i4-450bp:GFP) transgenic line. (b) *In situ* hybridization of a 14-somite Tg(*gata2a*-i4-450bp:GFP) embryo (flat mount)shows *gfp* mRNA expression in the posterior lateral plate mesoderm (arrowheads), visible over high background staining in the yolk. (c,d) Fluorescent image of a live Tg(*gata2a*-i4-450bp:GFP) embryo and *in situ* hybridization image of the same embryo probed for *gfp* mRNA at 22hpf reveal activity of the i4 enhancer in the forming dorsal aorta. GFP is also present in the heart. (e-h) At 27hpf, GFP protein was detected in the endothelial cells and is particularly strong in the dorsal aorta (arrowhead in panel g). *In situ* hybridization reveals high degree of overlap between *gfp* RNA and protein expression. (g-h) magnified trunks of embryos in panels e and f as indicated. (i-j) By 3dpf, GFP protein persists in the dorsal aorta (arrowhead) and other endothelial cells, but *gfp* mRNA is undetectable by *in situ* hybridization. (k) Schematic representation of the construct used to generate the Tg(*gata2a*-i4-1.1kb:GFP) transgenic line. (l-n) Confocal images of the trunk of a Tg(*gata2a*-i4- 1.1kb:GFP) embryo immunostained with anti-GFP antibody (l) and probed for *gata2a* mRNA (m) at 30hpf, showing an overlap of GFP and *gata2a* in the DA (red arrowheads) but lack of GFP expression in the *gata2a*$^{+}$ neural tube (yellow arrowhead). (n) Merged images from panels l-m. (o) Counting of the *gata2a*$^{+}$ cells represented in panels l-n in 9 embryos shows that >95% of *gata2a*$^{+}$ cells in the DA are also GFP^{+}. N=2. Error bars: ±SD.

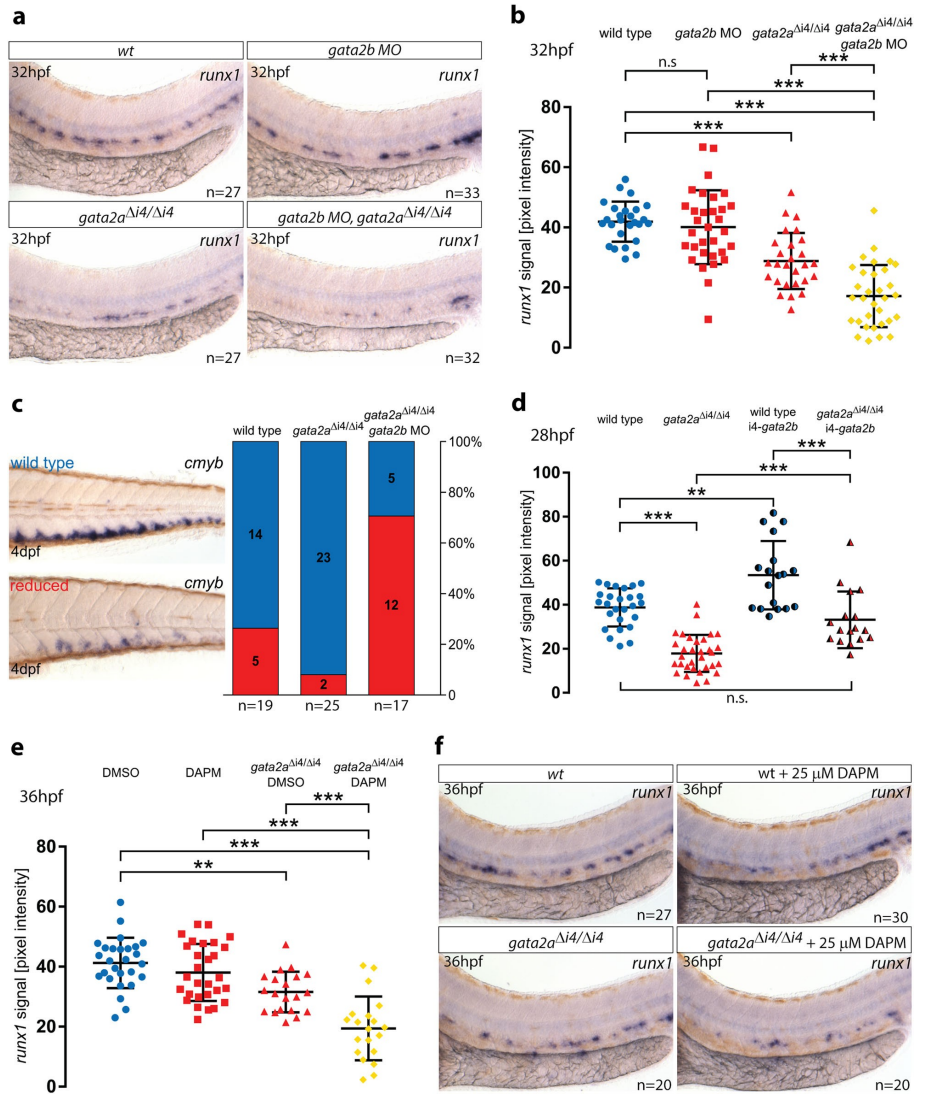




Supplementary Figure 3. Generation of *gata2a* $^{\Delta i4}$ zebrafish mutants with CRISPR/Cas9 and expression of endothelial markers in *gata2a*$^{\Delta i4/\Delta i4}$ mutants. (a) Two pairs of CRISPR sgRNAs (dark blue) were designed to flank the highly conserved 150bp region within the *gata2a*-i4 enhancer. Approximate binding positions of diagnostic PCR primers f, r1 and r2 are indicated. Light blue: Ets binding sites; purple: E-box binding sites; green: GATA binding sites (mapped computationally). (b) Two isolated mutant alleles (Δ78-292 and Δ73-303) carry deletions (red gaps) including the majority of highly conserved transcription factor binding sites. (c) Genotyping of embryos from the incross of two *gata2a*$^{+/\Delta i4}$ mutants with PCR relies on the identification of bands specific to wild type alleles (pink) and the mutant-specific band (orange). Primers f, r1 and r2 are the same as in panel a. The first lane from the left in each gel: 100bp marker. WT: wild type; HET: heterozygote; M: mutant. (d-f) qRT-PCR in sorted *kdrl*:GFP^{+} cells from wt (blue) and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (red) shows no differences in the levels of (d) *kdrl* (23hpf: t=1.576, d.f.=5, p>0.1; 30hpf: t=1.399, d.f.=4, p>0.2), (e) *dld* (23hpf: *t*=0.585, d.f.=5, *p*>0.5; 30hpf: *t*=1.097, d.f.=4, *p*>0.3) and (f) *dll4* (23hpf: *t*=1.892, d.f.=5, *p*>0.1; 30hpf:*t*=0.734, d.f.=4, *p*>0.5) in the endothelium of *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 23hpf and 30hpf, compared to wild type. (g) Expression of *efnb2a* was decreased in *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 23hpf (*t*=3.008, d.f.=5, *p*<0.05), but not at 30hpf (*t*=0.358, d.f.=4, *p*>0.7). n=4 for *gata2a*$^{\Delta i4/\Delta i4}$ at 23hpf, n=3 for other samples.

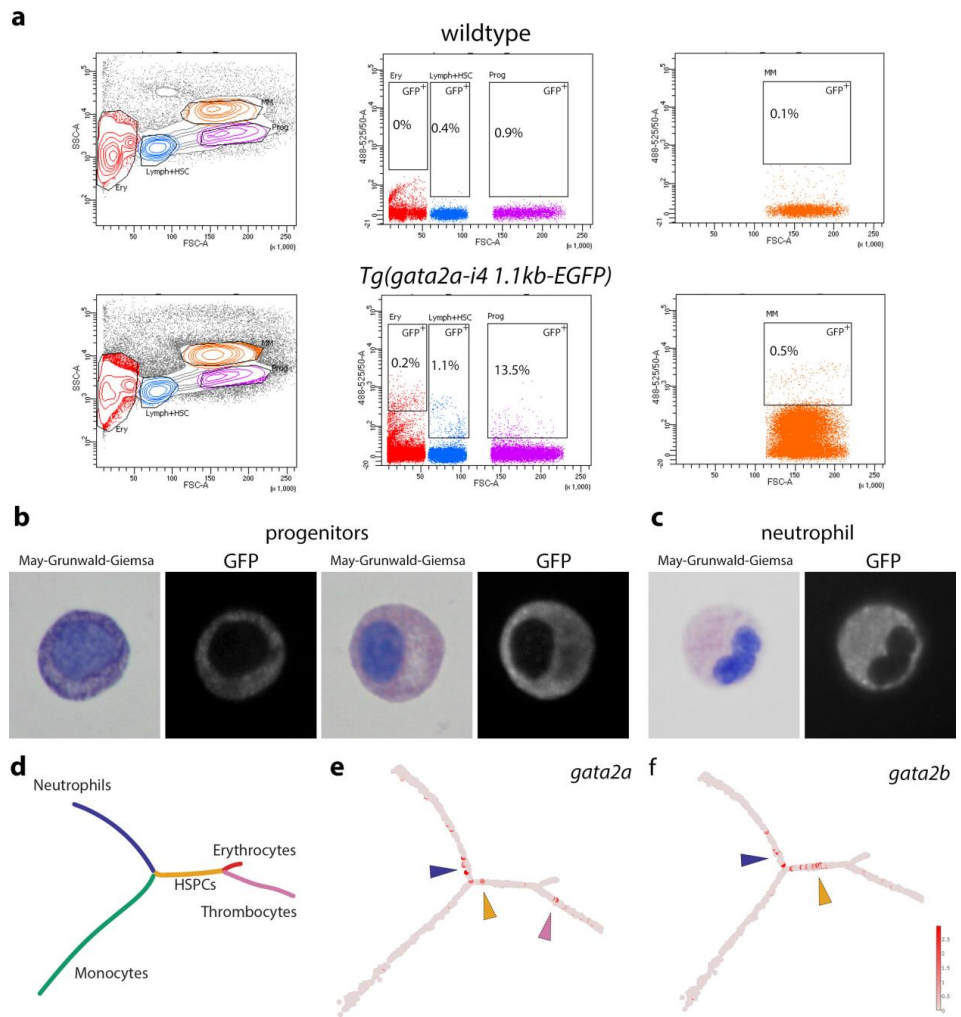


Supplementary Figure 4. Gene expression analyses in single and double *gata2a*$^{\Delta i4/\Delta i4}$ and gata2b$^{-/-}$mutants. (a-c) Analysis of *runx1* expression at 24hpf in *gata2a*$^{\Delta i4/\Delta i4}$ mutants. *Runx1* expression was significantly reduced in *gata2a*$^{\Delta i4/\Delta i4}$ compared to wild type or heterozygous siblings (μwt=17.3, μhet=14.9, μmut=6.4; F=8.752, d.f.=2, 77; p=0.0006; ANOVA). Representative embryos stained for *runx1* in the (b) wildtype and (c) *gata2a*$^{\Delta i4/\Delta i4}$ mutant embryos (shown in panel a as orange dots). Embryo numbers were n=20, wild type; n=40, *gata2a*$^{\Delta i4/+}$; n=20, *gata2a*$^{\Delta i4/\Delta i4}$. (d) Quantification of *gata2b* expression in wildtype and *gata2a*$^{\Delta i4/\Delta i4}$ mutant embryos at 26hpf. Expression of *gata2b* is significantly reduced in *gata2a*$^{\Delta i4/\Delta i4}$ mutants (μwt=42.6, μmut=25.1,; t=5.15, d.f.=34; p<0.0001; unpaired t test). The number of embryos analysed (n=14, wild type; n=22, *gata2a*$^{\Delta i4/\Delta i4}$) are shown inthe panel. (e) Quantification of *gata2b* expression in wildtype and *gata2a*$^{\Delta i4/\Delta i4}$ mutant embryos at 30hpf. We detected no differences in expression of *gata2b* between wild type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants (μwt=41.56, μmut=*41*.34,; F=2.54, d.f.=2, 54; p=0.99, ANOVA). The number of embryos analysed (n=12, wild type; n=27, *gata2a*$^{\Delta i4/+}$; n=18, *gata2a*$^{\Delta i4/\Delta i4}$) are shown in the panel. (f) Quantification of *cmyb* expression in embryos of all genotypes from *gata2a*$^{\Delta i4/+}$;gata2b$^{+/-}$ incrosses at33hpf. The expression levels in wildtype, single mutants and double mutant are shown in colour to better highlight the differences. Embryo numbers were n=22, wild type; n=57, *gata2a*$^{\Delta i4/+}$; n=29, *gata2a*$^{\Delta i4/\Delta i4}$; n=19 *gata2b*$^{+/-}$; n=42 *gata2a*$^{\Delta i4/+}$; *gata2b*$^{+/-}$; n=23 *gata2a*$^{\Delta i4/\Delta i4}$; *gata2b*$^{+/-}$; n=3, *gata2b*$^{-/-}$; n=13 *gata2a*$^{\Delta i4/+}$; *gata2b*$^{-/-}$; n=15 *gata2a*$^{\Delta i4/\Delta i4}$; *gata2b*$^{-/-}$. The data was analysed with a one way Welch's ANOVA test. Note that *cmyb* expression show no statistically significant differences between *gata2b*$^{-/-}$ and *gata2a*$^{\Delta i4/\Delta i4}$; *gata2b*$^{-/-}$ mutants (Dunnett's T3 test for multiple comparisons, t=1.616, d.f.=9.817, p=0.93). Error bars: mean±SD. (g) Representative image of *cmyb* expression in the trunk of 33hpf wildtype, (h) gata2b$^{-/-}$ and (i) *gata2a*$^{\Delta i4/\Delta i4}$; *gata2b*$^{-/-}$ mutants. Pixel intensity of the *in situ* hybridization staining was performed as described [2].



Supplementary Figure 5. Regulation of *runx1* and *gata2b* in the haemogenic endothelium by Notch signalling. (a,b) Zebrafish embryos were treated with the Notch inhibitor DAPM from 2-5somite stage until 30hpf at 0 (DMSO control), 25, 50 and 100 μM and expression of *runx1* and *gata2b* analysed by *in situ* hybridization. (a) Analysis of *runx1* expression at 30hpf upon treatment with increasing amounts of the Notch inhibitor DAPM. Expression of *runx1* was significantly reduced upon treatment with 50 M or 100μM DAPM, but not 25 μM DAPM (μ_{DMSO}=25.65, $\mu_{25\mu M}$ =22.6, $\mu_{50\mu M}$=10.91, $\mu_{100\mu M}$=8.44; F=14.60, d.f.=3, 17.94; p<0.0001; Welch's ANOVA). Multiple comparisons performed with a Dunnett's T3 test; significant p values shown in the panel. Embryo numbers were n=10, DMSO; n=9, for the 25 μM, 50 μM or 100 μM DAPM treatment. (b) Analysis of *gata2b* expression at 30hpf upon treatment with increasing amounts of the Notch inhibitor DAPM. *Gata2b* was significantly reduced upon treatment with all the DAPM concentrations tested (μ_{DMSO}=50.13, $\mu_{25\mu M}$=39.62, $\mu_{50\mu M}$=35.09, $\mu_{100\mu M}$=29.54; F=13.58, d.f.=3, 24.49; p<0.0001; Welch's ANOVA). Embryo numbers were n=12, DMSO; n=16, 25 μM DAPM; n=11, 50 μM DAPM; n=12, 100 μM DAPM treatment. (c) Quantification of *gata2b* expression at 36hpf in *gata2a*$^{\Delta i4/\Delta i4}$ mutants upon DAPM treatment. 100 μM DAPM treatment induced a significant decrease in *gata2b* expression in wild type and *gata2a*$^{\Delta i4/}$

$^{\Delta i4}$ mutants. ($\mu_{wt+DMSO}$=26.94, $\mu_{wt+25\mu M}$=22.07, $\mu_{wt+100\mu M}$=15.04; $\mu_{mut+DMSO}$=20.09, $\mu_{mut+25\mu M}$=22.32, $\mu_{mut+100\mu M}$=8.15; Welch's ANOVA). Embryo numbers were n=17, wt+ DMSO; n=14, wt+25□M DAPM; n=15, wt+100μM DAPM; n=11, *gata2a*$^{\Delta i4/\Delta i4}$ + DMSO; n=20, *gata2a*$^{\Delta i4/\Delta i4}$+25μM DAPM; n=16, *gata2a*$^{\Delta i4/\Delta i4}$ +100μM DAPM treatment. (d-e) Overexpression of a constitutively active NICD plasmid driven by the endothelial-specific *fli1a* promoter (*fli1a*-NICD:GFP[3]) using 30 and 50pg DNA. (d) *fli1a*-NICD:GFP overexpression increased expression of *runx1* (μ_{NI}=24.51, μ_{30pg}=37.97, μ_{50pg}=41.20; F=11.01, d.f.=2, 33.59; p<0.0001; Welch's ANOVA, n≥15 for all samples) and (e) *gata2b* in haemogenic endothelium at 30hpf (μ_{NI}=20.42, μ_{30pg}=41.09, μ_{50pg}=47.18; F=17.01, d.f.=2, 20.88; p<0.0001; Welch's ANOVA, n=15 for all samples). (f) Effect of *fli1a*-NICD:GFP overexpression (30pg DNA) on *runx1* expression between wild type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 26hpf. Endothelial-specific NICD overexpression can significantly rescue *runx1* expression in *gata2a*$^{\Delta i4/\Delta i4}$ mutants (μ_{wt}=48.96, μ_{mut}=29.35, $\mu_{wt+NICD}$=62.24, $\mu_{mut+NICD}$=38.21; F=25.67, d.f.=3, 12.39; p<0.0001; Welch's ANOVA, n=9, wt; n=6 for all other samples). (g) Effect of *fli1a*-NICD:GFP overexpression (30pg DNA) on *gata2b* expression between wild type and *gata2a*$^{\Delta i4/\Delta i4}$ mutants at 26hpf. Endothelial-specific NICD overexpression can significantly upregulate *gata2b* expression in *gata2a*$^{\Delta i4/\Delta i4}$ mutants (μ_{wt}=22.52, μ_{mut}=15.45, $\mu_{wt+NICD}$=56.21, $\mu_{mut+NICD}$=50.12; F=25.29, d.f.=3 ; p<0.0001; Kruskal-Wallis ANOVA, n=10, wt; n=21, *gata2a*$^{\Delta i4/\Delta i4}$; n=6 for all other samples). Multiple comparisons were performed with a Dunnett's T3 test or uncorrected Dunn's test where relevant; p values for each comparison are shown in the respective panels. Error bars: mean±SD; each dot represents the corrected pixel intensity of the *in situ* signal in one embryo. Orange dots correspond to the quantified images shownin each panel.



Supplementary Figure 6. Flow cytometry analysis of GFP expression in Tg(gata2a-i4-1.1Kb:GFP) adult zebrafish. (a) Plot showing the distribution of cell size (forward scatter) vs granularity (side scatter) to distinguish the various haematopoietic cell populations in the WKM as described [4]. The panels on the right show the gate settings to detect GFP+ cells in the lymphoid gate (Lymph+HSC), progenitor gate (Prog) and myeloid gate (MM). Gata2a-i4-driven expression of GFP was found in all the gates examined. Examples of (b) haematopoietic progenitors and (c) neutrophils expressing GFP, together with May-Grunwald/ Wright-Giemsa staining. (d) Monocle trajectory representing the differentiation path for 5 major cell types in the WKM based on gene expression differences (adapted from [5]). The colours represent each of the branches differentiating from the HSPC 'root'. (e,f) Expression summary for (e) *gata2a* and (f) *gata2b* from single cell RNA-seq data superimposed on the Monocle trajectory. The data is publicly available and was obtained from the BASiCz - Blood Atlas of Single Cells in zebrafish website (https://www.sanger.ac.uk/science/tools/basicz) [5]. (e) *gata2a* expression was detected in single cells in the neutrophil (blue arrowhead), HSPC (yellow arrowhead) and thrombocyte branch (pink arrowhead). (f) *gata2b* expression was detected in single cells mainly in the neutrophil (blue arrowhead) and HSPC (yellow arrowhead) branches. Expression levels are shown from red to pink (high to low). HSPCs – haematopoietic stem and progenitor cells.

Supplementary Table 1. Sequences of primers used in this study.

Name	Sequence (5'-3')	Purpose	Source
gata2a-i4-1.1kb f	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCAGCATCGGGATCCTATAA	Cloning	This study
gata2a-i4-1.1kb r	GGGGACCACTTTGTACAAGAAAGCTGGGTCACGAATCAAACGCTTTCAG	Cloning	This study
gata2a-i4-450bp f	GGGGACAAGTTTGTACAAAAAAGCAGGCTGATTGTTGAGAATGTTGTGGTGA	Cloning	This study
gata2a-i4-450bp r	GGGGACCACTTTGTACAAGAAAGCTGGGTCCAGTCGAGCATGACAAACA	Cloning	This study
gata2a$^{\Delta i4}$ f	TGGCTAAGTGACCGTCAGAG	Genotyping	This study
gata2a$^{\Delta i4}$ r1	TGAAACAAAACGCAGACGAC	Genotyping	This study
gata2a$^{\Delta i4}$ r2	GGGTTTGTTGAAGACGGAAA	Genotyping	This study
gfp F	ACGTAAACGGCCACAAGTTC	Amplify in situ hybridization probe	This study
gfp R	TGCTCAGGTAGTGGTTGTCG	Amplify in situ hybridization probe	This study
Kdrl F	CTCCTGTACAGCAAGGAATG	qRT-PCR (SYBR	6
Kdrl R	ATCTTTGGGCACCTTATAGC	qRT-PCR (SYBR)	6
efnB2a F	CCCATTTCCCCCAAAGACTA	qRT-PCR (SYBR)	7
efnB2a R	CTTCCCCATGAGGAGATGC	qRT-PCR (SYBR)	7
Dll4 F	ACGCATACAACCCTAACATGC	qRT-PCR (SYBR)	8
Dll4 R	CTCTGTCTGCTTCCCACTTTG	qRT-PCR (SYBR)	8
ephB4 F	CCTGATGAACACGAAAACGGA	qRT-PCR (SYBR)	9
ephB4 R	TGATAGGTCCGCACACTGTT	qRT-PCR (SYBR)	9
Flt4 F	ACAGAGGAGCCATGTTGACA	qRT-PCR (SYBR)	9
Flt4 R	GTCTGGCCTGAGAGTTGAGT	qRT-PCR (SYBR)	9
UbiC F	AAGAGACTCCCATACACCGC	qRT-PCR (SYBR)	9
UbiC R	ATTCTCAATGGTGTCGCTGG	qRT-PCR (SYBR)	9
Ef1a F	GAGAAGTTCGAGAAGGAAGC	qRT-PCR (SYBR)	10
Ef1a R	CGTAGTATTTGCTGGTCTCG	qRT-PCR (SYBR)	10
gata2a sgRNA1	GAAATTAATACGACTCACTATAGGGTGACTCCATGGAAAACCCTGGTTTTAGAG CTAGAAATAGC	Prepare guide RNA template (target region in red)	This study
gata2a sgRNA2	GAAATTAATACGACTCACTATAGGGGGCCCATGCTTTTGCATTTTGTTTTAGAG CTAGAAATAGC	Prepare guide RNA template (target region in red)	This study
gata2a sgRNA3	GAAATTAATACGACTCACTATAGGGGATGTCATTCGGGCCTGCCGGTTTTAGAG CTAGAAATAGC	Prepare guide RNA template (target region in red)	This study
gata2a sgRNA4	GAAATTAATACGACTCACTATAGGGAGCACTATGTGTGAAGGGGCGTTTTAGAG CTAGAAATAGC	Prepare guide RNA template (target region in red)	This study
Universal reverse primer	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTT TAACTTGCTATT TCTAGCTCTAAAAC	Prepare guide RNA template	11
T7 Gata2b ex3_1	TAATACGACTCACTATAAATCCGTAGCAACCCGCATCGTTTTAGAGCTAGAAAT AGCAAG	Prepare guide RNA template	This study
gata2b ex_3 Fwd	CTGTCGATGACGCAACACTG	Genotyping	This study
gata2b ex_3 Rev	TGTCGTCATGTTTCCGAGCA	Genotyping	This study

Most primers were obtained from Sigma-Aldrich. Gata2b primers and guide template were obtained from IDT

Supplementary Table 2. Catalogue numbers of TaqMan® assays used in this work.

Gene name	TaqMan® assay catalogue number
cops2	Dr03114763_m1
dld	Dr03111908_m1
dll4	Dr03428646_m1
eef1a1l1	Dr03432748_m1
efnb2a	Dr03073975_m1
gata2a	Dr03086718_m1
gata2b	Dr03140570_m1
kdrl	Dr03432897_m1
lsm12b	Dr03139893_m1
rpl13a	Dr03101115_g1
rplp0	Dr03131546_m1
runx1	Dr03074179_m1

All TaqMan® assays were obtained from ThermoFisher Scientific.

SUPPLEMENTARY REFERENCES

1. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
2. Dobrzycki, T., Krecsmarik, M., Bonkhofer, F., Patient, R. & Monteiro, R. An optimised pipeline for parallel image-based quantification of gene expression and genotyping after in situ hybridisation. *Biol Open* **7**, bio031096, doi:10.1242/bio.031096 (2018).
3. Liu, Z. *et al.* Primary cilia regulate hematopoietic stem and progenitor cell specification through Notch signaling in zebrafish. *Nat Commun* **10**, 1839, doi:10.1038/s41467-019-09403-7 (2019).
4 Traver, D. *et al.* Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol* **4**, 1238-1246, doi:10.1038/ni1007 (2003).
5 Athanasiadis, E. I. *et al.* Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nature Communications* **8**, 2045, doi:10.1038/s41467-017- 02305-6 (2017).
6. Bertrand, J. Y. *et al.* Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108-111, doi:10.1038/nature08738 (2010).
7. Lee, Y. *et al.* FGF signalling specifies haematopoietic stem cells through its regulation of somitic Notch signalling. *Nat Commun* **5**, 5583, doi:10.1038/ncomms6583 (2014).
8. Monteiro, R. *et al.* Transforming Growth Factor beta Drives Hemogenic Endothelium Programming and the Transition to Hematopoietic Stem Cells. *Dev Cell* **38**, 358-370, doi:10.1016/j.devcel.2016.06.024 (2016).
9. Bonkhofer, F. *et al.* Blood stem cell-forming haemogenic endothelium in zebrafish derives from arterial endothelium. *Nat Commun* **10**, 3577, doi:10.1038/s41467-019-11423-2 (2019).
10. Bertrand, J. Y., Kim, A. D., Teng, S. & Traver, D. CD41+ cmyb+ precursors colonize the zebrafish pronephros by a novel migration route to initiate adult hematopoiesis. *Development* **135**, 1853-1862, doi:10.1242/dev.015297 (2008).
11. Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J. L. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell Rep* **4**, 220-228, doi:10.1016/j.celrep.2013.06.020 (2013).

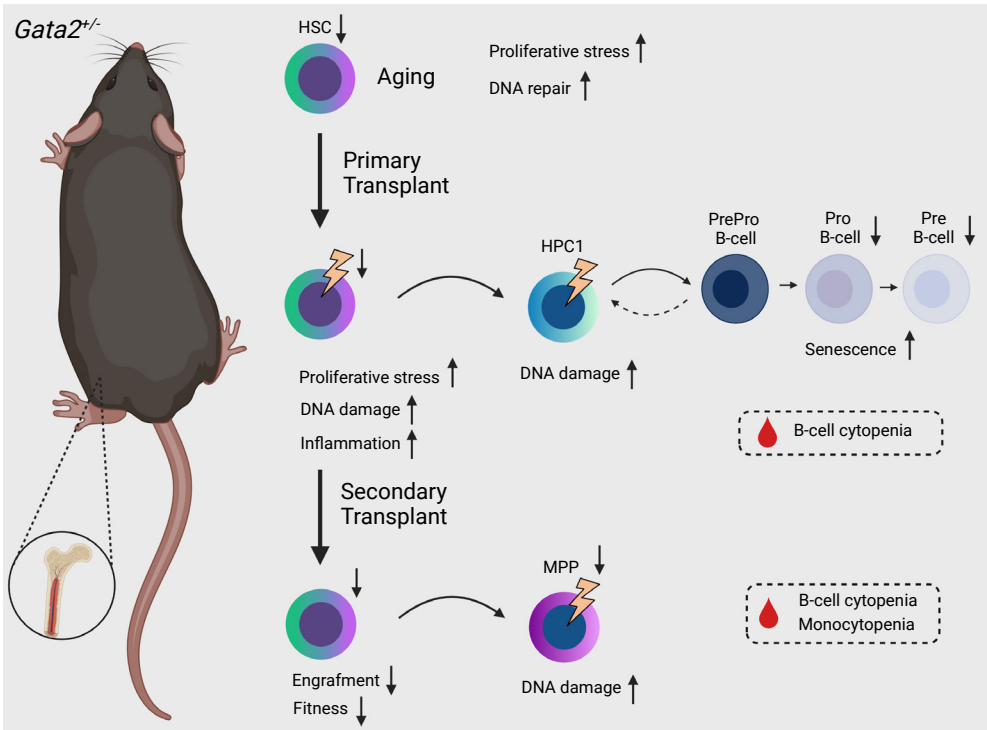
6

Gata2 haploinsufficiency promotes genome instability in aged hematopoietic stem and progenitor cells upon transplantation resulting in B-cell cytopenia and monocytopenia

Cansu Koyunlar[1,3], Emanuele Gioacchino[1,3], Hans de Looper[1], Mariëtte Ter Borg[1], Joke Zink[1], Remco Hoogenboezem[1], Paulette van Strien[1], Kirsten J Gussinklo[1], Eric Bindels[1], Elaine Dzierzak[2], Ivo Touw[1] and Emma de Pater[1].[3]
*

[1] Department of Hematology, Erasmus Medical Center Cancer Institute, Rotterdam, the Netherlands.
[2] Centre for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh, UK.
[3] Equal contribution.

Manuscript in preperation



ABSTRACT

The transcription factor *GATA2* is required for maintaining the functionality of hematopoietic stem cells (HSCs). Patients with heterozygous *GATA2* mutations develop bone marrow failure (BMF) syndromes and have 81% risk of developing myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) by the age of 40. However, underlying mechanisms of these defects remain incompletely understood. In this study we show that while transplantation of aged-*Gata2*$^{+/-}$ HSCs predominantly impairs B-lymphopoiesis, secondary transplantation results in reduced HSC fitness and pancytopenia, resembling the BMF phenotype of *GATA2* haploinsufficiency patients. To understand the underlying molecular mechanisms, we profiled the transcriptome of *Gata2*$^{+/-}$ HSCs during embryonic development, throughout aging and after transplantation. Our results suggest that the combination of proliferative stress and genome instability in aged-*Gata2*$^{+/-}$ HSCs are the driving factors of B-cell and monocyte differentiation defects in *Gata2* haploinsufficiency.

Highlights

- Aging and bone marrow transplantation impairs B-lymphopoiesis in germline *Gata2*$^{+/-}$ mice.
- Secondary transplantation of aged-*Gata2*$^{+/-}$ HSCs results in B-cell cytopenia and monocytopenia, resembling BMF phenotype of *GATA2* haploinsufficiency patients.
- *Gata2*$^{+/-}$ HSCs are proliferative throughout aging and accumulate DNA damage following transplantation.

INTRODUCTION

GATA2 is one of the key transcription factors required for effective hematopoiesis. In humans, germline heterozygous *GATA2* mutations cause GATA2 haploinsufficiency and bone marrow failure (BMF) syndromes with incomplete penetrance (*Dickinson et al., 2011; Hsu et al., 2011; Ostergaard et al., 2011*). 73% of patients manifest various cytopenias at diagnosis (49% monocytopenia, 39% neutropenia and profound B-cell cytopenia), 15% of patients suffer from lymphedema and more than 80% develop MDS/AML (*Donadieu et al., 2018; Nováková et al., 2016*). Strikingly, patients carrying germline *GATA2* mutations have a higher risk of developing clinical symptoms and MDS/AML at older age (*Donadieu et al., 2018; Spinner et al., 2014*). However, it is currently unclear how *GATA2* haploinsufficiency contributes to BMF and malignant transformation in these patients (*Spinner et al., 2014; Hahn et al., 2011*).

In mice, *Gata2* is required both for the generation and maintenance of HSCs. Germline homozygous deletion of *Gata2* (*Gata2*$^{-/-}$) is embryonically lethal due to hematopoietic failure at E10 (*Tsai et al., 1994*). On the other hand, *Gata2* heterozygous knockout mice (*Gata2*$^{+/-}$) are viable, breed normally and have normal lineage differentiation despite showing a reduction in hematopoietic stem and progenitor cells (HSPCs) during both embryogenesis (*de Pater et al., 2013; Ling et al., 2004; Tsai et al., 1994*) and adulthood (*Gao et al., 2013; Rodrigues et al., 2005*). Furthermore, conditional heterozygous deletion of *Gata2* in the hematopoietic cells results to increased proliferation and decreased lymphoid lineage differentiation ability in HSCs upon aging (*Abdelfattah et al., 2021*). However molecular mechanisms underlying these defects are incompletely understood.

In this study we show that transplantation of aged-*Gata2*$^{+/-}$ HSCs impairs B-lymphopoiesis and monocyte development. To understand the mechanisms underlying these functional defects, we explored the transcriptomic signatures of *Gata2*$^{+/-}$ HSCs during embryonic development, throughout aging. From embryonic stages onwards, *Gata2*$^{+/-}$ HSCs in steady state are marked by increased proliferative signatures. However, enforcing aged-*Gata2*$^{+/-}$ HSCs to restore the hematopoietic system in transplantation assays promotes genome instability and results in B-cell and monocyte differentiation defects. As BMF syndromes are characterized by defects in one or more lineages that result in cytopenias, our results illustrate the transcriptional attributes of *Gata2*$^{+/-}$ HSCs that possibly promote the progression of BMF in *Gata2* haploinsufficiency syndromes.

MATERIALS AND METHODS

Mouse breeding and maintenance

Gata2$^{+/-}$ mice have been previously described (*Tsai et al., 1994*). Mice were kept and bred under specific pathogen-free (SPF) conditions, and sacrificed by cervical dislocation. The studies were approved by the Animal Welfare/Ethics Committee of the EDC in accordance with legislation in the Netherlands.

Bone marrow sampling

Mouse BM was obtained by flushing femurs and tibias using a 25G (BD Microlance) needle with Phosphate buffer solution (PBS) supplemented with 5 IU/mL penicillin, 5 µg/mL streptomycin, and 10% fetal calf serum (FCS). The bone marrow cell suspension was filtered using a 70-micron nylon mesh, treated with red blood cell lysis buffer (BD Pharm Lyse™) and stored on ice until use. An aliquot of the BM cell suspension was diluted with phosphate-buffered saline (PBS) and loaded into a hemocytometer to count nucleated cells. BM cellularity was calculated as number of nucleated cells obtained from 1 femur / body weight (g) per mouse.

