Consequences of Gata2 Deficiency

Emanuele Gioacchino
Consequences of Gata2 Deficiency

Gevolgen van Gata2 Tekort

Thesis

to obtain the degree of Doctor from the

Erasmus University Rotterdam

by command of the
rector magnificus

Prof.dr. F.A. van der Duijn Schouten

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

Tuesday 14 September 2021 at 13.00hrs

by

Emanuele Gioacchino

born in Urbino (PU), Italy.

Copyright © E. Gioacchino 2021

All rights reserved. No part of this thesis may be reproduced, stored in retrieval systems or transmitted in any form by any means without permission from the author. The copyrights of the articles that have been published or accepted for publication have been transferred to the respective journals.

ISBN:
Layout: Egied Simons
Cover: Josephine Bergmann.
Printing: Ridderprint

The work described in this thesis was performed at the Department of Hematology at the Erasmus Medical Center, Rotterdam, the Netherlands. This work has been financially supported by grants and fellowships from the Dutch Cancer Society “Koningin Wilhelmina Fonds” KWO.

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

Promotors: Prof.dr. I.P. Touw

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

Co-promoter: dr. E.M. De Pater
General introduction
1.1 Hematopoiesis and stem cells

Hematopoiesis is the process of blood cell production. Blood cells originate from hematopoietic stem cells (HSCs), which are characterized by their ability to self-renew and to differentiate into all blood cell types (Ogawa, 1993; Till, J.E. McCulloch, 1961). Adult HSCs reside almost exclusively in the bone marrow (BM), replicate with a low frequency, and are responsible for the production of all differentiated blood cells via intermediate cells called hematopoietic progenitor cells (HPCs) (Akashi et al., 2000; Kondo et al., 1997; Sun et al., 2014). Unlike the HSCs, HPCs are highly proliferative cells with no or limited self-renewal; they possess a more restricted lineage differentiation capacity and ensure a continuous supply of blood cells, both under homeostatic conditions and in episodes of hematological stress, such as blood cell loss or bacterial infections. The cell cycle of hematopoietic stem and progenitor cells (HSPCs) is controlled by the synergistic activity of extrinsic signals from the BM niche and hematopoietic cytokines and cell intrinsic pathways, mainly directed through transcription factors (TF) (Laurenti and Göttgens, 2018).

Mature blood cells are specialized in performing specific tasks in the organism like immune response, wound healing, oxygenation. Most blood cells have a limited life-span, lasting from a few hours (platelets) to several weeks (red blood cells). The two main blood cell categories are myeloid and lymphoid. The myeloid compartment comprises the cells of the innate immune response (neutrophils, eosinophils, and basophils), macrophages, thrombocytes, mast cells and erythrocytes. The lymphoid compartment is made up of cells responsible for the adaptive immune response, including B and T cells. In humans more than a million differentiated blood cells are produced every second. This continuous blood production requires a tight regulation. The proper control of proliferation and/or differentiation is a key aspect for a balanced and healthy blood system. This is highlighted when HSPCs struggle to differentiate into specific cell lineages or maintain a long-term self-renewal capacity, causing respectively cytopenia or leukemia (Fuchs and Chen, 2013; Seita and Weissman, 2010).

The understanding of hematopoiesis starts with the interpretation of steady state conditions. Although the mechanisms that determine how HSCs are instructed to differentiate are still not fully understood, there is consensus over a tree-like model of hematopoiesis (Akashi et al., 2000; Kondo et al., 1997). HSCs reside at the top of this model, progressively giving rise to all blood lineages, and restricting their lineage potential in a hierarchical manner. The steps of differentiation are represented by branches, where a cell restricts its potential and differentiates. This model is mostly based on immunophenotypical purification of stem and progenitor cells. Depending on the set of markers or the assay used to define progenitors, various branching patterns have been proposed leading to slightly different models of hematopoiesis. Recent advances in single-cell RNA sequencing and cellular barcoding led to a more continuous and dynamic model (Rodriguez-Fraticelli et al., 2018).
1.1.1 Identification of adult murine HSCs

The identification of HSCs has been a key aspect in hematology for various purposes, such as transplantation, re-programming technology and for a biological understanding of hematopoiesis and hematopoietic disease. Being such a small population, HSC identification was a challenging aspect in hematology. These cells can be defined both functionally and phenotypically as was pointed out by Purton and Scadden: ‘HSCs are functionally defined by their ability to sustain multi-lineage engraftment for an extended period of time upon serial transplantation into irradiated recipient mice and are immunophenotypically defined by the expression of a set markers’ (Purton and Scadden, 2007). The transplantation of bone marrow HSC and MPPs subpopulations revealed functional differences in self-renewal, clonal lifespan and repopulation capacity in irradiated mice. Consequently, multipotent hematopoietic cells have been broadly subclassified in long-term HSC (LT-HSC), short-term HSC (ST-HSC) and multipotent progenitors (MPPs). LT-HSC can support hematopoiesis for more than 4 months in lethally irradiated recipients and maintain this competence at least in a secondary transplantation (Colvin et al., 2009; Lemischka et al., 1986; Lu et al., 2011; Morrison and Weissman, 1994; Nakauchi, Hiromitsu Sudo, Kazuhiro Ema, 2006; Oguro et al., 2013; Szilvassy et al., 1990; Yang et al., 2005).

The phenotypic definition of HSCs takes advantage of a specific combination of cell surface markers. Most methods for enriching HSCs include the exclusion of lineage positive cells (B cell, T cells, macrophages/monocytes, granulocytes and erythrocytes) and the inclusion of cells co-expressing Sca1 and c-Kit (Osawa et al., 1996; Spangrude et al., 1988; Zhao et al., 2000). The cells in this subgroup are called LSK (Lin Sca1·Kit+) and include all multipotent hematopoietic cells in the bone marrow with an HSC purity that was defined to be about 10% in steady state hematopoiesis for C57Bl/6 (Spangrude and Brooks; Spangrude et al., 1988; Uchida and Weissman, 1992). A more stringent purification method incorporates the inclusion of so-called SLAM (signaling lymphocyte activation molecule) family members, namely CD48 and CD150 (Kiel et al., 2005). The frequency of long-term reconstituting HSCs in Lin-Kit+Sca1·CD48·CD150+ (LSK SLAM) is between 20-40% and represents about 0.008% of C57Bl bone marrow cells (Kent et al., 2009; Kiel et al., 2005; Yilmaz et al., 2006).

1.1.2 Characteristics and regulation of adult HSCs

The behaviour of HSCs is governed by extrinsic and intrinsic factors. For instance, during transplantation HSCs are challenged to reconstitute an irradiated recipient and consequently boost their proliferation. Contrarily, steady state HSCs are mainly quiescent, (Cabezas-Wallscheid et al., 2017; Cheshier et al., 1999; Foudi et al., 2009; Pietras et al., 2011; Wilson et al., 2008) with low mitochondrial activity (Ito et al., 2016; Vannini et al., 2016) and a preference for glycolytic metabolism over oxidative phosphorylation (Simsek et al., 2010; Takubo et al., 2013). In steady state vs transplantation setting, the proliferation difference in HSCs is due to the environment, highlighting the importance of extrinsic regulation.

Extrinsic regulation of adult steady state HSPCs is mediated by non-hematopoietic cells (like osteoblasts, endothelial cells, adipocytes, fibroblast and mesenchymal stromal cells) and differentiated hematopoietic cells like macrophages, osteoclasts (derived from monocytes and macrophages (Udagawa et al., 1990)), and lymphocytes (Hirata et al., 2018). All of these cells form a microenvironment for HSCs, also known as the HSC niche, within the bone marrow. This niche protects HSCs by limiting byproducts of aerobic metabolism, such as reactive oxygen species (ROS), supporting the prevention of DNA damage (Eliasson and Jönsson, 2010), while also maintaining and directing their self-renewal and differentiation state by releasing signals such as cytokines, chemokines and extracellular matrix molecules. Parallel to extrinsic regulation, HSC behaviour is driven intrinsically by transcription factors, signal transduction molecules, non-coding RNA and epigenetic modifiers. Transcription factors are the most researched intrinsic regulators due to their relevance with HSCs self-renewal, differentiation (for example C/EBPα for granulocyte differentiation (Zhang et al., 1997)) or apoptosis (for example P53 (Fridman and Lowe, 2003)).

1.1.3 Lineage priming and stem cell heterogeneity

The standard hematopoietic tree model assumes that HSCs are a homogeneous population with the same self-renewal and differentiation potential. However, most experiments to define HSCs were obtained from expression profiling of pooled cell subsets. Those bulk immunophenotypical analyses put together cells that are functionally often dissimilar, leading to a misinterpretation of the HSC compartment. Nowadays, phenotypic HSCs are not considered a homogeneous population anymore, yet mostly a heterogeneous lineage primed population. Already in 1997, the concept of multilineage gene expression in a single hematopoietic multipotent progenitor was recognized (Hu et al., 1997) and defined as a state called “promiscuous”. This observation was based on single cell RT-PCR on a hematopoietic multipotential cell line reporting co-expression of granulocytic, monocytic and erythroblastic markers such as myeloperoxidase and beta-globin. The pattern of gene expression allowed to define a predictable functional outcome of these cells. These experiments supported the lineage priming model. Lately, a prediction model classified cells undergoing differentiation based on a gradual decrease in the number of genes that are expressed, confirming multilineage gene expression in HSCs (Gulati et al., 2020). In this model HSCs express multiple differentiation genes, representing their differentiation potential. During differentiation, one pathway gets upregulated causing the downregulation of the other differentiation pathways (Figure 1). Therefore, the expression level of lineage-restrictive transcription factors can be used to infer cell fate of HSPCs.
Figure 1. Multilineage priming allows HSCs to commit into different cell fates. Adapted from Nimmo et al (Nimmo et al., 2015). Multiple lineage affiliated genes are co-expressed in multipotent cells before lineage commitment. Commitment is associated with downregulation of genes associated with alternative lineages.

Even if bona fide HSCs and MPPs have the potential to differentiate into all hematopoietic cell types, subgroups have distinct differentiation behaviours (Cabezas-Wallscheid et al., 2014; Pietras et al., 2015). For example, a thrombocyte-biased HSC population has been proven both transcriptionally and functionally via transplantation and in vitro assays (Grover et al., 2016; Rodriguez-Fraticelli et al., 2018; Sanjuan-Pla et al., 2013; Yamamoto et al., 2013). Megakaryocytes are considered to arise independently of other lineages but the reason for this atypical way of differentiation is still doubtful. Megakaryocytic priming and independence in differentiation also occurs in zebrafish (Macauly et al., 2016). Since megakaryocytes and HSCs share common markers and niche (Cheng et al., 2020), it is still under debate whether the HSCs purification methods cause the inclusion of megakaryocytes. Besides HSCs, phenotypic MPPs are considered a heterogeneous lineage-biased population as well. MPP subgroups have been distinguished as being megakaryocytic, myeloid and lymphoid biased (Pietras et al., 2015).

The first functional evidence of a heterogeneous stem cell compartment arose from single cell transplantation experiments of phenotypic HSCs resulting in different outputs. Several single cell transplantation experiments had widely variable reconstitution kinetics and very rarely a single transplanted HSC gave rise to a balanced differentiated progeny (Dykstra et al., 2007; Morita et al., 2010; Müller-Sieburg et al., 2002; Yamamoto et al., 2013). These results are in contrast with the view of a homogeneous HSCs population and infer that every HSC possess a very specific, if not unique, self-renewal and differentiation characteristic. Transplantation experiments test the ability of an HSC to generate other cell types when placed in a new environment, also called potency, while the direction of an HSC in an unperturbed setting represents its cell fate. Cell fate, in this context, has been studied by ex vivo barcoding and transplantation (Aiuti et al., 2013; Gerrits et al., 2010; Jordan and Lemischka, 1990; Lemischka, 1993; Lemischka et al., 1986; Mazurier et al., 2004; Shi et al., 2002; Snodgrass and Keller, 1987) and more recently by in vivo barcoding experiments (Rodriguez-Fraticelli et al., 2018; Yu et al., 2016), confirming the presence of HSC subpopulations in steady state hematopoiesis. In the last years, with the aid of new technologies, like single cell RNA sequencing, a molecular and functional heterogeneity in the HSC compartment has been supported by multiple research groups. The technological advances in single cell transcriptome analysis now allow for a higher sensitivity and to detect multiple genes expressed by single HSCs. In summary, transplantation, single cell RNA sequencing and barcoding experiments of putative HSCs confirmed the presence of primed HSCs (Figure 2) by revealing a heterogeneity in the expression of transcription factors and revealing hematopoiesis as a less stepwise and more continuous process (Grover et al., 2016; Karamitros et al., 2018; Moignard et al., 2013; Velten et al., 2017).

Figure 2. HSC Lineage Biases and Transcriptional Lineage Priming
Adapted from Haas et al (Haas et al., 2018). Balanced HSCs show a roughly equivalent contribution toward the production of all mature blood cells (left). Lineage biases HSCs show an unbalanced production of differentiated cells (right). Transcriptional lineage priming is an indicator of the differentiation path that multipotent cells will take.

1.2 GATA2
More than 500 transcription factors are encoded in the human genome (Hamosh et al., 2002). Six GATA factors (GATA1 to 6) have been identified in vertebrates. These proteins are pivotal transcriptional regulators involved in both physiological and pathological states. Their name is derived from their ability to bind the consensus DNA sequence (A/T) GATA(A/G) using two zinc finger protein structural motifs (Ko and Engel, 1993). While GATA4 to 6 have been associated to cardiac, pulmonary, pancreatic and hepatic function
GATA2 haploinsufficiency has been associated with a series of disorders called GATA2 haploinsufficiency syndromes. GATA2 deficiency results in different manifestations with variable age of onset and outcomes. GATA2 deficiency is responsible for monocytopenia and mycobacterial infection (MonoMAC) syndrome and Embryer syndrome. MonoMAC is a complex phenotype involving susceptibility to various infections and cytopenias in monocytes, B-and NK cells (Vinh et al., 2010). Embryer syndrome involves lymphedema and myelodysplasia or leukemia (Hsu et al., 2011). Other GATA2 deficiencies include dendritic cell, monocyte, B and NK lymphoid (DCML) deficiency (Bigley and Collin, 2011; Dickinson et al., 2011), mild chronic neutropenia (Pasquet et al., 2013) and familial AML (Hahn et al., 2011). GATA2 haploinsufficiency has been associated with a probability to develop MDS-AML ranging from 39% at 20 years old to 80% at 40 years old (Donadieu et al., 2018). Out of all AML patients, 3.6% to 14.4% are reported to carry a GATA2 mutation (Fasan et al., 2013; Green et al., 2013; Greif et al., 2012; Hahn et al., 2011).

In myeloid malignancies, heterozygous GATA2 mutations are found throughout the gene, but often cluster in one of the two zinc finger domains, most likely affecting the ability of GATA2 to bind DNA and other proteins, thus causing haploinsufficiency (Collin et al., 2015). GATA2 mutations are diverse (missense, nonsense, deletion, uni/allelic or in regulatory elements) but no specific phenotype has been associated to a mutation type (except lymphedema in nonsense and deletion mutations)(Spinner et al., 2014). Mutations abrogating GATA2 (like frameshift mutations) are associated with an earlier onset of clinical presentation than less dramatic mutations (like substitution mutations), more likely to contribute in a transcriptional reduction (Dickinson et al., 2014). The WT allele in GATA2 haploinsufficient patients appears to retain functionality (Crispino and Horwitz, 2017), excluding the need for both alleles to be non-functional for the phenotype to occur. Besides a loss in GATA2 functionality, GATA2 over-expression can result in acute myeloid leukemia (Persons et al., 1999; Tipping et al., 2009; Zhang et al., 2008). Since both GATA2 over-expression and down-regulation can cause blood phenotypes, the correct levels of GATA2 expression are essential for healthy hematopoiesis. With the goal of understanding the effect of GATA2 de-regulation in mind, in chapter 3 and 4 we explore the effect of a zebrafish Gata2b and a mouse Gata2 heterozygous knockout.

1.2.1 Gata2 and ontogeny of murine hematopoiesis
While HSCs in adult mammals reside in the bone marrow, their origin can be traced back during embryogenesis. An understanding of GATA2 function during ontogeny is particularly important because of its essential role in embryonal hematopoiesis and because the pathogenic mutations are frequently present in the germline. GATA2 possess unique roles during the generation, amplification, and maintenance of embryonic HSCs (Ling et al., 2004; de Pater et al., 2013; Rodrigues et al., 2005; Tsai and Orkin, 1997). In mice the first hematopoietic cells appear at embryonic day (E)7.5 in the blood island of the yolk sac, during the so-called primitive hematopoiesis. This process is temporary and allows the developing organism to sustain growth via tissue remodelling, oxygen and metabolite distribution before the overtake of definitive hematopoiesis (Jagannathan-Bogdan and Zon, 2013). Although this process is transitory, it gives rise to permanent tissue resident macrophages (Palis et al., 1999; Tober et al., 2007). At this stage of embryonic development, Gata2 is expressed in the primitive streak and in endothelial cells of the paired dorsal aorta (Kaimakis et al., 2016; Minegishi et al., 1999; Robert-Moreno et al., 2005). A second wave of yolk-sac-derived primitive hematopoiesis causes the production of erythroid-myeid progenitor cells (EMP) by E8.5 (McGrath et al., 2015) and immune restricted progenitors (Böiers et al., 2013). These HSC-independent progenitor cells are the first hematopoietic cells with multilineage potential.

1.2.2 GATA2 in definitive hematopoiesis
The first HSCs are generated in the aorta-gonads-mesonephros (AGM) region (Dieterlen Lievre, 1975; Medvinsky and Dzierzak, 1996) and other arteries (de Bruijn et al., 2000) from specialized hemogenic endothelial cells (Ohneda et al., 1998; Zovein et al., 2008), through a conserved mechanism called endothelial to hematopoietic transition (EHT) at E10.5 (Bertrand et al., 2010; Boisset et al., 2010; Dzierzak and Bigas, 2018; Jaffredo et al., 1998; Kiss and Herbomel, 2010; Müller et al., 1994). During EHT, hemogenic endothelial cells activate a hematopoietic transcriptional program which requires the balanced expression of transcription factors, such as Runx1 and Gata2. These expression changes cause morphological alterations like the loosening of tight junctions and the rounding up of cells forming intra-aortic hematopoietic cluster cells (IAHCS). The HSPCs at this stage are marked by the co-expression of the endothelial marker CD31 and c-Kit (Sasaki et al., 2010; Yokomizo and Dzierzak, 2010). HSCs can also be found after E10.5 in vitelline and umbilical arteries, placenta, yolk sac, and embryonic head (De Bruijn et al., 2000; Gekas et al., 2005; Gordon-Keylock et al., 2013; Li et al., 2012; Medvinsky et al., 2008; Müller et al., 1994; Ottersbach and Dzierzak, 2005; Robin et al., 2011; Rybtsou et al., 2011). Gata2 is expressed during definitive hematopoiesis in the hemogenic endothelium and emerging HSCs and IAHCs at E10 (Eich et al., 2018).
Using a conditional KO, Gata2 has been deleted before the generation of HSCs using Vec-Cadherin-expressing endothelial cells (Vec-Cre;Gata2<sup>f/f</sup>). This mouse model had a deficiency in intra-aortic cluster cell generation and did not generate long-term repopulating HSCs, confirming a quantitative and functional role of Gata2 in HSC generation in hemogenic endothelium. Gata2 has a dose-dependent function in the generation of HSCs. Indeed, Gata2 homozygous knock out mouse embryos are anemic and die at E10.5, not generating HSCs (Tsai et al., 1994), while heterozygous knock out embryos survive to adulthood but have more subtle changes in the hematopoietic system. Gata2 haploinsufficient embryos have significantly reduced emerging HSPCs in the AGM. This decrease in HSPCs was confirmed at E11 by a Sca-1 transgenic line (Ling et al., 2004) and at E10 by c-kit immunostaining (de Pater et al., 2013). Besides, AGM and yolk sac explants from Gata2 haploinsufficient embryos have a reduced engraftment capacity in irradiated recipients (Ling et al., 2004).

Gata2 not only controls the generation of HSCs but also their maintenance. HSPCs generated during EHT colonize the fetal liver from E11, mainly in response to Cxcl12 (Chou and Lodish, 2010). Fetal liver colonization from HSPCs is followed by their expansion, reaching a maximum of about 1000 HSC at E14.5 to 15.5 (Ema and Nakauchi, 2000; Morrison et al., 1995). Stem cells at this stage can be distinguished using SLAM markers (Kim et al., 2006). Vascular invasion in the developing bones allows the seeding of HSPCs from circulation in the BM. Gata2 has been deleted after the generation of HSCs using a conditional KO in Vav-expressing hematopoietic cells (Vav-Cre;Gata2<sup>f/f</sup>). Vav-Cre;Gata2<sup>f/f</sup> mice had decreased colony formation and their phenotypic HSCs had concurrent decrease in number and increase in apoptosis. Furthermore, transplantation studies showed that LT-HSCs are absent in Vav-CRE;Gata2<sup>f/f</sup> embryos (de Pater et al., 2013). HSCs migrate to the BM from E17.5 and this site remains the main source of hematopoietic cells in healthy adult mice. While fetal liver HSCs are expanding, and cycle frequently, bone marrow HSCs are mostly quiescent (BM) (Christensen et al., 2004; Wilson et al., 2008). Gata2 haploinsufficient mice are viable and become adults without any apparent anomaly, including normal blood values (Rodrigues et al., 2005). However, their bone marrow has a diminished frequency of CD34- LSK, and LSK cells have an increased proliferation and a propensity for apoptosis.

While GATA2 mutations in the zinc finger domains alter binding, enhancer mutations effect its expression. Two enhancers are essential for the murine GATA2 activity, one 9.5 kb downstream (+9.5) and one 77 Kb upstream (-77) of the transcriptional start site. The deletion of the +9.5 enhancer causes lethality at E13.5, abrogating HSCs generation (Gao et al., 2013), the correspondent human enhancer is mutated in 10% of all the MonoMAC syndrome patients (Hsu et al., 2013). While a deletion of the -77 enhancer allows HSC generation but causes lethality at approximately E15.5 due to a block in myelo-erythropoiesis and HSPC expansion (Johnson et al., 2015). Although it is currently challenging to entirely weigh the various criteria regulating Gata2 dosage, balanced expression of functional GATA2 is necessary for hematopoiesis and its de-regulation is linked to leukemogenesis. Indeed, both high (Luesink et al., 2012; Menendez-Gonzalez et al., 2019) or low Gata2 expression (see haploinsufficiency) predispose to AML.

### 1.3 Zebrafish in hematological research

Zebrafish (*Danio rerio*) became the emerging model for hematological research in the last decades, particularly for the convenience of molecular manipulation and the possibility to visualize embryogenesis in vivo, due to its transparency at this stage. The external fertilization and transparency of zebrafish allows to observe embryogenesis without interferences in the growth process. These advantages, together with the available transgenic reporter lines, allowed researchers to visualize embryonic growth with unprecedented ease. For example, the first in vivo EHT events (Bertrand et al., 2010; Kissa and Herbomel, 2010) and the HSPCs and stromal cell interaction in the caudal hematopoietic tissue (CHT) were visualized using zebrafish (Tamplin et al., 2015). The external fertilization allows to modify the genetic material (for example using CRISPR gene editing) directly at one-cell stage. In chapter 2 and chapter 3 we took advantage of this opportunity to generate a gata2b knockout. Despite an evolutionary divergence, zebrafish and mammals conserved most of the molecular and physiological processes directing hematopoiesis, like the generation and amplification of HSCs. Differently from mammals, zebrafish erythrocytes and thrombocytes retain their nucleus. However, zebrafish hematopoietic cells maintain a clear resemblance to human cells in a way that all hematopoietic lineages can be distinguished (Figure 3).
Zebrafish belong to the infraclass of teleost, which underwent an extra whole-genome duplication compared to other vertebrates (Meyer and Schartl, 1999). While some of the duplicated genes became pseudogenes, others acquired new functions related to the original gene in a way that more than 70% of human genes have at least one zebrafish orthologue (Howe et al., 2013). The extra genome duplication in teleost caused the generation of two Gata2 orthologues in zebrafish called Gata2a and Gata2b (Gillis et al., 2009). These two paralogues genes share just 57% of sequence identity but retained a well conserved region coding for the two zinc fingers, compared to humans and mice (Butko et al., 2015).