Peripheral blood count

Blood was sampled from cheek puncture and collected in a Microtainer Brand Tube with EDTA (ethylenediaminetetraacetic acid). Samples were analyzed on Hemavet 850 (Drew scientific).

Colony-forming unit assay

MethoCult GF M3434 (Stem Cell Technologies, Vancouver, BC, Canada) was defrosted overnight at 4°C and used according to manufacturer's instructions to enumerate colony-forming units (CFU). For optimal colony growth, 100 freshly sorted cells were plated in 1,1ml of MethoCult placed in 10mm style Falcon petri dish (Corning Incorporated) and colonies scored after 10 days of culture at 37°C with 5% CO2. After colony count, MethoCult was removed from the plates using PBS with 10% FCS at 37°C, cells counted and 104 cells re-plated for 1st, 2nd and 3rd re-plating. Growth of primitive erythroid progenitor cells (BFU-E), granulocyte-macrophage progenitor cells (CFU-GM, CFU-G and CFU-M), and multi-potential granulocyte, erythroid, macrophage, megakaryocyte progenitor cells (CFU-GEMM) were scored using an inverted microscope.

LSK SLAM staining

BM cells were labeled with lineage antibodies: PE Rat Anti Mouse CD45R/B220 Clone RA3-6B2 (cat 553089), PE Rat Anti-Mouse CD3 Molecular Complex Clone 17-A2 (Cat 555275), PE Rat Anti-CD11b Clone M1/70 (Cat 553311), PE Rat Anti-Mouse TER-119 (Cat 553673), PE

Rat Anti-Mouse Ly-6G and Ly-6C (Gr1) Clone RB6-8C5 (Cat 553128) (all from BD Bioscience). To identify murine hematopoietic stem and progenitor cells, BM cells were co-stained with APC Rat Anti Mouse CD117 Clone:2B8 Cat:553356 (BD Bioscience), BB700 Rat anti Mouse Ly-6A (Sca1) Clone D7 Cat 742089 (BD Bioscience), Alexa Fluor 700 anti-mouse CD48 Cat 103426 (BioLegend), Rat anti-mouse CD150 Cat 115914 (BioLegend). All LSK SLAM FACS antibodies incubations were performed in PBS + 10% FCS for 20 min on ice. Fetal liver LSK SLAM analyses were performed according to Kim et al. (*Kim et al., 2006*).

Cell proliferation, apoptosis and γH2AX assays

Apoptosis was analyzed using FITC Annexin V Apoptosis Detection Kit I (BD Biosciences) according to the recommendation of the manufacturer. To detect death from apoptotic cells, Annexin-V was used together with 4,6-diamidino-2-phenylindole (DAPI) (Molecular Probes). For Cell cycle analysis, $2x10^6$ freshly isolated BM cells are stained with surface markers, were fixed while vortexing using 2% PFA and incubated for 1 hour at 4°C in the dark. After washing, cells were permeabilized by re-suspending the cell pellet in PBS/0.2% Triton with 1:500 DAPI and incubated overnight at 4°C in the dark. After a washing step, cells were stained using Ki67 FITC (Thermofisher, catalog #11-5698-82) 1:25 in PBS 10% FCS with 1:500 DAPI and incubated for 2 hours at 4°C in the dark vortexing every 30 minutes. After washing, cells were resuspended in PBS with 1:500 DAPI and analyzed (Szade et al., 2016). γH2AX levels were assessed in fixed and permeabilized cells with Cytofix/Cytoperm Fixation/Permeabilization Solution Kit (BD Biosciences) by incubating cells with Alexa Fluor 488 mouse anti-γH2AX (N1-431, Cat 560445, BD Biosciences), diluted in 1X Perm/Wash buffer (BD Biosciences) 1:200. All FACS events were recorded using a BD LSR II Flow Cytometer or a BD FACSAria III and analyzed with FlowJo 7.6.5 software (Tree Star). Cells were sorted with a BD FACSAria III.

Flow cytometry for lineage differentiation cellular senescence analysis

PB and BM cells were labeled with lineage antibodies: BV510 Hamster Anti-mouse CD3e (BD Biosciences, Cat: 740113), APC anti-mouse CD115 (Sony, Cat: 1277550), APC/Cy7 anti-mouse CD45.2 (Sony, Cat: 1149115), 7-AAD (BD Biosciences, Cat: 559925), Alexa Fluor 700 anti-mouse Gr-1 (Sony, Cat: 1142105), PE/Cy7 anti-mouse/human CD11b (Sony, Cat: 1106080), eFluor 450 B220 (eBioscience, Cat: 48-0452-80), PE anti-mouse CD45.1 (Biolegend, Cat: 110707). All antibody incubations were performed in PBS + 10% FCS for 20 min on ice. After washing, cells were resuspended in PBS with 1:1000 DAPI or 1:100 7AAD and analyzed using a BD LSR II Flow Cytometer. Cellular senescence is assessed using CellEvent Senescence Detection Green Kit (Thermofisher) according to manufacturer's protocol prior to staining with lineage antibodies.

RNA extraction and RNA quality control

Total sample RNA isolation was performed according to the standard protocol of RNA isolation with Trizol and GenElute LPA (Sigma). Quality and quantity of the total RNA was checked on a 2100 Bioanalyzer (Agilent) using the Agilent RNA 6000 Pico Kit.

RNA Sequencing and gene set enrichment analysis (GSEA)

cDNA was prepared using SMARTer procedure with SMARTer Ultra Low RNA kit (Clontech) for Illumina Sequencing. SMARTer Ultra Low RNA Kit (Clontech) for Illumina Sequencing was used for cDNA based. The gene expression values were measured as FPKM (Fragments per kilobase of exon per million fragments mapped). Fragment counts were determined per gene with HTSeq-count, utilizing the strict intersection option, and subsequently used for differential expression analysis using the DESeq2 package, with standard parameters, in the R environment. GSEA (Gene Set Enrichment Analysis) was performed on the FPKM values using the curated C2 collection of gene-sets.

Serial bone marrow transplantation

$3x10^6$ freshly isolated nucleated BM $CD45.2^+$ cells were transplanted via tail vein injection into lethally irradiated (10.5Gy) 7-10 weeks old $CD45.1^+$ recipient mice (B6.SJL-PtprcaPepcb/BoyCrl, Charles River). The donor cell chimerism was determined in PB every month after transplantation.

Statistics

Data are presented as mean ± SEM. All statistical analysis was carried out in GraphPad Prism 8.0.1 (GraphPad Software Inc., San Diego, CA). Normally distributed data were analyzed using an unpaired t-test, not normally distributed data were analyzed using Mann-Whitney test. A p value less than 0.05 was considered significant.

RESULTS

$Gata2^{+/-}$ mice have reduced numbers of HSCs, but normal myeloid and erythroid potential.

Previous studies have shown that *Gata2* is predominantly expressed in HSCs and that *Gata2* haploinsufficiency results in a diminished HSC compartment (*de Pater et al., 2013; Ling et al., 2004; Tsai et al., 1994; Guo et al., 2013; Rodrigues et al., 2005*). In mice, following their emergence in the developing embryo, HSCs migrate to fetal liver (FL), where they gain the expression of cell surface marker combination $Lin^-Sca1^+cKit^+CD48^-CD150^+$ (LSK SLAM), and robustly expand in numbers. HSCs then relocate and reside in the BM to establish the rest of the adult hematopoiesis. To study the effect of Gata2 haploinsufficiency, we analyzed the

embryonic and adult HSC compartment in *Gata2*$^{+/-}$ mice at embryonic day (E)14 FL and in the adult mice BM using LSK SLAM markers (Figure 1A). Consistent with earlier studies (*Guo et al., 2013; Rodrigues et al., 2005*), we found reduced numbers of phenotypic HSCs both in FL and BM of *Gata2*$^{+/-}$ mice (Figure 1B-C).

We next sought to investigate if this reduction affects myeloid and erythroid differentiation capacity of HSCs and multipotent progenitors (MPP or LSK CD48^{-}CD150^{-}) of adult *Gata2*$^{+/-}$ BM in CFU-C with equal input of HSCs or MPPs per dish. No significant differences in the number or type of colonies were detected for the first plating or following serial replating experiments (Figure 1D-E and S1A-B) indicated an unaffected myeloid and erythroid differentiation potential from phenotypic HSCs and MPPs in adult *Gata2*$^{+/-}$ BM.

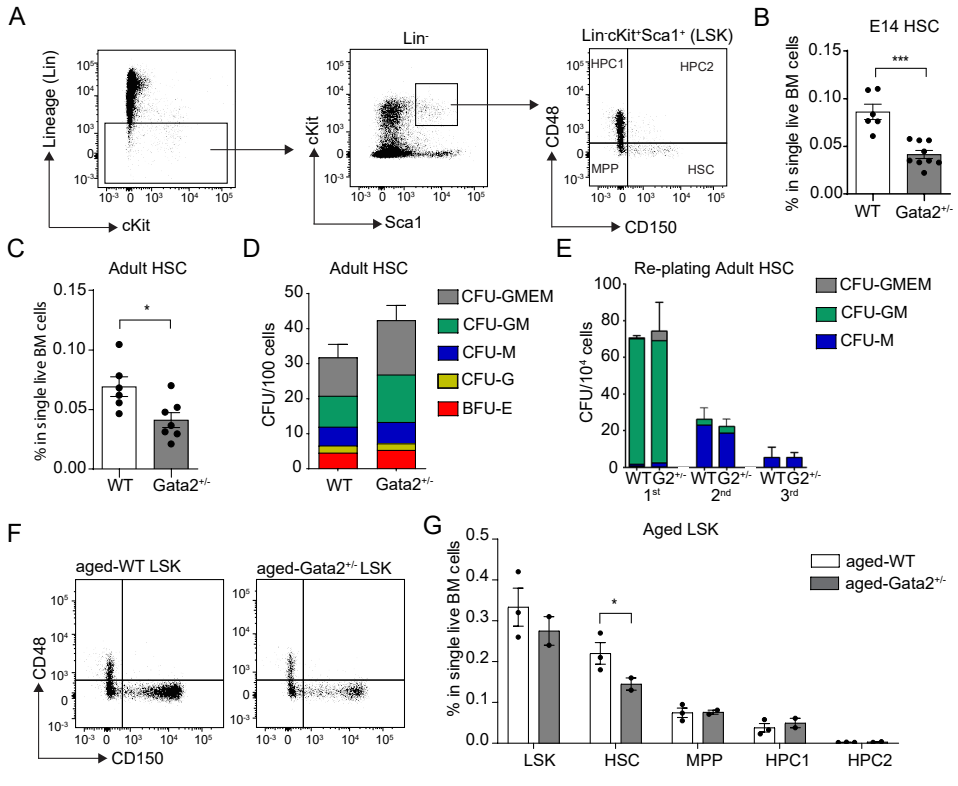


Figure 1. *Gata2*$^{+/-}$ mice have decreased number of HSCs at E14 and throughout aging.
A) Sorting strategy for BM HSC, MPP, HPC1 and HPC2 populations in FACS experiments. B) Quantification of HSC proportions in WT and *Gata2*$^{+/-}$ E14 FL. C) Quantification of HSC proportions in adult WT and *Gata2*$^{+/-}$ BM. D) Quantification of the number of BFU-E, CFU-G, CFU-M and CFU-GM and CFU-GMEM colonies obtained from CFU-C assay of adult WT and *Gata2*$^{+/-}$ HSCs and E) from serial re-plating experiments. F) Representative image of LSK compartments obtained from FACS experiments and compared between aged-WT (left) and aged-*Gata2*$^{+/-}$ (right) BM. G) Quantification of LSK, HSC, MPP, HPC1 and HPC2 proportions in aged-WT and aged-*Gata2*$^{+/-}$ BM. *P < 0.05, ***P < 0.001.

Gata2$^{+/-}$ *mice do not show lineage differentiation defects during aging.*
Despite the reduction in HSCs, myeloid and erythroid lineage differentiation was not altered in *Gata2*$^{+/-}$ mice in CFU assays (Figure 1C-E). In addition, normal peripheral blood (PB) values of *Gata2*$^{+/-}$ mice indicated that terminal lineage differentiation was intact (Figure S1C). In mice, however, aging reduces the lineage differentiation potential of *Gata2*$^{+/-}$ HSCs resulting in lymphoid lineage defects (*Abdelfattah et al., 2021*). Additionally, in GATA2 haploinsufficiency patients, the propensity to develop MDS/AML increases with older age to about 80% at the age of 40 (*Donadieu et al., 2018*). Therefore, we investigated the contribution of aging to the onset of BMF in *Gata2*$^{+/-}$ mice. As reduced BM functionality often manifests with cytopenia, we examine the PB values of WT and *Gata2*$^{+/-}$ mice on a monthly basis up to 15 months of age. However, we did not observe any terminal differentiation defects in PB cells in *Gata2*$^{+/-}$ mice during aging (Figure S2A). Next, we analyzed the proportion of HSCs and MPPs as well as the hematopoietic progenitor cells-1 (HPC1 or LSK CD48^{+}CD150^{-}) and hematopoietic progenitor cells-2 (HPC2 or LSK CD48^{-}CD150^{+}) compartments in aged-WT and aged-*Gata2*$^{+/-}$ BM. Although the proportion of HSCs in both WT and *Gata2*$^{+/-}$ mice were expanded (~ 2-fold) during aging, HSCs in aged-*Gata2*$^{+/-}$ mice remained reduced compared to aged-WT (Figure 1F-G). On the other hand, the proportion of MPP, HPC1 and HPC2 in the BM of aged-*Gata2*$^{+/-}$ mice were not changed compared to aged-WT (Figure 1F-G). These results show that Gata2 heterozygosity in mice continuously affects the HSC pool but not the MPP, HPC1 and HPC2 compartments or mature PB cells after 15 months of aging.

Aged-Gata2$^{+/-}$ *mice developed B-cell cytopenia after bone marrow transplantation.*
Transplantation of whole BM into lethally irradiated recipients forces HSCs to re-establish the hematopoietic system, an ultimate criteria for HSC functionality (*Rossi et al., 2012*). To explore the effects of aging on the lineage reconstitution ability of *Gata2*$^{+/-}$ HSCs, we transplanted 3x10^{6} aged-WT or aged-*Gata2*$^{+/-}$ nucleated BM cells to 10 weeks old WT, lethally irradiated recipients (Figure 2A). After transplantation, we tracked the lineage contribution of CD45.2 expressing donor cells in recipient mice. Two months after transplantation and onwards, aged-*Gata2*$^{+/-}$ mice showed a reduction of the percentage of CD45.2^{+} cells in the peripheral blood (PB) (Figure 2B). Next to reduced donor cell chimerism, we observed a decreased white blood cell (WBC) count in aged-*Gata2*$^{+/-}$ mice during monthly PB analysis (Figure 2C and S2B). Although morphologically normal, the PB smears showed a decreased proportion of lymphocytes and a relatively increased proportion of monocytes of aged-*Gata2*$^{+/-}$ mice (Figure 2D-H). Furthermore, flow cytometry analysis showed normal neutrophil (CD11b^{+}Gr1^{hi}), monocyte (CD11b^{+}CD115^{+}), other myeloid cells (eosinophils and basophils or CD11b^{+}CD115^{-}Gr1$^{int/low}$) and T-cell (CD11b^{-}CD3^{+}) numbers, but decreased B-cell numbers (CD11b^{-}B220^{+}) showing lymphopenia in the PB after transplantation of aged-*Gata2*$^{+/-}$ BM (Figure 2I-J and S3A). On the other hand, the proportion of myeloid cells and T-cells were comparable between transplanted aged-WT and aged-*Gata2*$^{+/-}$ BM (Figure 2J).

To explore the effect of *Gata2* haploinsufficiency on immature and lineage biased BM compartments, we isolated nucleated BM cells 4 months after primary transplantation and found that both aged-WT and aged-*Gata2*$^{+/-}$ HSCs had fully engrafted (Figure S3B). Moreover, while the proportion of HSCs remained reduced in aged-*Gata2*$^{+/-}$ BM after transplantation, other LSK compartments such as HPC1, HPC2 and MPP were intact (Figure 2K). To investigate BM lineage differentiation, we used the same marker combination as in PB analysis and additionally used erythroid lineage markers CD71 and Ter119 to analyze proerythroblast (CD71^{+}Ter119^{int}) and erythroblast (CD71^{+}Ter119^{hi}) populations (Figure S3A and E). Upon sacrifice of primary transplant recipients, we detected a minor reduction in BM B-cells and as expected overall lineage differentiation was not significantly altered in aged-*Gata2*$^{+/-}$ BM after transplantation (Figure S3C-D and F).

These results show that, although BM repopulation and differentiation ability of HSCs was preserved, aging and transplantation result in B-cell cytopenia, which is also the most profound phenotypic consequence of *GATA2* haploinsufficiency in patients.

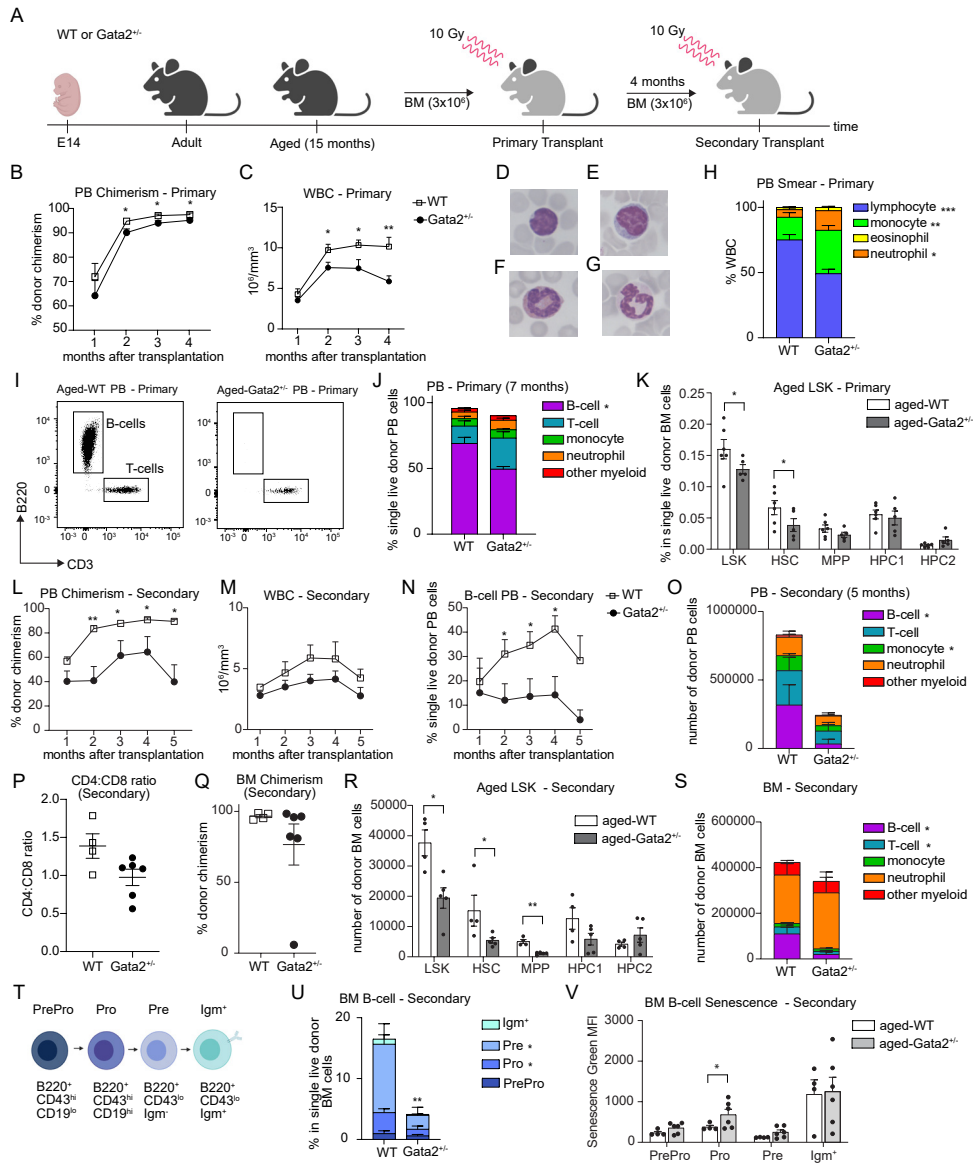


Figure 2. Transplantation of aged-*Gata2*$^{+/-}$ HSCs results in bone marrow failure.
A) Illustration of the experimental setup. B) Proportion of CD45.2-expressing donor cells in the PB of recipients after primary transplantation of aged-WT and aged-*Gata2*$^{+/-}$ BM cells. C) WBC count in the PB of recipients after primary transplantation of aged-WT and aged-*Gata2*$^{+/-}$ BM cells. D) Representative image of a lymphocyte, E) monocyte, F) eosinophil and G) neutrophil in PB smears. H) Quantification of WBC types in the PB smears of recipients 4 months after primary transplantation of aged-WT and aged-*Gata2*$^{+/-}$. 100 WBCs per PB smear were counted. I) Representative image of B-cell and T-cell compartments obtained from flow cytometry experiments and compared between aged-WT (left) and aged-*Gata2*$^{+/-}$ (right) BM. J) Quantification of lymphoid and myeloid cell proportions in the PB of aged-WT and aged-*Gata2*$^{+/-}$ mice 7 months after primary transplantation. K) Quantification of LSK, HSC, MPP, HPC1 and HPC2 proportions in aged-WT and aged-*Gata2*$^{+/-}$ BM after primary transplantation. L) Proportion of CD45.2-expressing donor cells in the PB of recipients after secondary transplantation of aged-WT and aged-*Gata2*$^{+/-}$ BM cells. M) WBC count in the PB of recipients after secondary transplantation of aged-WT and aged-*Gata2*$^{+/-}$ BM cells. N) Proportion of donor B-cells in the PB of recipients after secondary transplantation. O) Quantification of the number of donor lymphoid and myeloid cells in the PB of aged-WT and aged-*Gata2*$^{+/-}$ mice 5 months after secondary transplantation. P) Quantification of CD4:CD8 ratio in CD3^{+} T-cell compartment in PB of aged-WT and aged-*Gata2*$^{+/-}$ mice 5 after secondary transplantation. Q) Proportion of CD45.2-expressing aged-WT and aged-*Gata2*$^{+/-}$ donor cells in the BM of secondary recipients. R) Quantification of the number of LSK, HSC, MPP, HPC1 and HPC2 cells in aged-WT and aged-*Gata2*$^{+/-}$ after secondary transplantation. S) Quantification of the number of donor lymphoid and myeloid cells in the BM of aged-WT and aged-*Gata2*$^{+/-}$ mice after secondary transplantation. T) Illustration of the B-cell differentiation trajectory in BM. U) Proportion of donor PrePro, Pro, Pre and Igm^{+} B-cells in the BM of aged-WT and aged-*Gata2*$^{+/-}$ mice after secondary transplantation. V) Quantification of the MFI of senescence signals in PrePro, Pro, Pre and Igm^{+} B-cells in the BM of aged-WT and aged-*Gata2*$^{+/-}$ mice after secondary transplantation. *P < 0.05, **P < 0.01, ***P < 0.001.

Aged-Gata2$^{+/-}$ *HSCs show reduced functionality after secondary transplantation.*

To test the self-renewability of aged-*Gata2*$^{+/-}$ HSCs, we performed secondary BM transplantation assays with 3x10^{6} aged-WT or aged-*Gata2*$^{+/-}$ cells harvested from the BM of primary transplanted mice after 4 months of follow-up (Figure 2A). During the monthly PB analysis, we observed a severe reduction of CD45.2^{+} cells in the aged-*Gata2*$^{+/-}$ mice (Figure 2L). Next to the reduced PB chimerism, B-cell cytopenia was also consistent in the aged-*Gata2*$^{+/-}$ mice after secondary transplantation (Figure 2M-N). Interestingly,

also monocytes were now significantly decreased in aged-*Gata2*$^{+/-}$ mice 5 months after secondary transplantation indicating BM involvement (Figure 2O). Furthermore, PB defects were restricted to WBCs whereas other PB values were not altered in aged-*Gata2*$^{+/-}$ mice (Figure S4A). In a subgroup of *GATA2* patients, an inverted CD4:CD8 ratio was described due to reduction in CD4^{+} T-helper cells (Dickinson et al., 2014). Although total CD3^{+} T-cell compartment was not reduced in the aged-*Gata2*$^{+/-}$ PB, CD4:CD8 ratio was reduced in aged-*Gata2*$^{+/-}$ mice. In addition, for 2 mice the ratio was inverted (CD4:CD8 < 1) suggesting a conserved phenotypic consequence of *GATA2* haploinsufficiency between human and mice on T-cell lineage differentiation (Figure 2P).

Next, we analyzed the BM of aged-WT and aged-*Gata2*$^{+/-}$ mice after secondary transplantation. BM cellularity was comparable in both groups 5 months after secondary transplantation (Figure S4B), but BM chimerism was reduced in 3 out of 6 aged-*Gata2*$^{+/-}$ transplanted mice indicating a reduced HSC engraftment ability (Figure 2Q). In addition, we found a significant reduction in the number of MPPs and HSCs in aged-*Gata2*$^{+/-}$ mice after secondary transplantation, resulting in a diminished LSK compartment (Figure 2R). Because we observed a reduction in B-cells, monocytes and CD4^{+} T-cells in PB, we then analyzed donor cell lineage contribution in BM and found reduced number of B-cells and T-cells in the BM of aged-*Gata2*$^{+/-}$ mice after secondary transplantation (Figure 2S). Notably, the number of myeloid and erythroid cells in the BM of aged-*Gata2*$^{+/-}$ mice was not altered despite the monocytopenia found in PB (Figure 2S). Furthermore, we detected an increased proportion of proerythroblast and erythroblast cells in the spleen of aged-*Gata2*$^{+/-}$ mice after secondary transplantation (Figure S4E). However, we did not observe splenomegaly in these mice (Figure S4F).

To examine if impaired B-cell differentiation resulted in a differentiation block in a specific stage of BM B-cell development, we performed flow cytometry analysis using B-cell lineage markers. In these experiments, we analyzed the proportion of PrePro (B220^{+}CD43^{hi}CD19^{low}), Pro (B220^{+}CD43^{hi}CD19^{hi}), Pre (B220^{+}CD43^{low}IgM^{-}) and IgM^{+} (B220^{+}CD43^{low}IgM^{+}) B-cells (Figure 2T and S4C) and found reduced B-cell differentiation in aged-*Gata2*$^{+/-}$ BM with significant reductions in PreB- and ProB-cell compartments (Figure 2U). This suggests that *Gata2* haploinsufficiency causes a differentiation block in the ProB-cells of the BM. Interestingly, the block is only partial and even though IgM^{+} class switched B-cells were reduced, they remain detectable as previously described in *Gata2* haploinsufficiency patients (Figure 2U) (*Dickinson et al., 2014*).

Finally, we asked whether the B-cell differentiation block is caused by increased senescence as this has been shown to predominantly affect transitional B-cells but not terminally differentiated IgM^{+} B-cells (*Bagnara et al., 2015; Frasca et al., 2018*). To investigate this, we used a fluorescence-based cellular senescence detection kit and measured mean fluorescent intensity (MFI) of senescence signals throughout the B-cell differentiation trajectory in the BM. Strikingly, we found an increased senescence signal in ProB-cells, possibly explaining the differentiation block observed in B-cells lineage in aged-*Gata2*$^{+/-}$ BM (Figure 2V).

These results indicate that aging and two consecutive transplants cause HSC exhaustion in *Gata2*$^{+/-}$ BM leading to reduced HSC engraftment ability and differentiation defects predominantly in B-cells as well as in CD4^{+} T-cells and monocytes indicative of BMF progression.

Transcriptome analysis revealed an increased proliferative signature in Gata2$^{+/-}$ *HSCs.*
In order to underpin the molecular defect of *Gata2*$^{+/-}$ induced BMF, we aimed to investigate transcriptional profiles of *Gata2*$^{+/-}$ HSCs. Therefore, we sorted phenotypic HSCs from WT or *Gata2*$^{+/-}$ mice at key developmental time points E14, adult, aged and 4 months after primary transplantation to performed RNA sequencing (RNA-seq) experiments. Hierarchical clustering of transcriptomes by principal component (PC) analysis indicated an evident separation of *Gata2*$^{+/-}$ HSCs by the first principal component (PC1) throughout development and aging and by the second principal component (PC2) after transplantation (Figure 3A-C and S5A).

To understand the biological processes affected in *Gata2*$^{+/-}$ HSCs, we performed hallmark pathway analysis using gene expression signatures and curated gene-sets. Strikingly, proliferation related hallmark gene-sets such as *E2F Targets*, *G2-M Checkpoint* and *Myc Targets V1* were enriched in *Gata2*$^{+/-}$ HSCs at all stages (Figure 3D-L and S5B-D). Some genes within these gene-sets such as *Chek1*, *Cdkn2c*, *Mcm4*, *Mcm5*, *Mybl2* were enriched in *Gata2*$^{+/-}$ HSCs compared to WT HSCs at all time-points indicating *Gata2*$^{+/-}$ HSCs comprise a unique proliferative transcriptional profile at embryonic stages and preserve these signatures throughout aging and after transplantation (Figure S5M-Q). Furthermore, hallmark gene-sets *Unfolded Protein Response* and *mTORC1 Signaling* were upregulated from E14 and onwards suggesting an increased cellular stress in *Gata2*$^{+/-}$ HSCs originated from embryonic stages (Figure S5E-L) (*Reiling and Sabatini, 2008*).

Network analysis using differentially expressed genes (P < 0.05) (Cytoscape; (*Otasek et al., 2019; Shannon et al., 2003*) showed increase proliferative signatures in *Gata2*$^{+/-}$ HSCs at all stages (Figure 3M, 4A and S6A-B). Interestingly, while we observed overrepresentation of gene-sets related to cell cycle, proliferation and DNA repair, gene-sets related to mRNA translation were downregulated in both E14 and aged-*Gata2*$^{+/-}$ HSCs (Figure 3M and S6A). Additionally, a group of gene-sets associated with cell quiescence were downregulated in aged-*Gata2*$^{+/-}$ HSCs, further indicating that *Gata2*$^{+/-}$ HSCs have left quiescence and are more proliferative (Figure 3M). To test this, we performed cell cycle analysis of adult- and aged-*Gata2*$^{+/-}$ LK, LSK and HSC populations. Both adult- and aged-*Gata2*$^{+/-}$ HSCs indeed showed a significant loss of quiescent G_0 phase cells and relevant acquisition of cells in the G_1 phase of cell cycle resulting in overall loss of quiescence in the LSK compartment (Figure 3N-P, S7B and G). Increased proliferation in *Gata2*$^{+/-}$ was specific for HSCs and LSK cells and the

proportion of cycling cells in the more mature LK compartment was not changed between WT and *Gata2*$^{+/-}$ (Figure S7A and F).

In cytoscape analysis we detected an enrichment of gene-sets related to DNA repair pathways in adult- and aged-*Gata2*$^{+/-}$ HSCs (Figure 3M and S6B). We hypothesized that the loss of quiescence and activation of DNA repair gene-sets in *Gata2*$^{+/-}$ HSCs could be a sign of impaired genome integrity. Therefore, we assessed DNA damage in adult- and aged-*Gata2*$^{+/-}$ LSK compartment. We used Ser139-phosphorylated H2AX histone (γH2AX) which accumulates in response to DNA damage and marks double strand breaks (DSBs) as a hallmark of proliferative stress (*Mah et al., 2010; Rogakou et al., 1999*). In flow cytometry DNA damage assay, we measured the γH2AX mean fluorescent intensity (MFI). However, we did not observe aberrant γH2AX signals in adult or aged *Gata2*$^{+/-}$ HSCs (Figure S7C and H).

Previous studies have pointed out a correlation between the cell cycle arrest at S phase and increased apoptosis (*Krek et al., 1995; Zhang et al., 2000*). Because we found an increased proportion of LSK cells in the S phase of cell cycle in *Gata2*$^{+/-}$ mice, we next analyzed the proportion of apoptotic cells in LSK and HSC compartments. However, we did not find a dramatic increase of the apoptotic cells in these populations in *Gata2*$^{+/-}$ mice (Figure S7D-E).

These results suggest that although *Gata2*$^{+/-}$ HSCs are more proliferative and their transcriptome indicates an increased cellular stress, they do not become apoptotic nor accumulate DNA damage during aging.

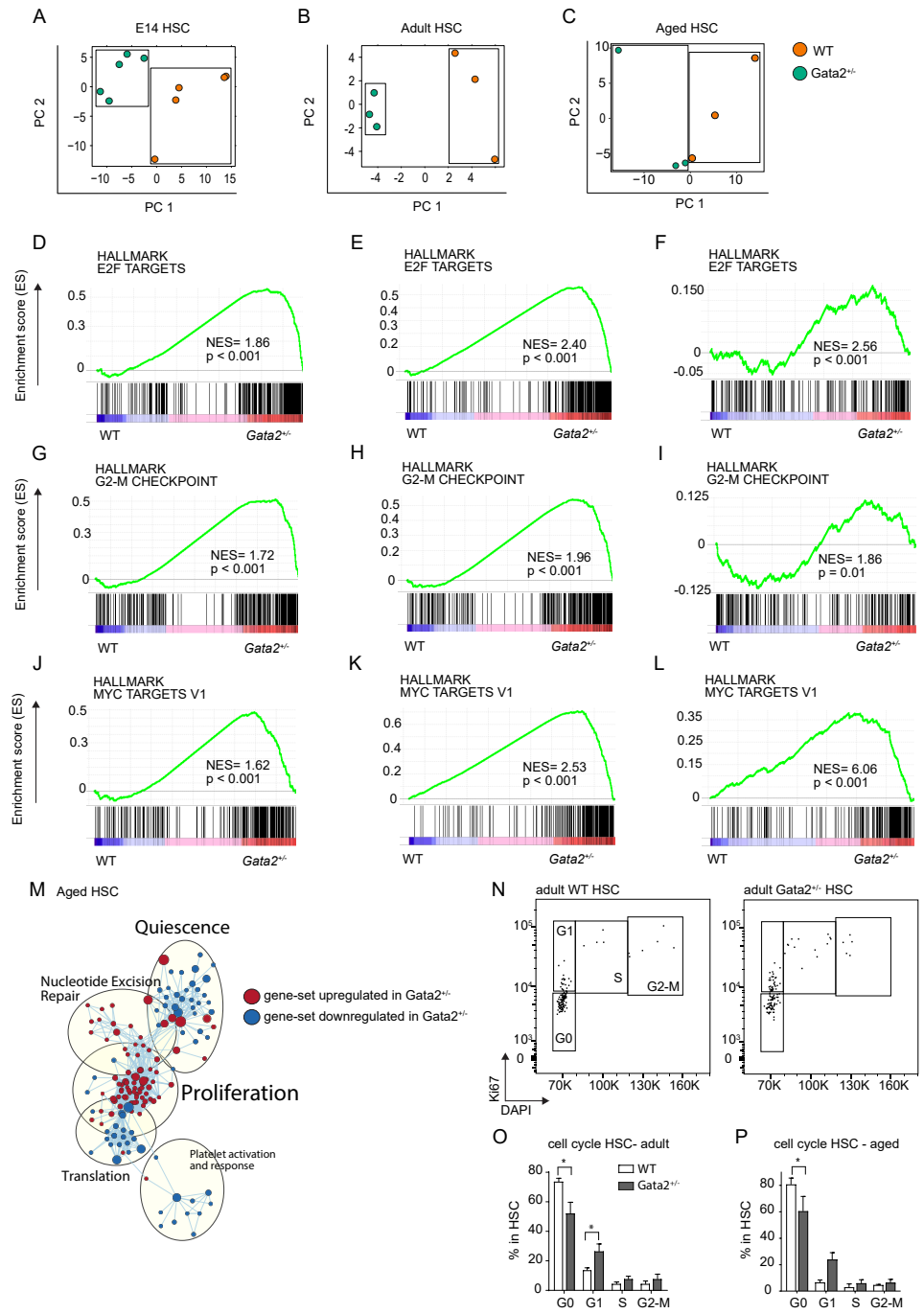


Figure 3. *Gata2*$^{+/-}$ HSCs are marked by proliferative transcriptomic signatures.
A) PCA of E14 B) Adult and C) Aged WT (orange) and *Gata2*$^{+/-}$ (green) HSCs. Each dot represents the transcriptome of and individual sample. D) Hallmark GSEA of *E2f* targets in E14 E) Adult and F) Aged WT and *Gata2*$^{+/-}$ HSCs. G) Hallmark GSEA of G_2-M checkpoint in E14 H) Adult and I) Aged WT and *Gata2*$^{+/-}$ HSCs. J) Hallmark GSEA of *Myc* targets in E14 K) Adult and L) Aged WT and *Gata2*$^{+/-}$ HSCs. M) Network analysis comparing aged-WT and aged-*Gata2*$^{+/-}$ HSCs. Red dots show upregulated and blue dots show downregulated gene-sets in *Gata2*$^{+/-}$ HSCs compared to WT. N) Representative image of Ki67 cell cycle analysis between WT (left) and *Gata2*$^{+/-}$ (right) HSCs. O) Proportions of individual cell cycle stages compared between adult and P) aged WT and *Gata2*$^{+/-}$ HSCs. *P < 0.05.