1.3.1 Ontogeny of zebrafish hematopoiesis

Primitive hematopoiesis in zebrafish begins in the lateral plate mesoderm (LPM). At around 15hpf the LPM forms the inner cell mass (ICM) which becomes the site of primitive hematopoiesis, generating erythroid and myeloid precursors (Detrich et al., 1995). Like in mammals, the first zebrafish definitive blood cells are derived from endothelial cells that undergo EHT. EHT occurs in the ventral side of the dorsal aorta and can be tracked in vivo using an endothelial reporter like Tg(fli:gfp). During EHT, flat hemogenic endothelial cells round up, shed in the space between the dorsal aorta and posterior cardinal vein (like shown in Figure 2I of chapter 2) and re-enter into circulation through the posterior cardinal vein. After generation, HSPCs enter the circulation and seed the caudal hematopoietic tissue (CHT) which functions as an intermediate hematopoietic site, analogous to the fetal liver of mammals (Figure 4), to expand the HSPC pool. The CHT is a highly vascular tissue that induces expansion of HSPCs with the aid of different cytokines (Murayama et al., 2006). After amplification, HSPCs relocate to the thymus, for T-cell differentiation, and kidney starting from 4dpf (Figure 4). The zebrafish kidney marrow, similarly to the mammal’s bone marrow, is the main site of definitive hematopoiesis and all hematopoietic lineages are distinguishable (Figure 3) (Davidson and Zon, 2004).
In the kidney marrow, HSPCs are marked by low levels of Tg(CD41:GFP) fluorescent (GFP™) signal (Ma et al., 2011) or can be distinguished based on their light scatter, together with the remaining main hematopoietic populations (Figure 5) (Traver et al., 2003). In chapter 2 and chapter 3 we took advantage of both methods for identifying and characterizing HSPCs.

Figure 5. Hematopoietic lineage differentiation in zebrafish. Whole kidney marrow populations classification based on FSC-A/SSC-A with corresponding representative images of sorted cells after May–Grünwald-Giemsa staining. The cells within the smaller quadrants depict additional representative cells within the same gate using the same magnification.

1.3.2 Role of Gata2a and Gata2b in hematopoiesis

The expression of Gata2 factors in zebrafish hematopoiesis starts with the expression of Gata2a at around 10.3hpf in the posterior lateral plate mesoderm. Instead Gata2b starts to be expressed at 16-18hpf in endothelial cells, based on qPCR and whole-mount in situ hybridization data. Gata2a expression is found mostly in vascular structures like DA and PCV; while Gata2b in sites of HSCs emergence and colonization (Butko et al., 2015). In adult zebrafish, Gata2b is expressed in the thymus and kidney and, using lineage tracing, Gata2b expressing cells have been proven to give rise to most blood cells in the KM. To recapitulate, while Gata2a is expressed more broadly in vasculature, Gata2b is expressed mostly in the hematopoietic system.

In support of a functional activity in the expression sites, homozygously mutated zebrafish in gata2a zinc finger nuclease (ZFN) reported shunts and other circulatory problems, particularly in the formation of the dorsal aorta (Zhu et al., 2011). Dobrzycki et al., identified the i4 enhancer in gata2a as the equivalent of the +9.5 Gata2 enhancer in mouse. The expression of the gata2a i4 enhancer in endothelial cells was needed to drive Gata2a expression. gata2a-i4 enhancer homozygous mutants had lower expression of Runx1 and Gata2b in the hemogenic endothelium at 28hpf and a notch-driven recovery of these transcription factors starting at 48hpf resulting in normal numbers of HSPCs by 5dpf (Dobrzycki et al., 2020). Gata2b morphants instead have a normal vascular development and arterial specification (Butko et al., 2015) but defects in Runx1 and C-myb expression respectively at 25 and 36hpf. Alongside, Gata2a did not rescue Gata2b morphants. These results indicate an evolutionary differentiation in the activity of Gata2a and Gata2b, and Gata2b as an uncharted transcription factor in zebrafish adult hematopoiesis.

Scope and outline

After years of human genetic analysis, substantial evidence emerged that germline mutations in GATA2 strongly predispose to the development of MDS and AML, with a penetrance of over 80% by the age of 40 (Donadieu et al., 2018). Furthermore, germline mutated GATA2 patients can present with a plethora of phenotypes like monocytopenia, DC-, NK and B cell-lymphopenia, neutropenia or lymphedema. In this dissertation we focus on dissecting the role of GATA2 gene deletions in both zebrafish and mice. To this end, in chapter 2 we generated a zebrafish gata2b mutant line and identified the hematopoietic consequences of the full loss of this transcription factor. Whereas in chapter 3 we used the same gata2b mutant zebrafish line as chapter 2 but analysed its heterozygous loss. These models allowed a direct comparison of the effects of 3 different levels of Gata2b in the zebrafish hematopoietic system. Whereas both the heterozygous and homozygous gata2b mutants revealed a transcriptional hyperproliferation in HSPCs, they exhibited distinct abnormalities in kidney marrow cells. gata2b homozygous deletion results in lineage differentiation defects resulting in neutropenia and defective B cell differentiation, whereas heterozygous deletion of gata2b results in erythroid and myeloid dysplasia, both phenotypes observed in patients. Indicating that different dosages of Gata2b result in specific cellular defects.

In chapter 4 we used a mouse model to study Gata2 haploinsufficiency. HSPCs from Gata2 heterozygous mice (Gata2+/-) had a transcriptional loss of quiescence both in embryogenesis and in adulthood. This proliferative stress, combined with the observation, in GATA2 deficient patients, of a drastic increase in MDS/AML incidence with age (Donadieu et al., 2018; Spinner et al., 2014) led us to investigate the effect of aging in Gata2+/- mice. In chapter 4 we revealed that after aging and transplantation of Gata2 heterozygous bone marrow recipients mice would develop leukopenia, recapitulating an aspect of the immunodeficiency reported in GATA2 patients. Finally, in chapter 5 we summarized the most
important findings and interpreted the meaning and relevancy of the results from previous chapters. In perspective to the current available literature we evaluated the limitations of our experiments and made recommendations for future research.

REFERENCES:


2

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

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

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

*Corresponding author: Email: e.depater@erasmusmc.nl

Key points:
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

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.

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 sample genotype (n=6) using TRIzol Reagent (Life Technologies) and cDNA was synthesized using SuperScript III Reverse Transcriptase kit (Invitrogen). gata2a (FWD primer: 5’-CAAACTCCACACGTCAACAG-3’, REV primer: 5’-CCCTCACACGATCGTTA3’-3’) and gata2b (FWD primer: 5’-TACCAATGTGACTGCTTAC-3’, REV primer: 5’-GAGGAGGATGGTTTGTCG-3’) expression levels were normalized to elfa (FWD primer: 5’-CCGCTAGCATTACCCTCC-3’, REV primer: 5’-CCGCTAGCATTACCCTCC-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 described27. The cmyb and runx1 probes were a kind gift from Roger Patient and quantified as described previously28. 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, MgCl2, 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 pCRIII-TOPO. Plasmid was linearized with HindIII and antisense probe transcribed with the DIG labelling kit (Sigma-Aldrich). Sense probe was used as negative control.

Transgenic lines, confocal imaging and adult KM FACS analysis

Embryos were anesthetized using tricine (3-amino benzoic acid ethylester) 160mg/L and selected for reporter expression. Tg(fli:eGFP)39 and Tg(CD41:GFP)30; Tg(flt1:RFP)30 embryos were imaged in 0.25% agarose with tricine and imaged using a Leica SP5 confocal microscope pre-warmed at 28°C. Tg(mpptg1.1:GFP)32, Tg(mp: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).

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 target of Gata2, was significantly reduced in mutant embryos (Figure S1A, B)[26,27]. Hereafter, we refer to this mutant as *gata2b⁻/⁻*.

**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 endothelio-to-hematopoietic transition (EHT)[28,29,30-32]. In mice, Gata2 is expressed in the endothelium, including the hemogenic endothelium (HE) of the dorsal aorta, and deletion of *Gata2* results in a reduction in HSC generation[14]. *cmyb* and *runx1* are two bona fide marker genes for HE at 26 hours post fertilization (hpf) in zebrafish[15,31]. We quantified *cmyb* and *runx1* expression by measuring pixel intensity of the in situ hybridization staining compared to background[15]. 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:eGFP)* 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[13,14], but Gata2b deficient zebrafish have intact HE and EHT (Figure 1E-J). High maternal expression of *gata2b* has been reported previously[33], 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[29]. Thus, we analysed *runx1* expression in *gata2a* mutants (*gata2a⁻/⁻*, lacking a conserved endothelial enhancer)[33]. *runx1* expression at 26 hpf was reduced in *gata2a⁻/⁻* 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.

![Image 548x278 to 585x315](image)

**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* mutants. C) Alignment of sequencing data of WT and mutants. D) *Gata2b* mutation leading to a STOP codon abrogating the protein before the DNA and protein binding zinc 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 *gata2b⁻/⁻* embryos at 26 hpf in the AGM region. L) Quantitation of signal intensity relative to background cells in WT and *gata2b⁻/⁻* embryos at 26 hpf, where each dot represents one embryo (41 ± 4.8 and 23.5 ± 2.0, n = 13 WT and n = 21 *gata2b⁻/⁻*). *** = p < 0.001, error bars represent standard error of mean (SEM).
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. 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) marker and the arterial Tgf(RFP) marker as definitive HSPCs are derived from arteries. CD41:GFP: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:RFP cells were significantly reduced in the CHT in gata2b-/embryos compared to WT (Figure 2D and Table S1).

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. 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: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 G0 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 (scRNAseq) (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+ expressing cells present in the kidney marrow pool as these cells were shown to contain transplantable HSCs (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+ cells within the total progenitor population.

We identified 20 different cell clusters were identified using the nearest neighbor algorithm in the R Seurat package. Most progenitors that were sequenced previously characterized differentiation markers (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 (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
myco, 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-Q).

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 the leosteis (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 the leosteis (Figure S3J) 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 the leosteis (Figure S3A). The B-cell clusters were overrepresented by 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) violin plots representing the expression levels of genes within the different clusters and each dot represents expression in one cell. G) mcl1 proto-oncogene (mcl1a), H) GATA binding protein 2b (gata2b), I) green fluorescent protein (GFP), indicating CD41-GFP 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 greater 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) quantification 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) and 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\(^{30,32,33,35}\). 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\(^{13,31}\) 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).

Because GATA2 is also required for HSC maintenance in mice\(^{12,13}\), we asked whether Gata2b deficiency resulted in a block in HSPC differentiation and thus an accumulation of HSPCs. In zebrafish, CD41:GFP\(^{30}\) expression marks the HSPC population most stringently\(^{10}\). Although CD41:GFP\(^+\) cell numbers and percentages were reduced during embryonic development (Figure 2D, I), at 20 dpf these percentages normalized resulting in comparable numbers of CD41:GFP\(^+\) 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\(^{30}\) and Tg(lck:GFP), marking T-cells\(^{30}\) were used to assess 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\(^{51,52}\) (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.
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 S5G 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.7,14 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 pseudotime 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-/- zebrasfish 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 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-/- zebrasfish. Percentages represent the average of all zebrasfish 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-/- zebrasfish kidney marrow by fluorescence microscopy. E) Forward and side scatter profile of Tg(mpx:GFP) expression in WT and gata2b-/- zebrasfish kidney marrow in green. F) Quantification of Tg(mpx:GFP)^+ cells expressed as percentage in single viable cells. Each dot represents kidney marrow analysis of one zebrasfish.

G) Representative figure of sorted Tg(mpx:GFP)^+ cells from WT and gata2b-/- zebrasfish kidney marrow after MGG staining.