Transplantation induces DNA damage accumulation in aged-Gata2$^{+/-}$ HSPCs.

Aged-*Gata2*$^{+/-}$ HSCs showed reduced HSC engraftment and defects in B-cell lineage differentiation only after BM transplantation. In addition, network analysis indicated that transcriptomic signatures related to proliferation and DNA repair remained enriched in aged-*Gata2*$^{+/-}$ HSCs after transplantation (Figure 4A). To investigate this further, we performed cell cycle analysis in LSK, HSC and LK compartments isolated 4 months after transplantation and found an increased proportion of G_2-M phase cells in aged-*Gata2*$^{+/-}$ HSCs and LSK cells (Figure 4B-D). This was not the case for the less immature LK population (Figure S7I). These results suggest that different mechanisms might be important for the regulation of cycling *Gata2*$^{+/-}$ HSCs before and after BM transplantation, where they occupy G_1 phase and G_2-M phase of the cell cycle respectively. Because we found an increased proportion of cells in the G_2-M phase and network analysis showed that gene-sets related *DNA damage checkpoint* and *Homology directed repair (HDR)* pathways were upregulated in aged-*Gata2*$^{+/-}$ HSCs after transplantation, we sought to understand if genomic stability was altered after the secondary injury of transplantation (Figure 4A). In the γH2AX assay we detected a significant increase in γH2AX signals in HSCs and HPC1 compartments of *Gata2*$^{+/-}$ BM. This suggested the genome integrity in aged-*Gata2*$^{+/-}$ HSPCs was impaired due to DSBs accumulation following the transplantation (Figure 4E-G). To investigate if increased genome instability in aged-*Gata2*$^{+/-}$ HSCs was caused by dysregulation of HDR pathway genes, we analyzed the expression of *Brca1*, *Brca2*, *Rad51*, *53bp1*, *Rpa1*, *Rpa2*, *Pcna* and RNAse H2 subunits *Rnaseh2a, Rnaseh2b* and *Rnaseh2c*. We found that, both in WT and *Gata2*$^{+/-}$ mice, expression levels of these genes were upregulated throughout aging (Figure 4H). Importantly, the expression of *Rnaseh2a* was significantly downregulated in aged-*Gata2*$^{+/-}$ HSCs compared to aged-WT HSCs (Figure S7C). This was particularly interesting since the loss of *Rnaseh2* was associated with increased genome instability (*Lockhart et al., 2019*).

Because the engraftment ability of HSCs was reduced, we next analyzed proliferation and DNA damage in aged-*Gata2*$^{+/-}$ LK and LSK compartments after secondary transplantation and found that cell proportions in individual cell cycle stages were comparable for LK, LSK and HSC populations between aged-WT and aged-*Gata2*$^{+/-}$ (Figure S7K-M). Interestingly, we found an increased γH2AX signals in aged-*Gata2*$^{+/-}$ MPPs but not in other LSK populations or LK cells upon secondary transplantation (Figure S7N).

Taken together, these results suggest that the combination of proliferative stress, aging and a secondary injury like transplantation induce genomic instability in *Gata2*$^{+/-}$ HSPCs.

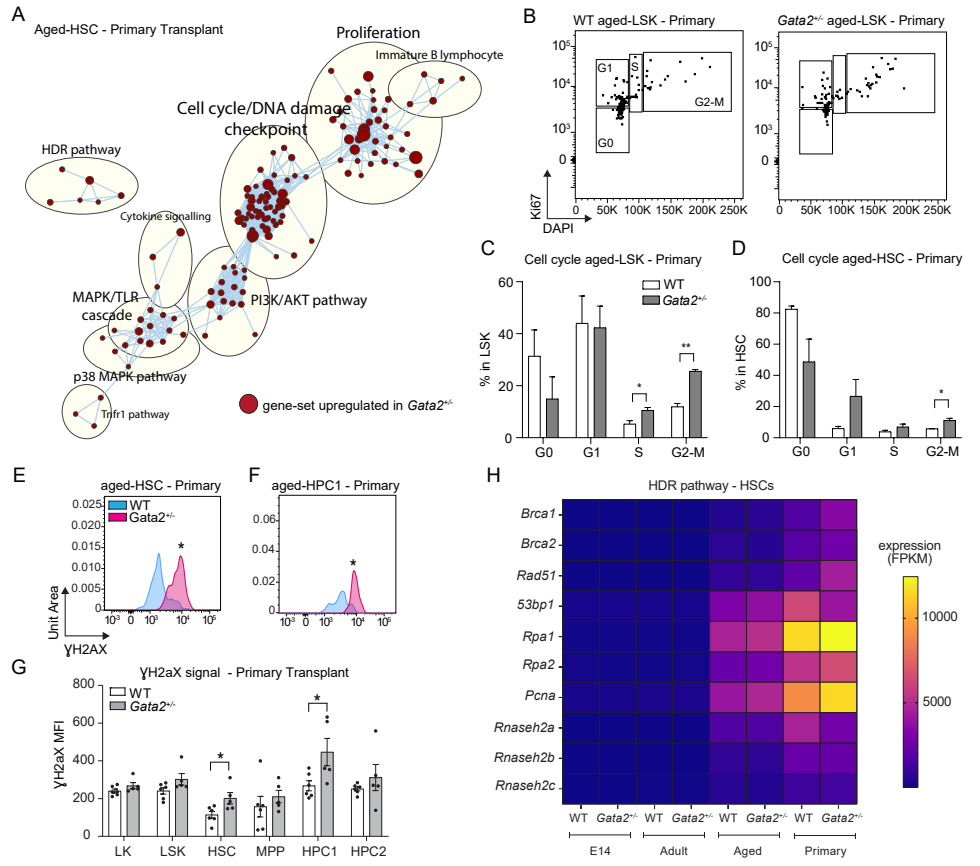


Figure 4. Transplantation of aged-*Gata2*$^{+/-}$ HSCs induces genome instability in HSPCs. A) Network analysis comparing aged-WT and aged-*Gata2*$^{+/-}$ HSCs after transplantation. Red dots show upregulated gene-sets in *Gata2*$^{+/-}$ HSCs compared to WT. B) Representative image of Ki67 cell cycle analysis between aged-WT (left) and aged-*Gata2*$^{+/-}$ (right) LSK cells after primary transplantation. C) Proportions of individual cell cycle stages compared between aged-WT and aged-*Gata2*$^{+/-}$ LSK cells and D) HSCs. E) Representative image of γH2AX signals compared between aged-WT (blue) and aged-*Gata2*$^{+/-}$ (pink) HSCs and F) HPC1 cells after primary transplantation. G) Quantification of γH2AX signals in aged-WT and aged-*Gata2*$^{+/-}$ LK, LSK, HSC, MPP, HPC1 and HPC2 compartments of the BM. H) Heatmap of the FPKM values of *Brca1*, *Brca2*, *Rad51*, *53bp1*, *Rpa1*, *Rpa2*, *Pcna*, *Rnaseh2a, Rnaseh2b* and *Rnaseh2c* compared between WT and *Gata2*$^{+/-}$ HSCs at timepoints E14, adult, aged and transplanted-aged. *P < 0.05, **P < 0.01.

TNFα and inflammatory signaling pathways are upregulated in aged-Gata2$^{+/-}$ HSPCs after transplantation.

Previous studies have shown that inflammatory signals could be activated in response to increased genome instability (*Ioannidou et al., 2016*). Interestingly, network analysis of aged-*Gata2*$^{+/-}$ HSCs showed an enrichment for TNFα signaling and downstream pathways such as *p38 MAPK pathway*, *MAPK/TLR cascade* and *PI3K/Akt pathway* as well as *Cytokine signaling* indicating unique acquisition of inflammatory signatures in aged-*Gata2*$^{+/-}$ HSCs after transplantation (Figure 4A). In addition, we observed an enrichment for hallmark

gene-sets *TNFα signaling via NFkB and inflammatory response* in aged-*Gata2*$^{+/-}$ HSCs after transplantation (Figure S8A-B). Remarkably, these gene-sets were downregulated in adult- and aged-*Gata2*$^{+/-}$ HSCs before transplantation (Figure S8D-G). Within these two gene-sets, expression of genes such as *Il1r2*, *Il3ra*, *Il13ra1*, *Il5ra*, *Fosb*, *Atf3*, *Cxcl2*, *Cxcr2*, *Ccl9*, *Map3k1*, *Mapk13*, *Atf3*, *Nfkbiz*, *Cebpb*, *Mxd1*, *Cdk2*, *Klf2*, *Klf6* and *Jun* were enriched in aged-*Gata2*$^{+/-}$ HSCs compared to aged-WT HSCs (Figure S8C). Notably, some of these genes were described in the stress related MAPK pathway and upstream TNFα signaling (*Sabio and Davis., 2014*).

It has been shown that, in the LSK compartment, HPC1 population predominantly gives rise to B-cell lineage while partially contributing to myeloid differentiation in the BM (*Oguro et al., 2014).* Because B-cell differentiation was impaired in aged-*Gata2*$^{+/-}$ mice after transplantation and we found an increased DNA damage in aged-*Gata2*$^{+/-}$ HPC1 population, we further investigated the transcriptome of HPC1 cells sorted from aged-WT and aged-*Gata2*$^{+/-}$ 4 months after primary transplantation. After RNA-seq experiments, we performed network analysis and hallmark GSEA to investigate the effect of *Gata2* haploinsufficiency in aged-HPC1 population. In the aged-HPC1 compartment, similarly to aged-HSCs, *TNFα signaling via NFkB* and *Inflammatory response* pathways were upregulated in *Gata2*$^{+/-}$ mice (Figure S8H-J). Within these gene-sets, upregulation of *Tnfsf13b*, *Tnfaip3*, *Nfkbia*, *Fos*, *Fosb*, *Jun*, *Jund*, *Dusp1*, *Klf2*, *Klf6*, *Il2rb* and *Il10ra* in aged-*Gata2*$^{+/-}$ HPC1 further confirmed the acquisition of inflammatory signatures in this population upon transplantation (Figure S8K).

These results suggest that activation of inflammatory signaling pathways in aged-*Gata2*$^{+/-}$ HSCs and HPC1 cells are co-occurrent with the onset of BMF phenotype in these mice. Because increased inflammation by TNFα signaling was associated with B-cell senescence (*Ratliff et al., 2013*), similar mechanisms could possibly explain the senescent phenotype of ProB-cells found in aged-*Gata2*$^{+/-}$ BM.

DISCUSSION

In this study, we examined the cellular characteristics of *Gata2*$^{+/-}$ mice and investigated the transcriptional profile of *Gata2*$^{+/-}$ HSCs from embryonic development throughout the onset of multi-lineage differentiation defects in aged mice. It has been described that B-cell cytopenia and monocytopenia are the common defects of *GATA2* haploinsufficiency patients (*Hsu et al., 2015; Nováková et al., 2015*). Here we showed that, upon aging and transplantation, *Gata2*$^{+/-}$ mice mimic the clinical phenotypes of the *GATA2* patients and are failed to produce sufficient numbers of B-cells and monocytes. Our study revealed that B-cell differentiation defect in aged-*Gata2*$^{+/-}$ mice was caused by increased senescence in ProB-cells.

Transcriptome analysis showed that, throughout embryonic development and aging, *Gata2*$^{+/-}$ HSCs lose quiescence and are marked by increased proliferative signatures. However, BM and PB lineage differentiation remained intact and reduced functionality in aged-*Gata2*$^{+/-}$ HSCs became evident only after they were challenged to re-establish the hematopoietic system in the primary transplantation assays. Simultaneously with the B-cell lineage differentiation defects, we found an increased genome instability in aged-*Gata2*$^{+/-}$ HSCs as well as the HPC1 population. Interestingly, HPC1 population was shown to substantially contribute to B-lymphopoiesis (*Oguro et al., 2014*). This suggested the HPC1 population, in addition to HSCs, is one of the first cellular components of BM that is affected from *Gata2* haploinsufficiency. Furthermore, we found an inverted CD4:CD8 ratio within CD3^{+} T-cell compartment, recapitulating the phenotype of a subgroup of *GATA2* patients (*Dickinson et al., 2014*). Both B-cell and T-cell compartments are affected from *Gata2* haploinsufficiency suggested a defect in CLPs. This was confirmed in a previous study where aging significantly reduced the number of CLPs in *Gata2*$^{+/-}$ mice (*Abdelfattah et al., 2021*).

Although mimicked the BMF phenotype, aged-*Gata2*$^{+/-}$ mice did not developed MDS or AML even after secondary transplantation (Supplementary Table). This can be explained by the lack of infectious and mutagen agents in our system that could potentially push the hematopoietic system further and elevate HSC exhaustion. However, our model recapitulates the immunodeficiency phenotype of *GATA2* patients that possibly contributes to the MDS and AML progression.

Gata2$^{+/-}$ HSCs lose quiescence and cell cycle analysis revealed that these cells mostly occupy the G_1 stage of the cell cycle during aging. Whether this is a compensatory mechanism due to a reduction in the number of HSCs and is an underlying factor for increased genome instability requires further investigation. However, HSCs are also required to leave the quiescent state to repair DSBs if cells are susceptible to genome instability (*Beerman et al., 2014*). If the latter is the case, simultaneous upregulation of proliferation and DNA repair pathways in *Gata2*$^{+/-}$ HSCs and preserved genome integrity together indicates that replication errors are being sufficiently repaired in these cells throughout aging. Interestingly, upon

transplantation, cycling cells of aged-*Gata2*$^{+/-}$ HSCs were found in the G_2-M stage of the cell cycle. A similar pattern was also shown in a previous study where aging shifts the cell cycle stage occupancy of *Gata2*$^{+/-}$ HSCs from G_1 to G_2-M stage of the cell cycle (*Abdelfattah et al., 2021*). In our model, concomitantly with this shift, genomic instability was induced in aged-*Gata2*$^{+/-}$ HSCs. This suggests the aging causes proliferative stress that is marked by increased G_2-M stage cells and impaired DNA damage tolerance and results in reduced functionality in *Gata2*$^{+/-}$ HSCs.

It has been shown that, in the S/G_2-M phase of cell cycle, DSBs undergo resection and cells preferentially use HDR pathway to repair DNA lesions (*Branzei and Foiani., 2008*). In line with this, we found an enrichment of HDR pathway genes together with the increase in the proportion of G_2-M cells in aged-*Gata2*$^{+/-}$ HSCs after transplantation. Although some HDR pathway related genes such as *Brca2* and *Rad51* were upregulated, *Rnaseh2* subunit *Rnaseh2a* was significantly downregulated in aged-*Gata2*$^{+/-}$ HSCs after transplantation. *Rnaseh2a* was shown to play a role in HDR pathway by interacting with *Brca2* to control DNA:RNA hybrids at DSBs in the S/G_2-M phase of the cell cycle. In addition, it was shown that *Rnaseh2a* is recruited at DSBs and localize in close proximity with γH2AX (*D'Alessandro et al., 2018*). Moreover, mutations in *Rnaseh2* were associated with increased genome instability (*Lockhart et al., 2019; Zimmermann et al., 2018*). Further studies are needed to understand whether dysregulation of DNA:RNA hybrids at DSBs contributes to genome instability in aged-*Gata2*$^{+/-}$ HSCs.

Inflammatory signatures such as *TNFα signaling via NFkB* were enriched in aged-*Gata2*$^{+/-}$ HSCs and HPC1 population uniquely after transplantation. It was shown that inflammation could be triggered as a result of DNA damage accumulation (*Ioannidou et al., 2016*). On the other hand, defects in B-cell differentiation were shown to activate pro-inflammatory signals and, in return, these signatures can activate senescence in B-cell precursors as a feedback mechanism (*Ratliff et al., 2013*). We proposed that, either up- or down-stream of DNA damage accumulation, the activation of inflammatory signatures could contribute to the B-cell differentiation block in aged-*Gata2*$^{+/-}$ BM upon transplantation. Additionally, increased TNFa signaling and the activation of stress related NFkB pathway were shown to promote differentiation ability, however, impair the self-renewability of HSCs (*Jahandideh et al., 2020*). This could explain the loss of self-renewal and consecutively the reduced repopulating ability of aged-*Gata2*$^{+/-}$ HSCs after secondary transplantation. In the future, investigating the link between DNA damage accumulation and activation of inflammatory signaling pathways in aged-*Gata2*$^{+/-}$ mice in the concept of HSC maintenance and B-cell differentiation could help us to identify the driving factors underlying these defects. In addition, identifying the type and the genomic location of mutations arise in aged-*Gata2*$^{+/-}$ HSCs could elucidate the downstream events that promote genome instability.

In conclusion, we identified unique transcriptomic signatures of *Gata2*$^{+/-}$ HSCs during development, aging and at the onset of multi-lineage differentiation defects. We showed that proliferative stress, increased genome instability and the activation of inflammatory responses are the attributes of aged-*Gata2*$^{+/-}$ HSCs leading to BMF progression in mice upon challenging these cells in transplantation assays. Furthermore, although phenotypic consequences of *Gata2* haploinsufficiency in mice were evident upon aging and transplantation, we showed that *Gata2*$^{+/-}$ HSCs acquire proliferative signatures as early as E14 FL stage. Therefore, it would be interesting to study the clonality of aged-*Gata2*$^{+/-}$ HSCs to determine in which developmental stage *Gata2*$^{+/-}$ HSCs become susceptible to genome instability.

REFERENCES

Abdelfattah, A., Hughes-Davies, A., Clayfield, L., Menendez-Gonzalez, J. B., Almotiri, A., Alotaibi, B., Tonks, A., & Rodrigues, N. P. (2021). Gata2 haploinsufficiency promotes proliferation and functional decline of hematopoietic stem cells with myeloid bias during aging. *Blood Adv*, *5*(20), 4285-4290.

Bagnara, D., Squillario, M., Kipling, D., Mora, T., Walczak, A. M., Da Silva, L., Weller, S., Dunn-Walters, D. K., Weill, J. C., & Reynaud, C. A. (2015). A Reassessment of IgM Memory Subsets in Humans. *J Immunol*, *195*(8), 3716-3724.

Beerman, I., Bhattacharya, D., Zandi, S., Sigvardsson, M., Weissman, I. L., Bryder, D., & Rossi, D. J. (2010). Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc Natl Acad Sci U S A*, *107*(12), 5465-5470.

Branzei, D., & Foiani, M. (2008). Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol*, *9*(4), 297-308.

D'Alessandro, G., Whelan, D. R., Howard, S. M., Vitelli, V., Renaudin, X., Adamowicz, M., Iannelli, F., Jones-Weinert, C. W., Lee, M., Matti, V., Lee, W. T. C., Morten, M. J., Venkitaraman, A. R., Cejka, P., Rothenberg, E., & d'Adda di Fagagna, F. (2018). BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. *Nat Commun*, *9*(1), 5376.

de Pater, E., Kaimakis, P., Vink, C. S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P. S., Camper, S. A., Speck, N., & Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. *J Exp Med*, *210*(13), 2843-2850.

Dickinson, R. E., Griffin, H., Bigley, V., Reynard, L. N., Hussain, R., Haniffa, M., Lakey, J. H., Rahman, T., Wang, X. N., McGovern, N., Pagan, S., Cookson, S., McDonald, D., Chua, I., Wallis, J., Cant, A., Wright, M., Keavney, B., Chinnery, P. F., Loughlin, J., Hambleton, S., Santibanez-Koref, M., & Collin, M. (2011). Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood*, *118*(10), 2656-2658.

Donadieu, J., Lamant, M., Fieschi, C., de Fontbrune, F. S., Caye, A., Ouachee, M., Beaupain, B., Bustamante, J., Poirel, H. A., Isidor, B., Van Den Neste, E., Neel, A., Nimubona, S., Toutain, F., Barlogis, V., Schleinitz, N., Leblanc, T., Rohrlich, P., Suarez, F., Ranta, D., Chahla, W. A., Bruno, B., Terriou, L., Francois, S., Lioure, B., Ahle, G., Bachelerie, F., Preudhomme, C., Delabesse, E., Cave, H., Bellanné-Chantelot, C., Pasquet, M., & French, G. s. g. (2018). Natural history of GATA2 deficiency in a survey of 79 French and Belgian patients. *Haematologica*, *103*(8), 1278-1287.

Frasca, D. (2018). Senescent B cells in aging and age-related diseases: Their role in the regulation of antibody responses. *Exp Gerontol*, *107*, 55-58.

Gao, X., Johnson, K. D., Chang, Y. I., Boyer, M. E., Dewey, C. N., Zhang, J., & Bresnick, E. H. (2013). Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *J Exp Med*, *210*(13), 2833-2842.

Guo, G., Luc, S., Marco, E., Lin, T. W., Peng, C., Kerenyi, M. A., Beyaz, S., Kim, W., Xu, J., Das, P. P., Neff, T., Zou, K., Yuan, G. C., & Orkin, S. H. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell*, *13*(4), 492-505.

Hahn, C. N., Chong, C. E., Carmichael, C. L., Wilkins, E. J., Brautigan, P. J., Li, X. C., Babic, M., Lin, M., Carmagnac, A., Lee, Y. K., Kok, C. H., Gagliardi, L., Friend, K. L., Ekert, P. G., Butcher, C. M., Brown, A. L., Lewis, I. D., To, L. B., Timms, A. E., Storek, J., Moore, S., Altree, M., Escher, R., Bardy, P. G., Suthers, G. K., D'Andrea, R. J., Horwitz, M. S., & Scott, H. S. (2011). Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*, *43*(10), 1012-1017.

Hsu, A. P., McReynolds, L. J., & Holland, S. M. (2015). GATA2 deficiency. *Curr Opin Allergy Clin Immunol*, *15*(1), 104-109.

Hsu, A. P., Sampaio, E. P., Khan, J., Calvo, K. R., Lemieux, J. E., Patel, S. Y., Frucht, D. M., Vinh, D. C., Auth, R. D., Freeman, A. F., Olivier, K. N., Uzel, G., Zerbe, C. S., Spalding, C., Pittaluga, S., Raffeld, M., Kuhns, D. B., Ding, L., Paulson, M. L., Marciano, B. E., Gea-Banacloche, J. C., Orange, J. S., Cuellar-Rodriguez, J., Hickstein, D. D., & Holland, S. M. (2011). Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*, *118*(10), 2653-2655.

Ioannidou, A., Goulielmaki, E., & Garinis, G. A. (2016). DNA Damage: From Chronic Inflammation to Age-Related Deterioration. *Front Genet*, *7*, 187.

Jahandideh, B., Derakhshani, M., Abbaszadeh, H., Akbar Movassaghpour, A., Mehdizadeh, A., Talebi, M., & Yousefi, M. (2020). The pro-Inflammatory cytokines effects on mobilization, self-renewal and differentiation of hematopoietic stem cells. *Hum Immunol*, *81*(5), 206-217.

Kim, I., He, S., Yilmaz, O. H., Kiel, M. J., & Morrison, S. J. (2006). Enhanced purification of fetal liver hematopoietic stem cells using SLAM family receptors. *Blood*, *108*(2), 737-744.

Krek, W., Xu, G., & Livingston, D. M. (1995). Cyclin A-kinase regulation of E2F-1 DNA binding function underlies suppression of an S phase checkpoint. *Cell*, *83*(7), 1149-1158.

Ling, K. W., Ottersbach, K., van Hamburg, J. P., Oziemlak, A., Tsai, F. Y., Orkin, S. H., Ploemacher, R., Hendriks, R. W., & Dzierzak, E. (2004). GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med*, *200*(7), 871-882.

Lockhart, A., Pires, V. B., Bento, F., Kellner, V., Luke-Glaser, S., Yakoub, G., Ulrich, H. D., & Luke, B. (2019). RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cell Rep*, *29*(9), 2890-2900 e2895.

Mah, L. J., El-Osta, A., & Karagiannis, T. C. (2010). gammaH2AX: a sensitive molecular marker of DNA damage and repair. *Leukemia*, *24*(4), 679-686.

Nováková, M., Žaliová, M., Suková, M., Wlodarski, M., Janda, A., Froňková, E., Campr, V., Lejhancová, K., Zapletal, O., Pospíšilová, D., Černá, Z., Kuhn, T., Švec, P., Pelková, V., Zemanová, Z., Kerndrup, G., van den Heuvel-Eibrink, M., van der Velden, V., Niemeyer, C., Kalina, T., Trka, J., Starý, J., Hrušák, O., & Mejstříková, E. (2016). Loss of B cells and their precursors is the most constant feature of GATA-2 deficiency in childhood myelodysplastic syndrome. *Haematologica*, *101*(6), 707-716.

Oguro, H., Ding, L., & Morrison, S. J. (2013). SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell*, *13*(1), 102-116.

Ostergaard, P., Simpson, M. A., Connell, F. C., Steward, C. G., Brice, G., Woollard, W. J., Dafou, D., Kilo, T., Smithson, S., Lunt, P., Murday, V. A., Hodgson, S., Keenan, R., Pilz, D. T., Martinez-Corral, I., Makinen, T., Mortimer, P. S., Jeffery, S., Trembath, R. C., & Mansour, S. (2011). Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet*, *43*(10), 929-931.

Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., & Demchak, B. (2019). Cytoscape Automation: empowering workflow-based network analysis. *Genome Biology*, *20*(1), 185. https://doi.org/10.1186/s13059-019-1758-4

Ratliff, B. B., Rabadi, M. M., Vasko, R., Yasuda, K., & Goligorsky, M. S. (2013). Messengers without borders: mediators of systemic inflammatory response in AKI. *J Am Soc Nephrol*, *24*(4), 529-536.

Reiling, J. H., & Sabatini, D. M. (2008). Increased mTORC1 signaling UPRegulates stress. *Mol Cell*, *29*(5), 533-535.

Rodrigues, N. P., Janzen, V., Forkert, R., Dombkowski, D. M., Boyd, A. S., Orkin, S. H., Enver, T., Vyas, P., & Scadden, D. T. (2005). Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood*, *106*(2), 477-484.

Rogakou, E. P., Boon, C., Redon, C., & Bonner, W. M. (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J Cell Biol*, *146*(5), 905-916.

Rossi, L., Lin, K. K., Boles, N. C., Yang, L., King, K. Y., Jeong, M., Mayle, A., & Goodell, M. A. (2012). Less is more: unveiling the functional core of hematopoietic stem cells through knockout mice. *Cell Stem Cell*, *11*(3), 302-317.

Sabio, G., & Davis, R. J. (2014). TNF and MAP kinase signalling pathways. *Semin Immunol*, *26*(3), 237-245.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, *13*(11), 2498-2504.

Spinner, M. A., Sanchez, L. A., Hsu, A. P., Shaw, P. A., Zerbe, C. S., Calvo, K. R., Arthur, D. C., Gu, W., Gould, C. M., Brewer, C. C., Cowen, E. W., Freeman, A. F., Olivier, K. N., Uzel, G., Zelazny, A. M., Daub, J. R., Spalding, C. D., Claypool, R. J., Giri, N. K., Alter, B. P., Mace, E. M., Orange, J. S., Cuellar-Rodriguez, J., Hickstein, D. D., & Holland, S. M. (2014). GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood*, *123*(6), 809-821.

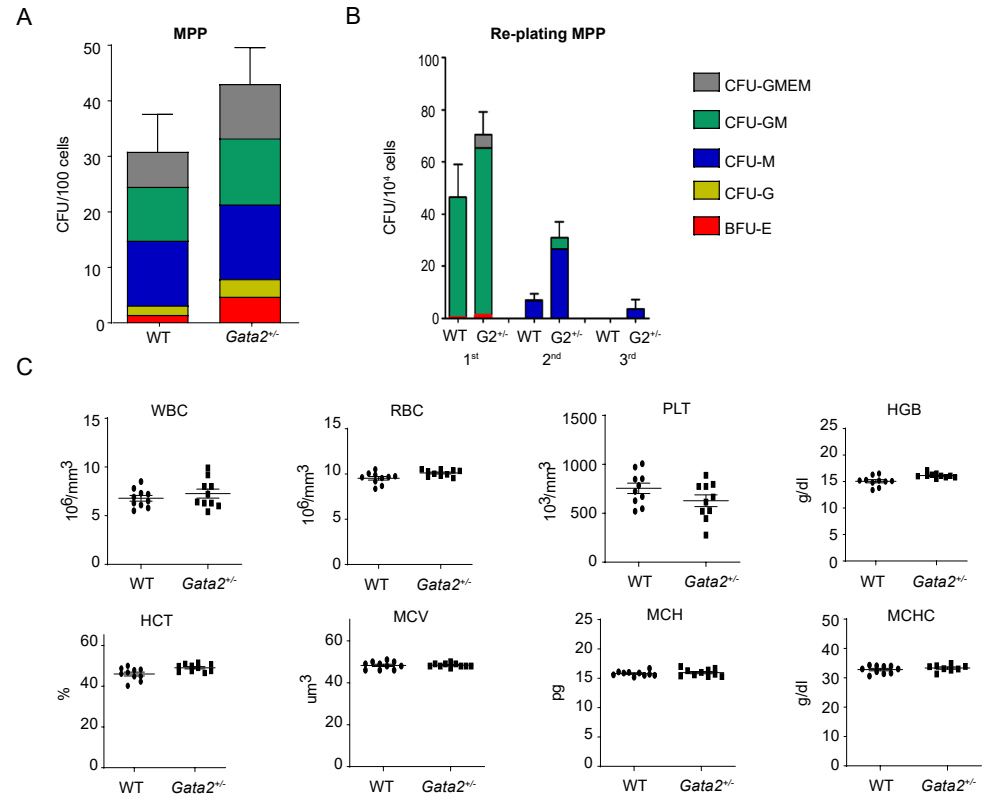
Szade, K., Bukowska-Strakova, K., Zukowska, M., Jozkowicz, A., & Dulak, J. (2016). Analysis of Cell Cycle Status of Murine Hematopoietic Stem Cells. *Methods Mol Biol*, *1516*, 91-99.

Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W., & Orkin, S. H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, *371*(6494), 221-226.

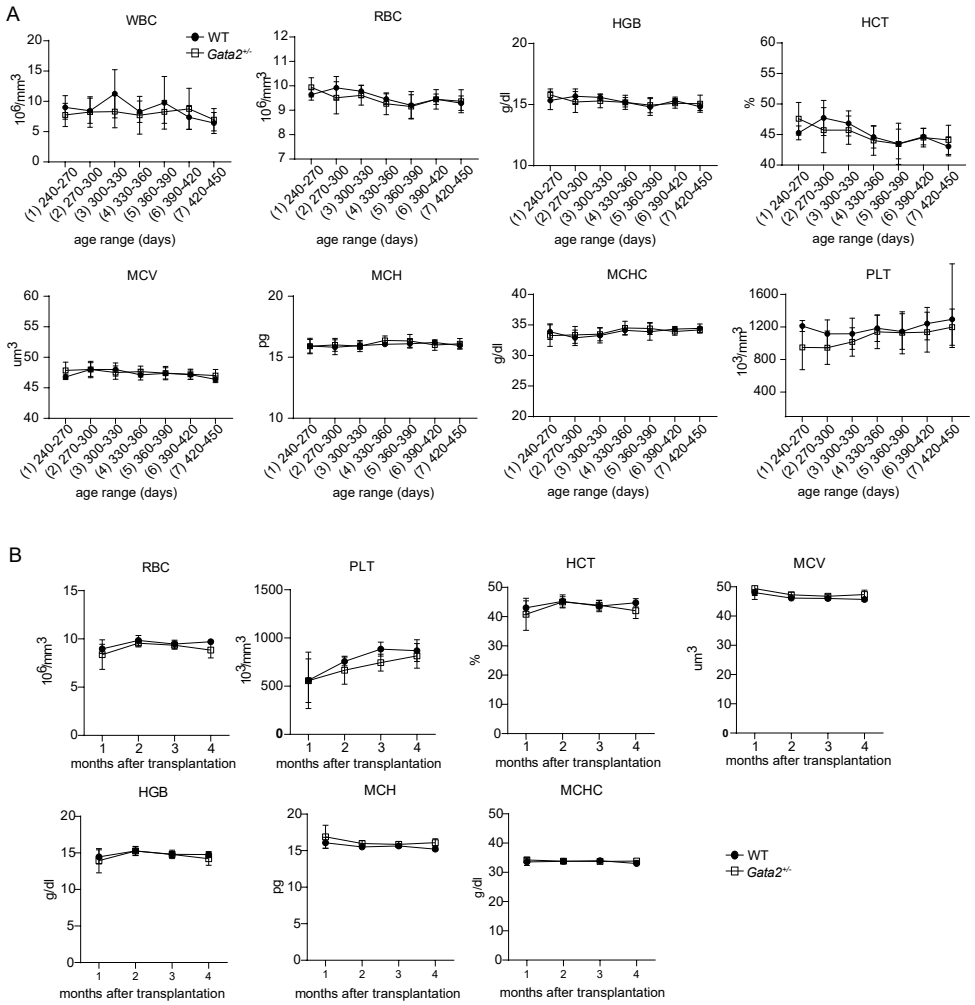
Zhang, Y., Rishi, A. K., Dawson, M. I., Tschang, R., Farhana, L., Boyanapalli, M., Reichert, U., Shroot, B., Van Buren, E. C., & Fontana, J. A. (2000). S-phase arrest and apoptosis induced in normal mammary epithelial cells by a novel retinoid. *Cancer Res*, *60*(7), 2025-2032.

Zimmermann, M., Murina, O., Reijns, M. A. M., Agathanggelou, A., Challis, R., Tarnauskaitė, Ž., Muir, M., Fluteau, A., Aregger, M., McEwan, A., Yuan, W., Clarke, M., Lambros, M. B., Paneesha, S., Moss, P., Chandrashekhar, M., Angers, S., Moffat, J., Brunton, V. G., Hart, T., de Bono, J., Stankovic, T., Jackson, A. P., & Durocher, D. (2018). CRISPR screens identify genomic ribonucleotides as a source of PARP-trapping lesions. *Nature*, *559*(7713), 285-289.

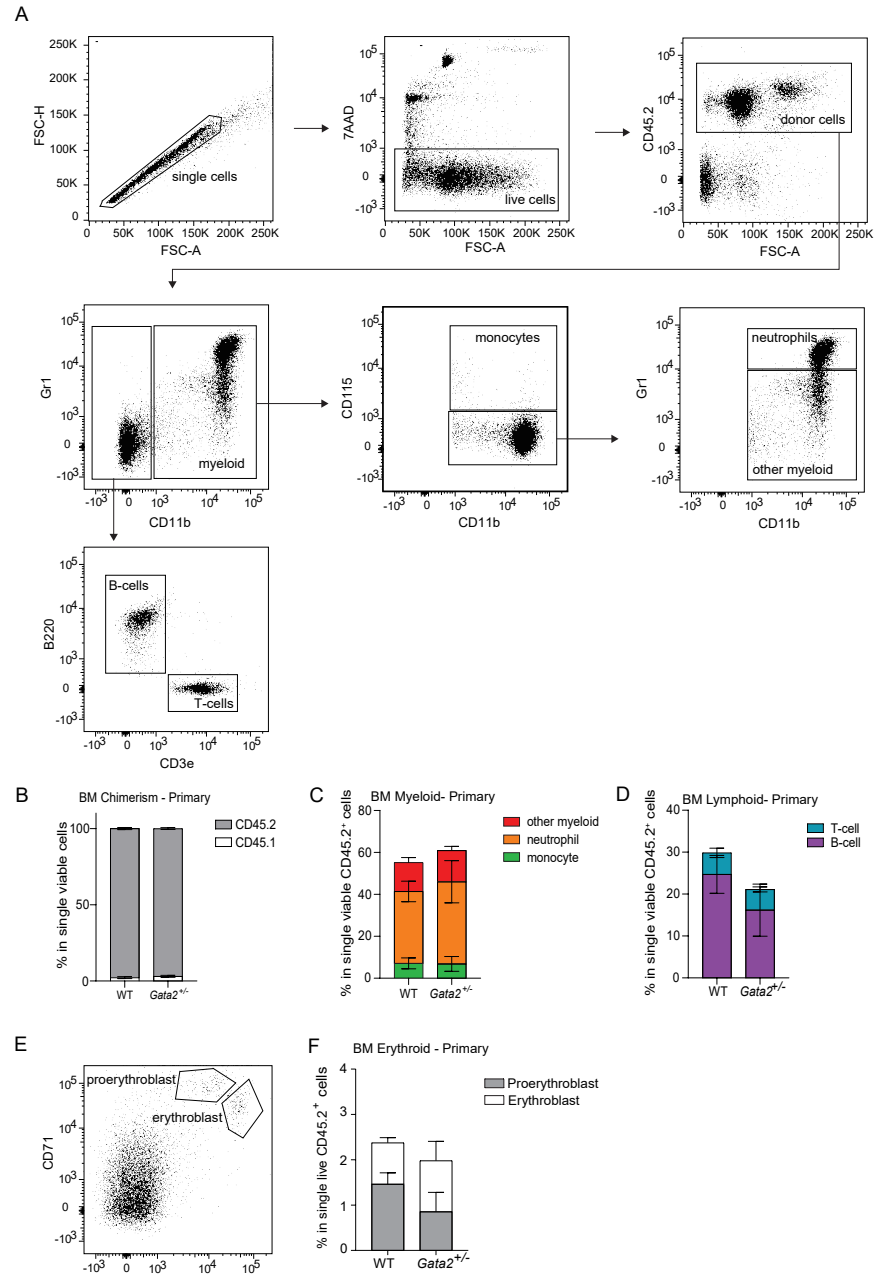
SUPPLEMENTARY INFORMATION



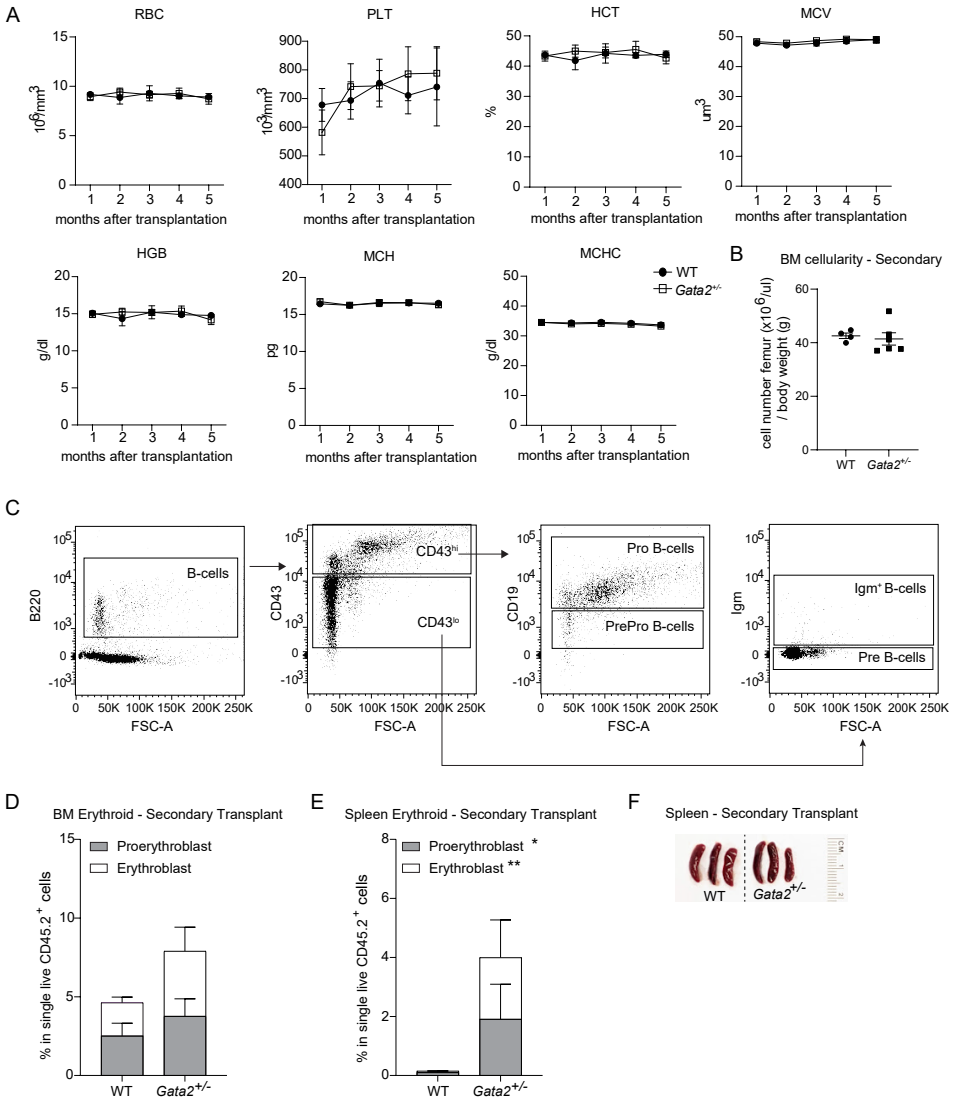
Supplementary figure 1. *Gata2*$^{+/-}$ mice have normal myeloid and erythroid differentiation.
A) Quantification of the number of BFU-E, CFU-G, CFU-M and CFU-GM and CFU-GMEM colonies obtained from CFU-C assay of adult WT and *Gata2*$^{+/-}$ MPPs and B) from serial re-plating experiments. C) PB analysis of WT and *Gata2*$^{+/-}$ mice.



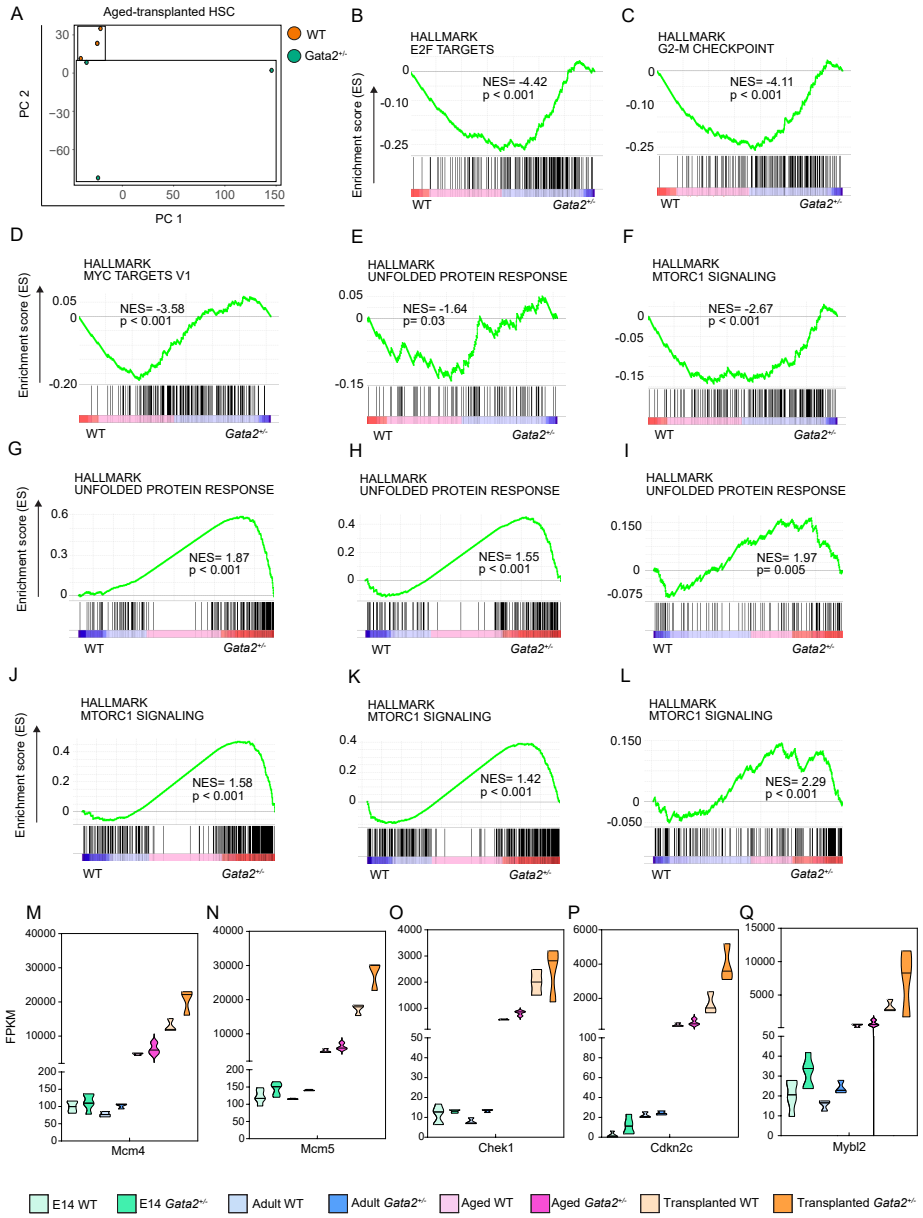
Supplementary figure 2. Peripheral blood values of *Gata2*$^{+/-}$ mice during aging and after transplantation.
A) PB analysis of WT and *Gata2*$^{+/-}$ mice during aging and B) after primary transplantation.



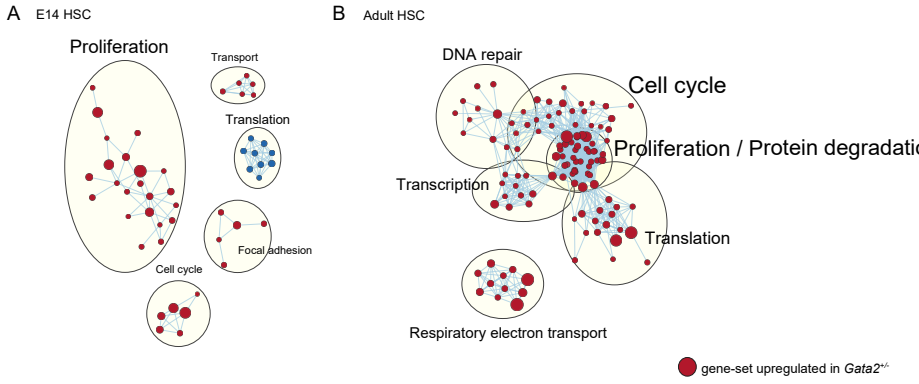
Supplementary figure 3. Bone marrow lineage differentiation in aged-*Gata2*$^{+/-}$ mice after primary transplantation.
A) Gating strategy for myeloid and lymphoid cells in PB and BM. B) Proportion of CD45.2-expressing donor cells and CD45.1-expressing recipient cells aged-WT and aged-*Gata2*$^{+/-}$ in the BM of primary recipients. C) Quantification of the number of donor myeloid and D) lymphoid cells in the BM of aged-WT and aged-*Gata2*$^{+/-}$ mice after primary transplantation. E) Gating strategy for proerythroblast and erythroblast cells in the BM. F) Proportion of proerythroblast and erythroblast cells in the BM aged-WT and aged-*Gata2*$^{+/-}$ mice after primary transplantation



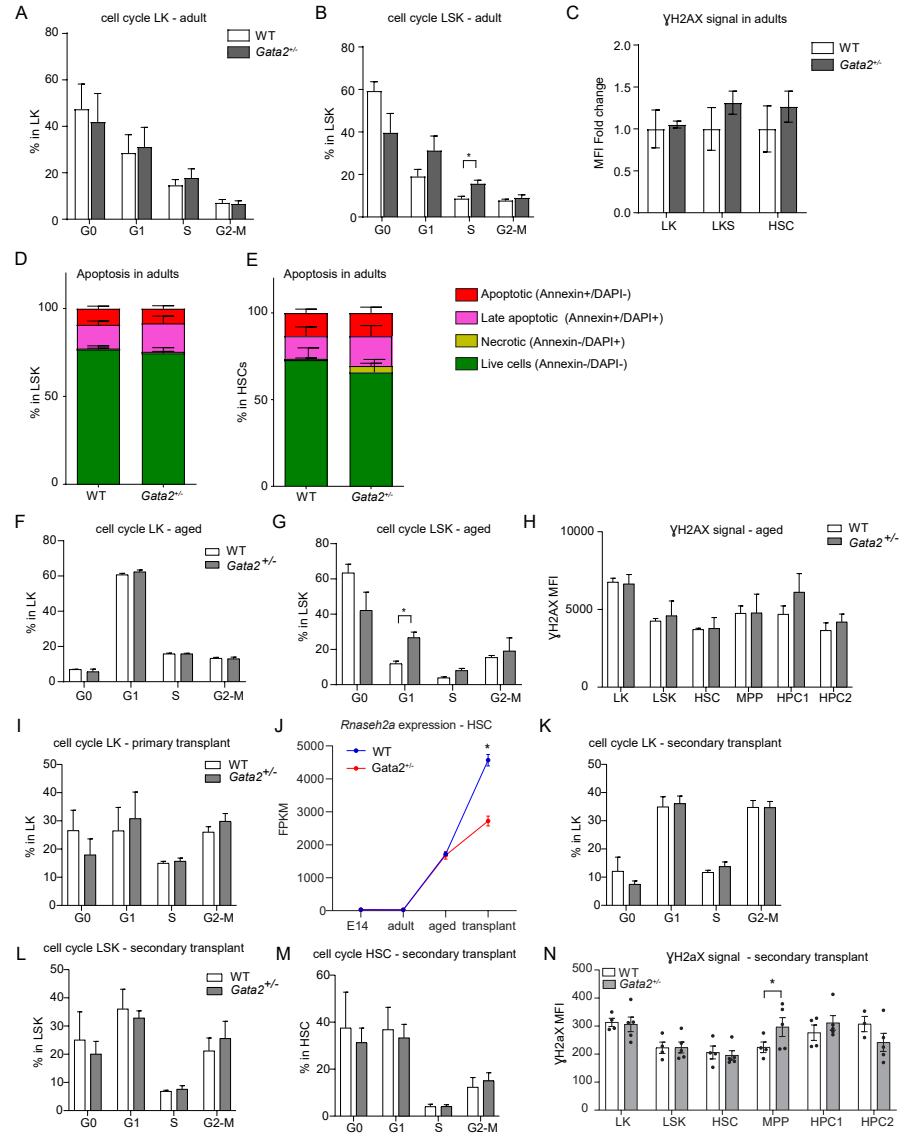
Supplementary figure 4. Aged-*Gata2*$^{+/-}$ mice have increased splenic erythropoiesis but normal spleen size after secondary transplantation.
A) PB analysis of aged-WT and aged-*Gata2*$^{+/-}$ mice after secondary transplantation. B) Quantification of BM cellularity in aged-WT and aged-*Gata2*$^{+/-}$ BM after secondary transplantation. C) Gating strategy for B-cell differentiation trajectory in the BM. D) Proportion of proerythroblast and erythroblast cells in the BM and E) spleen of aged-WT and aged-*Gata2*$^{+/-}$ mice after secondary transplantation. F) spleens obtained upon sacrificing the aged-WT and aged-*Gata2*$^{+/-}$ mice after secondary transplantation. *P < 0.05, **P < 0.01.



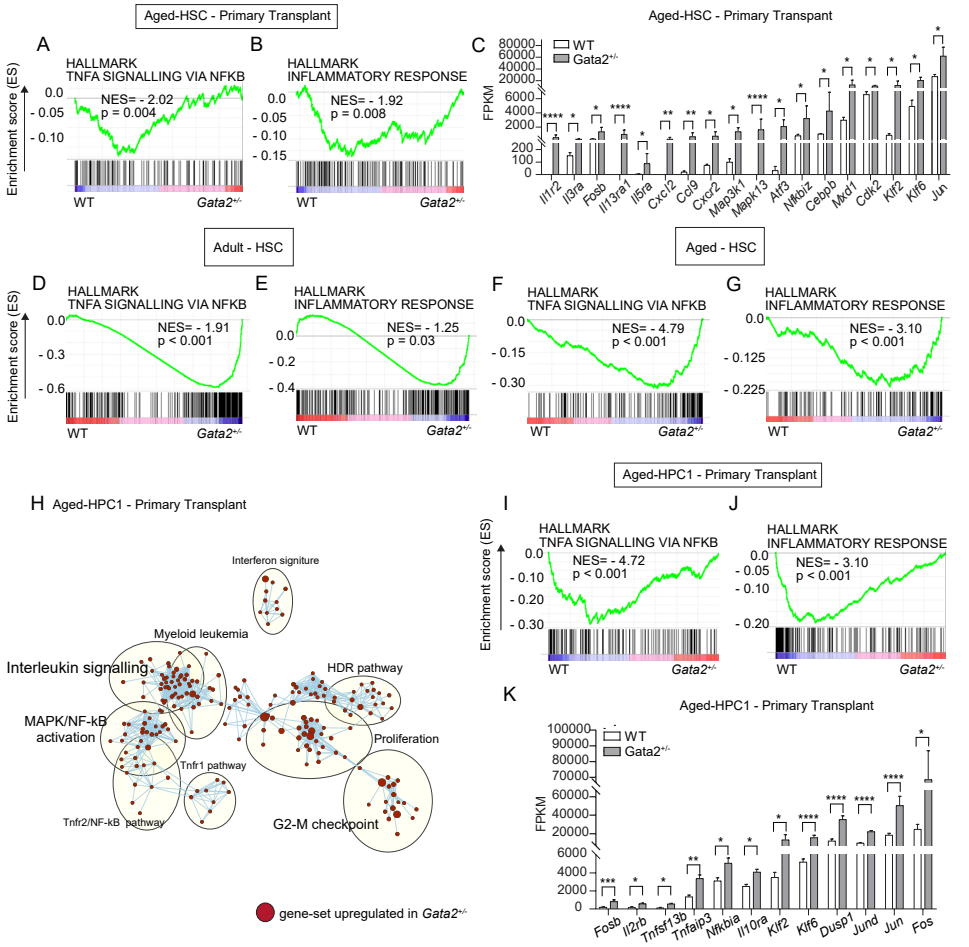
Supplementary figure 5. Unfolded protein response and MTORC1 signaling pathways are upregulated in *Gata2*$^{+/-}$ HSCs at E14, during aging and after transplantation.
A) PCA of aged-WT (orange) and aged-*Gata2*$^{+/-}$ (green) HSCs after primary transplantation. B) Hallmark GSEA of *E2f* targets, C) G_2-M checkpoint, D) *Myc* targets, E) Unfolded protein response and F) MTORC1 signaling in aged-WT and aged-*Gata2*$^{+/-}$ HSCs after primary transplantation. G) Hallmark GSEA of Unfolded protein response in E14 H) Adult and I) Aged WT and *Gata2*$^{+/-}$ HSCs. J) Hallmark GSEA of MTORC1 signaling in E14 K) Adult and L) Aged WT and *Gata2*$^{+/-}$ HSCs. M) Expression level (FPKM) of *Mcm4,* N) *Mcm5,* O) *Chek1,* P) *Cdkn2c* and Q) *Mybl2* compared between WT and *Gata2*$^{+/-}$ HSCs at timepoints E14, adult, aged and transplanted-aged.



Supplementary figure 6. Network analysis of adult- and aged-*Gata2*$^{+/-}$ HSCs.
A) Network analysis comparing WT and *Gata2*$^{+/-}$ HSCs at E14 and B) at the adult stage. Red dots show upregulated and blue dots show downregulated gene-sets in *Gata2*$^{+/-}$ HSCs compared to WT.



Supplementary figure 7. Cell cycle in adult- and aged-*Gata2*$^{+/-}$ HSCs.
A) Proportions of individual cell cycle stages compared between adult WT and *Gata2*$^{+/-}$ LK and B) LSK cells. C) γH2AX signals in adult WT and *Gata2*$^{+/-}$ LK, LSK and HSC populations. γH2AX MFI in *Gata2*$^{+/-}$ cells was calculated as fold-change. D) Proportion of live, apoptotic, late apoptotic and necrotic cells within LSK and E) HSC compartments of adult WT and *Gata2*$^{+/-}$ BM. F) Proportions of individual cell cycle stages compared between aged-WT and aged-*Gata2*$^{+/-}$ LK and G) LSK cells. H) Quantification of γH2AX signals in aged-WT and aged-*Gata2*$^{+/-}$ LK, LSK, HSC, MPP, HPC1 and HPC2 compartments of the BM. I) Proportions of individual cell cycle stages compared between aged-WT and aged-*Gata2*$^{+/-}$ LK population after primary transplantation. J) Expression level (FPKM) of *Rnaseh2a* compared between WT and *Gata2*$^{+/-}$ HSCs at timepoints E14, adult, aged and transplanted-aged. K) Proportions of individual cell cycle stages compared between aged-WT and aged-*Gata2*$^{+/-}$ LK cells, L) LSK cells and M) HSCs after secondary transplantation. N) Quantification of γH2AX signals in aged-WT and aged-*Gata2*$^{+/-}$ LK, LSK, HSC, MPP, HPC1 and HPC2 compartments of the BM after secondary transplantation. *P < 0.05.



Supplementary figure 8. Inflammatory signatures were upregulated in aged-*Gata2*$^{+/-}$ HSPCs upon transplantation. A) Hallmark GSEA of TNFa signaling via NFkB and B) Inflammatory response in aged-WT and aged-*Gata2*$^{+/-}$ HSCs after primary transplantation. C) Expression level (FPKM) of *Il1r2*, *Il3ra*, *Il13ra1*, *Il5ra*, *Fosb*, *Atf3*, *Cxcl2*, *Cxcr2*, *Ccl9*, *Map3k1*, *Mapk13*, *Atf3*, *Nfkbiz*, *Cebpb*, *Mxd1*, *Cdk2*, *Klf2*, *Klf6* and *Jun* compared between aged-WT and aged-*Gata2*$^{+/-}$ HSCs after primary transplantation. D) Hallmark GSEA of TNFa signaling via NFkB and E) Inflammatory response in E14 WT and *Gata2*$^{+/-}$ HSCs. F) Hallmark GSEA of TNFa signaling via NFkB and G) Inflammatory response in adult WT and *Gata2*$^{+/-}$ HSCs. H) Network analysis comparing aged-WT and aged-*Gata2*$^{+/-}$ HPC1 cells after primary transplantation. I) Hallmark GSEA of TNFa signaling via NFkB and J) Inflammatory response in aged-WT and aged-*Gata2*$^{+/-}$ HPC1 cells after primary transplantation. K) Expression level (FPKM) of *Tnfsf13b*, *Tnfaip3*, *Nfkbia*, *Fos*, *Fosb*, *Jun*, *Jund*, *Dusp1*, *Klf2*, *Klf6*, *Il2rb* and *Il10ra* compared between aged-WT and aged-*Gata2*$^{+/-}$ HPC1 cells after primary transplantation. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

Supplementary Table. Analysis of bone marrow smears after secondary transplantation.

	WT 1	WT 2	WT 3	WT 4		*Gata2*$^{+/-}$ 1	*Gata2*$^{+/-}$ 2	*Gata2*$^{+/-}$ 3	*Gata2*$^{+/-}$ 4	*Gata2*$^{+/-}$ 5
	n=500	n=500	n=500	n=500		n=500	n=500	n=500	n=200	n=200
cellularity	normal	normal	normal	normal		normal/increased	normal	normal	decreased	decreased
erytropoiesis	all stages of maturation	all stages of maturation	all stages of maturation	all stages of maturation		all stages of maturation	all stages of maturation	all stages of maturation	all stages of maturation	all stages of maturation
erythroid dsyplasia	none	nuclear budding < 1%	nuclear budding < 1%	nuclear budding: 6%		nuclear budding < 1%	nuclear budding: 4%	nuclear budding < 1%	nuclear budding < 1%	nuclear budding < 1%
granulopoiesis	all stages of maturation	all stages of maturation	all stages of maturation	all stages of maturation		all stages of maturation	all stages of maturation	all stages of maturation	all stages of maturation	all stages of maturation
myeloid dysplasia	none	none	none	none		none	none	none	none	none
macrophages	+	< +	< +	< +			< +		< +	< +
other			hemophagocytosis < +							
Differentiation										
blasts	1.4	1.4	1	0.6		2	0.4	0.6	2	2
promyelocytes	2.4	1.8	2.4	1		2.8	2.8	1.8	5.5	1.5
myelocytes	6	4.4	4.2	3.4		5	7.2	4.4	6.5	4.5
metamyelocytes	3.4	4.4	4.4	2.4		2.8	3	3	6	4
band forms	12.4	15.2	12.6	3.4		8	12	11.2	16.5	11
segmented granulocytes	40.8	32.8	24	12.4		21.2	28.4	22.8	24	45
eosinophils	2.2	1.8	0.6	0.8		0.2	2.2	0.6	1.5	1
basophils										
monocytes	2.8	3	3.8	1		0.8	3.4	1.2	4.5	4
lymphocytes	13.8	14	23	12		10	6.4	9.2	6	6.5
plasma cells	0.2	present, but scarce	0.6	2.6		0.6	0.2	0.4	1	0.5
erythroblasts	14.6	21.2	23.4	60.4		46.6	34	44.8	26.5	20
total	100	100	100	100		100	100	100	100	100
Megakaryocytes	+	++	<+	++		+++	+++	++	+	<+
dysplasia megakaryocytes				one nuclear lobe <+						
mast cells	none	none	none	<+		none	none	+++		<+

7

From basic biology to patient mutational spectra of GATA2 haploinsufficiencies: What are the mechanisms, hurdles and prospects of genome editing for treatment

Cansu Koyunlar1 and Emma de Pater1*

[1] Department of Hematology, Erasmus MC, Rotterdam, the Netherlands.
*Correspondence

Front. Genome Ed. (2020) 2:602182.

ABSTRACT

Inherited bone marrow failure syndromes (IBMFS) are monogenetic disorders that result in a reduction of mature blood cell formation and predisposition to leukemia. In children with myeloid leukemia the gene most often mutated is *Gata binding protein 2* (*GATA2*) and 80% of patients with *GATA2* mutations develop myeloid malignancy before the age of forty. Although GATA2 is established as one of the key regulators of embryonic and adult hematopoiesis, the mechanisms behind the leukemia predisposition in GATA2 haploinsufficiencies is ambiguous. The only curative treatment option currently available is allogeneic hematopoietic stem cell transplantation (allo-SCT). However, allo-SCT can only be applied at a relatively late stage of the disease as its applicability is compromised by treatment related morbidity and mortality (TRM). Alternatively, autologous-SCT (auto-SCT), which is associated with significantly less TRM, might become a treatment option if repaired hematopoietic stem cells would be available. Here we discuss the recent literature on leukemia predisposition syndromes caused by *GATA2* mutations, current knowledge on the function of *GATA2* in the hematopoietic system and advantages and pitfalls of potential treatment options provided by genome editing.

INTRODUCTION

IBMFS are a heterogeneous cluster of disorders manifested by an ineffective blood production and concurrent cytopenias that eventually result in a hypoplastic bone marrow. These syndromes constitute an increased propensity to develop hematological malignancies such as myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) (Dokal and Vulliamy, 2010; Wilson et al., 2014; Cook, 2018). Mutations in *GATA2* are the most common genetic defects in pediatric MDS (Spinner et al., 2014). GATA2 is one of the master regulators of blood production and patients that carry a mutation in one of the two alleles of *GATA2* often manifest with immunodeficiency syndromes and increased lifetime risk for MDS/AML (Wlodarski et al., 2016; Donadieu et al., 2018; McReynolds et al., 2018). Once malignant transformation becomes overt, survival rates are below 50% (Spinner et al., 2014). Due to the inherited mutation, allo-SCT is the only curative treatment option for these patients (Simonis et al., 2018; van Lier et al., 2020). Unfortunately, the use of allo-SCT is compromised by TRM and not applicable for patients who have not progressed to leukemia yet. Uncovering the *modus operandi* of GATA2 and other (epi)genetic factors in the complex network of blood regulation is essential to design noninvasive and preventive treatment options for IBMFS patients.

Genome editing strategies, especially the implementation of clustered regularly interspaced short palindromic repeat/associated protein 9 (CRISPR/Cas9) nuclease platforms, improve rapidly and progress towards efficient therapies for several genetic diseases (Cong et al., 2013; Mali et al., 2013; Anzalone et al., 2019). In this review, we will summarize clinical symptoms of GATA2 haploinsufficiency patients and results from Gata2 experimental models to inspect the function of GATA2 in leukemogenesis. Our aim is to explore the potential and pitfalls of genome editing methods to treat GATA2 deficiency syndromes in the light of current technologies.

The transcription factor GATA2

GATA2 is a zinc finger transcription factor that contains 2 first exons; a hematopoietic and neuronal cell specific distal first exon and a proximal first exon that is utilized ubiquitously. These two transcript variants encode the same protein (Minegishi et al., 1998; Pan et al., 2000). GATA2 binds a highly conserved (A/T)GATA(A/G) DNA sequence and other protein partners through two multifunctional zinc finger (ZF) domains; ZF1 and ZF2 that are encoded by exon 4 and exon 5 respectively (Evans and Felsenfeld, 1989; Alfayez et al., 2019). Two GATA2 protein isoforms can be formed, one lacking exon 5 and consequently lacking the ZF2 domain (Vicente et al., 2012)(Figure 1). To date, the functional consequence of this remains unclear.

Germline GATA2 mutations

In 2011, four different studies described germline heterozygous *GATA2* mutations in a total of 44 patients with various syndromes; monocytopenia and mycobacterial infection (MonoMAC) syndrome (Hsu et al., 2011), monocyte, B cell, NK cell and dendritic cell deficiencies (DCML)(Dickinson et al., 2011), Emberger Syndrome, which is characterized by primary lymphedema with a predisposition to AML (Ostergaard et al., 2011) and familial MDS/AML predisposition (Hahn et al., 2011). Distinct clinical perspectives discerned in these studies coalesce under the theme of the loss of one allele of *GATA2* resulting in the GATA2 haploinsufficiency syndrome, which can present with immunodeficiency, lymphedema and 80% predisposition to develop MDS/AML.

Taken together, 60% of patients present with a truncating mutation in *GATA2* before the ZF2 domain and 30% of patients present with a nonsynonymous mutation in ZF2. However, some patients develop MonoMAC syndrome without mutations in the coding region of *GATA2* but have reduced *GATA2* expression levels (Hsu et al., 2013). These patients harbor mutations in the intronic region, specifically in intron 4. Mutations in this region abrogate the function of a conserved +9.5 cis-element, that regulates *GATA2* transcription levels resulting in GATA2 haploinsufficiency (Hsu et al., 2013) and intron 4 mutations represent 10% of all GATA2 haploinsufficiency cases (Wlodarski et al., 2017)(Figure 1). *GATA*2 mutations are also present in a subset of patients with chronic neutropenia and aplastic anemia (AA) (Townsley et al., 2012; Pasquet et al., 2013). However, BM of AA patients with GATA2 mutations encompasses noticeably different types of altered hematopoietic populations than idiopathic AA patients, such as the complete loss of lymphoid progenitors and atypical megakaryocytes (Ganapathi et al., 2015).

Both familial and sporadic mutations in the coding and cis-regulatory regions of *GATA2* are found and are the underlying cause in 15% of advanced and 7% of all pediatric MDS cases (Wlodarski et al., 2016). Most of these mutations can be found in the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar). Currently, despite the improving definition of the phenotypic characteristics of GATA2 deficiency syndromes and high penetrance of myeloid malignancy, the mutational background and phenotypic outcome observed in these patients do not correlate, suggesting that additional events are important for disease progression (Collin et al., 2015; Wlodarski et al., 2016). Evidence for this is found in a cohort of pediatric MDS-GATA2 patients that acquired additional somatic mutations in *ASXL1, RUNX1, SETBP1, IKZF1* and *CRLF*2 genes, which resulted in an increased progression to AML. Furthermore, 72% of adolescents with MDS and monosomy 7 had an underlying *GATA2* mutation (Wlodarski et al., 2016; Fisher et al., 2017; Yoshida et al., 2020).