H) Quantification of Tg(mpx:GFP)^+ cells from WT and gata2b-/- zebrasfish kidney marrow based on the differentiation phenotype using MGG staining. I) Gating strategy for CD41:GFP^+ expressing cells in total kidney marrow in green. J) Forward- and side-scatter plot of WT and gata2b-/- kidney marrow cells and CD41:GFP^+ expressing cells in green. K) Quantification of the frequency of CD41:GFP^+ cells in single live cells of total kidney marrow. Each dot represents kidney marrow analysis of one zebrasfish. L) FSC/GFP scatter profile of Tg(IgM:GFP) WT KM.

M) Quantification 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 zebrasfish. 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 zebrasfish. 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.
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) mes1b, 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− cells in adult WT and gata2b−/− KM cells. R) Bar graph representing the quantitation of cell cycle of CD41:GFP− cells in adult WT and gata2b−/− KM cells. Bars represent mean ± SEM, each dot indicates analysis from one zebrafish. UMAP; uniform manifold approximation projection.
Co-expression of the myeloid and lymphoid genes is represented in turquoise indicated by red arrows in the WT and 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. However, later reports suggested that some transcription factors have a reinforcing activity for terminal differentiation and propose that microenvironmental upstream regulators are decisive for lineage commitment. 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 CD41:GFP⁺ 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. 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⁺ 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⁺ expressing cells were scattered among different HSPC populations. Transplantation data suggest that only a minority of these cells are bona fide HSCs. 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\textsuperscript{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 suffering from MonoMAC syndrome, which is characterized by neutropenia, monocytopenia, DC- and B-cell lymphopenia\textsuperscript{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


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

### Table 1. Adult hematopoietic cell quantitations

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>gata2Zb&lt;sup&gt;−/−&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scatter analysis in freq. of single viable cells</td>
<td>N = 26</td>
<td>N = 23</td>
</tr>
<tr>
<td>Erythrocytes in KM (3 to 12 months)</td>
<td>25.4 ± 2.8</td>
<td>25.9 ± 3.0</td>
</tr>
<tr>
<td>Progenitors in KM (3 to 12 months)</td>
<td>26.8 ± 2.0</td>
<td>24.4 ± 1.9</td>
</tr>
<tr>
<td>Lymphocytes and HSCs in KM (3 to 12 months)</td>
<td>18.1 ± 1.1</td>
<td>29.0 ± 1.9&lt;sup&gt;****&lt;/sup&gt;</td>
</tr>
<tr>
<td>Myeloid cells in KM (3 to 12 months)</td>
<td>20.3 ± 1.4</td>
<td>10.3 ± 0.9&lt;sup&gt;****&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

#### Reporter expression in freq. of single viable cells

<table>
<thead>
<tr>
<th></th>
<th>Tg(mpeg:GFP)</th>
<th>Tg(CD41::GFP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gfp+ in single live cells</td>
<td>Tg(mpeg:GFP)</td>
<td>Tg(CD41::GFP)</td>
</tr>
<tr>
<td>N = 8</td>
<td>17.2 ± 3.0</td>
<td>0.4 ± 0.1 (n=9)</td>
</tr>
<tr>
<td>N = 4</td>
<td>3.5 ± 1.9</td>
<td>3.1 ± 0.8 (n=4)</td>
</tr>
<tr>
<td>N = 6</td>
<td>5.1 ± 1.0 (n=7) **</td>
<td>3.4 ± 0.6 (n=4)</td>
</tr>
<tr>
<td>N = 7</td>
<td>2.1 ± 0.7 (n=6)</td>
<td>0.4 ± 0.1 (n=8)</td>
</tr>
</tbody>
</table>

*The data are means 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.*

### References


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 35 (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 46 (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 analyze 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 cebp\(a\) and cebp\(b\) were found significantly downregulated in gata2b mutants. 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 46 (Figure S2G).

Erythroid cells, thrombocytes, neutrophils, monocytes and their progenitors were identified by comparing our data to known single cell expression analysis and known lineage markers 46, 50, 62, 65. 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 hba1, hbg1, hbg2 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 (cebp\(a\)) 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 46, 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 ighz4\(^{\text{high}}\) 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 35 showed indeed the presence of several B-cell populations (Figure 4L-S, Figure S4G-L).

One cluster was marked by high expression of pigr\(l3.5\) which encodes a polymeric Ig receptor 47. The pigr\(l3.5\) 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) 47.

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 fl1a and meis1b are typical HSC markers 46, 50. The HSPC1 and HSPC2 populations meet these criteria by expressing low levels of lineage markers (Figure S3B-D). These populations showed high expression levels of proliferation markers like myca, pcna and mki67 (Figure 3I, Figure S2H) and some cells with high levels of the stem cell markers fl1a 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. 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.**

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>gata2b&lt;sup&gt;+/−&lt;/sup&gt;</th>
<th>gata2b&lt;sup&gt;−/−&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EHT events</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGM (32 to 40hpf)</td>
<td>3.3 ± 0.4</td>
<td>2.9 ± 0.3</td>
<td>2.2 ± 0.4</td>
</tr>
<tr>
<td>CD41:GFP+Flt1:RFP+ cells</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHT (52-54hpf)</td>
<td>8.5 ± 1.7</td>
<td>9.0 ± 1.5</td>
<td>11.9 ± 1.0</td>
</tr>
<tr>
<td>CHT (58-60hpf)</td>
<td>25.2 ± 3.2</td>
<td>28.7 ± 1.9</td>
<td>23.3 ± 2.9</td>
</tr>
<tr>
<td>CHT (76-58hpf)</td>
<td>55.5 ± 3.8</td>
<td>43.4 ± 5.6</td>
<td>36.5 ± 3.0</td>
</tr>
<tr>
<td><strong>cm瑶 expression intensity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGM (24 hpf)</td>
<td>33.0 ± 6.3</td>
<td>28.5 ± 4.5</td>
<td>30.6 ± 2.6</td>
</tr>
<tr>
<td>AGM (33 hpf)</td>
<td>32.9 ± 10.9</td>
<td>30.9 ± 7.9</td>
<td>28.5 ± 6.5</td>
</tr>
<tr>
<td>CHT (56 hpf)</td>
<td>10.5 ± 2.5</td>
<td>12.8 ± 0.5</td>
<td>4.2 ± 1.1</td>
</tr>
<tr>
<td>CHT (76 hpf)</td>
<td>21.3 ± 7.4</td>
<td>20.8 ± 1.125</td>
<td>16.142 ± 1.950</td>
</tr>
<tr>
<td><strong>Runx1 expression intensity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGM (26 hpf)</td>
<td>24.1 ± 9.6</td>
<td>23.7 ± 11.6</td>
<td>25.4 ± 9.3</td>
</tr>
<tr>
<td>AGM (36 hpf)</td>
<td>30.0 ± 7.1</td>
<td>23.1 ± 8.1</td>
<td>25.3 ± 6.8</td>
</tr>
<tr>
<td>mpeg:GFP+ cells</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHT (54hpf)</td>
<td>513.9±23.67</td>
<td>564.9±23.51</td>
<td>545.9±33.27</td>
</tr>
<tr>
<td>mpx:GFP+ cells</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHT (75hpf)</td>
<td>227.5±14.40</td>
<td>221.1±8.63</td>
<td>235.0±15.49</td>
</tr>
<tr>
<td><strong>LCK:GFP+ pixel number</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thymus (5dpf)</td>
<td>2093±186.2</td>
<td>2449±165.5</td>
<td>2354±206.8</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Population</th>
<th>WT 1</th>
<th>WT 2</th>
<th>gata2b+/−</th>
<th>gata2b−/−</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSPCs</td>
<td>588</td>
<td>146</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>HSCs</td>
<td>63</td>
<td>16</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Myeloid progenitor 1</td>
<td>182</td>
<td>41</td>
<td>44</td>
<td>56</td>
</tr>
<tr>
<td>Myeloid progenitor 2</td>
<td>109</td>
<td>13</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>Myeloid progenitor 3</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Erythroid progenitor</td>
<td>8</td>
<td>19</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>Monocytes</td>
<td>127</td>
<td>112</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Macrophages</td>
<td>38</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Erythroid progenitors</td>
<td>9</td>
<td>12</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Erythroid progenitors</td>
<td>32</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Myeloid progenitors 1</td>
<td>123</td>
<td>109</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>Myeloid progenitors 2</td>
<td>119</td>
<td>112</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>Myeloid progenitors 3</td>
<td>55</td>
<td>55</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
| Total               | 2504970 | 370 | 321 | 557 | 173 | 772 | 203 | 1818792 | 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 gata2b references against elfa on pooled WT and gata2b−/− embryos at 30 hpf (n = 3) shows reduced gata2b expression levels 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 expressing cells from the kidney marrow pool of cells. D-E-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.
Expression

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

Chapter 2
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<sup>+</sup> expressing cells from the kidney marrow pool of cells. D-E) Quality control parameters for each sample G) UMAP showing cluster analysis on aggregated data set of both WT and gata2b<sup>−/−</sup> 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<sup>−/−</sup>. 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<sup>−/−</sup> cells.

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<sup>−/−</sup> 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(mpeg:GFP) WT and Gata2b⁻/⁻ KM. Q) Quantitation of GFP⁺ cells in Tg(mpeg:GFP) WT, gata2b⁺/⁻ and gata2b⁻/⁻ KM as percentage of single viable cells. Error bars represent SE
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) Pointplot 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) ccr10a and J) xpb1.

SUPPLEMENTARY REFERENCES


5

General discussion
Hematopoietic stem cells (HSCs) provide the millions of differentiated blood cells we need in our everyday life. In blood we distinguish the lymphoid and myeloid lineage. The lymphoid lineage generates adaptive immune cells like T cells and B cells. The myeloid lineage generates erythrocytes for oxygen transport, megakaryocytes for coagulation and innate immune cells like neutrophils and monocytes. These lineages have to be produced at the right proportion to maintain these processes. Transcriptional regulation of HSCs is an essential mechanism to maintain a balanced production of blood cells, and the transcription factor GATA2 plays a key function in this process. GATA2 directs embryonic and adult hematopoiesis in vertebrates by occupying GATA-DNA motifs in numerous genes. Homozygous GATA2 mutations in mice are embryonically lethal, making it difficult to assess its function in the different aspects of hematopoiesis. Heterozygous germine GATA2 mutations in humans are the cause of a series of disorders called GATA2 deficiency syndromes. They lead to a spectrum of phenotypes including lymphedema, cytopenias and, in more than 80% of patients, MDS or leukemia. To date, there is no clear correlation between the patient phenotype and the type of GATA2 mutation. The exact role of GATA2 in HSCs generation, expansion, self-renewal and differentiation is still unknown.

In this thesis, we focus on the hematopoietic consequences of GATA2 deficiency in zebrafish (chapter 2-3) and mouse (chapter 4).

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. Being the predominant GATA2 orthologue required for the maintenance of hematopoietic stem cells, we targeted gata2b to generate a knockout zebrafish line adopted in chapters 2-3. Interestingly, the homozygous Gata2b mutant is viable, allowing us to assess the functionality of Gata2b in adult hematopoiesis.

In chapter 2 we characterize the effects of Gata2b homozygous knockout (gata2b/-) in the hematopoietic system. Gata2b is required for the proliferation of hematopoietic stem and progenitor cells (HSPCs) during embryonic HSPCs expansion in the caudal hematopoietic tissue. Adult gata2b/- HSPCs are more proliferative and their kidney marrow shows a lymphoid bias and a block in neutrophil differentiation. Single cell transcriptome analysis showed that the HSPCs were the origin of the increased lymphoid lineage output in gata2b/- kidney marrow cells, due to a failure to initiate myeloid gene expression to sufficient levels and a subsequent co-expression of both myeloid and lymphoid genes in gata2b/- HSPCs.

To evaluate the transcriptional and phenotypic consequences of half Gata2b dose in zebrafish hematopoiesis, in chapter 3 we assessed gata2b heterozygous zebrafish (gata2b+/−). We hypothesized that, with half Gata2b dose, zebrafish would still present reduced myeloid differentiation, although to a lesser extent than when knocking out gata2b homozygously.

Instead, we observed an unprecedented erythroid dysplasia in the kidney marrow, but a conservation of the major differentiation lineages. The HSPC compartment also showed transcriptional changes fitting with the transcriptional changes as found in MDS.