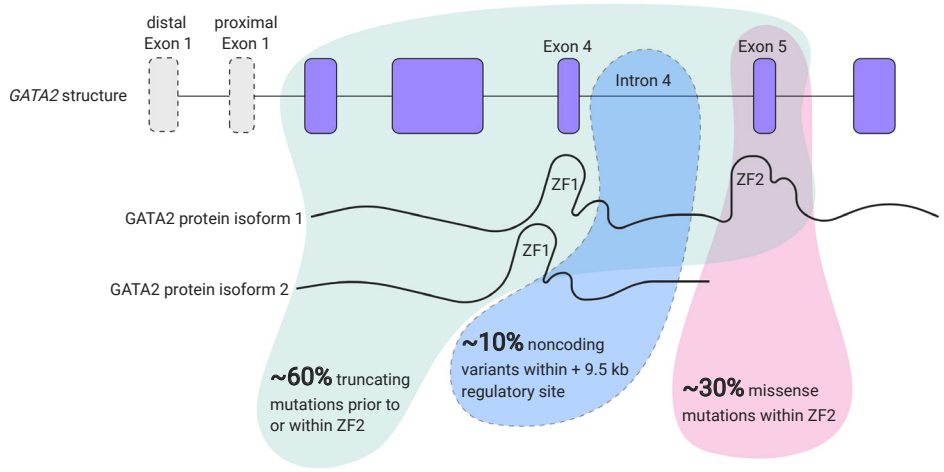


Figure 1. GATA2 locus organization and overview of mutation types found in GATA2 haploinsufficiency patients.

Somatic *GATA*2 mutations

Although truncating germline *GATA*2 mutations occur most often, a few somatic mutations are reported that phenocopy germline loss-of-function mutations (Sekhar et al., 2018; Alfayez et al., 2019). These cause a relatively milder form of the immunodeficiency phenotype observed in germline mutant *GATA2* patients, along with a common presentation of AML, atypical chronic myeloid leukemia and in some cases acute erythroid leukemia (Ping et al., 2017; Sekhar et al., 2018; Alfayez et al., 2019).

Somatic *GATA2* mutations are found both in ZF1 and ZF2 and all patients with somatic *GATA2* mutations harbor mutations in other genes, predominantly *CEBPA* with an incidence of 18-21% (Fasan et al., 2013; Hou et al., 2015; Theis et al., 2016). In one cohort of AML patients, ZF1 but not ZF2 mutations in *GATA2* closely associate with biallelic *CEBPA* mutations (Tien et al., 2018). This implies that ZF1 is crucial for GATA2 function in disease progression in combination with *CEBPA* mutations.

The function of the transcription factor Gata2 in mammalian hematopoiesis

The function of Gata2 in embryonic hematopoiesis

In mouse, homozygous deletion of Gata2 results in 67% lethality at E10.5 and none survive beyond E11.5, due to severe anemia. Chimeras of WT and *Gata2*$^{-/-}$ embryonic stem (ES) cells show that *Gata2*-null cells cannot contribute to hematopoiesis in adult blood, fetal liver, BM and thymus revealing a requirement for Gata2 in embryonic hematopoiesis (Tsai et al., 1994). Besides the embryonic lethality of *Gata2*-null embryos, the number and function of hematopoietic stem and progenitor cells from germline heterozygous *Gata2* mutant mice at E10.5-E12 is impaired (Ling et al., 2004).

Both in human and mouse embryos, *Gata2* is expressed in a specialized endothelial cell population called hemogenic endothelium (HE) and in the first transplantable HSCs that differentiate from HE (Marshall et al., 1999; Yokomizo and Dzierzak, 2010; Eich et al., 2018; Vink et al., 2020). Conditional deletion of *Gata2* in HE cells resulted in reduced hematopoietic cluster formation in the embryo and long-term repopulating HSCs were not formed. Conditional deletion of *Gata2* in HSCs induced apoptosis indicating that GATA2 is required both for HSC generation and maintenance (de Pater et al., 2013).

Gata2 expression is regulated by the enhancer activity of multiple conserved cis-regulatory elements. The disruption of the +9.5 element of *Gata2* impaired vascular integrity and formation of HSCs from HE in the mouse embryo (Lim et al., 2012; Gao et al., 2013).

Although both number and functionality of HSCs were reduced in embryonic Gata2 haploinsufficiency, it is yet to be discovered whether and how the propensity for MDS/AML observed in GATA2 haploinsufficiency patients is influenced by these early embryonic functions.

The function of Gata2 in adult bone marrow hematopoiesis

The function of GATA2 in adult hematopoiesis is still abstruse. In BM, *Gata2* is highly expressed in HSCs and downregulated during lineage commitment (Akashi et al., 2000; Miyamoto et al., 2002; Guo et al., 2013). HSCs in the BM of *Gata2*$^{+/-}$ mice are impaired in number and functionality as shown by serial transplantation assays (Rodrigues et al., 2005; Guo et al., 2013). In addition, *Gata2*-heterozygosity in BM HSCs is associated with a decreased proliferation ability together with increased quiescence and apoptosis (Ling et al., 2004; Rodrigues et al., 2005). Moreover, Gata2 haploinsufficiency reduces the function of granulocyte-macrophage progenitors but not of other myeloid committed progenitors (Rodrigues et al., 2008). However, *Gata2*$^{+/-}$ mice do not develop MDS/AML. This makes it difficult to study the contribution of GATA2 haploinsufficiency to leukemic progression in these models.

On the other hand, *Gata2* overexpression results in the self-renewal of myeloid progenitors and blocks lymphoid differentiation in mouse BM (Nandakumar et al., 2015). In addition, overexpression of *GATA2* in human ES cells (hESC) promotes proliferation in hESCs, but quiescence in hESC-derived HSCs (Zhou et al., 2019). Furthermore, increased *GATA2* expression is also observed in adult and pediatric AML patients with poor prognosis (Ayala et al., 2009; Luesink et al., 2012; Vicente et al., 2012; Menendez-Gonzalez et al., 2019). These findings indicate that, next to its tumor suppressor role, GATA2 might act as an oncogene when overexpressed.

Genome editing: a cure for GATA2 haploinsufficiencies?

GATA2 repair strategies

Allo-SCT is a powerful approach to treat malignancies in GATA2 haploinsufficiency patients (Simonis et al., 2018; van Lier et al., 2020). However, finding a matched donor and TRM compromises the use of allo-SCT and is therefore not suitable before the onset of malignancy (Bogaert et al., 2020). Regulation of *GATA2* expression is crucial in HSCs and in leukemia predisposition. This makes overexpression of WT *GATA2* using lenti-viral transgenic approaches not suitable as gene therapy method. An auto-SCT approach, after *ex vivo* correction of the underlying patient specific *GATA*2 mutation by genome editing tools is possibly a more effective treatment option for these patients (Figure 2).

Genome editing, since it was pioneered in the previous century, is developing meteorically as a revolutionary therapeutic tool for genetic defects, including hematological disorders (Xie et al., 2014; Hoban et al., 2016; De Ravin et al., 2017; Orkin and Bauer, 2019) CRISPR/Cas9, a part of the bacterial acquired immune system, was adapted as a breakthrough genome engineering technology and has since been extensively used to engineer eukaryotic cells in basic research and holds great potential for gene therapy (Gasiunas et al., 2012; Jinek et al., 2012; Cong et al., 2013; Barrangou and Doudna, 2016). CRISPR/Cas9 mediated genome editing relies on sequence specific guide RNAs that assemble with Cas9 protein to create double strand breaks (DSBs) in the targeted sequence. DSBs activates cell intrinsic repair mechanisms if the cell is to undergo proliferation and repaired by one of two mechanisms: non-homologous end joining (NHEJ) in which random insertions/deletions (InDels) are introduced or homology-directed repair (HDR) which uses the other DNA strand as template to restore its original sequence. This system can be hijacked by providing an exogenous repair template containing any desired sequence. Because HDR is rare, a selection cassette can be inserted for positive selection of the desired repair (Doudna and Charpentier, 2014).

Despite *GATA2* mutations predominantly occur prior or within the ZF2, variety of mutations are found in patients rather than identical mutations (https://www.ncbi.nlm.nih.gov/clinvar). These mutations would need to be restored at the endogenous locus, requiring HDR as repair mechanism. Therefore, optimizing an editing strategy by using a large HDR donor template could provide treatment for a substantial group of GATA2 patients. An efficient method for gene correction in HSCs with CRISPR/Cas9 and large HDR donor delivered by rAAV6 (adeno-associated viral vectors of serotype 6) was used to correct a *HBB* gene mutation causing sickle cell disease and has potential to correct *GATA2* mutations in HSCs using the same strategy (Dever et al., 2016; DeWitt et al., 2016; Bak et al., 2018) (Figure 2).

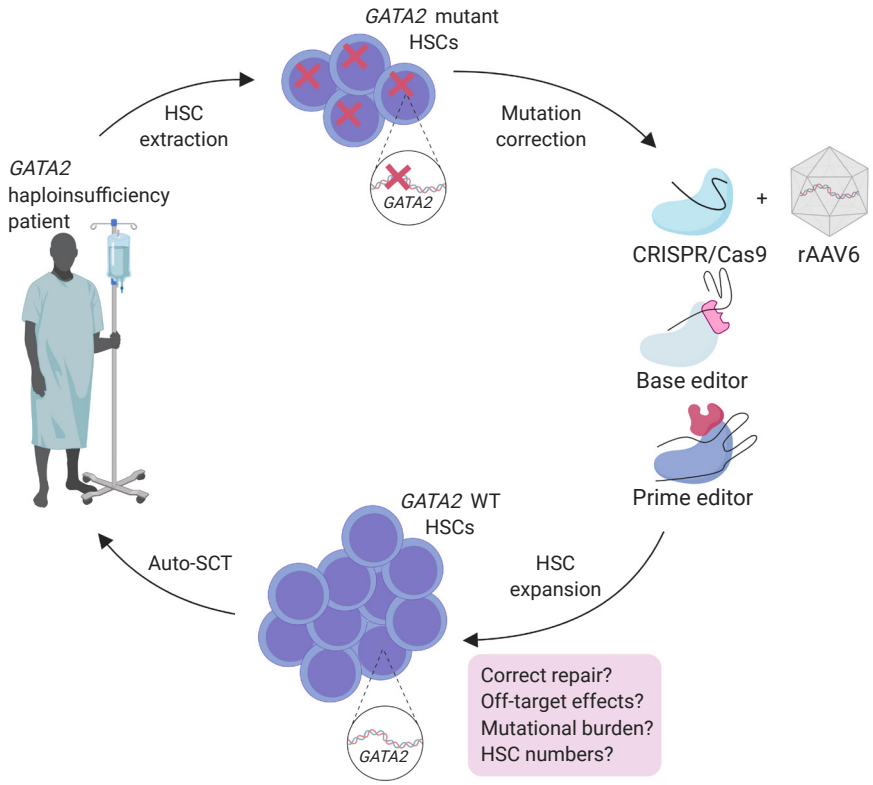


Figure 2. Treatment strategy for genome engineering of autologous HSC of GATA2 haploinsufficiency patients with safety considerations.

Hurdles

GATA2 haploinsufficiencies result in a diminished number of HSCs in both embryonic and adult stages. Additionally, HDR mediated repair works with low efficiencies and studies showed that it is more efficient in hematopoietic progenitor cells rather than long-term repopulating HSCs (Genovese et al., 2014; Hoban et al., 2015). Together this implicates the biggest hurdle to treat GATA2 haploinsufficiency patients would be to obtain sufficient number of corrected HSCs for auto-SCT. An enrichment method, possibly a reporter-based selection followed by an *ex vivo* expansion of GATA2-corrected HSCs, could potentially solve this problem. For this purpose, small molecule drugs promote *ex vivo* expansion of HSCs, like SR1 or UM171, could be used to obtain higher number of corrected HSCs prior to auto-SCT (Boitano et al., 2010; Fares et al., 2014).

Furthermore, in GATA2 haploinsufficiency patients, additional mutations in other genes could be the driver of leukemia which brings challenges to treat these patients by only correcting the mutant *GATA2* allele. Therefore, a preliminary genetic screening for additional mutations should be compulsory in GATA2 haploinsufficiency patients to

elucidate if correcting only the mutant *GATA2* allele would eliminate the disease phenotype of the patient.

Another hurdle when using genome editing tools for clinical applications is the off-target effects (OTEs) that might occur in undesired parts of the DNA. Detection of OTEs with whole genome sequencing are often challenging due to high background of random reads in combination with low sequence depth (<10 fold) (Kim et al., 2015). More screening strategies for OTEs, like GUIDE-seq (Genome wide, Unbiased Identification of DSB Enabled by sequencing), CIRCLE-seq (Circularization for In vitro Reporting of Cleavage Effects) and DISCOVER-seq (Discovery of In Situ Cas Off-targets and VERification by Sequencing), are shown to overcome these obstacles and could be used to efficiently identify OTEs that might result from *GATA2*-editing strategy before its clinical translation (Tsai et al., 2015; Tsai et al., 2017; Wienert et al., 2019).

Prospects

Fortunately, recent improvements of CRISPR/Cas9 genome editing may overcome some of these hurdles for patient applications. Base editing methods are developed by the addition of enzymes to Cas9 to provide single base pair changes without making DSBs (Komor et al., 2016; Gaudelli et al., 2017). Although base editing can correct point mutations that are also found in GATA2 patients, the off-target effects caused by the broad activity of cytidine deaminases used in this method should be considered carefully (Zuo et al., 2019; Yu et al., 2020). More recently Anzolone et al. described prime editing that introduces specific insertions, deletions and point mutations to a variety of genomic regions with high efficiency without DSBs. Prime editing was successfully used in human cells to correct mutations that cause sickle cell disease and Tay-Sachs disease and only 1-10% of prime-edited cells are found to have unwanted off-target InDels throughout the genome (Anzalone et al., 2019). These recent advances in genome editing techniques anticipate the improvement of a safer and more efficient correction of the patient mutations in HSCs prior to auto-SCT, therefore should be considered for the treatment of GATA2 haploinsufficiencies (Figure 2).

Currently, the minimum level of donor chimerism necessary to reverse the disease phenotype in GATA2 haploinsufficiency patients remains unclear (Hickstein, 2018). For sickle cell disease however, it was shown that clinical benefits might be observed when as few as 2-5 HSCs are engrafted (Walters et al., 2001; DeWitt et al., 2016). Interestingly, an asymptomatic germline GATA2 mutant individual acquired a somatic mutation reversing the harmful *GATA2* mutation. This resulted in a selective advantage of the corrected HSCs and prevented from developing malignancy (Catto et al., 2020). Together this implicates having a few mutation-corrected HSCs might already have clinical significance for GATA2 haploinsufficiency patients.

CRISPR/Cas9 technology has been approved in patient treatment for various type of malignancies including hematological diseases (https://clinicaltrials.gov). Currently, clinical

trials are performed where CRISPR/Cas9 is used to remove erythroid expression of the fetal hemoglobin repressor *BCL11A* in the treatment of hemoglobinopathies implicates a highly promising potential for genome editing to treat various hematological disorders (Orkin and Bauer, 2019; The Lancet, 2019).

Careful consideration of possible challenges discussed for GATA2 haploinsufficiency patients could lead to a beneficial clinical translation of genome editing to treat these patients in the near future.

DISCUSSION

Although GATA2 haploinsufficiency depletes the HSC compartment in humans and mice, the function of GATA2 haploinsufficiency in MDS/AML progression is poorly understood. A possibility could be that GATA2 haploinsufficiency provides a fertile ground for the emergence of additional mutations in HSCs and these acquired mutations promote leukemogenesis. Evidence that support this hypothesis is the inconsistent penetrance of leukemia in GATA2 haploinsufficiency patients that cannot be explained solely by the mutations in the *GATA2* locus and MDS/AML patients with germline *GATA2* mutation presented with additional mutations which are linked to hematological malignancies (Wlodarski et al., 2016; Fisher et al., 2017; Yoshida et al., 2020). In order to understand the concept of fertile ground as a driver of MDS/AML in GATA2 deficiency syndromes, more fundamental research is needed to reveal the clonal origin (embryonic and/or adult) of leukemogenic driver mutations to help us choose an appropriate time frame and strategy to treat these patients using genome editing. If leukemic driver mutations arise early during hematopoietic development, targeting leukemic clones will be challenging.

in vivo Gata2$^{+/-}$ models have not developed an MDS/AML phenotype (Ling et al., 2004; Rodrigues et al., 2005). This could be due to differences governing HSC mechanisms in these models or due to differences in lifespan, infection status, genetic background or a combination of these factors. Perhaps aged *Gata2*$^{+/-}$ models could provide more insight, since this would challenge the HSC compartment and increase the chances of additional events that would promote leukemogenesis to occur.

Base editing and prime editing are the recent promising and rigorous refinements of genome editing technologies which could provide and improve a patient specific mutation correction for *GATA*2 mutations or any other gene mutations that predispose to hematological malignancies when potentials and risks of these tools are tested sufficiently prior to the actual patient treatments. In addition to their potential for gene therapy discussed in this review, CRISPR base and prime editing technologies are also fantastic tools for basic research to introduce additional predicted leukemia driver mutations to HSCs in GATA2 haploinsufficiency models in order to identify their potential role in malignant transformation.

Author Contributions

CK and EP wrote the manuscript.

Funding

EP was supported by grant SK_10321 (Dutch Cancer foundation/Alpe d'Huzes).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

We thank Dr. IP Touw and Dr. T Cupedo for careful reading of the manuscript. Figures are created with BioRender.com.

REFERENCES

Akashi, K., T. Reya, D. Dalma-Weiszhausz and I. L. Weissman (2000). "Lymphoid precursors." Curr Opin Immunol **12**(2): 144-150.

Alfayez, M., S. A. Wang, S. A. Bannon, D. P. Kontoyiannis, S. M. Kornblau, J. S. Orange, E. M. Mace and C. D. DiNardo (2019). "Myeloid malignancies with somatic GATA2 mutations can be associated with an immunodeficiency phenotype." Leuk Lymphoma **60**(8): 2025-2033.

Anzalone, A. V., P. B. Randolph, J. R. Davis, A. A. Sousa, L. W. Koblan, J. M. Levy, P. J. Chen, C. Wilson, G. A. Newby, A. Raguram and D. R. Liu (2019). "Search-and-replace genome editing without double-strand breaks or donor DNA." Nature **576**(7785): 149-157.

Ayala, R. M., J. Martínez-López, E. Albízua, A. Diez and F. Gilsanz (2009). "Clinical significance of Gata-1, Gata-2, EKLF, and c-MPL expression in acute myeloid leukemia." Am J Hematol **84**(2): 79-86.

Bak, R. O., D. P. Dever and M. H. Porteus (2018). "CRISPR/Cas9 genome editing in human hematopoietic stem cells." Nat Protoc **13**(2): 358-376.

Barrangou, R. and J. A. Doudna (2016). "Applications of CRISPR technologies in research and beyond." Nat Biotechnol **34**(9): 933-941.

Bogaert, D. J., G. Laureys, L. Naesens, D. Mazure, M. De Bruyne, A. P. Hsu, V. Bordon, E. Wouters, S. J. Tavernier, B. N. Lambrecht, E. De Baere, F. Haerynck and T. Kerre (2020). "GATA2 deficiency and haematopoietic stem cell transplantation: challenges for the clinical practitioner." Br J Haematol **188**(5): 768-773.

Boitano, A. E., J. Wang, R. Romeo, L. C. Bouchez, A. E. Parker, S. E. Sutton, J. R. Walker, C. A. Flaveny, G. H. Perdew, M. S. Denison, P. G. Schultz and M. P. Cooke (2010). "Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells." Science **329**(5997): 1345-1348.

Catto, L. F. B., G. Borges, A. L. Pinto, D. V. Clé, F. Chahud, B. A. Santana, F. S. Donaires and R. T. Calado (2020). "Somatic genetic rescue in hematopoietic cells in GATA2 deficiency." Blood **136**(8): 1002-1005.

Collin, M., R. Dickinson and V. Bigley (2015). "Haematopoietic and immune defects associated with GATA2 mutation." Br J Haematol **169**(2): 173-187.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.

Cook, J. R. (2018). 5 - Bone Marrow Failure Syndromes. Hematopathology (Third Edition). E. D. Hsi. Philadelphia, Content Repository Only!**:** 167-183.e161.

de Pater, E., P. Kaimakis, C. S. Vink, T. Yokomizo, T. Yamada-Inagawa, R. van der Linden, P. S. Kartalaei, S. A. Camper, N. Speck and E. Dzierzak (2013). "Gata2 is required for HSC generation and survival." J Exp Med **210**(13): 2843-2850.

De Ravin, S. S., L. Li, X. Wu, U. Choi, C. Allen, S. Koontz, J. Lee, N. Theobald-Whiting, J. Chu, M. Garofalo, C. Sweeney, L. Kardava, S. Moir, A. Viley, P. Natarajan, L. Su, D. Kuhns, K. A. Zarember, M. V. Peshwa and H. L. Malech (2017). "CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-linked chronic granulomatous disease." Sci Transl Med **9**(372).

Dever, D. P., R. O. Bak, A. Reinisch, J. Camarena, G. Washington, C. E. Nicolas, M. Pavel-Dinu, N. Saxena, A. B. Wilkens, S. Mantri, N. Uchida, A. Hendel, A. Narla, R. Majeti, K. I. Weinberg and M. H. Porteus (2016). "CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells." Nature **539**(7629): 384-389.

DeWitt, M. A., W. Magis, N. L. Bray, T. Wang, J. R. Berman, F. Urbinati, S. J. Heo, T. Mitros, D. P. Muñoz, D. Boffelli, D. B. Kohn, M. C. Walters, D. Carroll, D. I. Martin and J. E. Corn (2016). "Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells." Sci Transl Med **8**(360): 360ra134.

Dickinson, R. E., H. Griffin, V. Bigley, L. N. Reynard, R. Hussain, M. Haniffa, J. H. Lakey, T. Rahman, X. N. Wang, N. McGovern, S. Pagan, S. Cookson, D. McDonald, I. Chua, J. Wallis, A. Cant, M. Wright, B. Keavney, P. F. Chinnery, J. Loughlin, S. Hambleton, M. Santibanez-Koref and M. Collin (2011). "Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency." Blood **118**(10): 2656-2658.

Dokal, I. and T. Vulliamy (2010). "Inherited bone marrow failure syndromes." Haematologica **95**(8): 1236-1240.

Donadieu, J., M. Lamant, C. Fieschi, F. S. de Fontbrune, A. Caye, M. Ouachee, B. Beaupain, J. Bustamante, H. A. Poirel, B. Isidor, E. Van Den Neste, A. Neel, S. Nimubona, F. Toutain, V. Barlogis, N. Schleinitz, T. Leblanc, P. Rohrlich, F. Suarez, D. Ranta, W. A. Chahla, B. Bruno, L. Terriou, S. Francois, B. Lioure, G. Ahle, F. Bachelerie, C. Preudhomme, E. Delabesse, H. Cave, C. Bellanné-Chantelot, M. Pasquet and G. s. g. French (2018). "Natural history of GATA2 deficiency in a survey of 79 French and Belgian patients." Haematologica **103**(8): 1278-1287.

Doudna, J. A. and E. Charpentier (2014). "The new frontier of genome engineering with CRISPR-Cas9." Science **346**(6213): 1258096.

Eich, C., J. Arlt, C. S. Vink, P. Solaimani Kartalaei, P. Kaimakis, S. A. Mariani, R. van der Linden, W. A. van Cappellen and E. Dzierzak (2018). "In vivo single cell analysis reveals Gata2 dynamics in cells transitioning to hematopoietic fate." J Exp Med **215**(1): 233-248.

Evans, T. and G. Felsenfeld (1989). "The erythroid-specific transcription factor Eryf1: a new finger protein." Cell **58**(5): 877-885.

Fares, I., J. Chagraoui, Y. Gareau, S. Gingras, R. Ruel, N. Mayotte, E. Csaszar, D. J. Knapp, P. Miller, M. Ngom, S. Imren, D. C. Roy, K. L. Watts, H. P. Kiem, R. Herrington, N. N. Iscove, R. K. Humphries, C. J. Eaves, S. Cohen, A. Marinier, P. W. Zandstra and G. Sauvageau (2014). "Cord blood expansion. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal." Science **345**(6203): 1509-1512.

Fasan, A., C. Eder, C. Haferlach, V. Grossmann, A. Kohlmann, F. Dicker, W. Kern, T. Haferlach and S. Schnittger (2013). "GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis." Leukemia **27**(2): 482-485.

Fisher, K. E., A. P. Hsu, C. L. Williams, H. Sayeed, B. Y. Merritt, M. T. Elghetany, S. M. Holland, A. A. Bertuch and M. M. Gramatges (2017). "Somatic mutations in children with GATA2-associated myelodysplastic syndrome who lack other features of GATA2 deficiency." Blood Adv **1**(7): 443-448.

Ganapathi, K. A., D. M. Townsley, A. P. Hsu, D. C. Arthur, C. S. Zerbe, J. Cuellar-Rodriguez, D. D. Hickstein, S. D. Rosenzweig, R. C. Braylan, N. S. Young, S. M. Holland and K. R. Calvo (2015). "GATA2 deficiency-associated bone marrow disorder differs from idiopathic aplastic anemia." Blood **125**(1): 56-70.

Gao, X., K. D. Johnson, Y. I. Chang, M. E. Boyer, C. N. Dewey, J. Zhang and E. H. Bresnick (2013). "Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo." J Exp Med **210**(13): 2833-2842.

Gasiunas, G., R. Barrangou, P. Horvath and V. Siksnys (2012). "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria." Proc Natl Acad Sci U S A **109**(39): E2579-2586.

Gaudelli, N. M., A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson and D. R. Liu (2017). "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage." Nature **551**(7681): 464-471.

Genovese, P., G. Schiroli, G. Escobar, T. D. Tomaso, C. Firrito, A. Calabria, D. Moi, R. Mazzieri, C. Bonini, M. C. Holmes, P. D. Gregory, M. van der Burg, B. Gentner, E. Montini, A. Lombardo and L. Naldini (2014). "Targeted genome editing in human repopulating haematopoietic stem cells." Nature **510**(7504): 235-240.

Guo, G., S. Luc, E. Marco, T. W. Lin, C. Peng, M. A. Kerenyi, S. Beyaz, W. Kim, J. Xu, P. P. Das, T. Neff, K. Zou, G. C. Yuan and S. H. Orkin (2013). "Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire." Cell Stem Cell **13**(4): 492-505.

Hahn, C. N., C. E. Chong, C. L. Carmichael, E. J. Wilkins, P. J. Brautigan, X. C. Li, M. Babic, M. Lin, A. Carmagnac, Y. K. Lee, C. H. Kok, L. Gagliardi, K. L. Friend, P. G. Ekert, C. M. Butcher, A. L. Brown, I. D. Lewis, L. B. To, A. E. Timms, J. Storek, S. Moore, M. Altree, R. Escher, P. G. Bardy, G. K. Suthers, R. J. D'Andrea, M. S. Horwitz and H. S. Scott (2011). "Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia." Nat Genet **43**(10): 1012-1017.

Hickstein, D. (2018). "HSCT for GATA2 deficiency across the pond." Blood **131**(12): 1272-1274.

Hoban, M. D., G. J. Cost, M. C. Mendel, Z. Romero, M. L. Kaufman, A. V. Joglekar, M. Ho, D. Lumaquin, D. Gray, G. R. Lill, A. R. Cooper, F. Urbinati, S. Senadheera, A. Zhu, P. Q. Liu, D. E. Paschon, L. Zhang, E. J. Rebar, A. Wilber, X. Wang, P. D. Gregory, M. C. Holmes, A. Reik, R. P. Hollis and D. B. Kohn (2015). "Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells." Blood **125**(17): 2597-2604.

Hoban, M. D., D. Lumaquin, C. Y. Kuo, Z. Romero, J. Long, M. Ho, C. S. Young, M. Mojadidi, S. Fitz-Gibbon, A. R. Cooper, G. R. Lill, F. Urbinati, B. Campo-Fernandez, C. F. Bjurstrom, M. Pellegrini, R. P. Hollis and D. B. Kohn (2016). "CRISPR/Cas9-Mediated Correction of the Sickle Mutation in Human CD34+ cells." Mol Ther **24**(9): 1561-1569.

Hou, H. A., Y. C. Lin, Y. Y. Kuo, W. C. Chou, C. C. Lin, C. Y. Liu, C. Y. Chen, L. I. Lin, M. H. Tseng, C. F. Huang, Y. C. Chiang, M. C. Liu, C. W. Liu, J. L. Tang, M. Yao, S. Y. Huang, B. S. Ko, S. C. Hsu, S. J. Wu, W. Tsay, Y. C. Chen and H. F. Tien (2015). "GATA2 mutations in patients with acute myeloid leukemia-paired samples analyses show that the mutation is unstable during disease evolution." Ann Hematol **94**(2): 211-221.

Hsu, A. P., K. D. Johnson, E. L. Falcone, R. Sanalkumar, L. Sanchez, D. D. Hickstein, J. Cuellar-Rodriguez, J. E. Lemieux, C. S. Zerbe, E. H. Bresnick and S. M. Holland (2013). "GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome." Blood **121**(19): 3830-3837, S3831-3837.

Hsu, A. P., E. P. Sampaio, J. Khan, K. R. Calvo, J. E. Lemieux, S. Y. Patel, D. M. Frucht, D. C. Vinh, R. D. Auth, A. F. Freeman, K. N. Olivier, G. Uzel, C. S. Zerbe, C. Spalding, S. Pittaluga, M. Raffeld, D. B. Kuhns, L. Ding, M. L. Paulson, B. E. Marciano, J. C. Gea-Banacloche, J. S. Orange, J. Cuellar-Rodriguez, D. D. Hickstein and S. M. Holland (2011). "Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome." Blood **118**(10): 2653-2655.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science **337**(6096): 816-821.

Kim, D., S. Bae, J. Park, E. Kim, S. Kim, H. R. Yu, J. Hwang, J. I. Kim and J. S. Kim (2015). "Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells." Nat Methods **12**(3): 237-243, 231 p following 243.

Komor, A. C., Y. B. Kim, M. S. Packer, J. A. Zuris and D. R. Liu (2016). "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage." Nature **533**(7603): 420-424.

Lim, K. C., T. Hosoya, W. Brandt, C. J. Ku, S. Hosoya-Ohmura, S. A. Camper, M. Yamamoto and J. D. Engel (2012). "Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning." J Clin Invest **122**(10): 3705-3717.

Ling, K. W., K. Ottersbach, J. P. van Hamburg, A. Oziemlak, F. Y. Tsai, S. H. Orkin, R. Ploemacher, R. W. Hendriks and E. Dzierzak (2004). "GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells." J Exp Med **200**(7): 871-882.

Luesink, M., I. H. Hollink, V. H. van der Velden, R. H. Knops, J. B. Boezeman, V. de Haas, J. Trka, A. Baruchel, D. Reinhardt, B. A. van der Reijden, M. M. van den Heuvel-Eibrink, C. M. Zwaan and J. H. Jansen (2012). "High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia." Blood **120**(10): 2064-2075.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013). "RNA-guided human genome engineering via Cas9." Science **339**(6121): 823-826.

Marshall, C. J., R. L. Moore, P. Thorogood, P. M. Brickell, C. Kinnon and A. J. Thrasher (1999). "Detailed characterization of the human aorta-gonad-mesonephros region reveals morphological polarity resembling a hematopoietic stromal layer." Dev Dyn **215**(2): 139-147.

McReynolds, L. J., K. R. Calvo and S. M. Holland (2018). "Germline GATA2 Mutation and Bone Marrow Failure." Hematol Oncol Clin North Am **32**(4): 713-728.

Menendez-Gonzalez, J. B., M. Vukovic, A. Abdelfattah, L. Saleh, A. Almotiri, L. A. Thomas, A. Agirre-Lizaso, A. Azevedo, A. C. Menezes, G. Tornillo, S. Edkins, K. Kong, P. Giles, F. Anjos-Afonso, A. Tonks, A. S. Boyd, K. R. Kranc and N. P. Rodrigues (2019). "Gata2 as a Crucial Regulator of Stem Cells in Adult Hematopoiesis and Acute Myeloid Leukemia." Stem Cell Reports **13**(2): 291-306.

Minegishi, N., J. Ohta, N. Suwabe, H. Nakauchi, H. Ishihara, N. Hayashi and M. Yamamoto (1998). "Alternative promoters regulate transcription of the mouse GATA-2 gene." J Biol Chem **273**(6): 3625-3634.

Miyamoto, T., H. Iwasaki, B. Reizis, M. Ye, T. Graf, I. L. Weissman and K. Akashi (2002). "Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment." Dev Cell **3**(1): 137-147.

Nandakumar, S. K., K. Johnson, S. L. Throm, T. I. Pestina, G. Neale and D. A. Persons (2015). "Low-level GATA2 overexpression promotes myeloid progenitor self-renewal and blocks lymphoid differentiation in mice." Exp Hematol **43**(7): 565-577 e561-510.

Orkin, S. H. and D. E. Bauer (2019). "Emerging Genetic Therapy for Sickle Cell Disease." Annu Rev Med **70**: 257-271.

Ostergaard, P., M. A. Simpson, F. C. Connell, C. G. Steward, G. Brice, W. J. Woollard, D. Dafou, T. Kilo, S. Smithson, P. Lunt, V. A. Murday, S. Hodgson, R. Keenan, D. T. Pilz, I. Martinez-Corral, T. Makinen, P. S. Mortimer, S. Jeffery, R. C. Trembath and S. Mansour (2011). "Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome)." Nat Genet **43**(10): 929-931.

Pan, X., N. Minegishi, H. Harigae, H. Yamagiwa, M. Minegishi, Y. Akine and M. Yamamoto (2000). "Identification of human GATA-2 gene distal IS exon and its expression in hematopoietic stem cell fractions." J Biochem **127**(1): 105-112.

Pasquet, M., C. Bellanné-Chantelot, S. Tavitian, N. Prade, B. Beaupain, O. Larochelle, A. Petit, P. Rohrlich, C. Ferrand, E. Van Den Neste, H. A. Poirel, T. Lamy, M. Ouachée-Chardin, V. Mansat-De Mas, J. Corre, C. Récher, G. Plat, F. Bachelerie, J. Donadieu and E. Delabesse (2013). "High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia." Blood **121**(5): 822-829.

Ping, N., A. Sun, Y. Song, Q. Wang, J. Yin, W. Cheng, Y. Xu, L. Wen, H. Yao, L. Ma, H. Qiu, C. Ruan, D. Wu and S. Chen (2017). "Exome sequencing identifies highly recurrent somatic GATA2 and CEBPA mutations in acute erythroid leukemia." Leukemia **31**(1): 195-202.

Rodrigues, N. P., A. S. Boyd, C. Fugazza, G. E. May, Y. Guo, A. J. Tipping, D. T. Scadden, P. Vyas and T. Enver (2008). "GATA-2 regulates granulocyte-macrophage progenitor cell function." Blood **112**(13): 4862-4873.

Rodrigues, N. P., V. Janzen, R. Forkert, D. M. Dombkowski, A. S. Boyd, S. H. Orkin, T. Enver, P. Vyas and D. T. Scadden (2005). "Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis." Blood **106**(2): 477-484.

Sekhar, M., R. Pocock, D. Lowe, C. Mitchell, T. Marafioti, R. Dickinson, M. Collin and M. Lipman (2018). "Can somatic GATA2 mutation mimic germ line GATA2 mutation?" Blood Adv **2**(8): 904-908.

Simonis, A., M. Fux, G. Nair, N. J. Mueller, E. Haralambieva, T. Pabst, J. Pachlopnik Schmid, A. Schmidt, U. Schanz, M. G. Manz and A. M. S. Müller (2018). "Allogeneic hematopoietic cell transplantation in patients with GATA2 deficiency-a case report and comprehensive review of the literature." Ann Hematol **97**(10): 1961-1973.

Spinner, M. A., L. A. Sanchez, A. P. Hsu, P. A. Shaw, C. S. Zerbe, K. R. Calvo, D. C. Arthur, W. Gu, C. M. Gould, C. C. Brewer, E. W. Cowen, A. F. Freeman, K. N. Olivier, G. Uzel, A. M. Zelazny, J. R. Daub, C. D. Spalding, R. J. Claypool, N. K. Giri, B. P. Alter, E. M. Mace, J. S. Orange, J. Cuellar-Rodriguez, D. D. Hickstein and S. M. Holland (2014). "GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity." Blood **123**(6): 809-821.

The Lancet, H. (2019). "CRISPR-Cas9 gene editing for patients with haemoglobinopathies." Lancet Haematol **6**(9): e438.

Theis, F., A. Corbacioglu, V. I. Gaidzik, P. Paschka, D. Weber, L. Bullinger, M. Heuser, A. Ganser, F. Thol, B. Schlegelberger, G. Göhring, C. H. Köhne, U. Germing, P. Brossart, H. A. Horst, D. Haase, K. Götze, M. Ringhoffer, W. Fiedler, D. Nachbaur, T. Kindler, G. Held, M. Lübbert, M. Wattad, H. R. Salih, J. Krauter, H. Döhner, R. F. Schlenk and K. Döhner (2016). "Clinical impact of GATA2 mutations in acute myeloid leukemia patients harboring CEBPA mutations: a study of the AML study group." Leukemia **30**(11): 2248-2250.

Tien, F. M., H. A. Hou, C. H. Tsai, J. L. Tang, Y. C. Chiu, C. Y. Chen, Y. Y. Kuo, M. H. Tseng, Y. L. Peng, M. C. Liu, C. W. Liu, X. W. Liao, L. I. Lin, C. T. Lin, S. J. Wu, B. S. Ko, S. C. Hsu, S. Y. Huang, M. Yao, W. C. Chou and H. F. Tien (2018). "GATA2 zinc finger 1 mutations are associated with distinct clinico-biological features and outcomes different from GATA2 zinc finger 2 mutations in adult acute myeloid leukemia." Blood Cancer J **8**(9): 87.

Townsley, D. M., A. Hsu, B. Dumitriu, S. M. Holland and N. S. Young (2012). "Regulatory Mutations in GATA2 Associated with Aplastic Anemia." Blood **120**(21): 3488-3488.

Tsai, F. Y., G. Keller, F. C. Kuo, M. Weiss, J. Chen, M. Rosenblatt, F. W. Alt and S. H. Orkin (1994). "An early haematopoietic defect in mice lacking the transcription factor GATA-2." Nature **371**(6494): 221-226.

Tsai, S. Q., N. T. Nguyen, J. Malagon-Lopez, V. V. Topkar, M. J. Aryee and J. K. Joung (2017). "CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets." Nat Methods **14**(6): 607-614.

Tsai, S. Q., Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee and J. K. Joung (2015). "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases." Nat Biotechnol **33**(2): 187-197.

van Lier, Y. F., G. J. de Bree, R. E. Jonkers, J. Roelofs, I. J. M. Ten Berge, C. E. Rutten, E. Nur, T. W. Kuijpers, M. D. Hazenberg and S. S. Zeerleder (2020). "Allogeneic hematopoietic cell transplantation in the management of GATA2 deficiency and pulmonary alveolar proteinosis." Clin Immunol **218**: 108522.

Vicente, C., A. Conchillo, M. A. García-Sánchez and M. D. Odero (2012). "The role of the GATA2 transcription factor in normal and malignant hematopoiesis." Crit Rev Oncol Hematol **82**(1): 1-17.

Vink, C. S., F. J. Calero-Nieto, X. Wang, A. Maglitto, S. A. Mariani, W. Jawaid, B. Göttgens and E. Dzierzak (2020). "Iterative Single-Cell Analyses Define the Transcriptome of the First Functional Hematopoietic Stem Cells." Cell Rep **31**(6): 107627.

Walters, M. C., M. Patience, W. Leisenring, Z. R. Rogers, V. M. Aquino, G. R. Buchanan, I. A. Roberts, A. M. Yeager, L. Hsu, T. Adamkiewicz, J. Kurtzberg, E. Vichinsky, B. Storer, R. Storb, K. M. Sullivan and D. Multicenter Investigation of Bone Marrow Transplantation for Sickle Cell (2001). "Stable mixed hematopoietic chimerism after bone marrow transplantation for sickle cell anemia." Biol Blood Marrow Transplant **7**(12): 665-673.

Wienert, B., S. K. Wyman, C. D. Richardson, C. D. Yeh, P. Akcakaya, M. J. Porritt, M. Morlock, J. T. Vu, K. R. Kazane, H. L. Watry, L. M. Judge, B. R. Conklin, M. Maresca and J. E. Corn (2019). "Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq." Science **364**(6437): 286-289.

Wilson, D. B., D. C. Link, P. J. Mason and M. Bessler (2014). "Inherited bone marrow failure syndromes in adolescents and young adults." Ann Med **46**(6): 353-363.

Wlodarski, M. W., M. Collin and M. S. Horwitz (2017). "GATA2 deficiency and related myeloid neoplasms." Semin Hematol **54**(2): 81-86.

Wlodarski, M. W., S. Hirabayashi, V. Pastor, J. Starý, H. Hasle, R. Masetti, M. Dworzak, M. Schmugge, M. van den Heuvel-Eibrink, M. Ussowicz, B. De Moerloose, A. Catala, O. P. Smith, P. Sedlacek, A. C. Lankester, M. Zecca, V. Bordon, S. Matthes-Martin, J. Abrahamsson, J. S. Kühl, K. W. Sykora, M. H. Albert, B. Przychodzien, J. P. Maciejewski, S. Schwarz, G. Göhring, B. Schlegelberger, A. Cseh, P. Noellke, A. Yoshimi, F. Locatelli, I.
Baumann, B. Strahm, C. M. Niemeyer and M. D. S. Ewog (2016). "Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents." Blood **127**(11): 1387-1397; quiz 1518.

Xie, F., L. Ye, J. C. Chang, A. I. Beyer, J. Wang, M. O. Muench and Y. W. Kan (2014). "Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac." Genome Res **24**(9): 1526-1533.

Yokomizo, T. and E. Dzierzak (2010). "Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos." Development **137**(21): 3651-3661.

Yoshida, M., K. Tanase-Nakao, H. Shima, R. Shirai, K. Yoshida, T. Osumi, T. Deguchi, M. Mori, Y. Arakawa, M. Takagi, T. Miyamura, K. Sakaguchi, H. Toyoda, H. Ishida, N. Sakata, T. Imamura, Y. Kawahara, A. Morimoto, T. Koike, H. Yagasaki, S. Ito, D. Tomizawa, N. Kiyokawa, S. Narumi and M. Kato (2020). "Prevalence of germline GATA2 and SAMD9/9L variants in paediatric haematological disorders with monosomy 7." Br J Haematol.

Yu, Y., T. C. Leete, D. A. Born, L. Young, L. A. Barrera, S. J. Lee, H. A. Rees, G. Ciaramella and N. M. Gaudelli (2020). "Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity." Nat Commun **11**(1): 2052.

Zhou, Y., Y. Zhang, B. Chen, Y. Dong, Y. Zhang, B. Mao, X. Pan, M. Lai, Y. Chen, G. Bian, Q. Zhou, T. Nakahata, J. Zhou, M. Wu and F. Ma (2019). "Overexpression of GATA2 Enhances Development and Maintenance of Human Embryonic Stem Cell-Derived Hematopoietic Stem Cell-like Progenitors." Stem Cell Reports **13**(1): 31-47.

Zuo, E., Y. Sun, W. Wei, T. Yuan, W. Ying, H. Sun, L. Yuan, L. M. Steinmetz, Y. Li and H. Yang (2019). "Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos." Science **364**(6437): 289-292.

8

General discussion

8.1. SUMMARY

Hematopoietic stem cells (HSCs) are the sole precursors of the hematopoietic system that can self-renew and differentiate into all mature blood cell types. The transcription factor (TF) *GATA2* is highly expressed in HSCs and regulates the HSC generation during embryonic development and HSC function throughout adulthood. In patients, heterozygous *GATA2* mutations cause GATA2 haploinsufficiency syndromes that present with a broad spectrum of phenotypes such as B-cell-, monocyte-, NK cell- and dendritic cell deficiencies, primary lymphedema, and more than 80% risk of developing myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). Although *GATA2* mutations evidently lead to hematopoietic defects in patients, the mechanisms underlying GATA2 haploinsufficiency phenotypes are relatively unknown. In this thesis, we examine the multifaceted role of *GATA2* throughout the formation and function of HSCs, the building blocks of the hematopoietic system, in both *Gata2*-mutant zebrafish and mouse models.

The first HSCs are generated through endothelial-to-hematopoietic transition (EHT) events during embryonic development. In mouse embryos, EHT events produce intra-aortic hematopoietic clusters (IAHCs), where HSCs are formed through a multistep maturation process characterized as pro-HSCs → pre-HSC1 → pre-HSC2/HSC. The maturation of HSCs through EHT is accompanied by progressive downregulation of endothelial transcriptional programming and upregulation of hematopoietic transcriptional programming. Although HSCs are severely reduced in *Gata2*$^{+/-}$ mouse embryos, the mechanism behind this reduction is incompletely understood. In **chapter 2**, we investigate the effect of Gata2 haploinsufficiency on the subpopulations of IAHCs and show that *Gata2*$^{+/-}$ embryos can produce pro-HSCs hence the hematopoietic programming is not abrogated. However, the maturation of pro-HSCs into pre-HSCs is significantly reduced in *Gata2*$^{+/-}$ embryos indicating *Gata2* is essential for the completion of EHT. To understand the mechanism hampering embryonic HSC maturation in *Gata2*$^{+/-}$, we look into the transcriptome of IAHC subpopulations and show that Gata2 haploinsufficiency reduces the expression of the endothelial repressor *Gfi1b* resulting in an incomplete repression of endothelial programming during EHT. Furthermore, we show that the ectopic expression of *gfi1b* in the hematopoietic cells rescues the number of embryonic HSCs in *gata2b*-deficient zebrafish suggesting that *Gata2* regulates *Gfi1b* and this regulation is crucial for the formation of HSCs during EHT.

Zebrafish have two orthologues of mammalian *Gata2*, i.e., *gata2a* and *gata2b*, that are sub-functionalized in the lymphoid-vascular system and hematopoietic system respectively, allowing us to investigate the role of *Gata2* in different cellular components. By generating homozygous *gata2b* knockout (**chapter 3**) and heterozygous *gata2b* knockout (**chapter 4**) zebrafish, we dissect the role of *Gata2* in the hematopoietic system without interfering with its role in the lymphoid-vascular system and circulation. In **chapter 3**, we show that homozygous deletion of *gata2b* (*gata2b*$^{-/-}$) attenuates the expansion of HSCs during

embryonic stages and impairs balanced lineage differentiation from the hematopoietic stem and progenitor cells (HSPCs) during adulthood. Using single-cell RNA sequencing (scRNA-seq) we show that the most immature HSPCs of *gata2b*$^{-/-}$ kidney marrow (KM) co-express myeloid and lymphoid specific genes resulting in reduced numbers of neutrophils and an incomplete B-cell differentiation. In **chapter 4**, we reveal that the heterozygous *gata2b* knockout (*gata2b*$^{+/-}$) HSPCs have normal lineage output; however, we observe erythro-myeloid dysplasia in the *gata2b*$^{+/-}$ KM. Furthermore, using scRNA-seq, we identify an aberrant *gata1a* expression in a subgroup of erythroid progenitors suggesting an impaired ‚GATA switch' process contributes to the erythroid dysplasia in *gata2b*$^{+/-}$ KM. In **chapter 5**, we investigate the function of a conserved enhancer region located in the 4th intron (i4) of *gata2a* that corresponds to the +9.5 enhancer region of mammalian *Gata2* locus. Complete deletion of this *gata2a* enhancer locus (*gata2a*$^{i4/i4}$) downregulates both *gata2a* and *gata2b* and temporarily impairs HSPC emergence during EHT. Despite both the expression level of *gata2b* and the number of HSPCs are restored through the activation of Notch signaling by 48 hours post fertilization (hpf), adult *gata2a*$^{i4/i4}$ zebrafish have an increase susceptibility to infections, edema, neutropenia and hypocellular KM. Each *Gata2*-mutant zebrafish model we characterize (**chapter 3-5**) resembles the phenotype of a subgroup of *GATA2* patients, suggesting *GATA2* dosage and cell type-specific expression pattern of *GATA2* contributes to the phenotypic diversity observed in GATA2 haploinsufficiency syndromes.

The risk of developing MDS/AML in GATA2 patients increases from 6% at the age 10 to 81% at the age 40 indicating aging deteriorates the phenotypic consequences of GATA2 haploinsufficiency syndromes. Although the ability of *Gata2*$^{+/-}$ HSCs to differentiate into lymphoid-lineage is reduced, the contribution of aging to this functional decline is unexplored. In **chapter 6**, we investigate the effect of Gata2 haploinsufficiency during the aging of mice and show that aged-*Gata2*$^{+/-}$ HSCs have decreased reconstitution ability upon their transplantation resulting in B-cell cytopenia and monocytopenia, which resembles the phenotype of a subgroup of GATA2 patients. To understand the mechanisms leading to the functional decline of aged-*Gata2*$^{+/-}$ HSCs, we investigate the transcriptome of *Gata2*$^{+/-}$ HSCs from embryonic development and throughout aging. Our results show that *Gata2*$^{+/-}$ HSCs lose quiescence during embryonic stages and remain proliferative throughout aging. Furthermore, the accumulation of double-strand breaks (DSBs) and the aberrant inflammatory transcriptomic signatures acquired in aged-*Gata2*$^{+/-}$ HSCs upon their transplantation suggest that Gata2 haploinsufficiency results in genome instability in aged-HSCs.

Due to the lack of mechanistic insights behind leukemia development in GATA2 haploinsufficiency syndromes, the only treatment option for GATA2 patients is the allogeneic HSC transplantation (allo-HSCT). However, not every GATA2 patient is suitable for or successfully responds to allo-HSCT and this treatment option often becomes life-threatening. In the last part of the thesis (**chapter 7)**, we explore the advantages and pitfalls

of current genome editing technologies to propose novel strategies for correcting the mutant *GATA2* allele in the own HSCs of GATA2 patients, in order to provide autologous HSC transplantation (auto-HSCT) as a treatment for GATA2 haploinsufficiency syndromes in the future.

In conclusion, this thesis elucidates the multifaceted role of *GATA2* in HSC biology. By dissecting the role of *Gata2* spatiotemporally in various *Gata2*-mutant mouse and zebrafish models we show that both formation of HSCs during embryonic development and crucial functions in HSCs throughout adulthood and aging, such as the quiescence, cell-fate commitment, lineage differentiation and reconstitution is regulated by *GATA2*-dependent mechanisms.

8.2. GENERAL DISCUSSION

8.2.1. Understanding the biology of GATA2 haploinsufficiency syndromes

In humans, heterozygous germline mutations in *GATA2* present with various phenotypes, ranging from cytopenias and lymphedema to the development of MDS/AML (Figure 1) (*Donadieu et al., 2018; Hsu et al., 2011;2015; Dickinson et al., 2011; Ostergaard et al., 2011; Hahn et al., 2011*). GATA2 haploinsufficiency syndromes show incomplete penetrance, and the phenotypic outcomes are not correlated to the type or location of *GATA2* mutations (*Wlodarski et al., 2016*). Moreover, about 50% of *GATA2* mutation carriers develop symptoms by the age of twenty, and only about 5% of *GATA2* mutation carriers remain symptom-free by the age of sixty (*National Institute of Allergy and Infectious Diseases, 2016*), suggesting that GATA2 haploinsufficiency syndromes are progressive and complex diseases. The variety of phenotypes in the *Gata2*-mutant mouse and zebrafish models examined in this thesis also proves that *Gata2* has multiple dose-sensitive functions in the hematopoietic system. Elucidating the still incomplete roles of *Gata2* at the cellular level in suitable model systems will help to obtain a more complete picture of GATA2 haploinsufficiency syndromes in man. In this General Discussion chapter, I will discuss the implications of the work presented in this thesis in combination with the recent progress made by other investigators to achieve this goal.

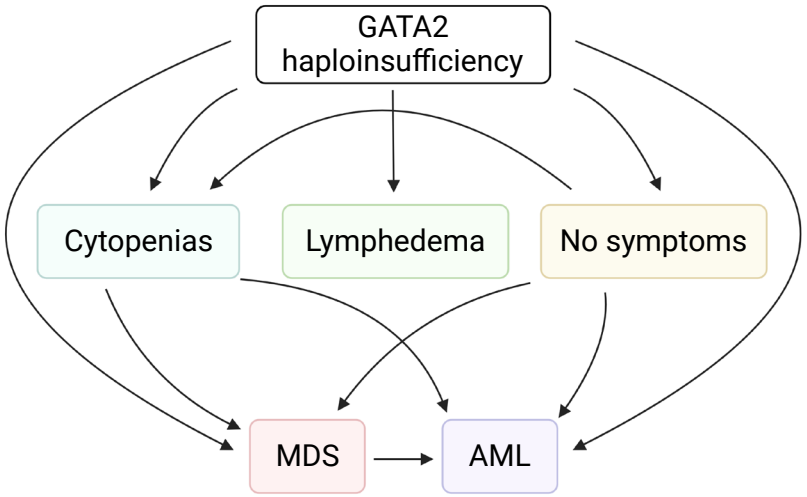


Figure 1

Figure 1. Germline heterozygous *GATA2* mutations result in a variety of phenotypes in humans. GATA2 haploinsufficiency syndromes are complex diseases and manifest with cytopenias, lymphedema and MDS/AML. All *GATA2* mutation carriers have a lifetime risk of MDS/AML, although some are initially asymptomatic. Arrows represent possible phenotypic progression trajectories of GATA2 haploinsufficiency syndromes.

8.2.1.1. Focusing on cell type-specific roles of *GATA2* in the hematopoietic system

Transcription factors (TFs) are intrinsic drivers of cell type-specific transcriptional programs, and they often function in a dose-sensitive manner (*Ni et al., 2019*). The dose-sensitive functions of *Gata2* were previously explored in various mouse models, which showed that Gata2 haploinsufficiency reduces the number and functionality of HSCs during embryonic development and adulthood (*Guo et al., 2013; de Pater et al., 2013; Rodrigues et al., 2005; Ling et al., 2004*). However, these models did not reflect the phenotypic spectrum of GATA2 patients, suggesting that GATA2 haploinsuffciency syndromes are difficult to tackle in a single model. Here, to address the possibility whether this is due to an incomplete representation of chronic and complex aspects of Gata2 haploinsufficiency in these models, we performed serial transplantations of aged-*Gata2*$^{+/-}$ HSCs in mice. Furthermore, we investigated the mechanisms underlying Gata2 deficiencies by collectively evaluating the dose- and cell type-specific roles of *Gata2* observed in different *Gata2*-mutant mouse and zebrafish models.

Hematopoietic stem cells: the journey from formation to regulation of reserves

HSCs are the source of the entire hematopoietic system and are generated in the aorta-gonad-mesonephros (AGM) region through EHT events during embryonic development (*Boisset et al, 2010; Kissa and Herbomel, 2010*). Previously, it has been shown in mice that deletion of *Gata2* (*Gata2*$^{-/-}$) abrogates trans-differentiation of hemogenic endothelial cells (HECs) and reduces the formation of hematopoietic stem and progenitor cells (HSPCs)

during EHT. Although HSPCs were formed in *Gata2*$^{+/-}$ embryos, both the number and their capacity to generate all hematopoietic lineages were reduced (*de Pater et al., 2013; Ling et al., 2004; Tsai et al., 1994*).

In mice, the embryonic HSC pool consists of HSC precursors and mature HSCs *(Rybtsov et al., 2014; 2011; Taoudi et al., 2008*). We showed that although the formation of HSC precursors is normal, the maturation of these precursors to HSCs is severely reduced in *Gata2*$^{+/-}$ embryos (Figure 2) (**chapter 2**). Nonetheless, *Gata2*$^{+/-}$ HSCs that were able to complete maturation produced similar numbers and types of colonies as the WT HSCs in colony-forming unit-culture (CFU-C). Therefore, our study showed that the reduced reconstitution ability of embryonic *Gata2*$^{+/-}$ HSCs previously shown was due to incomplete HSC maturation rather than reduced functionality. These results revised the role of *Gata2* during EHT and showed that *Gata2* is required not only for the trans-differentiation of HECs during the formation of HSC precursors, but also for the maturation of nascent HSCs.

HSCs were diminished in both mouse *Gata2*$^{+/-}$ fetal liver (FL) (**chapter 6**) and zebrafish *gata2b*$^{-/-}$ caudal hematopoietic tissue (CHT) (**chapter 3**). This observation raised the question whether *Gata2* has a direct effect on the embryonic expansion of HSCs or whether this is a secondary effect due to reduced HSC formation in AGM. We addressed this by exploiting the sub-functions of *gata2a* and *gata2b* in zebrafish. We showed that EHT events are reduced in *gata2a*$^{i4/i4}$ AGM (**chapter 4**), but not in *gata2b*$^{-/-}$ AGM (**chapter 3**). However, the number of HSCs was reduced in *gata2b*$^{-/-}$ CHT, showing that *Gata2* regulates both formation and expansion of HSCs during embryonic development (Figure 2).

During adulthood, one of the most crucial functions of HSCs that ensures longevity is to maintain the balance between quiescence and proliferation (*Wilson et al., 2009*). We showed that HSCs in *Gata2*$^{+/-}$ mice become more proliferative during FL stage and remain proliferative throughout adulthood and aging compared to wild type controls (**chapter 6**). Furthermore, adult *gata2b*$^{-/-}$ (**chapter 3**) and *gata2b*$^{+/-}$ (**chapter 4**) zebrafish HSCs were also more proliferative than wild type, further supporting that increased proliferation in HSCs was a consequence of Gata2 deficiency. The number of HSCs reduced in *Gata2*$^{+/-}$ mice (marked by LSK SLAM) hinted at a possibility of a compensatory mechanism in HSCs to meet the mature blood cell demand by proliferating more frequently. However, the number of adult HSCs were not reduced in *gata2b*$^{-/-}$ or *gata2b*$^{+/-}$ zebrafish (marked by CD41^{int}) even though they had increased proliferation similar to HSCs in *Gata2*$^{+/-}$ mice. It has been also shown that the overexpression of *Gata2* in human and mouse HSCs results in reduced proliferation (*Tipping et al., 2009*). Hence, our findings suggested that proliferation in *Gata2*-mutant HSCs was cell-autonomously induced due to the loss of *Gata2*.

CD41 is currently the only marker used to detect the transplantable HSC population in zebrafish (*Lin et al., 2005*), and it is possible that the CD41^{int} population contains both long-term and short-term HSCs. Investigating additional zebrafish HSC markers in future studies may allow us to understand how the long-term HSC pool is affected by Gata2 deficiency in

zebrafish. Irrespective of this, these findings demonstrated that the loss of *Gata2* in both mice and zebrafish reduces the number of HSCs and promotes proliferation in HSCs throughout the organism's entire lifespan. Previous studies have shown that prolonged proliferation of HSCs might result in reduced engraftment ability, functional decline and consecutively lead to the development of hematological malignancies (*Kirschner et al., 2017; Bowie et al., 2006; Lin et al., 2011*). Whether *Gata2*-mutant HSCs fail to undergo transition to the quiescent state or whether the proliferative state is a cause or consequence of an underlying cellular defect remains to be investigated.

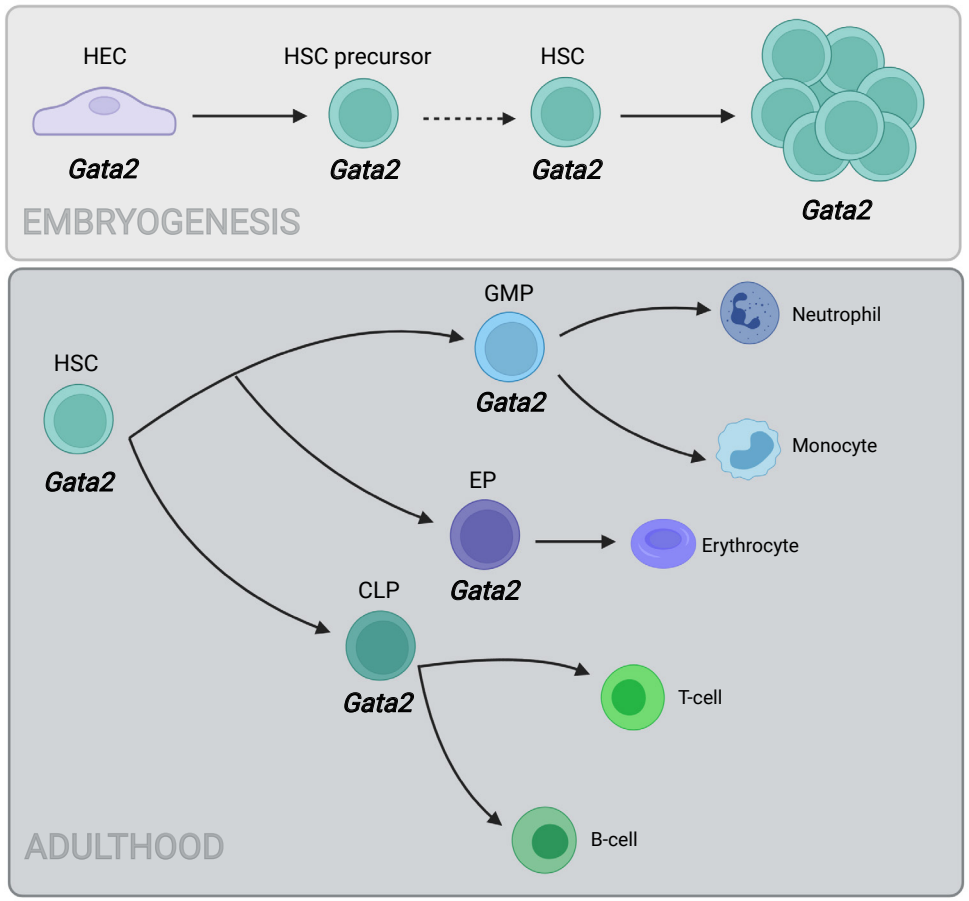


Figure 2. *Gata2* has many roles in the hematopoietic system during embryonic development and adulthood. *Gata2* regulates trans-differentiation of HECs, maturation of HSC precursors into HSCs and the expansion of HSCs during embryonic development. *Gata2* is required for the maintenance and function of HSCs, GMPs, EPs and CLPs during adulthood. Hemogenic endothelial cells, HEC; Hematopoietic stem cell, HSC; Granulocyte-monocyte progenitor; GMP; Erythroid progenitor, EP; Common lymphoid progenitor, CLP.

Hematopoietic progenitor cells: a decision point for cell-fate

HSCs differentiate into myeloid, lymphoid and erythroid lineage-committed intermediate hematopoietic progenitor cell (HPC) types to produce mature blood cells (*Orkin, 2000)*. Because the complete deletion of *gata2b* is not lethal in zebrafish, we were able to reveal various dose-dependent functions of *Gata2* in HPCs by comparing the *gata2b*$^{-/-}$ and *gata2b*$^{+/-}$ zebrafish models. We showed that *gata2b*$^{-/-}$ zebrafish had reduced erythro-myeloid progenitors and increased lymphoid progenitors in the KM (Figure 2) (**chapter 3**). In a parallel study, another homozygous mutation that abrogates *gata2b* expression in zebrafish also resulted in reduced erythro-myeloid progenitors and an increased lymphoid-bias in HPCs (*Avagyan et al., 2021*). Thus, these independent studies both suggest that *Gata2* acts as a decision factor between erythro-myeloid and lymphoid lineage-fates and favors the formation of erythro-myeloid HPCs. Strikingly however, erythro-myeloid progenitors were present in *gata2b*$^{+/-}$ zebrafish, but these progenitors showed dysplastic features (**chapter 4**). These results indicated that haploid-dose of *gata2b* is sufficient for the formation but insufficient for the normal maturation of erythro-myeloid progenitors. Notably, erythro-myeloid dysplasia is also among the phenotypic outcomes observed in GATA2 haploinsufficiency patients (*Ganapathi et al., 2015*), supporting the relevance of the zebrafish model for studying human disease.

It was previously shown that the depletion of B-cell precursors and B-cell cytopenia is the most common phenotypic outcome of pediatric GATA2 patients (*Nováková et al., 2016*). Although the number of lymphoid progenitors appeared to be less affected by Gata2 deficiency compared to the erythro-myeloid progenitors in both our study (**chapter 3**) and the parallel study (*Avagyan et al., 2021*), B-cell differentiation was incomplete in both of these models. As *Gata2* expression is not found in the mature B-cells, these results hinted at a possible defect in B-cell progenitors. We focused on the origin of these defects in **chapter 6,** and our results suggested that ProB-cells are the bottleneck of B-cell differentiation defects in aged-*Gata2*$^{+/-}$ due to increased senescence. Moreover, both B-cell and T-cell progenitors reduced in aged-*Gata2*$^{+/-}$ mice suggested a defect in the common lymphoid progenitors (CLPs), which was also confirmed in another study where authors showed that Gata2 haploinsufficiency depletes the CLPs in aged mice (*Abdelfattah et al., 2021*). Furthermore, although *GATA2* mutations are more frequently associated with erythro-myeloid malignancies in patients (*Ganapathi et al., 2015*), they are also found in patients with B-cell and T-cell acute lymphoblastic leukemia (*Wang et al., 2021; Esparza et al., 2019; Koegel et al., 2015*), highlighting that *GATA2*-dosage is also crucial for the regulation of lymphoid progenitors (Figure 2).

Mature blood cells: development of cytopenias

In humans, GATA2 haploinsufficiency syndromes often result in cytopenias in one or more lineages. It has been shown that B-cell cytopenia (as mentioned above), NK-cell cytopenia

and monocytopenia are observed in more than 75% of GATA2 patients, whereas GATA2 mutations have been found in 10% of those with congenital neutropenia (*Hsu et al., 2015; Spinner et al., 2014*). We found that mature neutrophils and B-cells were reduced in *gata2b*$^{-/-}$ zebrafish (**chapter 3**). Additionally, in the parallel study it was shown that *gata2b*$^{-/-}$ zebrafish have reduced monocytes and B-cells (*Avagyan et al., 2021*). Furthermore, aged-*Gata2*$^{+/-}$ mice presented with monocytopenia and B-cell cytopenia (**chapter 6**). These results revealed that B-cell cytopenia is the most common phenotypic outcome in our *Gata2*-mutant models, similar to GATA2 patients (*Nováková et al., 2016*). Our observation of B-cell cytopenia as the first phenotypic outcome in aged-*Gata2*$^{+/-}$ mice also suggested that B-cell cytopenia might be representing the early stages of GATA2 haploinsufficiency syndromes. On the other hand, the observation of neutropenia in one and monocytopenia in the other *gata2b*$^{-/-}$ zebrafish model suggested that Gata2 deficiency leads to defects in granulocyte-monocyte progenitors (GMPs) (Figure 2), which is in line with previous studies showing that Gata2 haploinsufficiency reduces the functionality of GMPs (*Rodrigues et al., 2008*).

In aged-*Gata2*$^{+/-}$ mice, monocytopenia developed only after the secondary transplantation of aged-*Gata2*$^{+/-}$ HSCs. Although monocytopenia is often found in GATA2 patients, one study of 90 GATA2 patients suggested that monocytopenia is only observed in the GATA2 patients who have already progressed to MDS (*Nováková et al., 2016*). A similar observation was made in another study of 57 GATA2 patients in which patients with advanced MDS showed monocytopenia (*Spinner et al., 2014*), suggesting that monocytopenia could be a sign of disease progression in GATA2 patients.

Although *gata2b* has been shown to be the main mammalian ortholog of *Gata2* expressed in hematopoietic cells in zebrafish (*Butko et al., 2015*), we and others showed that adult *gata2a*$^{i4/i4}$ zebrafish develop monocytopenia, neutropenia and edema (**chapter 5** and *Mahony et al., 2021*). These results demonstrated that the activity of zebrafish gata2a *i4* enhancer, which corresponds to +9.5 enhancer of mammalian *Gata2*, is required for myeloid lineage differentiation. Moreover, *gata2a* is the main orthologue in zebrafish that is expressed during lympho-vascular development (*Butko et al., 2015*). The fact that edema was found uniquely in *gata2a*$^{i4/i4}$ zebrafish suggested that the activity of this conserved cis-element is also required in lymphatic endothelial cells.

As bone marrow failure (BMF) syndromes are characterized by defects in one or more lineages that result in cytopenias, our studies have shed light on the initiating cellular events that lead to BMF in GATA2 haploinsufficiency syndromes.

8.2.1.2. Transcription factors regulating cell-fate programs in conjunction with Gata2

Cell type-specific programs are driven by the collective operation of TF networks in which the dosage of TFs also affects the function of other TFs acting in the same network. By investigating the transcriptome and chromatin accessibility of *Gata2*-mutant HSPCs, we were able to reveal *Gata2*-dependent TFs during the regulation of cell fate-specific programs.

The master regulator of the endothelial-to-hematopoietic transcriptional switch

EHT is accompanied by the progressive downregulation of endothelial transcriptional programming and simultaneous upregulation of a hematopoietic transcriptional program (*Oatley et al., 2020; Baron et al., 2018; Zhou et al., 2016; Swiers et al., 2013*). *Gata2* is one of the ‚heptad' TFs that regulates hematopoietic transcriptional programming during EHT (Figure 3) (*Goode et al., 2016; Solaimani Kartalaei et al., 2015; Wilson et al., 2010*). However, whether *Gata2* regulates the endothelial programming during EHT has not been shown previously. We found that endothelial repressor *Gfi1b* is downregulated in embryonic *Gata2*$^{+/-}$ HSPCs resulting in an incomplete repression of endothelial programming (Figure 2) (**chapter 2**). Moreover, ectopic expression of *Gfi1b* in the hematopoietic cells of *gata2*$^{-/-}$ zebrafish restored the embryonic HSCs, thus proving that *Gata2* activates *Gfi1b* during EHT and that this regulatory mechanism is conserved between mouse and zebrafish. It was previously predicted that *Gata2* binds to +16 and +17 kb regions of *Gfi1b* locus (*Moignard et al., 2013*). We showed that indeed these chromatin regions were less open in *Gata2*$^{+/-}$ HSPCs. Therefore, our study provided the first experimental proof showing that *Gata2* regulates *Gfi1b* through +16 and +17 kb enhancer regions to repress endothelial transcriptional programming during HSC formation (Figure 3).