The clear dichotomy in phenotypes between the heterozygous Gata2b mutant and the homozygous Gata2b mutant show that, depending on its dose, Gata2 has distinct functions in the maintenance of HSPCs and the subsequent differentiation into the different hematopoietic lineages. This could possibly explain the variable phenotypes found in GATA2 patients.

Mice, like humans, possess one copy of the Gata2 gene. Gata2 knock out (Gata2−/−) mouse embryos succumb before the generation of HSCs, while Gata2 heterozygous knock-out mice (Gata2+/-) do not develop noticeable hematological phenotypes. We hypothesized that a transcriptional analysis of HSPCs would reveal unnoticeable peculiarities caused by Gata2 heterozygosity. In chapter 4 we found that Gata2−/− HSPCs are hyperproliferative and to determine the long term consequences of this condition, we let Gata2−/− mice age up to 14-15 months. Gata2−/− aged HSPCs remain hyperproliferative and caused a failure to produce sufficient numbers of white blood cells after bone marrow transplantation.

The onset of leukopenia after transplantation of aged Gata2−/− HSCs reflects a major aspect of human GATA2 deficiency and provides a consistent methodology for reproducing a GATA2 associated disorder in mice. Further experiments will be necessary to better characterize the leukopenia identified and to test if other stress factors can contribute to a similar outcome in a reduced time-span. Since aging is necessary for the leukopenia to occur, it is would be interesting addressing its role in the phenotype development such as bone marrow niche-mediated inflammation.

Collectively our findings provide insights into the biology of GATA2 and the hematopoietic consequences following its deficiency. They establish the human disease relevance of the Gata2−/− mouse model and spotlight an essential role for Gata2 in HSPCs lineage output/proliferative stress in human GATA2 deficiency. These findings pave the way for further research on the synergism between HSPCs reduction, loss of quiescence and altered lineage output caused by Gata2 deficiency.
## 5.2 GENERAL DISCUSSION

**One gene, multiple roles in hematopoietic cell development and functions**

Transcription factors like GATA2 have key roles in directing the expression of lineage specific target genes in the hematopoietic system. Gata2 has several different functions in the hematopoietic system. Gata2 is required for HSCs generation (Tsai et al., 1994), survival (de Pater et al., 2013), proper proliferation and differentiation (this thesis). In humans, around 90% of germline GATA2 pathogenic mutations are either missense or truncating, and considered to cause GATA2 haploinsufficiency (Cortés-Lavaud et al., 2015; Leubolt et al., 2020; Wlodarski et al., 2017). Treatment of the diverse clinical presentations remains limited and the only potential cure is allogeneic HSC transplantation (Hirabayashi et al., 2017). Both hereditary and sporadic forms of GATA2 deficiency exist. Their clinical features include immunodeficiency, pulmonary, vascular/lymphatic dysfunctions and a strong propensity to develop MDS and AML (Spinner et al., 2014)(Fig.1).

**Figure 1. Effects of GATA2 deficiency.** Adapted from McReynolds et al (McReynolds et al., 2018). GATA2 deficiency patients have lymphedema secondary to effects on the vasculature, likely due to GATA2 activity at the endothelial cell level. Loss of GATA2 protein leads to hematopoietic stem and progenitor cell (HSPC) loss and dysfunction. This depletion can support cytopenias—B cells, dendritic cells, natural killer cells, and monocytes. The resultant immunodeficiency drives infections. Overall, these factors lead to bone marrow stress, hypocellularity and the ultimate development of bone marrow failure (BMF), myelodysplastic syndrome (MDS), and acute myeloid leukemia (AML). The combination of continued HSC depletion, infections, and cytopenias stress the bone marrow creating an environment supportive for the development of MDS, which may lead to AML or CMML.

Since first described, GATA2 haploinsufficiency provides management difficulties due to the challenges in delineating the genotype to phenotype relationship. Given the heterogeneity in the mutational spectrum, age of onset and clinical manifestations, a standardized analysis of the effects of GATA2 mutations remains difficult. For instance, even members of the same family (identical GATA2 mutation) can present with different clinical outcomes, with disease latencies varying from 14 to 74 years of age (Mutsaers et al., 2013). The reasons for this wide array of phenotypes remain unexplained at present. Possibly, chronic pathogen challenges of the immune system (highly frequent in GATA2 patients) could provide a pathophysiologic basis for myeloid neoplasms (Kristinsson et al., 2011) in association with the GATA2 haploinsufficiency. Indeed, pancytopenia and bone marrow suppression are known complications of chronic infectious diseases (Achi et al., 2013), and inflammatory stimulation and dysregulation of interferon signaling contribute to depletion of hematopoietic stem and progenitors (Essers et al., 2009; Lin et al., 2014; Matatall et al., 2016; Sato et al., 2009). On the other hand, mechanisms controlling the rate of allelic expression might have a role in disease progression. Our data suggests that different dosages of Gata2b have different effects on the hematopoietic system. In chapter 2 we show that the lack of wild type gata2b expression in gata2b homozygous knockout zebrafish causes a shift in differentiation, while the heterozygous expression of the wild type gata2b causes erythroblastopenia (chapter 3). The distinct phenotypes demonstrate a gata2b dose-dependent effect on zebrafish hematopoiesis. Indeed, a recent study showed that leukemic blasts of AML patients display GATA2 allele specific expression more often than any other myeloid or cancer-related gene, and GATA2 mutated patients preferably express the mutant allele (Mulet-Lazaro 2021). These results suggest that the ratio of mutated vs wildtype allele expression could cause dose dependent effects.

**Interplay of different GATA family members in controlling hematopoiesis**

All members of the GATA family of transcription factors (GATA1 to GATA6) possess zinc-fingers domains responsible for binding the consensus DNA sequence (A/T)GATA(A/G). Given the simplicity of this sequence, GATA binding sites can be found scattered through the entire genome. Transcription factors can replace each other on these DNA binding sites but do not bind with the same affinity to all GATA sites. Distinct activities of the GATA members are predominantly determined by their temporal expression at distinct stages of hematopoietic development (Katsumura & Bresnick, 2017). GATA1, GATA2 and GATA3 are the hematopoietic regulators in this family. GATA1 is mostly involved in the development of erythroid (Fujiwara et al., 1996) and megakaryocyte progenitors (Martin et al., 1990), and required for their differentiation (Pevny et al., 1991; Shvidranski et al., 1997). While GATA3 has an essential function in T cell differentiation (Ho et al., 2009), together with Gata2, it can partially rescue the erythroid phenotype in Gata1 deficient mice (Ferreira et al., 2007; Takahashi et al., 2000; Tsai et al., 1998). A strict regulation has been attributed to Gata1 and
Gata2 during erythroid differentiation since these transcription factors regulate the same binding sites in different steps of differentiation, in a process called GATA switch (Grass et al., 2006). These findings suggest that Gata2 is positively regulated by its protein product via binding its promoter (Grass et al., 2003; Vicente et al., 2012), while it gets repressed by GATA1, in partly by disrupting its autoregulation (Bresnick & Johnson, 2019). The erythroid dysplasia observed in gata2b+/− mice (chapter 3) might be suggestive of an aberrant Gata1 compensation in zebrafish. The finding that Gata1 is not differentially expressed in bulk or scRNA sequencing of zebrafish gata2b+/− HSPCs may argue against this hypothesis. On the other hand, it is possible to have Gata1 compensation even if the level of its expression remains unaltered. An increased association rate of Gata1 to gene promoters in response to gata2b heterozygosity (Hasegawa et al., 2016) can make transcriptional upregulation not necessary for compensation. Consequently, we do not exclude that altered Gata factors binding to DNA can be causative of the phenotype.

When one gene copy is not enough
For most genes, a single copy is enough for normal development and most heterozygous knockout mice result in no detectable phenotype (Seidman & Seidman, 2002). When a single copy of the wild type allele in diploid organisms is not sufficient to produce the standard phenotype (Morrill & Amon, 2019), haploinsufficiency occurs. About 300 human genes are known to be haploinsufficient (Dang et al., 2008) and, when mutated, can become the source of multiple health issues. Why only some genes cause haploinsufficiency and the mechanism behind their regulation, is not clear. The most recent hypothesis called “dose-stabilizing” assumes that haploinsufficient genes have a narrow gene dosage range where gene products are limiting for fitness when under-expressed, while higher gene products are toxic, possibly because they interfere with cellular functions or because of a detrimental production and degradation of the excess product (Morrill & Amon, 2019).

In humans, GATA2 heterozygosity mainly causes haploinsufficiency (Cortés-Lavaud et al., 2015). Like other haploinsufficient genes, GATA2 balanced expression is necessary for proper homeostasis and consequently more than 90% of GATA2 mutant patients develop a life-threatening hematologic disease by 40 years of age (Donadieu et al., 2018). However, the age of onset and clinical manifestations widely vary and a small percentage of carriers remains asymptomatic. Using reverse genetics in inbred animal models we standardized the effects of Gata2 knockout. Whereas in zebrafish Gata2b heterozygous fish show erythroid dysplasia (chapter 3), in mice, Gata2 heterozygosity (Gata2+/−) causes the development of a hematologic phenotype only under stress conditions. However, some genes can result in haploinsufficiency only under specific conditions. For instance, in yeast there is little overlap between haploinsufficient genes based on growth medium differences (Deutschbauer et al., 2005). Possibly, Gata2+/− mice did not meet the conditions to cause a phenotype like human GATA2 heterozygous patients. Based on the observation that GATA2 patients’ phenotypes are strongly correlated with age, in chapter 4 we promoted Gata2 haploinsufficiency by ageing Gata2+/− mice and after bone marrow transplantation, observing a hematologic phenotype. These results support the notion that a Gata2 knockout background might require further challenges for haploinsufficiency to occur.

Proliferative and differentiation aspects of Gata2 dosage in hematopoiesis
HSCs are multipotent cells which rarely enter cell cycle (Orford & Scadden, 2008). However, a combination of different factors shapes their potency and dormancy. Both of these aspects are affected by Gata2 mutations in this thesis. In accordance to the dose-stabilizing hypothesis of haploinsufficiency, GATA2 overexpression and downregulation can both cause adverse consequences to the hematopoietic system. GATA2 can be found overexpressed in the bone marrow of acute myeloid leukemia (AML) patients and in leukemic blasts (Bullinger et al., 2004; Luesink et al., 2012; Vicente et al., 2012), while GATA2 gain of function mutations can be observed in chronic myeloid leukemia (Zhang et al., 2008). GATA2 overexpression in vitro can cause cell cycle arrest and reduced colony forming capacity (Ezoe et al., 2002; Heyworth et al., 1999) and in vivo enforced GATA2 expression causes inhibition of cell cycle and differentiation (Persons et al., 1999; Tipping et al., 2009). If increasing Gata2 expression promotes quiescence, predictably a decrease in Gata2 expression would promote proliferation. Whereas a finding reported decreased cell cycle activity in Gata2+/− HSPCs (Rodrigues et al., 2005), in chapter 4, we identify that, starting from embryonic development, Gata2+/− phenotypic HSPCs are quantitatively scarce but lose quiescence compared to WT. These contrasting results have not been clarified yet and might be caused by multiple factors like the HSPC population taken in analysis, the growth conditions of the animals, or the different assays used. Consistent with the loss of quiescence, in chapter 2 and 3, by knocking out Gata2b in zebrafish we observe a consistent transcriptional increase in proliferation. In chapter 2, Gata2b deficient cells almost entirely lose a quiescent HSPC subcluster and gain a proliferative subcluster while in chapter 3 we confirm a pivotal role for gata2b in cell cycle, observing DNA replication and ribosome biogenesis as the main upregulated transcriptionally characteristics representing gata2b+/− HSPCs.

When considering the mechanisms governing Gata2 deficient proliferation, we hypothesized that the reduced HSPC population observed in Gata2+/− mouse embryos (de Pater et al., 2013; Gao et al., 2013, this thesis) and adults (Rodrigues et al., 2005, this thesis) could trigger a loss in quiescence as a compensatory mechanism. Gata2b deficient zebrafish instead, after HSCs amplification, re-acquire WT HSPCs numbers throughout adulthood, excluding that the reduction in HSPCs is a direct signal inducing their hyperproliferation in zebrafish. While the Gata2+/− proliferative changes observed in vitro suggest intrinsic signalling as an important factor in this regulation, extrinsic signalling might as well play an important role. Stress factors such as infections, cytopenias, ageing, chemotherapy, and bone marrow transplantation can force them out of dormancy and contribute actively to
hematopoiesis (Essers et al., 2009; Singh et al., 2020; Wilson et al., 2008). For example, cytokine signalling is a well-studied regulator of HSCs dormancy and could be involved in the proliferative phenotype observed in Gata2 deficiency. As cytokine signalling can direct lineage choice (Mossadegh-Keller et al., 2013; Rieger et al., 2009), it would be interesting to explore the niche signalling in this context.

Like proliferation, differentiation has been observed to depend on Gata2 dosage. For example, GATA2 knockout accelerates adipocyte differentiation from mesenchymal stromal cells (Kamata et al., 2014; Tong et al., 2000). While increasing Gata2 expression promotes quiescence, a lower level of overexpression promotes myeloid self-renewal and a block in lymphoid differentiation (Nandakumar et al., 2015). In humans, a lymphoid/myeloid lineages loss has been observed previously in GATA2 deficient patients (Bigley & Collin, 2011), and recently appreciated transcriptionally in GATA2 deficient CD34+ cells (Wu et al., 2020). In gata2b knockout zebrafish (chapter 2), we observe repercussions involving analogous processes, with a lymphoid biased differentiation and a block in myeloid differentiation. These results are consistent with an essential function for Gata2b in maintaining analogous differentiation and the lympho-myeloid commitment balance. In chapter 3, gata2b heterozygosity does not cause major shifts in differentiation, but supports the generation of erythroid dysplastic and (occasionally) myeloid dysplastic cells. In chapter 4 we show that mouse Gata2 heterozygous HSCs fail to produce sufficient white blood cells. In conclusion, previous research and the data presented in this thesis support the notion that Gata2 contributes to the regulation of HSC proliferation and coordinates their commitment toward lymphoid and myeloid lineages.

From embryonic to chronic Gata2 deficiency: Consequences for the development of malignancy

While acute stress can cause a burst of cycling HSCs, a chronic stress exposes HSCs to sustained proliferative signals which can result in loss of fitness and exhaustion (Baldridge et al., 2010; Ruzankina & Brown, 2007). The chronic hyperproliferation that we observe in Gata2 deficient HSPCs is possibly the cause of the reduced self-renewal capacity of the Gata2+/− HSCs. Considering the germline nature of Gata2 pathogenic mutations and of the mutants used in this thesis, we tested whether embryonic defects might affect adult hematopoiesis. Besides, to assess whether longer time periods were required for haploinsufficiency to emerge, we aged heterozygous gata2b zebrafish and Gata2 mice.

One of the main advantages of using zebrafish as an animal model is its transparency and the possibility to observe its embryogenesis in vivo. In chapter 2 we show that the generation of HSCs through endothelial to hematopoietic transition (EHT) tends to be reduced both in gata2b knockout heterozygous and homozygous zebrafish. The number of EHT events is directly proportional to the expression of the WT gata2b allele, indicating a supportive function for this transcription factor in the generation of hematopoietic stem cells. Instead, we show that the expression of the other Gata2 orthologue, Gata2a, is necessary for the specification of hemogenic endothelium. In chapter 4 we confirm that Gata2+/− mice have an obvious loss of HSPCs during embryogenesis (Ling et al., 2004) and conclude that Gata2 causes a reduction in HSPC generation during embryogenesis concomitant with higher proliferative signatures, already starting at the fetal liver stage.

Supported by the fact that GATA2 haploinsufficient patients’ phenotypes worsen with age, we aged Gata2+/− mice to assess the consequences of the accumulation of defects. The combination of transcriptional and functional analysis pointed to a loss of quiescence and proliferative stress. However, nor previous research (Ling et al., 2004; Rodrigues et al., 2005) nor our results highlighted anomalies in the blood of adult Gata2+/−. Repetitive BM transplantation causes a continuous call for stem cell function and, just like a loss in quiescence, can cause additive cell-autonomous effects, such as epigenetic dysregulation, mutations, and telomere erosion (Ermolaeva et al., 2018; Mejia-Ramirez & Florian, 2020). However, Gata2+/− repetitive bone marrow transplantation (BMT) of Gata2+/− young adults did not support the development of a blood disorder in recipients (Rodrigues et al., 2005). These results indicate that proliferative stress, on its own, is not enough to cause a disorder in Gata2+/−. Instead, Gata2 heterozygous HSCs are exhausted only after ageing and transplantation. Ageing can cause an additional HSPC extrinsic mis-regulation like a chronically inflamed, oxidative stressed microenvironment (Kovtunyk et al., 2016; Tower, 2012), which can induce maladaptive transformations in HSPCs (Raaijmakers et al., 2010; Walkley et al., 2007; Zambetti et al., 2016). GATA2 is expressed in both hematopoietic and non-hematopoietic cells, including mesenchymal, endothelial, adipocyte, eosinophil, erythroid progenitor and megakaryocytic cells (Minami et al., 2004; Tong et al., 2000; Tsai & Orkin, 1997). These cells comprise a relevant part of the HSCs microenvironment but play a still unexplored role in the pathogenesis of germline GATA2 deficiency. It would be interesting to explore whether the aging microenvironment alone or the fact that the environment is also Gata2 haploinsufficient, contribute to the HSC exhaustion phenotype.

Conclusion and perspectives:
The work described in this dissertation was aimed at better understanding the consequences of GATA2 deficiency. For this purpose, we adopted Gata2 deficient zebrafish and mouse models, with a particular focus on the transcriptional consequences in Gata2 deficient HSPCs.

In chapter 2, Gata2b homozygous knockout (gata2b−−) zebrafish revealed a role for this transcription factor in proliferation and lineage commitment of early hematopoietic progenitor cells. Adult gata2b−− kidney marrow shows a lymphoid bias and a block in neutrophil differentiation. In chapter 3, Gata2b homozygous knockout (gata2b−−) zebrafish exposed the necessity of unaltered levels of Gata2b to avoid the development of dysplasia. Gata2b heterozygosity causes dysplasia in erythroid progenitors of zebrafish kidney marrow.
Taken together, the findings in chapter 2 and 3 highlight the importance of adequate levels of Gata2b in directing HSCs commitment and differentiation of myeloid, lymphoid and erythroid lineage cells. Moreover, we reveal that Gata2a is required during embryogenesis for the specification of hemogenic endothelium, upstream of Gata2b, while Gata2b acquires significant functions during HSPCs expansion and adulthood.

Different dosages of Gata2b result in distinct phenotypes, including cytopenia, loss of differentiation and dysplasia, suggesting a possible relation of GATA2 levels to the clinical outcome of GATA2 deficient patients. Variation of WT GATA2 expression levels occur frequently in GATA2 patients via allele-specific expression (ASE), reaching a 95% chance when associated with CEBPA double mutations (Mulet-Lazo et al. 2021). GATA2 is the most recurrent gene with ASE in AML and the variations in its expression levels are crucial aspects for understanding the mechanism responsible for the manifestations of GATA2 deficiency. We addressed this aspect analysing the consequences of different gata2b expression levels which requires further stress factors, like an infection challenge or secondary mutations, on the onset of leukopenia is the first observable mouse blood phenotype derived from GATA2 deficiency in rodents. In patients, possibly the combination of low HSCs numbers cytopenic model represents a valuable method for recapitulating characteristics of human haploinsufficient HSCs and reflects a major aspect of human GATA2 deficiency. This

The lack of gross morphologic abnormalities or MDS/AML in our fish does not undervalue our research. Instead it highlights how GATA2 de-regulation represents a predisposition which requires further stress factors, like an infection challenge or secondary mutations, to cause the development of severe cytopenias and MDS/AML. Future experiments in this direction are likely necessary to determine the relation of GATA2 predisposition with environmental factors for the phenotype occurrence, particularly considering the variable penetrance observed in patients.

In chapter 4 we examined the effects of Gata2 haploinsufficiency at different time points in mice and found that the fitness of aged Gata2+/− HSCs is impaired, resulting in a failure to produce sufficient numbers of WBC after bone marrow transplantation. The onset of leukopenia is the first observable mouse blood phenotype derived from GATA2 haploinsufficient HSCs and reflects a major aspect of human GATA2 deficiency. This cytopenic model represents a valuable method for recapitulating characteristics of human GATA2 deficiency in rodents. In patients, possibly the combination of low HSCs numbers and proliferative stress supports the cytopenias and marrow dysplasia, leading to clonal hematopoiesis and malignant transformation. In our model, the increased proliferative stress and aging likely promote leukopenia. Aging is associated with the accumulation of hematopoiesis and malignant transformation. In our model, the increased proliferative stress supports the cytopenias and marrow dysplasia, leading to clonal GATA2 deficiency in rodents. In patients, possibly the combination of low HSCs numbers and proliferative stress supports the cytopenias and marrow dysplasia, leading to clonal hematopoiesis and malignant transformation. In our model, the increased proliferative stress supports the cytopenias and marrow dysplasia, leading to clonal hematopoiesis and malignant transformation.

REFERENCES:
Ferreira, R., Wai, A., Shimizu, R., Gillemans, N., Rotter, R., Von Lindern, M., Ohnedas, K., Grosved, F., Yamamoto, M.,


Addendum
Hematopoietic stem cells (HSCs) provide the millions of differentiated blood cells we need in our everyday life. In blood we distinguish the lymphoid and myeloid lineage. The lymphoid lineage generates adaptive immune cells like T cells and B cells. The myeloid lineage generates erythrocytes for oxygen transport, megakaryocytes for coagulation and innate immune cells like neutrophils and monocytes. These lineages have to be produced at the right proportion to maintain these processes. Transcriptional regulation of HSCs is an essential mechanism to maintain a balanced production of blood cells, and the transcription factor GATA2 plays a key function in this process. GATA2 directs embryonic and adult hematopoiesis in vertebrates by occupying GATA-DNA motifs in numerous genes. Homozygous GATA2 mutations in mice are embryonically lethal, making it difficult to assess its function in the different aspects of hematopoiesis. Heterozygous germline GATA2 mutations in humans are the cause of a series of disorders called GATA2 deficiency syndromes. They lead to a spectrum of phenotypes including lymphedema, cytopenias and, in more than 80% of patients, MDS or leukemia. To date, there is no clear correlation between the patient phenotype and the type of GATA2 mutation. The exact role of GATA2 in HSCs generation, expansion, self-renewal and differentiation is still unknown.

In this thesis, we focus on the hematopoietic consequences of GATA2 deficiency in zebrafish (chapter 2-3) and mouse (chapter 4).

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. Being the predominant GATA2 orthologue required for the maintenance of hematopoietic stem cells, we targeted gata2b to generate a knockout zebrafish line adopted in chapters 2-3. Interestingly, the homozygous Gatat2b mutant is viable, allowing us to assess the functionality of Gata2b in adult hematopoiesis.

In chapter 2 we characterize the effects of Gata2b homozygous knockout (gata2b-/-) in the hematopoietic system. Gata2b is required for the proliferation of hematopoietic stem and progenitor cells (HSPCs) during embryonic HSPCs expansion in the caudal hematopoietic tissue. Adult gata2b-/- HSPCs are more proliferative and their kidney marrow shows a lymphoid bias and a block in neutrophil differentiation. Single cell transcriptome analysis showed that the HSPCs were the origin of the increased lymphoid lineage output in gata2b-/- kidney marrow cells, due to a failure to initiate myeloid gene expression to sufficient levels and a subsequent co-expression of both myeloid and lymphoid genes in gata2b-/- HSPCs.

To evaluate the transcriptional and phenotypic consequences of half Gata2b dose in zebrafish hematopoiesis, in chapter 3 we assessed gata2b heterozygous zebrafish (gata2b+/-). We hypothesized that, with half Gata2b dose, zebrafish would still present reduced myeloid differentiation, although to a lesser extent than when knocking out gata2b homozygously. Instead, we observed an unprecedented erythroid dysplasia in the kidney marrow, but a
conservation of the major differentiation lineages. The HSPC compartment also showed transcriptional changes fitting with the transcriptional changes as found in MDS.