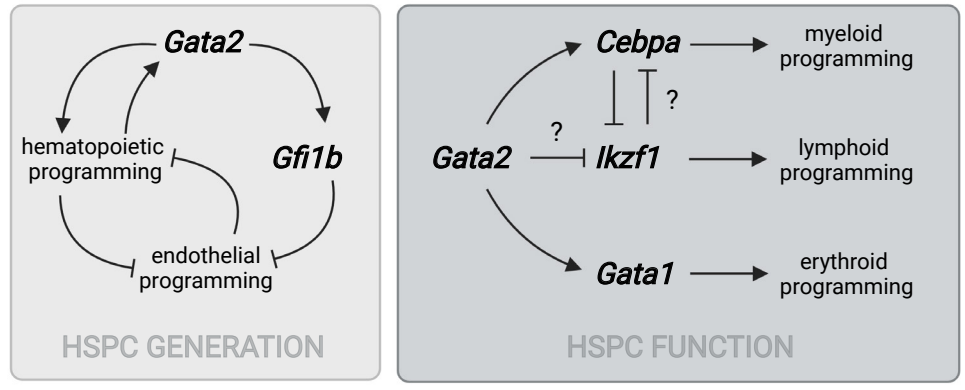


Figure 3. Downstream transcription factor partners of *Gata2* regulating cell-fate programs. *Gata2* activates hematopoietic programming and regulates *Gfi1b* to repress endothelial programming during the generation of HSPCs. *Gata2* regulates the expression of *Cebpa*, *Ikzf1* and *Gata1* in adult HSPCs during myeloid, lymphoid and erythroid lineage commitment and differentiation.

Myeloid versus lymphoid lineage-fate choice in hematopoietic stem and progenitor cells (HSPCs)

We showed that *Gata2* deficiency depletes myeloid progenitors at the expense of lymphoid progenitors suggesting *Gata2* acts as a key factor for lineage-fate choice in the immature HSPCs (**chapter 3**). Exploring the mechanisms behind this, we found that *gata2*$^{-/-}$ HSPCs co-express myeloid and lymphoid transcripts and that the expression level of myeloid genes in *gata2*$^{-/-}$ HSPCs are below the threshold to maintain myeloid differentiation. Among

the differentially expressed genes, we found that myeloid TF *cebpa* was downregulated in *gata2b*$^{-/-}$ HSPCs. Moreover, two parallel studies showed that chromatin accessibility at *cebpa* locus was reduced in *gata2b*$^{-/-}$ myeloid progenitors (*Avagyan et al., 2021*) and that *cebpa* was downregulated in *gata2a*$^{i4/i4}$ HSPCs resulting in myeloid differentiation defects (*Mahony et al., 2021*). Because *CEBPA* is one of the key TFs required in myeloid progenitors and deletion of *Cebpa* was shown to block myeloid differentiation (*Pundhir et al., 2018; Avellino et al., 2016; Zhang et al., 2004*), these results suggested that *Gata2* is regulating *Cebpa* to promote myeloid commitment and differentiation in HSPCs (Figure 3).

On the other hand, our experiments in zebrafish showed that the lymphoid TF *ikzf1* was upregulated in *gata2b*$^{-/-}$ HSPCs (**chapter 3**), and moreover, *gata2b*$^{-/-}$ lymphoid progenitors showed increased chromatin accessibility at the *ikzf1* locus (*Avagyan et al., 2021*). *IKZF1* is crucial for the development of lymphoid progenitors and B-cell differentiation although overexpression of *Ikzf1* blocks terminal B-cell differentiation (*Rahmani et al., 2019*). While our results suggested that increased *ikzf1* expression in *gata2b*$^{-/-}$ HSPCs might be the underlying cause of lymphoid-bias and incomplete B-cell differentiation, it remains to be investigated whether *ikzf1* expression increases as a result of decreased myeloid transcriptional programming or whether *gata2b* plays a role in direct suppression of *ikzf1* in HSPCs (Figure 3).

Taken together, these results demonstrated that *Gata2* regulates the function of TFs that drive myeloid and lymphoid transcriptional programs in HSPCs and that the loss of *Gata2* leads to defects of myeloid and lymphoid lineage commitment and differentiation.

Gata2-to-Gata1 switch in erythroid progenitors

Lineage trajectory analysis showed that *gata2b*$^{+/-}$ erythroid progenitors have a complete opposite expression pattern for *gata1a* compared to WT erythroid progenitors (**chapter 4**). It has been previously shown that *Gata1* and *Gata2* occupy similar regions on the chromosome to drive distinct cell-fate programs and that erythroid differentiation is established through *Gata2*-to-*Gata1* switch in erythroid progenitors (*Moriguchi and Yamamoto, 2014; Doré et al., 2012; Bresnick et al., 2010; Pal et al., 2004; Grass et al., 2003*). We and others showed that the erythroid progenitors are formed in *Gata2*$^{+/-}$ mice (**chapter 6**) and *gata2b*$^{+/-}$ zebrafish (**chapter 4** and *Avagyan et al., 2021*), however, *gata2b*$^{+/-}$ zebrafish uniquely developed erythroid dysplasia, which was not observed in *gata2b*$^{-/-}$ zebrafish (**chapter 3**). These results suggested that haploinsufficiency of *Gata2* has a unique effect on the expression of *Gata1* in erythroid progenitors, thereby contributing to erythro-myeloid dysplasia (Figure 3).

8.2.1.3. What is the contribution of aging to the functional decline of *Gata2*$^{+/-}$ HSCs?

The risk of developing MDS/AML in GATA2 patients increases with age (*Donadieu et al., 2018; Spinner et al., 2014*), implying that aging contributes to the phenotypic consequences of GATA2 haploinsufficiency. Although *Gata2*$^{+/-}$ mice have reduced number of HSCs and

Gata2$^{+/-}$ HSCs are hampered in their engraftment ability compared to WT HSCs in serial transplantation assays, *Gata2*$^{+/-}$ mice did not show terminal differentiation defects (*Guo et al., 2013; Rodrigues et al., 2005*). On the other hand, we (**chapter 6**) and others (*Abdelfattah et al., 2021*) showed that, after 15 months of aging, the lineage differentiation ability of *Gata2*$^{+/-}$ HSCs severely reduced, resulting in peripheral blood cytopenias. These results suggested that Gata2 haploinsufficiency initially affects HSCs, and that the development of cytopenias are therefore consequences of impaired HSC functionality over time. We asked whether differences between aged-*Gata2*$^{+/-}$ HSCs compared to their pre-aging states at the transcriptome level might be held responsible for this. We found that aged-*Gata2*$^{+/-}$ HSCs are transcriptionally marked by increased DNA damage, dysregulation of DNA damage repair genes and increased inflammatory signatures. Similarly, in a recent study authors showed that zebrafish *gata2a* regulates the expression of genes associated with DNA damage repair and that adult *gata2a*$^{i4/i4}$ HSCs acquire DNA damage (*Mahony et al., 2021*). These results suggested that *Gata2* has a role in the genome maintenance of HSCs and that aging enhances the susceptibility to genome instability in *Gata2*$^{+/-}$ HSCs.

Genome instability in HSCs that are induced by DNA damage repair defects and the acquisition of somatic mutations are associated with age-related functional reduction in HSCs and the development of hematological malignancies (*Moehrle and Geiger, 2016; Kenyon and Gerson, 2007*). In addition, germline *GATA2* mutation carriers that acquire somatic mutations in other genes, such as *ASXL1*, *RUNX1*, *SETBP1*, *IKZF1*, and *CRLF2*, have been shown to progress to MDS/AML (*Wlodarski et al., 2016; Fisher et al., 2017; Yoshida et al., 2020*), suggesting that *Gata2* haploinsufficiency causes a vulnerability for secondary genomic events to occur, which likely are related to the acquisition of secondary mutations involved in the leukemic progression of GATA2 haploinsufficiency syndromes.

8.2.2. Conclusion and perspectives

This thesis dealt with the role of *Gata2* in the hematopoietic system and demonstrated that *Gata2* has multiple functions; from the ontogenesis of HSCs to their function and maintenance, as well as cell fate programming of downstream progenitors and terminal differentiation of mature blood cells. The results, revealing various functions of *Gata2* and the transcriptional profile of *Gata2*-mutant hematopoietic cells, have provided a better understanding of the mechanisms underlying Gata2 deficiency syndromes. However, many questions related to complex cell type- and age-dependent functions remain to be addressed, as will be briefly touched upon below.

Development of hematological malignancies in GATA2 haploinsufficiency syndromes

Although *Gata2*-mutant zebrafish and mouse models examined in this thesis presented with cytopenia and erythro-myeloid dysplasia, MDS/AML was not observed in these models. A possible explanation for this could be that these model systems are kept and bred in specific-

pathogen free conditions and consequently lack the environmental insults to challenge the hematopoietic system. Future work is needed to assess whether infectious or inflammatory stimuli induce MDS/AML progression in *Gata2*-mutant models. In addition, other differences between these model systems and humans, such as mechanisms regulating HSCs and the hematopoietic system, lifespan, genetic background, or a combination of these factors, may contribute differentially to the development of the MDS/AML. For instance, unlike humans, mice do not spontaneously develop MDS/AML as they age. Moreover, it has been shown that some genetic defects causing MDS in humans do not result in MDS in mice; and even if these genetic perturbations result in MDS, mice do not reflect the diversity of features found in humans with MDS (*Zhou et al., 2015*).

Increased proliferation in Gata2-mutant HSCs: cause or consequence?

We showed that *Gata2*$^{+/-}$ HSCs acquire proliferative signatures at embryonic stages and maintain these signatures throughout adulthood and aging (Figure 4). Prolonged proliferation has been shown to promote genome instability in HSCs, as DNA becomes vulnerable to errors at each replication (*Walter et al., 2015; Kamminga et al., 2005*). Therefore, our results showing that *Gata2*$^{+/-}$ HSCs accumulate DNA damage after aging suggested that proliferative stress may be the underlying cause of genome instability. Conversely, HSCs prone to accumulation of DNA damage have been shown to proliferate to activate proliferation-dependent DNA damage repair pathways (*Beerman et al., 2014*). If the latter is true, Gata2 haploinsufficiency could be causing a vulnerability in the genome for additional mutations to occur, and therefore, *Gata2*$^{+/-}$ HSCs might be proliferating to activate DNA damage repair mechanisms. Irrespective of the above proposed possibility, the causal relationship between proliferation and genome instability in *Gata2*$^{+/-}$ HSCs should continue to be explored in future studies as it may also contribute to the development of hematological malignancies. Finally, exploring the clonal evolution of the acquired mutations in *Gata2*$^{+/-}$ HSCs can help to understand if specific HSC clones are responsible for the functional decline of aged-*Gata2*$^{+/-}$ HSCs.

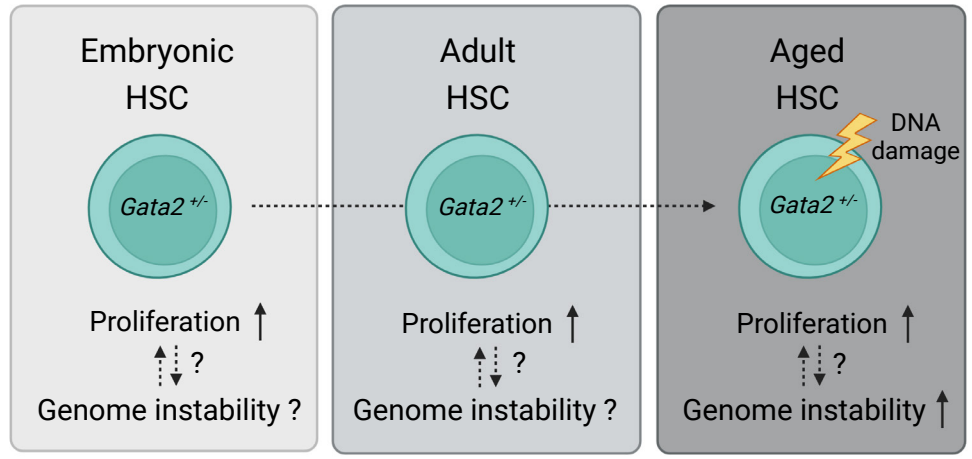


Figure 4. Contribution of embryonic GATA2 haploinsufficiency to adult phenotype. *Gata2*$^{+/-}$ HSCs are proliferative from embryonic stages to aging and aged-*Gata2*$^{+/-}$ HSCs accumulate DNA damage. Does proliferative stress from the embryonic stages contribute to genome instability of aged-*Gata2*$^{+/-}$ HSCs?

Contribution of embryonic GATA2 haploinsufficiency to adult phenotype

GATA2 haploinsufficiency syndromes present with a variable penetrance, although all patients carry congenital *GATA2* mutations (*Wlodarski et al., 2016*). Therefore, it is currently unclear from which developmental stage the hematopoietic defects originate in GATA2 patients. For instance, if increased proliferation in *Gata2*$^{+/-}$ HSCs is associated with genome instability, this would imply that *Gata2*$^{+/-}$ HSCs are prone to acquire secondary mutations from the time of embryonic development (Figure 4). Investigating the clonal evolution of *Gata2*$^{+/-}$ HSCs in inducible *Gata2*-knockout and -rescue models can help to understand the origin of adult hematopoietic defects in *Gata2*-mutants.

In conclusion, we show that *GATA2* regulates various hematopoietic cell-fate programs during embryonic development and adulthood. Because GATA2 haploinsufficiency syndromes are complex diseases, our comprehensive approach exploring multiple *Gata2*-mutant model systems allowed us to unveil the diverse dose-sensitive functions of *GATA2* and the transcriptional programs regulated by *GATA2*, providing a roadmap for understanding the biology of GATA2 haploinsufficiency syndromes.

REFERENCES

Abdelfattah, A., Hughes-Davies, A., Clayfield, L., Menendez-Gonzalez, J. B., Almotiri, A., Alotaibi, B., Tonks, A., & Rodrigues, N. P. (2021). Gata2 haploinsufficiency promotes proliferation and functional decline of hematopoietic stem cells with myeloid bias during aging. *Blood Adv*, *5*(20), 4285-4290.

Avagyan, S., Weber, M. C., Ma, S., Prasad, M., Mannherz, W. P., Yang, S., Buenrostro, J. D., & Zon, L. I. (2021). Single-cell ATAC-seq reveals GATA2-dependent priming defect in myeloid and a maturation bottleneck in lymphoid lineages. *Blood Adv*, *5*(13), 2673-2686.

Avellino, R., Havermans, M., Erpelinck, C., Sanders, M. A., Hoogenboezem, R., van de Werken, H. J., Rombouts, E., van Lom, K., van Strien, P. M., Gebhard, C., Rehli, M., Pimanda, J., Beck, D., Erkeland, S., Kuiken, T., de Looper, H., Gröschel, S., Touw, I., Bindels, E., & Delwel, R. (2016). An autonomous CEBPA enhancer specific for myeloid-lineage priming and neutrophilic differentiation. *Blood*, *127*(24), 2991-3003.

Baron, C. S., Kester, L., Klaus, A., Boisset, J. C., Thambyrajah, R., Yvernogeau, L., Kouskoff, V., Lacaud, G., van Oudenaarden, A., & Robin, C. (2018). Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. *Nat Commun*, *9*(1), 2517.

Beerman, I., Bhattacharya, D., Zandi, S., Sigvardsson, M., Weissman, I. L., Bryder, D., & Rossi, D. J. (2010). Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc Natl Acad Sci U S A*, *107*(12), 5465-5470.

Boisset, J. C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., & Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*, *464*(7285), 116-120.

Bowie, M. B., McKnight, K. D., Kent, D. G., McCaffrey, L., Hoodless, P. A., & Eaves, C. J. (2006). Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. *J Clin Invest*, *116*(10), 2808-2816.

Bresnick, E. H., Lee, H. Y., Fujiwara, T., Johnson, K. D., & Keles, S. (2010). GATA switches as developmental drivers. *J Biol Chem*, *285*(41), 31087-31093.

Butko, E., Distel, M., Pouget, C., Weijts, B., Kobayashi, I., Ng, K., Mosimann, C., Poulain, F. E., McPherson, A., Ni, C. W., Stachura, D. L., Del Cid, N., Espín-Palazón, R., Lawson, N. D., Dorsky, R., Clements, W. K., & Traver, D. (2015). Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. *Development*, *142*(6), 1050-1061.

de Pater, E., Kaimakis, P., Vink, C. S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P. S., Camper, S. A., Speck, N., & Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. *J Exp Med*, *210*(13), 2843-2850.

Dickinson, R. E., Griffin, H., Bigley, V., Reynard, L. N., Hussain, R., Haniffa, M., Lakey, J. H., Rahman, T., Wang, X. N., McGovern, N., Pagan, S., Cookson, S., McDonald, D., Chua, I., Wallis, J., Cant, A., Wright, M., Keavney, B., Chinnery, P. F., Loughlin, J., Hambleton, S., Santibanez-Koref, M., & Collin, M. (2011). Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood*, *118*(10), 2656-2658.

Donadieu, J., Lamant, M., Fieschi, C., de Fontbrune, F. S., Caye, A., Ouachee, M., Beaupain, B., Bustamante, J., Poirel, H. A., Isidor, B., Van Den Neste, E., Neel, A., Nimubona, S., Toutain, F., Barlogis, V., Schleinitz, N., Leblanc, T., Rohrlich, P., Suarez, F., Ranta, D., Chahla, W. A., Bruno, B., Terriou, L., Francois, S., Lioure, B., Ahle, G., Bachelerie, F., Preudhomme, C., Delabesse, E., Cave, H., Bellanné-Chantelot, C., Pasquet, M., & French, G. s. g. (2018). Natural history of GATA2 deficiency in a survey of 79 French and Belgian patients. *Haematologica*, *103*(8), 1278-1287.

Doré, L. C., Chlon, T. M., Brown, C. D., White, K. P., & Crispino, J. D. (2012). Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. *Blood*, *119*(16), 3724-3733.

Esparza, O., Xavier, A. C., Atkinson, T. P., Hill, B. C., & Whelan, K. (2019). A unique phenotype of T-cell acute lymphoblastic leukemia in a patient with GATA2 haploinsufficiency. *Pediatr Blood Cancer*, *66*(6), e27649.

Fisher, K. E., Hsu, A. P., Williams, C. L., Sayeed, H., Merritt, B. Y., Elghetany, M. T., Holland, S. M., Bertuch, A. A., & Gramatges, M. M. (2017). Somatic mutations in children with GATA2-associated myelodysplastic syndrome who lack other features of GATA2 deficiency. *Blood Adv*, *1*(7), 443-448.

Ganapathi, K. A., Townsley, D. M., Hsu, A. P., Arthur, D. C., Zerbe, C. S., Cuellar-Rodriguez, J., Hickstein, D. D., Rosenzweig, S. D., Braylan, R. C., Young, N. S., Holland, S. M., & Calvo, K. R. (2015). GATA2 deficiency-associated bone marrow disorder differs from idiopathic aplastic anemia. *Blood*, *125*(1), 56-70.

Goode, D. K., Obier, N., Vijayabaskar, M. S., Lie, A. L. M., Lilly, A. J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., Challinor, M., Wallace, K., Gilmour, J., Assi, S. A., Cauchy, P., Hoogenkamp, M., Westhead, D. R., Lacaud, G., Kouskoff, V., Göttgens, B., & Bonifer, C. (2016). Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Dev Cell*, *36*(5), 572-587.

Grass, J. A., Boyer, M. E., Pal, S., Wu, J., Weiss, M. J., & Bresnick, E. H. (2003). GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A*, *100*(15), 8811-8816.

Guo, G., Luc, S., Marco, E., Lin, T. W., Peng, C., Kerenyi, M. A., Beyaz, S., Kim, W., Xu, J., Das, P. P., Neff, T., Zou, K., Yuan, G. C., & Orkin, S. H. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell*, *13*(4), 492-505.

Hahn, C. N., Chong, C. E., Carmichael, C. L., Wilkins, E. J., Brautigan, P. J., Li, X. C., Babic, M., Lin, M., Carmagnac, A., Lee, Y. K., Kok, C. H., Gagliardi, L., Friend, K. L., Ekert, P. G., Butcher, C. M., Brown, A. L., Lewis, I. D., To, L. B., Timms, A. E., Storek, J., Moore, S., Altree, M., Escher, R., Bardy, P. G., Suthers, G. K., D'Andrea, R. J., Horwitz, M. S., & Scott, H. S. (2011). Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*, *43*(10), 1012-1017.

Hsu, A. P., McReynolds, L. J., & Holland, S. M. (2015). GATA2 deficiency. *Curr Opin Allergy Clin Immunol*, *15*(1), 104-109.

Hsu, A. P., Sampaio, E. P., Khan, J., Calvo, K. R., Lemieux, J. E., Patel, S. Y., Frucht, D. M., Vinh, D. C., Auth, R. D., Freeman, A. F., Olivier, K. N., Uzel, G., Zerbe, C. S., Spalding, C., Pittaluga, S., Raffeld, M., Kuhns, D. B., Ding, L., Paulson, M. L., Marciano, B. E., Gea-Banacloche, J. C., Orange, J. S., Cuellar-Rodriguez, J., Hickstein, D. D., & Holland, S. M. (2011). Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*, *118*(10), 2653-2655.

Kamminga, L. M., van Os, R., Ausema, A., Noach, E. J., Weersing, E., Dontje, B., Vellenga, E., & de Haan, G. (2005). Impaired hematopoietic stem cell functioning after serial transplantation and during normal aging. *Stem Cells*, *23*(1), 82-92.

Kenyon, J., & Gerson, S. L. (2007). The role of DNA damage repair in aging of adult stem cells. *Nucleic Acids Res*, *35*(22), 7557-7565.

Kirschner, K., Chandra, T., Kiselev, V., Flores-Santa Cruz, D., Macaulay, I. C., Park, H. J., Li, J., Kent, D. G., Kumar, R., Pask, D. C., Hamilton, T. L., Hemberg, M., Reik, W., & Green, A. R. (2017). Proliferation Drives Aging-Related Functional Decline in a Subpopulation of the Hematopoietic Stem Cell Compartment. *Cell Rep*, *19*(8), 1503-1511.

Kissa, K., & Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*, *464*(7285), 112-115.

Koegel, A. K., Hofmann, I., Moffitt, K., Degar, B., Duncan, C., & Tubman, V. N. (2016). Acute lymphoblastic leukemia in a patient with MonoMAC syndrome/GATA2 haploinsufficiency. *Pediatr Blood Cancer*, *63*(10), 1844-1847.

Lin, H. F., Traver, D., Zhu, H., Dooley, K., Paw, B. H., Zon, L. I., & Handin, R. I. (2005). Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. *Blood*, *106*(12), 3803-3810.

Lin, K. K., Rossi, L., Boles, N. C., Hall, B. E., George, T. C., & Goodell, M. A. (2011). CD81 is essential for the re-entry of hematopoietic stem cells to quiescence following stress-induced proliferation via deactivation of the Akt pathway. *PLoS Biol*, *9*(9), e1001148.

Ling, K. W., Ottersbach, K., van Hamburg, J. P., Oziemlak, A., Tsai, F. Y., Orkin, S. H., Ploemacher, R., Hendriks, R. W., & Dzierzak, E. (2004). GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med*, *200*(7), 871-882.

Mahony, C. B., Noyvert, B., Vrljicak, P., Ott, S., Higgs, M., & Monteiro, R. (2021). Gata2a maintains cebpa and npm1a in haematopoietic stem cells to sustain lineage differentiation and genome stability. *bioRxiv*, 2021.2007.2019.452890.

Moehrle, B. M., & Geiger, H. (2016). Aging of hematopoietic stem cells: DNA damage and mutations? *Exp Hematol*, *44*(10), 895-901.

Moignard, V., Macaulay, I. C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F. J., Kinston, S., Joshi, A., Hannah, R., Theis, F. J., Jacobsen, S. E., de Bruijn, M. F., & Göttgens, B. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat Cell Biol*, *15*(4), 363-372.

Moriguchi, T., & Yamamoto, M. (2014). A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation. *Int J Hematol*, *100*(5), 417-424.

Ni, Z., Zhou, X. Y., Aslam, S., & Niu, D. K. (2019). Characterization of Human Dosage-Sensitive Transcription Factor Genes. *Front Genet*, *10*, 1208.

Nováková, M., Žaliová, M., Suková, M., Wlodarski, M., Janda, A., Froňková, E., Campr, V., Lejhancová, K., Zapletal, O., Pospíšilová, D., Černá, Z., Kuhn, T., Švec, P., Pelková, V., Zemanová, Z., Kerndrup, G., van den Heuvel-Eibrink, M., van der Velden, V., Niemeyer, C., Kalina, T., Trka, J., Starý, J., Hrušák, O., & Mejstříková, E. (2016). Loss of B cells and their precursors is the most constant feature of GATA-2 deficiency in childhood myelodysplastic syndrome. *Haematologica*, *101*(6), 707-716.

Oatley, M., Bölükbası Ö, V., Svensson, V., Shvartsman, M., Ganter, K., Zirngibl, K., Pavlovich, P. V., Milchevskaya, V., Foteva, V., Natarajan, K. N., Baying, B., Benes, V., Patil, K. R., Teichmann, S. A., & Lancrin, C. (2020). Single-cell transcriptomics identifies CD44 as a marker and regulator of endothelial to haematopoietic transition. *Nat Commun*, *11*(1), 586.

Orkin, S. H. (2000). Diversification of haematopoietic stem cells to specific lineages. *Nat Rev Genet*, *1*(1), 57-64.

Ostergaard, P., Simpson, M. A., Connell, F. C., Steward, C. G., Brice, G., Woollard, W. J., Dafou, D., Kilo, T., Smithson, S., Lunt, P., Murday, V. A., Hodgson, S., Keenan, R., Pilz, D. T., Martinez-Corral, I., Makinen, T., Mortimer, P. S., Jeffery, S., Trembath, R. C., & Mansour, S. (2011). Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet*, *43*(10), 929-931.

Pal, S., Cantor, A. B., Johnson, K. D., Moran, T. B., Boyer, M. E., Orkin, S. H., & Bresnick, E. H. (2004). Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A*, *101*(4), 980-985.

Pundhir, S., Bratt Lauridsen, F. K., Schuster, M. B., Jakobsen, J. S., Ge, Y., Schoof, E. M., Rapin, N., Waage, J., Hasemann, M. S., & Porse, B. T. (2018). Enhancer and Transcription Factor Dynamics during Myeloid Differentiation Reveal an Early Differentiation Block in Cebpa null Progenitors. *Cell Rep*, *23*(9), 2744-2757.

Rahmani, M., Fardi, M., Farshdousti Hagh, M., Hosseinpour Feizi, A. A., Talebi, M., & Solali, S. (2019). An investigation of methylation pattern changes in the IKZF1 promoter in patients with childhood B-cell acute lymphoblastic leukemia. *Blood Res*, *54*(2), 144-148.

Rodrigues, N. P., Boyd, A. S., Fugazza, C., May, G. E., Guo, Y., Tipping, A. J., Scadden, D. T., Vyas, P., & Enver, T. (2008). GATA-2 regulates granulocyte-macrophage progenitor cell function. *Blood*, *112*(13), 4862-4873.

Rodrigues, N. P., Janzen, V., Forkert, R., Dombkowski, D. M., Boyd, A. S., Orkin, S. H., Enver, T., Vyas, P., & Scadden, D. T. (2005). Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood*, *106*(2), 477-484.

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O., & Medvinsky, A. (2014). Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43(-) embryonic precursor. *Stem Cell Reports*, *3*(3), 489-501.

Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., Frampton, J., Zhao, S., & Medvinsky, A. (2011). Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J Exp Med*, *208*(6), 1305-1315.

Solaimani Kartalaei, P., Yamada-Inagawa, T., Vink, C. S., de Pater, E., van der Linden, R., Marks-Bluth, J., van der Sloot, A., van den Hout, M., Yokomizo, T., van Schaick-Solernó, M. L., Delwel, R., Pimanda, J. E., van, I. W. F., & Dzierzak, E. (2015). Whole-transcriptome analysis of endothelial to hematopoietic stem cell transition reveals a requirement for Gpr56 in HSC generation. *J Exp Med*, *212*(1), 93-106.

Spinner, M. A., Sanchez, L. A., Hsu, A. P., Shaw, P. A., Zerbe, C. S., Calvo, K. R., Arthur, D. C., Gu, W., Gould, C. M., Brewer, C. C., Cowen, E. W., Freeman, A. F., Olivier, K. N., Uzel, G., Zelazny, A. M., Daub, J. R., Spalding, C. D., Claypool, R. J., Giri, N. K., Alter, B. P., Mace, E. M., Orange, J. S., Cuellar-Rodriguez, J., Hickstein, D. D., & Holland, S. M. (2014). GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood*, *123*(6), 809-821.

Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K. D., Yoshimoto, M., Yoder, M. C., Frampton, J., Schroeder, T., Enver, T., Göttgens, B., & de Bruijn, M. (2013). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat Commun*, *4*, 2924.

Taoudi, S., Gonneau, C., Moore, K., Sheridan, J. M., Blackburn, C. C., Taylor, E., & Medvinsky, A. (2008). Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-cadherin+CD45+ pre-definitive HSCs. *Cell Stem Cell*, *3*(1), 99-108.

Tipping, A. J., Pina, C., Castor, A., Hong, D., Rodrigues, N. P., Lazzari, L., May, G. E., Jacobsen, S. E., & Enver, T. (2009). High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood*, *113*(12), 2661-2672.

Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W., & Orkin, S. H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, *371*(6494), 221-226.

Walter, D., Lier, A., Geiselhart, A., Thalheimer, F. B., Huntscha, S., Sobotta, M. C., Moehrle, B., Brocks, D., Bayindir, I., Kaschutnig, P., Muedder, K., Klein, C., Jauch, A., Schroeder, T., Geiger, H., Dick, T. P., Holland-Letz, T., Schmezer, P., Lane, S. W., Rieger, M. A., Essers, M. A., Williams, D. A., Trumpp, A., & Milsom, M. D. (2015). Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature*, *520*(7548), 549-552.

Wang, H., Cui, B., Sun, H., Zhang, F., Rao, J., Wang, R., Zhao, S., Shen, S., & Liu, Y. (2021). Aberrant GATA2 Activation in Pediatric B-Cell Acute Lymphoblastic Leukemia. *Front Pediatr*, *9*, 795529.

Wilson, A., Laurenti, E., & Trumpp, A. (2009). Balancing dormant and self-renewing hematopoietic stem cells. *Curr Opin Genet Dev*, *19*(5), 461-468.

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P. M., Kinston, S., Ouwehand, W. H., Dzierzak, E., Pimanda, J. E., de Bruijn, M. F., & Göttgens, B. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, *7*(4), 532-544.

Wlodarski, M. W., Hirabayashi, S., Pastor, V., Starý, J., Hasle, H., Masetti, R., Dworzak, M., Schmugge, M., van den Heuvel-Eibrink, M., Ussowicz, M., De Moerloose, B., Catala, A., Smith, O. P., Sedlacek, P., Lankester, A. C., Zecca, M., Bordon, V., Matthes-Martin, S., Abrahamsson, J., Kühl, J. S., Sykora, K. W., Albert, M. H., Przychodzien, B., Maciejewski, J. P., Schwarz, S., Göhring, G., Schlegelberger, B., Cseh, A., Noellke, P., Yoshimi, A., Locatelli, F., Baumann, I., Strahm, B., Niemeyer, C. M., & Ewog, M. D. S. (2016). Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. *Blood*, *127*(11), 1387-1397; quiz 1518.

Yoshida, M., Tanase-Nakao, K., Shima, H., Shirai, R., Yoshida, K., Osumi, T., Deguchi, T., Mori, M., Arakawa, Y., Takagi, M., Miyamura, T., Sakaguchi, K., Toyoda, H., Ishida, H., Sakata, N., Imamura, T., Kawahara, Y., Morimoto, A., Koike, T., Yagasaki, H., Ito, S., Tomizawa, D., Kiyokawa, N., Narumi, S., & Kato, M. (2020). Prevalence of germline GATA2 and SAMD9/9L variants in paediatric haematological disorders with monosomy 7. *Br J Haematol*, *191*(5), 835-843.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M. L., Dayaram, T., Owens, B. M., Shigematsu, H., Levantini, E., Huettner, C. S., Lekstrom-Himes, J. A., Akashi, K., & Tenen, D. G. (2004). Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. *Immunity*, *21*(6), 853-863.

Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., Li, Z., Ni, Y., Mu, X., Wen, L., Cheng, T., Lan, Y., Yuan, W., Tang, F., & Liu, B. (2016). Tracing haematopoietic stem cell formation at single-cell resolution. *Nature*, *533*(7604), 487-492.

Zhou, T., Kinney, M. C., Scott, L. M., Zinkel, S. S., & Rebel, V. I. (2015). Revisiting the case for genetically engineered mouse models in human myelodysplastic syndrome research. *Blood*, *126*(9), 1057-1068.

A

Addendum

ENGLISH SUMMARY

Hematopoietic stem cells (HSCs) are the sole precursors of the hematopoietic system that can self-renew and differentiate into all mature blood cell types. The transcription factor (TF) *GATA2* is highly expressed in HSCs and regulates the HSC generation during embryonic development and HSC function throughout adulthood. In patients, heterozygous *GATA2* mutations cause GATA2 haploinsufficiency syndromes that present with a broad spectrum of phenotypes such as B-cell-, monocyte-, NK cell- and dendritic cell deficiencies, primary lymphedema, and more than 80% risk of developing myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). Although *GATA2* mutations evidently lead to hematopoietic defects in patients, the mechanisms underlying GATA2 deficiency phenotypes are relatively unknown. In this thesis, we examine the multifaceted role of *GATA2* throughout the formation and function of HSCs, the building blocks of the hematopoietic system, in both *Gata2*-mutant zebrafish and mouse models.

The first HSCs are generated through endothelial-to-hematopoietic transition (EHT) events during embryonic development. In mouse embryos, EHT events produce intra-aortic hematopoietic clusters (IAHCs), where HSCs are formed through a multistep maturation process characterized as pro-HSCs → pre-HSC1 → pre-HSC2/HSC. The maturation of HSCs through EHT is accompanied by progressive downregulation of endothelial transcriptional programming and upregulation of hematopoietic transcriptional programming. Although HSCs are severely reduced in *Gata2*$^{+/-}$ mouse embryos, the mechanism behind this reduction is incompletely understood. In **chapter 2**, we investigate the effect of Gata2 haploinsufficiency on the subpopulations of IAHCs and show that *Gata2*$^{+/-}$ embryos can produce pro-HSCs hence the hematopoietic programming is not abrogated. However, the maturation of pro-HSCs into pre-HSCs is significantly reduced in *Gata2*$^{+/-}$ embryos indicating *Gata2* is essential for the completion of EHT. To understand the mechanism hampering embryonic HSC maturation in *Gata2*$^{+/-}$, we look into the transcriptome of IAHC subpopulations and show that Gata2 haploinsufficiency reduces the expression of the endothelial repressor *Gfi1b* resulting in an incomplete repression of endothelial programming during EHT. Furthermore, we show that the ectopic expression of *gfi1b* in the hematopoietic cells rescues the number of embryonic HSCs in *gata2b*-deficient zebrafish suggesting that *Gata2* regulates *Gfi1b* and this regulation is crucial for the formation of HSCs during EHT.