The clear dichotomy in phenotypes between the heterozygous Gata2b mutant and the homozygous Gata2b mutant show that, depending on its dose, Gata2 has distinct functions in the maintenance of HSPCs and the subsequent differentiation into the different hematopoietic lineages. This could possibly explain the variable phenotypes found in GATA2 patients.

Mice, like humans, possess one copy of the Gata2 gene. Gata2 double knock out (Gata2+/−) mouse embryos succumb before the generation of HSCs, while Gata2 heterozygous knock-out mice (Gata2+/−) do not develop noticeable hematological phenotypes. We hypothesized that a transcriptional analysis of HSPCs would reveal unnoticeable peculiarities cause by Gata2 heterozygosity. In chapter 4 we found that Gata2+/− HSPCs are hyperproliferative and to determine the long term consequences of this condition, we let Gata2+/− age up to 14-15 months. Gata2+/− aged HSPCs remain hyperproliferative and cause a failure to produce sufficient numbers of white blood cells after bone marrow transplantation.

The onset of leukopenia after transplantation of aged Gata2+/− HSCs reflects a major aspect of human GATA2 deficiency and provides a consistent methodology for reproducing a GATA2 associated disorder in mice. Further experiments will be necessary to better characterize the leukopenia identified and to test if other stress factors can contribute to a similar outcome in a reduced time-span. Since aging is necessary for the leukopenia to characterize the leukopenia identified and to test if other stress factors can contribute to the human disease relevance of the bone marrow niche-mediated inflammation.

Collectively our findings provide insights into the biology of GATA2 and the hematopoietic consequences following its deficiency. They establish the human disease relevance of the Gata2+/− mouse model and spotlight an essential role for Gata2 in HSPCs lineage output/proliferative stress in human GATA2 deficiency. These findings pave the way for further research on the synergism between HSPCs reduction, loss of quiescence and altered lineage output caused by Gata2 deficiency.

**NEDERLANDSE SAMENVATTING**

Hematopoïetische stamcellen (HSC’s) leveren de miljoenen gedifferentieerde bloedcellen die we in ons dagelijks leven nodig hebben. In bloed kunnen we een onderscheid maken tussen bloedcellen van lymfoïde en myeloïde afkomst. De lymfoïde tak levert adaptieve immuuncellen zoals T-cellen en B-cellen. De myeloïde tak genereert erytrocyten voor zuurstoftransport, megakaryocyt voor coagulatie, en aangebroken immuuncellen zoals neutrofielen en monocyten. Deze verschillende bloedcellen moeten in de juiste verhouding worden geproduceerd om deze processen in stand te houden. Transcriptionele regulatie van HSC’s is een essentieel mechanisme om een gebalanceerde productie van bloedcellen te behouden, en de transcriptiefactor GATA2 vervult een sleutelrol in dit proces.

GATA2 stuurt embryonale en volwassen hematopoëse in gewervelden door de GATA-DNA-motiveven van talrijke genen te bezetten. Homozygote GATA2-mutaties zijn in muizen embryonaal dodelijk, waardoor het moeilijk is om de functie van GATA2 in de verschillende aspecten van hematopoëse vast te stellen. Heterozygote kiembaan GATA2-mutaties in mensen zijn de oorzaak van een reeks aandoeningen die als GATA2-deficiëntiesyndromen worden aangeduid. Ze leiden tot een spectrum van fenotypes, waaronder lymfoedeem, cytopenieën, en bij meer dan 80% van de patiënten, MDS of leukemie.

Tot op heden is er geen duidelijk verband tussen het fenotype van de patiënt en het type GATA2-mutatie. De exacte rol van GATA2 bij het genereren, vermenigvuldigen, zelf-vernieuwen en differentiëren van HSC’s is nog onbekend. In dit proefschrift richten we ons op de hematopoïetische gevolgen van GATA2-deficiëntie bij zebravis (hoofdstuk 2-3) en muizen (hoofdstuk 4).

Zebravissen hebben twee orthologen van GATA2: Gata2a en Gata2b. Eerdere studies hebben aangetoond dat gata2b dominant tot expressie komt in HSPC’s, terwijl gata2a voornamelijk tot expressie komt in het bloedvatnestelsel. Aangezien dat de eerste de GATA2 ortholoog is die nodig is voor het behoud van hematopoïetische stamcellen, hebben we ons gericht op gata2b om een knock-out zebravissen te genereren waar gebruik van is gemaakt in hoofdstukken 2-3. Opvallend is dat de homozygote Gata2b-mutant levensvatbaar is, waardoor we de functie van Gata2b in volwassen-hematopoëse kunnen bepalen. In hoofdstuk 2 karakteriseren we de effecten van Gata2b homozygote knock-out (gata2b−/−) in het hematopoïetische systeem. Gata2b is nodig voor de proliferatie van hematopoïetische stam- en voorlopercellen (HSPC’s) tijdens embryonale HSPC-expansie in het caudale hematopoïetische weefsel. Volwassen gata2b−/− HSPC’s zijn meer proliferatief en hun niermerg vertoont een lymfoïde neiging en een blokkade in neutrofielendifferentiatie.

Eencellige transcriptoomanalyse toonde aan dat de HSPC’s de bron waren van de verhoogde lymfoïdecellen productie in gata2b−/− niermergcellen, als gevolg van het niet voldoende induceren van myeloïde genexpressie en een daaropvolgende co-expressie van zowel myeloïde als lymfoïde genen in gata2b−/− HSPC’s.
Om de transcriptionele en fenotypische gevolgen van een halve dosis Gata2b voor de hematopoëse van de zebravis te evalueren, hebben we in hoofdstuk 3 gata2b heterozygote zebravis (gata2b +/-) onderzocht. Onze hypothese was dat zebravissen met een halve Gata2b-dosis ook een verminderde myeloïde differentiatie zouden vertonen, zij het in mindere mate dan bij het homozygote deficiëntie van gata2b. In plaats daarvan observeerden we een ongekende erytroïde dysplasie in het niermerg, maar een behoud van de belangrijkste differentiatielijnen.

Het HSPC-compartiment vertoonde ook transcriptionele veranderingen die overeenkomen met de transcriptionele veranderingen gevonden in humane MDS. Dit zou mogelijk de variabele fenotypes kunnen verklaren die worden aangetroffen bij GATA2-patiënten.

Muisen bezitten, net als mensen, één kopie van het Gata2-gen. Gata2 dubbele knock-out (Gata2-/-) muizenembryo’s bezwijken vóór de generatie van HSC’s, terwijl Gata2 heterozygote knock-out muizen (Gata2 +/-) geen merkbare hematologische fenotypes ontwikkelen. We hypothesiseerden dat een transcriptionele analyse van HSPC’s onopgemerkte eigenaardigheden, die door Gata2 heterozygotie werden veroorzaakt, zou onthullen.

In hoofdstuk 4 ontdekten we dat Gata2 +/- HSPC’s hyperproliferatief zijn en om de langetermijngevolgen van deze aandoening te bepalen, hebben we Gata2 +/- muizen tot 14-15 maanden oud laten worden. Gata2 +/- verouderde HSPC’s blijven hyperproliferatief en veroorzaken een onvoldoende productie van wittebloedcel aantallen na beenmergtransplantatie. Het begin van leukopenie na transplantatie van verouderde Gata2 +/- HSC’s weerspiegelt een belangrijk aspect van menselijke GATA2-deficiëntie en biedt een consistente methodologie voor het reproduceren van een GATA2-geassocieerde aandoening in muizen. Verdere experimenten zullen nodig zijn om de geïdentificeerde leukopenie beter te karakteriseren en om uit te zoeken of andere stressfactoren kunnen bijdragen aan een vergelijkbaar fenotype in een kortere tijdspanne. Aangezien veroudering noodzakelijk is om leukopenie te laten optreden, zou het interessant zijn om de rol hiervan in de ontwikkeling van het fenotype, zoals bij née-gemediaerde ontsteking van het beenmerg, aan te pakken.

Gezamenlijk bieden onze bevindingen inzicht in de biologie van GATA2 en de hemato-poëtische gevolgen die na deficiëntie hiervan ontstaan. Ze stellen de relevantie van het Gata2 +/- muismodel voor de menselijke ziekte en de rol hiervan in de ontwikkeling van het fenotype, zoals bij née-gemediaerde ontsteking van het beenmerg, aan.

Deze bevindingen maken de weg vrij voor verder onderzoek naar de expressie van HSPCs en het effect van veroudering op het fenotype en de expressie van Gata2-deficiëntie in hetbeenmergtransplantaties. Dit zou kunnen leiden tot een betere inzicht in de rol van veroudering in de ontwikkeling van het fenotype en de expressie van Gata2-deficiëntie in hetbeenmergtransplantaties.
Emanuele Gioacchino was born on the 25th of May 1989 in Urbino, Italy. After receiving his high school diploma from “Istituto tecnico Mazzocchi” (Ascoli Piceno, Italy) in 2008, he studied Biology at University of New Orleans, University of Manchester, and University of Tor Vergata (Rome), obtaining the bachelor degree from the latter in 2012. He pursued his master degree in Molecular Biology at the University of Tor Vergata, completing an internship in the research group of Prof. Castagnoli (Rome) and an internship in the research group of Louise Purton (St Vincent’s Institute of medical research, Melbourne). In 2015 he defended his thesis titled “Impact of Hoxa1 expression on MDS” and graduating cum laude. In 2016 he moved to the Netherlands where he was appointed PhD candidate in the research group of Dr. De Pater at the department of Hematology of Erasmus Medical Center (Rotterdam). Here he focused on the role of Gata2 deficiency in mouse and zebrafish hematopoiesis.

Scholarships and Awards

### PHD PORTFOLIO

**Name:** Emanuele Gioacchino  
**PhD period:** May 2016-May 2021  
**Erasmus MC department:** Hematology  
**Promoter:** Prof.dr. I.P. Touw  
**Research school:** Molecular Medicine (MolMed)  
**Supervisor:** Dr. Emma de Pater

<table>
<thead>
<tr>
<th>1. PhD Training</th>
<th>Year</th>
<th>ECTS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General courses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory Animal Science (Art.9)</td>
<td>2016</td>
<td>4.2</td>
</tr>
<tr>
<td>Biomedical Research Techniques</td>
<td>2016</td>
<td>1.5</td>
</tr>
<tr>
<td>Biomedical English writing course</td>
<td>2019</td>
<td>2.5</td>
</tr>
<tr>
<td>Annual course on Molecular Medicine</td>
<td>2017</td>
<td>0.7</td>
</tr>
<tr>
<td>Biomedical Research Techniques</td>
<td>2016</td>
<td>1.5</td>
</tr>
<tr>
<td>Photoshop and Illustrator workshop</td>
<td>2019</td>
<td>0.3</td>
</tr>
</tbody>
</table>

| **In-depth Courses and Workshops** | | |
| Molecular aspects of hematological disorders (2x) | 2016-2017 | 1.4 |
| BD Flow Cytometry Course | 2017 | 0.2 |
| Species-Specific Laboratory fish and claw frogs course (Hubrecht institute) | 2016 | 0.6 |
| Galaxy for NGS | 2017 | 1 |
| Microscopic Image analysis: from Theory to Practice | 2017 | 0.8 |
| Basic course on R | 2019 | 1.8 |

| **Scientific Meetings Department of Hematology** | | |
| Work discussion (Weekly) | 2016-2020 | 8 |
| Journal club/literature discussion (bi-monthly) | 2016-2020 | 7 |
| PhD lunch with invited speaker (Monthly) | 2016-2020 | 2.5 |
| Erasmus Hematology Lectures (Monthly) | 2016-2020 | 2 |
| Hematology floor meetings (Weekly) | 2016-2020 | 8 |

| **National/International conferences** | | |
| Molecular Medicine Day (2x) (Rotterdam) | 2017-2018 | 0.6 |
| Dutch hematology congress (3x) (Arnhem) | 2017-2019 | 0.9 |
| International society of experimental hematology (Frankfurt) | 2017 | 0.3 |
| European Hematology Association (Amsterdam) | 2019 | 0.3 |

| **Presentations/posters** | | |
| Departmental work discussion (Oral 7x) (Rotterdam) | 2016-2020 | 3.5 |
| Journal clubs (Oral 4x) (Rotterdam) | 2016-2020 | 2 |
2. Teaching, Supervision & Organization Activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Year</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization of PhD day</td>
<td>2017</td>
<td>0.2</td>
</tr>
<tr>
<td>Organization and supervision PhD lunch with invited speakers (3x)</td>
<td>2017-2019</td>
<td>0.6</td>
</tr>
</tbody>
</table>

**Total** 57.4

---

**LIST OF PUBLICATIONS**


Emanuele Gioacchino, Cansu Koyunlar, Hans de Looper, Joke Peulen, Mariëtte Ter Borg, Remco Hoogenboezem, Dennis Bosch, Paulette van Strien, Kirsten J Gussinklo, Eric Bindels, Elaine Dzierzak, Ivo Touw and Emma de Pater. GATA2 haploinsufficiency reduces fitness of aged hematopoietic stem cells. Manuscript in submission.
ACKNOWLEDGMENTS

The PhD experience has transformed me during the years and drastically made me grow as a man and researcher. Schopare. I would like to briefly thank all of you for making these years memorable.

First of all a heartfelt thanks to my supervisor Dr. Emma de Pater. Dear Emma, you have guided my research under every aspect, from zebrafish handling to microscopy, writing, presenting and more. I still smile remembering your joy after the PCR confirming germline transmission of the gata2b mutation. You spent so much time training me and correcting my thesis that we almost understand each other without introducing the topic. As the first PhD student of your group it has been interesting observing the responsibility and developments associated with the creation of a new research group. I really appreciated your encouragement during my underwhelming reaction to results and I admired your capacity to deal with your family and your job in an apparent effortless manner. I wish you satisfying progresses at the transplantation lab and all the best for the continuation of your research group in the future.

My warmest thanks goes to my promotor Prof. dr. Ivo Touw. Your enthusiasm for science has been inspiring for me and probably for a generation of students. Thank you for all your support, key comments and suggestions for my dissertation. Your active presence during the Friday morning seminars was much better than a coffee. Thanks to you I had to (secretly) learn how to make the original carbonara. I still hope for you to try it once.

I would like to thank the members of the doctoral reading committee, Prof. dr. Ruud Delwel, Prof. dr. J.N.J. Sjaak Philipsen, Prof. dr. Anna Bigas for scrupulously and efficiently reading my dissertation. A big thanks to all the PIs in the Hematology Department, including Mark, Rebecca, Tom, Mojca, Peter, Frank, Moniek. Your questions and suggestions during our meetings helped me better understand how to approach science and how to reason my research.

This thesis would not be complete without expressing a huge gratitude to the dj, artist, parapnymph and soon to be Dr. Koyunlar. Dear Cansu, the first thing I knew about you was that you like to play the piano. Soon after I discovered your passion for scientific research, art, music and one of my best amigos and lovers Raz. You have been my closest colleague but also one of my best friends and confidant. The joy you transmit every day is obvious to everybody and helped me to overcome the Dutch toneless days. Thank you for all the GATA2 discussions, the experiments and laughter we had together in the last 4 years. You are evolving fast as a scientist and will defend your thesis very soon. I’m sure you will make your daddy proud!

Doctor Hans you were the first person teaching me and showing me the lab tools. Since then you have been a solid reference point for our experiments and always with extreme calm and reliability. A huge thanks for doing the hard work in the generation of the
zebrafish gata2b mutant and a myriad of thanks for all your help and lighting the mood of the lab together with Sensei. Dr Eric you are possibly the most hardworking person in the department. I want to especially thank you for helping me and Cansu with the single cell RNA sequencing. Your offstage presence in the laboratory is of vital importance and allowed plenty of students (me included) to conclude their experiments.

**Joke:** Our dear fishwoman. Thank you for taking good care and respect of our little friends. You had to learn to handle many aspects in our group. I am particularly thankful for the analysis of the kidney marrow cells, it must have been tedious and not a joke. Good luck with your future career.

Thank you Paulette for helping me with the mouse stuff I couldn’t do and suerte being a great double mama.

One, two, three and to the four Dr D is at the door. Dennis Keep It Real.

I owe a huge amount of gratitude to the people that helped me during these years and taught me how to use softwares and machines. Thank you Roger and Remco for your help in the bioinformatics support. You were always very busy but always found some time to help me in the bulk and single cell RNA sequencing. Another big thanks to Michael for your flow cytometry tricks and your patience with mine and Cansu’s sample analysis. Thank you Madelon for the script you provided us and the support in the single cell analysis. I have always been a big fan of your clothing style. To Claire for helping with flow cytometry and antibodies. I still owe you a tiramisu. Good luck with your new family and your beautiful Nora.

Thank you Patty Patricia for giving me a very good impression of the Department after my first interview, for your tips and for your contagious laughter. Good luck in Utrecht!

Dear Eline, my trusty presence to my right side. We have been sitting next to each other for a loong time. I hope one day to be as productive as you at work. Thanks for all the suggestions you gave me, short but to the point!

A big thanks to all the remaining members of the Hematology Department who helped me one way or another like Tessa, Paola, Andrea, Bella, Inge, Jaqueline, Helene and the other students and teachers.

To Egied, thank you for your help in the dissertation design and the patience with my continuous modifications.

**Dr. Mr. Keane,** what would I be doing without you? It has been an honor to teach you the “care-free, bordering on the lackadaisical, attitude” and how to play basketball. I’m very grateful for the graphpad and flojo tricks you taught me in the asian office when Ping was too annoyed to repeat twice, and for the endless suggestions you always provided. Thanks to your gift of the gab I met most of my friends in Rotterdam. I think the 43 shared whatsapp groups summarize well how many extra lab activities we have been through. You will always be a honor member of my community. With your new job you became more serious and less of a party king. Materialism is a false prophet. Keep fit and eat your vitamins. I hope you and Nina will take care of each other like you did in the past and will settle in a more latin country.

**Dr. Mista P.i.n.g,** you are not just a friend but represent for me hard work, perfectionism and honor. You are a nice character and someway always give me joy. I’m grateful for your tips and vision of the scientific world and for the FACS and flojo tricks I was eager to learn from you. My honor for being your host in Italy and at Bababoom. Claudio and Jacopo keep on asking me about you! Even if we have opposite views on priorities we found an unexpected connection which I hope will last even if you won’t invite me to your house anymore (I’m again sorry about the plant;)). Talking about plants... Sap Mr. Onnoooool! Your bush has been a point of reference both during festivals and during my lab days. You have been always very kind to the newbies like me and were available for helping with confocal, stress assays and a bunch of other technical issues. Some time has passed since our first out-the-lab activity in Rochussentraat. Our shared interests in music, travelling, nature and fun gave us plenty of opportunities to have some good memories outside the lab which I hope to repeat as soon as possible. What about when a guy started crying for a drum solo at the jazz festival in Tivoli? Thanks for being such a kind, solid and happy column in my years in Roffa. I am sure I will keep on seeing you and your bush clones in the future!

Multumesc dragostea mea Raz, together with Ivasito you made me enter into a new world in Rotterdam and I will always be grateful for it. I wish I could spend more time with that serious funny face of yours and get lost in your beard.

A big thanks to all the bisons Alex and Tania, Alicia, the Frank, Ian my man, Enrique, Irene and family, la bella e bona Isabona, Kasia, Lena, Lennon, Pablo, Jana, Rodrigo etc. and to the Italian crew Enrico, Francesca, Agnese, Matteo, Livia, Simon, Giorgia, Marta, Roberta, Elena, Pazzo, Carlos and Joanna, Geoffrey. All of you made me feel so loved and exponentially boosted my well-being, thank you chicos y chicas!

**Francesca** grazie per storare tutti i miei beni materiali e ... per i caffe’. Sei sempre bella e carica e mi fai sentire a casa da dodicesimo piano. I know you expect me to write a long ringraziamento per you. Sorry to disappoint tu, but I tuoi rimangono longer than Francois.

Statte tranquilla e daje forte. Te vulem ben.

A big hug to all my ex flatmates Thomas, Matteo, Lina and Gizem. Good times damn... good times! Thomas desaparecido fatti sentire cazzo che ci manchi. Matteo thank you in advance for hiring me if I can’t find a job. E’ stato un piacere spiegarti come fare le rovescite. Thanks to Dr Sebasito for introducing me to roffa underground scene and taking care of our whatever disease we get. Without you and Lina Gini Pina I would have done 2 vaccine chip doses instead of one. Muchas gracias Pina for all the love you transmitted in these years and for ensuring that me and Thomas were going to sleep on time. Grazie to my short lasting landlord Gizem for hosting me and, just in a few days, for having recharged me of energies and self-trust.

To all the desaparecidos like my dear Ivasito, Teresa, Roberto, Monica, Adrian, Mira.
You are not forgotten! Your presence is still very vivid in Rotterdam for many of us.

Thanks to Lorenzo for not caring at all about my work and dissertation. You showed me how it is possible to be more careless than me. This made you a valuable paraparanymph.

We were meant to be together from the start and for a variety of reasons we got to know each other pretty well. It’s sometimes scary how many passions we share and how many things you manage to do better than me at the end. We are so similar and so different at the same time but we are both big fans of lafiga. Suerte with your last experiments y con la guapasita elenita.

Cristina you have been next to me in the most tough years of my research and we had a beautiful experience together. I hope you find the right connection with nature and the best compromise with commodity. Lemme know what you find.

Thanks to the Almondestraat crew who gave me a house when I was jobless and allowed me to finish my thesis with a roof over my head. You can be destroyed but not forgotten.

Ciao to Pietro, Gaia, Francisco, David, Izzy and the rest it has been some good saunas and some good parties. 217 respect.

The odd case of discovering a passion after 4 years. Francois, I love you as much as a straight man can love another man. More words would only underrepresent the connection we surprisingly found my dear paraparanymph.

Grazie a tutti gli italiani d’Italia come il Fiorellin, Tota and family, lu scem di Luzi and family, il magico Teo, la stramamma Feta e il Mauri, la strasabrina Sabrina, la neomamma Farciola, i polish Zu e Polska, il Cicchen, il Garcia lorca Paolo e Maria. Avete provato un tutti i modi a farmi fare solo festa ma alla fine sono riuscito a districarmi pure con gli studi. Vi voglio bbe’!

Obviously grazie al faro Anna, la Maiorca e la mia cara ammmericana Giulia. Grazie per avermi fatto laureare, senza di voi starei ancora facendo l’ esame di informatica a Tor Vergata.

Thanks to el mi amor Josephine who supported and encouraged me during the last years. Besides making the amazing gata2b animation and my thesis cover you assured to always make me feel loved and still teach me how to be kind. I hope you will soon appreciate how smart and interesting you are and how lucky you make me feel. We already had quite some adventures together and more will come! No matter what happens and where it will happen we can count on each other. T’amo tanto.

E chiaramente il ringraziamento più grande va alla mia famiglia che ancora crede quando dico che faccio lo scienziato. Un grazie alla bella sorella Gioel per avermi sempre dato il buon esempio e un obiettivo da raggiungere. Lo sai solo tu quanto ti voglio bene. Mo non esprimo tutte le emozioni sennó mi richiedi la tesina per approfondire. Un abbraccio a Kirsten che e’ tanto una brava uagliona e che trova la mia simpatia quando gli altri non la colgono.

A michela, giuliano e andrea che continuano ad ospitarmi a Roma facendomi sentire a casa e che sono rimasti sempre al mio fianco. Alla nonna Adina pe tutt lu magna’che c’ ha fatto e tutto l’ amore che ci hai dato e che continuiamo ad usare. M capisc sol tu.

Un grazie infinito ai miei genitori che facendomi sentire di essere fieri di me mi hanno trasmesso una enorme fiducia in me stesso e la capacità di andare oltre molti ostacoli.

Grazie a “il padre” il Papà che ci ha sempre supportato negli studi e aiutato in mille altri modi. Hai sempre trovato il modo di dedicarti il tuo tempo e di giocare con noi. Il patre esemplare!

E ovviamente il grazie più grande va a MAMMA’, o core mio e la cosa piu’ cara che ho. Sei sempre andata oltre i tuoi limiti per assicurarti che stessimo bene. Anche se alle volte non hai i riconoscimenti che meriti ricorda che i tuoi figli ti amano tanto.