Zebrafish have two orthologues of mammalian *Gata2*, i.e., *gata2a* and *gata2b*, that are subfunctionalized in the lymphovascular system and hematopoietic system respectively, allowing us to investigate the role of *Gata2* in different cellular components. By generating homozygous *gata2b* knockout (**chapter 3**) and heterozygous *gata2b* knockout (**chapter 4**) zebrafish, we dissect the role of *Gata2* in the hematopoietic system without interfering with its role in the lymphovascular system and circulation. In **chapter 3**, we show that

homozygous deletion of *gata2b* (*gata2b*$^{-/-}$) attenuates the expansion of HSCs during embryonic stages and impairs balanced lineage differentiation from the hematopoietic stem and progenitor cells (HSPCs) during adulthood. Using single-cell RNA sequencing (scRNA-seq) we show that the most immature HSPCs of *gata2b*$^{-/-}$ kidney marrow (KM) co-express myeloid and lymphoid specific genes resulting in reduced numbers of neutrophils and an incomplete B-cell differentiation. In **chapter 4**, we reveal that the heterozygous *gata2b* knockout (*gata2b*$^{+/-}$) HSPCs have normal lineage output; however, we observe erythro-myeloid dysplasia in the *gata2b*$^{+/-}$ KM. Furthermore, using scRNA-seq, we identify an aberrant *gata1a* expression in a subgroup of erythroid progenitors suggesting an impaired 'GATA switch' process contributes to the erythroid dysplasia in *gata2b*$^{+/-}$ KM. In **chapter 5**, we investigate the function of a conserved enhancer region located in the 4th intron (i4) of *gata2a* that corresponds to the +9.5 enhancer region of mammalian *Gata2* locus. Complete deletion of this *gata2a* enhancer locus (*gata2a*$^{i4/i4}$) downregulates both *gata2a* and *gata2b* and temporarily impairs HSPC emergence during EHT. Despite both the expression level of *gata2b* and the number of HSPCs are restored through the activation of Notch signaling by 48 hours post fertilization (hpf), adult *gata2a*$^{i4/i4}$ zebrafish have an increase susceptibility to infections, oedema, neutropenia and hypocellular KM. Each *Gata2*-mutant zebrafish model we characterize (**chapter 3-5**) resembles the phenotype of a subgroup of *GATA2* patients, suggesting *GATA2* dosage and cell type-specific expression pattern of *GATA2* contributes to the phenotypic diversity observed in GATA2 deficiency syndromes.

The risk of developing MDS/AML in GATA2 patients increases from 6% at the age 10 to 81% at the age 40 indicating aging deteriorates the phenotypic consequences of GATA2 deficiency syndromes. Although the ability of *Gata2*$^{+/-}$ HSCs to differentiate into lymphoid-lineage is reduced, the contribution of aging to this functional decline is unexplored. In **chapter 6**, we investigate the effect of Gata2 haploinsufficiency during the aging of mice and show that aged-*Gata2*$^{+/-}$ HSCs have decreased reconstitution ability upon their transplantation resulting in B-cell cytopenia and monocytopenia, which resembles the phenotype of a subgroup of GATA2 patients. To understand the mechanisms leading to the functional decline of aged-*Gata2*$^{+/-}$ HSCs, we investigate the transcriptome of *Gata2*$^{+/-}$ HSCs from embryonic development and throughout aging. Our results show that *Gata2*$^{+/-}$ HSCs lose quiescence during embryonic stages and remain proliferative throughout aging. Furthermore, the accumulation of double-strand breaks (DSBs) and the aberrant inflammatory transcriptomic signatures acquired in aged-*Gata2*$^{+/-}$ HSCs upon their transplantation suggest that Gata2 haploinsufficiency results in genome instability in aged-HSCs.

Due to the lack of mechanistic insights behind leukemia development in GATA2 deficiency syndromes, the only treatment option for GATA2 patients is the allogeneic HSC transplantation (allo-HSCT). However, not every GATA2 patient is suitable for or successfully responds to allo-HSCT and this treatment option often becomes life-threatening. In the last part of the thesis (**chapter 7)**, we explore the advantages and pitfalls of current genome editing technologies to propose novel strategies for correcting the mutant *GATA2* allele in the own HSCs of GATA2 patients, in order to provide autologous HSC transplantation (auto-HSCT) as a treatment for GATA2 deficiency syndromes in the future.

In conclusion, this thesis elucidates the multifaceted role of *GATA2* in HSC biology. By dissecting the role of *Gata2* spatiotemporally in various *Gata2*-mutant mouse and zebrafish models we show that both formation of HSCs during embryonic development and crucial functions in HSCs throughout adulthood and aging, such as the quiescence, cell-fate commitment, lineage differentiation and reconstitution is regulated by *GATA2*-dependent mechanisms.

NEDERLANDSE SAMENVATTING

Hematopoëtische stamcellen (HSC's) zijn de enige voorlopers van het hematopoëtische systeem die zichzelf kunnen vernieuwen en differentiëren tot alle volwassen bloedcellen. De transcriptiefactor (TF) *GATA2* komt sterk tot expressie in HSC's en reguleert de HSC-generatie tijdens de embryonale ontwikkeling en HSC-functie gedurende de volwassenheid. Bij patiënten veroorzaken heterozygote *GATA2*-mutaties GATA2-haplo-insufficiëntiesyndromen die zich presenteren met een breed spectrum van fenotypes zoals B-cel-, monocyt-, NK-cel- en dendritische cel deficiënties, primair lymfoedeem en een risico van meer dan 80% op het ontwikkelen van myelodysplastisch syndroom (MDS) en acute myeloïde leukemie (AML). Hoewel *GATA2*-mutaties duidelijk leiden tot hematopoëtische defecten bij patiënten, zijn de mechanismen die ten grondslag liggen aan fenotypes van *GATA2*-deficiëntie relatief onbekend. In dit proefschrift onderzoeken we de veelzijdige rol van *GATA2* in de vorming en functie van HSC's, de bouwstenen van het hematopoëtische systeem, in zowel *Gata2*-mutante zebravis- als muismodellen.

De eerste HSC's worden gegenereerd door endotheliale naar hematopoëtische transitie (EHT) gebeurtenissen tijdens de embryonale ontwikkeling. In muizenembryo's produceren EHT-gebeurtenissen intra-aortische hematopoëtische clusters (IAHC's), waar HSC's worden gevormd door een meerstaps rijpingsproces dat wordt gekenmerkt als pro-HSC's → pre-HSC1 → pre-HSC2/HSC. De rijping van HSC's door EHT gaat gepaard met progressieve neerwaartse regulatie van endotheliale transcriptionele programmering en opregulatie van hematopoëtische transcriptionele programmering. Hoewel HSC's ernstig zijn verminderd in *Gata2*$^{+/-}$ muizenembryo's, wordt het mechanisme achter deze reductie niet volledig begrepen. In **hoofdstuk 2** onderzoeken we het effect van *Gata2* haplo-insufficiëntie op de subpopulaties van IAHC's en laten we zien dat *Gata2*$^{+/-}$ embryo's pro-HSC's kunnen produceren en daarom wordt de hematopoëtische programmering niet opgeheven. De rijping van pro-HSC's tot pre-HSC's is echter significant verminderd in *Gata2*$^{+/-}$ embryo's, wat aangeeft dat *Gata2* essentieel is voor de voltooiing van EHT. Om het mechanisme te begrijpen dat de embryonale HSC-rijping in *Gata2*$^{+/-}$ belemmert, kijken we naar het transcriptoom van IAHC-subpopulaties en laten we zien dat *Gata2*-haplo-insufficiëntie de expressie van de endotheliale repressor *Gfi1b* vermindert, wat resulteert in een onvolledige onderdrukking van endotheelprogrammering tijdens EHT. Verder laten we zien dat de ectopische expressie van *gfi1b* in de hematopoëtische cellen het aantal embryonale HSC's in *gata2b*-deficiënte zebravissen redt, wat suggereert dat *Gata2 Gfi1b* reguleert en dat deze regulatie cruciaal is voor de vorming van HSC's tijdens EHT.

Zebravissen hebben twee orthologen van *Gata2*, d.w.z. *gata2a* en *gata2b*, die elk hun functie vervullen in respectievelijk het lymfovasculaire systeem en het hematopoëtische systeem,

waardoor we de rol van *Gata2* in verschillende cellulaire componenten kunnen onderzoeken. Door homozygote *gata2b* knock-out (**hoofdstuk 3**) en heterozygote *gata2b* knock-out (**hoofdstuk 4**) zebravis te genereren, ontleden we de rol van *Gata2* in het hematopoëtische systeem zonder zijn rol in het lymfovasculaire systeem en de bloedsomloop te verstoren. In **hoofdstuk 3** laten we zien dat homozygote deletie van *gata2b* (*Gata2*$^{-/-}$) de expansie van HSC's tijdens de embryonale stadia afzwakt en de evenwichtige afstammingsdifferentiatie van de hematopoëtische stam- en progenitorcellen (HSPC's) tijdens de volwassenheid schaadt. Met behulp van single-cell RNA-sequencing (scRNA-seq) laten we zien dat de meest onrijpe HSPC's van *gata2b*$^{-/-}$ niermerg myeloïde en lymfoïde-specifieke genen tot co-expressie brengen, wat resulteert in een verminderd aantal neutrofielen en een onvolledige B-celdifferentiatie. In **hoofdstuk 4** onthullen we dat de heterozygote *gata2b* knock-out (*gata2b*$^{+/-}$) HSPC's een normale afstammingsoutput hebben; we nemen echter erythro-myeloïde dysplasie waar in de *gata2b*$^{+/-}$ niermerg. Verder identificeren we met behulp van scRNA-seq een afwijkende gata1a-expressie in een subgroep van erytroïde voorlopers, wat suggereert dat een verstoord 'GATA-switch'-proces bijdraagt aan de erytroïde dysplasie in gata2b$^{+/-}$ niermerg. In **hoofdstuk 5** onderzoeken we de functie van een geconserveerd enhancer-gebied in het 4e intron (i4) van *gata2a* dat overeenkomt met het +9.5-enhancer-gebied van de *Gata2*-locus van zoogdieren. Volledige deletie van deze *gata2a*-enhancer-locus (*gata2a*$^{i4/i4}$) reguleert zowel *gata2a* als *gata2b* neer en schaadt tijdelijk de opkomst van HSPC tijdens EHT. Ondanks dat zowel het expressieniveau van gata2b als het aantal HSPC's wordt hersteld door de activering van Notch-signalering 48 uur na de bevruchting, hebben volwassen *gata2a*^{i4}-zebravissen een verhoogde gevoeligheid voor infecties, oedeem, neutropenie en hypocellulaire niermerg. Elk *Gata2*-mutant zebravismodel dat we karakteriseren (**hoofdstuk 3-5**) lijkt op het fenotype van een subgroep van *GATA2*-patiënten, wat suggereert dat de *GATA2*-dosering en het celtype-specifieke expressiepatroon van *GATA2* bijdraagt aan de fenotypische diversiteit die wordt waargenomen bij *GATA2*-deficiëntiesyndromen.

Het risico op het ontwikkelen van MDS/AML bij *GATA2*-patiënten neemt toe van 6% op 10-jarige leeftijd tot 81% op 40-jarige leeftijd, wat aangeeft dat veroudering de fenotypische gevolgen van *GATA2*-deficiëntiesyndromen verslechtert. Hoewel het vermogen van *Gata2*$^{+/-}$ HSC's om te differentiëren in lymfoïde lijn is verminderd, is de bijdrage van veroudering aan deze functionele achteruitgang onontgonnen. In **hoofdstuk 6** onderzoeken we het effect van *Gata2*-haplo-insufficiëntie tijdens het ouder worden van muizen en laten we zien dat verouderde *Gata2*$^{+/-}$ HSC's een verminderd reconstitutievermogen hebben na hun transplantatie, wat resulteert in B-cel cytopenie en monocytopenie, wat lijkt op het fenotype van een subgroep van *GATA2* patiënten. Om de mechanismen te begrijpen die leiden tot de functionele achteruitgang van verouderde-*Gata2*$^{+/-}$ HSC's, onderzoeken we het transcriptoom van *Gata2*$^{+/-}$ HSC's van embryonale ontwikkeling en tijdens veroudering.

Onze resultaten laten zien dat *Gata2*$^{+/-}$ HSC's hun rust verliezen tijdens de embryonale stadia en proliferatief blijven tijdens het ouder worden. Bovendien suggereren de accumulatie van dubbelstrengs breuken (DSB's) en de afwijkende inflammatoire transcriptomische handtekeningen die zijn verkregen in verouderde *Gata2*$^{+/-}$ HSC's na hun transplantatie dat *Gata2*-haplo-insufficiëntie resulteert in genoominstabiliteit in verouderde HSC's.

Vanwege het gebrek aan mechanistische inzichten achter de ontwikkeling van leukemie bij *GATA2*-deficiëntiesyndromen, is de enige behandelingsoptie voor *GATA2*-patiënten de allogene HSC-transplantatie (allo-HSCT). Niet elke *GATA2*-patiënt is echter geschikt voor of reageert succesvol op allo-HSCT en deze behandeloptie wordt vaak levensbedreigend. In het laatste deel van het proefschrift (**hoofdstuk 7**) onderzoeken we de voordelen en valkuilen van de huidige technologieën voor genoombewerking om nieuwe strategieën voor te stellen voor het corrigeren van het mutante *GATA2*-allel in de eigen HSC's van *GATA2*-patiënten, om zo autologe HSC-transplantatie mogelijk te maken als een behandeling voor *GATA2*-deficiëntiesyndromen in de toekomst.

Concluderend, dit proefschrift verheldert de veelzijdige rol van *GATA2* in de HSC-biologie. Door de rol van *Gata2* spatiotemporeel te ontleden in verschillende Gata2-mutante muis- en zebravismodellen, laten we zien dat zowel de vorming van HSC's tijdens de embryonale ontwikkeling als cruciale functies in HSC's gedurende de volwassenheid en veroudering, zoals de rust, cel-lotbetrokkenheid, afstammingsdifferentiatie en reconstitutie wordt gereguleerd door *GATA2*-afhankelijke mechanismen.

LIST OF MOST IMPORTANT ABBREVIATIONS

AA	Aplastic anemia
AGM	Aorta-gonad-mesonephros
Allo-SCT	Allogeneic hematopoietic stem cell transplantation
Auto-SCT	Autologous hematopoietic stem cell transplantation
AML	Acute myeloid leukemia
BM	Bone marrow
BMF	Bone marrow failure
CFU	Colony-forming unit
CH	Clonal hematopoiesis
CHT	Caudal hematopoietic tissue
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
DA	Dorsal aorta
DC	Dendritic cell
DCML	Dendritic cell, monocyte, B and neutral kille lymphoid cell deficiency
Dpf	Day post fertilization
DSB	Double strand break
E	Embryonic day
EC	Endothelial cell
EHT	Endothelial-to-hematopoietic transition
ES	Enrichment score
FACS	Fluorescence activated cell sorting
FC	Fold change
FCS	Fetal calf serum
FDR	False discovery rate
FISH	Fluorescent in situ hybridization
FL	Fetal liver
FPKM	Fragments per kilobase of exon per million fragments mapped
FWD	Forward
GFP	Green fluorescent protein
GMP	Granulocyte-monocyte progenitors
GO	Gene ontology
GSEA	Gene set enrichment analysis
γH2AX	Ser139-phosphorylated H2AX histone
HCT	Hematocrit
HDR	Homology directed repair
HE	Hemogenic endothelium

HEC	Hemogenic endothelial cell
hESC	Human embryonic stem cell
HGB	Hemoglobin
HPC	Hematopoietic progenitor cell
Hpf	Hour post fertilization
HPV	Human papillomavirus
HSC	Hematopoietic stem cell
HSPC	Hematopoietic stem and progenitor cell
IAHC	Intra-aortic hematopoietic cluster
InDel	Insertion/deletion
ISH	In situ hybridization
KM	Kidney marrow
MCH	Mean corpuscular hemoglobin
MCHC	Mean corpuscular hemoglobin concentration
MCV	Mean corpuscular volume
MDS	Myelodysplastic syndrome
MFI	Mean fluorescent intensity
MGG	May-Grünwald Giemsa
MO	Morpholino oligonucleotide
MonoMAC	Monocytopenia and Mycobacterium Avium infection
MPP	Multipotent progenitor
NC	Notochord
NHEJ	Non-homologous end joining
NK	Natural killer cell
NT	Neural tube
OTE	Off-target effect
PB	Peripheral blood
PBS	Phosphate buffer solution
PC	Principal component
PCA	Principal component analysis
PCV	Posterior cardinal vein
PFA	Paraformaldehyde
PLM	Posterior lateral mesoderm
PLT	Platelet
PTU	1-phenyl-2-thiourea
RBC	Red blood cell
REV	Reverse
RFP	Red fluorescent protein
RNP	Ribonucleoprotein
RT	Room temperature
SBC	Single bulging cell
Seq	Sequencing
SPF	Specific pathogen-free
TF	Transcription factor
TRM	Treatment related morbidity and mortality
WBC	White blood cell
WT	Wild type
ZF	Zinc finger

CURRICULUM VITAE

Cansu Koyunlar was born on the 1st of January 1991 in Karsiyaka, Turkey. After receiving her high school diploma from Ayranci Anatolian High School (Ankara, Turkey) Science/ Mathematics program in 2009, she studied Biological Sciences at Hacettepe University (Ankara, Turkey) and graduated in 2014 with a GPA of 3.43/4.00. During her undergraduate education, she did an internship in Nephrogenetics Laboratory at the Hacettepe University Medical School (2012-2014). Later, she pursued her master degree in the research group of Prof.dr. Pervin Dincer at the Medical Biology Department of Hacettepe University Medical School. In 2017, she defended her thesis titled “Examination of desmin expression on *desma* and *desmb* knockout zebrafish models” and graduated with a GPA of 3.94/4.00. In 2017, she moved to the Netherlands, where she was appointed as a PhD candidate at the Department of Hematology of Erasmus Medical Center (Rotterdam). Here, she worked in the research group of dr. de Pater, and focused on the role of the transcription factor Gata2 during embryonic and adult hematopoiesis.

LIST OF PUBLICATIONS

Cansu Koyunlar*, Emanuele Gioacchino*, Joke Zink, Hans de Looper, Madelon de Jong, Tomasz Dobrzycki, Christopher B. Mahony, Remco Hoogenboezem, Dennis Bosch, Paulina M. H. van Strien, Martin E. van Royen, Pim J. French, Eric Bindels, Kirsten J. Gussinklo, Rui Monteiro, Ivo P. Touw, Emma de Pater**.** Essential role for Gata2 in modulating lineage output from hematopoietic stem cells in zebrafish. *Blood Adv* (2021) 5(13):2687-2700.

Gülsüm Kayman Kürekçi, Ecem Kural Mangit, **Cansu Koyunlar**, Seyda Unsal, Berk Saglam, Bora Ergin, Merve Gizer, Ismail Uyanik, Niloufar Boustanabadimaralan Düz, Petek Korkusuz, Beril Talim, Nuhan Purali, Simon M. Hughes & Pervin R. Dincer. Knockout of zebrafish desmin genes does not cause skeletal muscle degeneration but alters calcium flux. *Sci Rep* (2021) 11, 7505.

Cansu Koyunlar and Emma de Pater. From Basic Biology to Patient Mutational Spectra of GATA2 Haploinsufficiencies: What Are the Mechanisms, Hurdles, and Prospects of Genome Editing for Treatment. *Front. Genome Ed.* (2020) 2:602182.

Tomasz Dobrzycki, Christopher B. Mahony, Monika Krecsmarik, **Cansu Koyunlar**, Rossella Rispoli, Joke Peulen-Zink, Kirsten Gussinklo, Bakhta Fedlaoui, Emma de Pater, Roger Patient & Rui Monteiro. Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis. *Commun Biol* (2020) 3, 71.

Ekim Z. Taskiran, Emine Korkmaz, Safak Gucer, Can Kosukcu, Figen Kaymaz, **Cansu Koyunlar,** Elizabeth C. Bryda, Moumita Chaki, Dongmei Lu, Komal Vadnagara, Cengiz Candan, Rezan Topaloglu, Franz Schaefer, Massimo Attanasio, Carsten Bergmann and Fatih Ozaltin. Mutations in ANKS6 cause a nephronophthisis-like phenotype with ESRD. *JASN* (2014) 25 (8) 1653-1661.

***** Equal contribution

PHD PORTFOLIO

Name PhD Student: **Cansu Koyunlar**
Erasmus MC Department: **Hematology**
Research School: **Molecular Medicine (MolMed)**

PhD Period: **June 2017 - February 2022**
Promoter: **Prof. Dr. I.P. Touw**
Supervisor: **Dr. Emma de Pater**

1. PhD Training

	Year	ECTS
Courses and workshops		
Laboratory Animal Science Course (Art.9)	2017	3.0
Species-Specific Small Rodents Course	2017	0.6
Species-Specific Fish Course (KNAW)	2017	0.6
Course and Master Class on Molecular Aspects of Hematological Disorders	2017	0.7
Annual Course on Molecular Medicine	2018	0.7
Introduction in GraphPad Prism 6	2018	0.3
Workshop on Molecular Aspects of Hematological Malignancies	2018	0.7
Basic and Translational Oncology	2018	1.8
Basic Course on R	2019	2.0
Workshop on Photoshop and Illustrator CC	2020	0.3
Data Analysis in Python Basic	2021	1.5
Course on Research Integrity	2021	0.3
Presentations		
13th Dutch Hematology Congress, Arnhem, Netherlands (Oral)	2019	1.0
24th European Hematology Association Congress, Amsterdam, Netherlands (Oral)	2019	1.0
International Society for Experimental Hematology Virtual Scientific Meeting (Poster)	2020	1.0
European Hematology Association Virtual Congress (Oral)	2021	1.0
Journal Club (Oral, 3x) (Rotterdam)	2017-2020	1.5
AIO/PostDoc meeting (Oral, 5x) (Rotterdam)	2017-2021	2.5
Work discussion (Oral, 10x) (Rotterdam)	2017-2021	5.0
Single-cell meeting (Oral) (Rotterdam)	2021	1.0
Zebrafish users meeting (Oral) (Rotterdam)	2021	1.0
Attendance in national/international meetings		
Molecular Medicine Day	2018	1.0
5th Applied Synthetic Biotechnology in Europe Virtual Scientific Meeting	2020	1.0

	Year	ECTS
Attendance in scientific meetings at Department of Hematology		
Work discussion (Weekly)	2017-2022	3.0
Journal Club (Bi-monthly)	2017-2020	1.0
PhD lunch with seminar speaker (Monthly)	2017-2020	1.0
Erasmus Hematology Lectures (Monthly)	2017-2020	1.0
Virtual Erasmus Hematology Lecture Series (Weekly)	2021-2022	1.0
2. eaching, Supervision & Organization Activities		
Supervising master's theses		
Master Student Infection & Immunity (3x)	2018-2021	9.0
Supervising practicals		
Organization and supervision of PhD lunch with seminar speakers	2019-2020	1.0
Total		46.5
Awards		
European Hematology Association Congress Participation Grant	2021	

WORD OF THANKS

I have learned so much from everyone I have met, discussed science, and spent time with over these years, and it has made me the person I am today. Thank you all for making my PhD experience a wonderful one beyond my imagination!

My warmest thanks to my supervisor **dr. Emma de Pater**. Dear **Emma**, I feel extremely happy to have met you and for completing my PhD under your supervision. Thank you for believing in me and giving me the opportunity to be a member of your team. You are an amazing scientist, a wonderful person and the best mentor! I have always felt your full support over the years. Thank you for training me, for your vision, making our research and this thesis possible and for feeding us regularly with candies! I will not only miss the excitement and joy in our scientific discussions, but also your positive energy, even on blanket days! I really hope to work with you again in the future, and cannot wait to celebrate when you become the coolest professor!

A heartfelt thanks to my promoter **Prof.dr. Ivo Touw**. Dear **Ivo**, no doubt you are one of the most influential scientists to me. Your passion for science has inspired me and I have learned so much from you! I benefited from your knowledge immensely over the years and am deeply grateful that you always provided me with the needed perspective on my research by pinpointing the strengths and weaknesses in a very constructive way. Thank you very much for your contribution to my scientific improvement, to this thesis and thank you for being my mentor!

I would like to thank **Prof.dr. Ruud Delwel**, **Prof.dr. Sjaak Philipsen** and **Prof.dr. Nancy Speck** for their time on reading my dissertation.

Big thanks to **dr. Rui Monteiro**, **dr. Miriam Erlacher** and **dr. Juncal Fernandez Orth** for our fruitful collaborations over the years. It was wonderful to have the chance to work with you on our projects.

A warm thanks to the PIs of the Hematology Department. Dear **Ruud**, **Marc**, **Ruben**, **Tom**, **Peter**, **Frank**, **Mathijs**, **Mojca**, **Jan**, **Bob** and **Moniek** and **Rebecca**, over the years your suggestions during our meetings and during many spontaneous conversations continuously improved the quality of our research and encouraged me to improve myself as a researcher. I am extremely proud and happy for completing my PhD in this department, amongst this wonderful group of scientists.

Huge thanks to the members of de Pater group! We had a wonderful team together and it was an absolute blast to work with you. I hope to catch up with you at our annual BBQs!

I owe a huge amount of gratitude to the members of the Hematology Department, not only for their experimental and theoretical support but also for bringing joy to my daily life over the years.

Dear **Emanuele**, I feel extremely lucky that I got to work side by side with one of my best friends throughout these years. I think I cannot thank you enough in any given space or with any words. I feel absolutely grateful not only for your support in the lab, but also for introducing me to the wonderful Rotterdam family as soon as I arrived, being there for me for all the good and hard times. We had an amazing work/life balance as a team: discussing how Gata2 works during the day (sometimes in the evening too) and discussing where to hang out after work. I really miss seeing you every day!

Dear **Hans**, **Dennis**, **Joke** and **Mariette**, thank you all not only for your extremely vital support and hard work on our projects, but also for all the fun we had in the lab. I absolutely miss spending time in the lab with you guys!

Dear **Hans**, you have trained me extensively and have undoubtedly given me one of the greatest support for almost all the experiments in the lab! I am deeply grateful that you always took time to teach me how to think when planning experiments, rather than just giving me an answer, which helped me to build confidence to become more independent in the lab. Also, thank you for brewing the most amazing beers! To date, I still cannot forget the taste of the 'Beermuda triangle'!

Dear **Dennis**, thank you so much for your support, especially when genotyping ridiculous amounts of embryos! As you know, I love your music taste and am very happy that you always shared cool songs and documentaries with me. I still remember when you sent me 'Take a walk' by Masta Ace when I was going for a walk on a beautiful sunny day! I want to organize an old school hip hop party with you!

Dear **Joke**, thank you so much for all the fish work and for always fish-nerding with me! I am proud of you being (almost) able to celebrate birthdays now, after finding you hiding in the dissection room the very first year haha! Please continue sending me whenever the results are CUTE. #exciting #neversaygoodbye

Dear **Mariette**, thank you so much for taking care of the mouse work and for your support on long experiment days! I could not manage this without you. Since I found out that you wrote those cool stories, I am impressed by you even more!

Many thanks to **Disha**, **Jonathan** and **Fabienne** for their contribution to our projects while completing their graduate internship in our group. Your fresh ideas gave me motivation, and it was a pleasure working with you in the lab. In return, I learned a lot from you! I wish you the best for your future careers and I am sure you will shine!

Dear **Onno**, the kindest and the most positive spirit on earth! Thank you for being this lovely person you are, supporting me as everyone else around you! I feel very lucky that I got to see you every day in the lab, so that I can get the daily dose of positivity, next to my coffee, to get through PhD days. Also, thank you for your support on stainings and microscopy. But next to that, I am really grateful for our friendship that continuously nourishes me every day. From the very beginning, I knew that I wanted you and Thomas to stand next to me when defending my thesis, and I cannot wait for this to become reality!

Dear **Ping**, I love talking to you! Whenever I felt stuck, both during my PhD and personal life, I always appreciated you listening to me and giving your best to help me out. You are the trustworthy person that I can always count on when I need an honest opinion, and definitely are my motivational gru! When you graduated and were not in the office anymore, I kept looking at your printed face on the office wall and whenever I needed to take a risky decision I asked myself: What would Ping do? I feel really happy that I can still meet you regularly in Utrecht!

Dear **Keano**, your smile gives me so much energy and I always felt extremely happy when I entered the office and said 'Keanoooooo' and you said 'Cansuuuuuu' back! Thank you for being you and for being a wonderful friend! Although it was hard for me to get used to not being around you guys when you all graduated, I always felt your support whenever I got to see you! Apart from being no doubt among my best friends in Rotterdam, you, **Ping** and **Emanuele** made my office life the most amazing one, and my whole PhD experience unforgettable!

Dear **Mira**, thank you for being the warmest and most welcoming person when I just started in the lab. I am really grateful for our long lunch breaks and cannot forget the moment when our hamburgers touched! But, no doubt, the actual turning point of our friendship was when you cleaned my hair in that little park after finding Mr. Pancake! Thank you for being the kindest, discovering Rotterdam with me, and pretending not to know Dutch, so that we could be tourists together. I am sure you will do great when finishing up your PhD. I cannot wait to celebrate your graduation!

Dear **Sensei Eric**, our and many others' research in the department would not be possible without your hard work! Thank you so much not only for being a support to us in the lab, but also for your vibrant energy! I absolutely miss working in the same bench corridor with

you and to hear you singing! PS: I hope that one day I could have a Nature paper with you, but I guess for now you can have my bench!

Dear **Remco**, **Roger** and **Chiel**, thank you all so much for your time to provide the bioinformatics support! Without your help and hard work, our research and this thesis would not be possible! Dear **Marije**, thank you for teaching me the ATAC-Seq protocol. It was really fun to spend time with you in the lab! Dear **Micheal** and **Claire**, thank you so much for your help with flow cytometry. Thank you for your patience when I was (always) late for my sorting time! I miss the long chats we had during those sorting experiments. Dear **Jaqueline**, thank you for helping me setting up the mouse peripheral blood staining, for sharing the antibodies with me and for all of our discussions about the projects! I wish you the best of luck finishing up your dissertation and cannot wait to be there for your graduation! Dear **Madelon**, thank you so much for the script you provided for the single cell and pseudotime analysis. I am sure you will do wonderful finishing up your dissertation and will rock the world! Dear **Lanpeng**, thank you for helping me setting up the senescence staining and providing me with the antibodies. Best of luck for the future! Dear **Egied**, thank you so much for helping me with the preparation of the thesis in such a short time! Dear **Eline**, it was wonderful to share the office with you and our life has significantly changed when we had our own coffee machine! Thank you so much for the chats we had! I have no doubt that you will do great in the future and I wish you all the best finishing up your thesis! Dear **Paola**, my cutie, thank you so much for all the positive energy and for our motivational talks about the future in science. I am sure you will do great finishing up your thesis and will do wonderful things in the future, wherever you would like to be! Your new colleagues will be very lucky to have you around.

A big thanks to **Paulette**, **Patricia**, **Leonie**, **Stanley**, **Iris**, **Zoltan**, **Dorien**, **Sabrin**, **Natalie**, **Mark**, **Cathelijne**, **Devine**, **Martijn**, **Sjoerd**, **Andrea**, **Isabel**, **Maurice**, **Elwin**, **Tessa**, **Leenke** and to all the remaining members of the Hematology Department. Thank you all for your help and support and for making these years unforgettable! I wish you all the best and wish to catch up whenever I have the chance!

I owe a tremendous gratitude to my friends in Rotterdam, who were the biggest support over the years and became my family.

Dear **Thomas**, my B, C, H, W, P, O and every letter of the alphabet. Thank you for being the best friend ever. Thank you for always being there for me, for your honesty, for understanding me sometimes without even using words. I always felt comfort with you, since the day we met, not only on my best days but even on my worst days. I trust you with all my heart, I feel that I have grown with you, and am grateful for your existence every day. Thank you for your help, not only for my thesis and graduation, but also for whenever I needed someone to lean on. Thank you for all of our adventures, and I cannot wait for the future prnndrn brnndrn days of our lives!

Dear **Raz**, we shared a wonderful time together, full of adventures and fun that has transformed us into who we are now. I am truly grateful for our experience, for your support during the whole period of my PhD, and for everything we shared over the years. I want to see you happy, doing what you love and seeing you achieve the success you deserved!

Dear **Lina**, we not only share the same heart structure, but also a HUMONGOUS heartfelt friendship! I am extremely grateful for your support in everything I have experienced over the years, and I would have felt incomplete without you! I cannot wait to see you give the most wonderful lectures and impress everyone in the room! PS: I love you! Dear **Sebo**, thank you for being the amazing person you are and an amazing friend. Your positive energy always brings positivity and happinesses to me whenever I see you. I love that you love scifi and am really impressed by your musical talent. PS: We seriously need to merge our birthdays and make that massive party happen! Together **Lina & Sebo** are my favorite couple. You guys are one of these rare couples that makes me appreciate that love can bring up better than the best of two amazing people.

Dear lovely **Ala**, thank you for your amazing genuine spirit. You are a beautiful person and I am really happy to call you my friend. You always give your full support to me and I believe that you are a super woman who can achieve anything she wants! PS: I really hope one day we can go to that amazing Polish festival you love!

Dear **Livia**, thank you so much for all the fun times and adventures we had. I cannot forget the time we spent together in Rome, and I do miss seeing you often. Then whenever I meet you again, I always feel that you actually never left. Thank you so much for the positive vibes and for your support!

A massive thanks to the Rotterdam family for being amazing people, and for all the fun times we had! Dear **Ivas**, **Theresa**, **Enrico**, **Giorgia**, **Frank**, **Lorenzo**, **Elena**, **Anna**, **Gizem**, **Gunce**, **Matteo**, **Francesca**, **Roberto**, **Kasia**, **Tania**, **Alex**, **Lena**, **Roberto**, **Ian**, **Irene**, **Isabona**, **Pablo**, **Rodrigo**, **Theodora**, **Peter**, **Jana**, **Agnese**, **Christina**, **Adrian**, **Francois, Monica** and all the remaining members of the family, thank you for being the warmest and loveliest people ever. Love you all!

This part would be incomplete without mentioning the **AvO lab** members. Thank you all for being the most welcoming and the absolute best group of people! Thank you for your support during the last few months, and making my time at the lab and Utrecht wonderful! I am extremely happy to be in your team!

My warmest gratitude for my **parents** and my **brother** who raised me and supported me no matter what. Thank you for your unconditional love, being the coolest, believing in me and encouraging me to be myself. I feel extremely lucky to have you in my life and will always love you with all my heart!