Gata2 in embryonic hematopoiesis

Polynikis Kaimakis

ISBN:

The work presented in this thesis was performed at hte Department of Cell Biology at the Erasmus MC, Rotterdam, The Netherlands. The department is a member of Medisch Genetisch Centrum Zuid-West Nederland (MGC).

The research has been funded by ZonMW (Dutch Medical Research Council, (911-09-036), FES NIRM (Dutch Innovation Grant) and NIH (RO37 DK54077).

Cover Design: ενπνοή Μπούρας Χαράλαμπος
info@enpnoi.gr

Gata2 in Embryonic Hematopoiesis

Gata2 in embryonale hematopoiese

Thesis

To obtain the degree of Doctor from the
Erasmus University of Rotterdam
by command of the
rector magnificus
Prof.dr. H.G. Schimidt
and in accordance with the decision of the Doctorate Board

The public defence shall be held on
Wednesday 23 of January at 9:30 hours

by

Polynikis Kaimakis

born in Thessaloniki, Greece

Doctoral committee

Promoter:	**Prof.dr E.A. Dzierzak**
Other members:	**Prof.Dr J.N.J. Philipsen**
	Prof.Dr. D.N. Meijer
	Dr.R.A. Poot

Ιθάκη - Κ.Π. Καβάφης

Σα βγεις στον πηγαιμό για την Ιθάκη,
να εύχεσαι νάναι μακρύς ο δρόμος,
γεμάτος περιπέτειες, γεμάτος γνώσεις.
Τους Λαιστρυγόνας και τους Κύκλωπας,
τον θυμωμένο Ποσειδώνα μη φοβάσαι,
τέτοια στον δρόμο σου ποτέ σου δεν θα βρεις,
αν μέν' η σκέψις σου υψηλή, αν εκλεκτή
συγκίνησις το πνεύμα και το σώμα σου αγγίζει.
Τους Λαιστρυγόνας και τους Κύκλωπας,
τον άγριο Ποσειδώνα δεν θα συναντήσεις,
αν δεν τους κουβανείς μες στην ψυχή σου,
αν η ψυχή σου δεν τους στήνει εμπρός σου.

Να εύχεσαι νάναι μακρύς ο δρόμος.
Πολλά τα καλοκαιρινά πρωιά να είναι
που με τι ευχαρίστησι, με τι χαρά
θα μπαίνεις σε λιμένας πρωτοειδωμένους·
να σταματήσεις σ' εμπορεία Φοινικικά,
και τες καλές πραγμάτειες ν' αποκτήσεις,
σεντέφια και κοράλλια, κεχριμπάρια κ' έβενους,
και ηδονικά μυρωδικά κάθε λογής,
όσο μπορείς πιο άφθονα ηδονικά μυρωδικά·
σε πόλεις Αιγυπτιακές πολλές να πας,
να μάθεις και να μάθεις απ' τους σπουδασμένους.

Πάντα στον νου σου νάχεις την Ιθάκη.
Το φθάσιμον εκεί είν' ο προορισμός σου.
Αλλά μη βιάζεις το ταξείδι διόλου.
Καλλίτερα χρόνια πολλά να διαρκέσει·
και γέρος πια ν' αράξεις στο νησί,
πλούσιος με όσα κέρδισες στον δρόμο,
μη προσδοκώντας πλούτη να σε δώσει η Ιθάκη.

Η Ιθάκη σ' έδωσε τ' ωραίο ταξείδι.
Χωρίς αυτήν δεν θάβγαινες στον δρόμο.
Άλλα δεν έχει να σε δώσει πια.

Κι αν πτωχική την βρεις, η Ιθάκη δεν σε γέλασε.
Έτσι σοφός που έγινες, με τόση πείρα,
ήδη θα το κατάλαβες η Ιθάκες τι σημαίνουν.

Ithaca - Constantine P. Cavafy

When you set sail for Ithaca,
wish for the road to be long,
full of adventures, full of knowledge.
The Lestrygonians and the Cyclopes,
an angry Poseidon — do not fear.
You will never find such on your path,
if your thoughts remain lofty, and your spirit
and body are touched by a fine emotion.
The Lestrygonians and the Cyclopes,
a savage Poseidon you will not encounter,
if you do not carry them within your spirit,
if your spirit does not place them before you.

Wish for the road to be long.
Many the summer mornings to be when
with what pleasure, what joy
you will enter ports seen for the first time.
Stop at Phoenician markets,
and purchase the fine goods,
nacre and coral, amber and ebony,
and exquisite perfumes of all sorts,
the most delicate fragances you can find.
To many Egyptian cities you must go,
to learn and learn from the cultivated.

Always keep Ithaca in your mind.
To arrive there is your final destination.
But do not hurry the voyage at all.
It is better for it to last many years,
and when old to rest in the island,
rich with all you have gained on the way,
not expecting Ithaca to offer you wealth.

Ithaca has given you the beautiful journey.
Without her you would not have set out on the road.
Nothing more does she have to give you.

And if you find her poor, Ithaca has not deceived you.
Wise as you have become, with so much experience,
you must already have understood what Ithacas mean.

(Poems 1897-1933)

... because that poem perfectly sums up the PhD process!

Table of Contents

Chapter 1: General Introduction p9

1.1 Hematopoiesis and Hematopoietic stem cells (HSCs) p10

1.2 Hematopoiesis in the mouse embryo p11

1.3 Molecular programming of HSC generation is distinct from programming of transient waves of hematopoietic cell generation. p13

1.3.1 Intrinsic regulators of hematopoiesis p15

1.3.1.a Scl p16

1.3.1.b Gata2 p18

1.3.1.c Runx1 p19

1.3.1.d Other Factors p21

1.3.2 Extrinsic regulation of HSC generation p22

1.4 Scope of the thesis p24

1.5 References p25

Chapter 2: Gata2 expressing cells in embryoid bodies (EB) show a hemangioblast phenotype resembling that of early embryonic hematopoiesis p31

Chapter 3: Changing levels of Gata2 expression affect hematopoietic development p45

Chapter 4: Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators p69

Chapter 5: General Discussion and future experiments p82

Summary Samenvating p91

Abbreviations p97

CV p99

PhD portfolio and publications p100

Acknowledgements p102

Chapter 1

General Introduction

Part of this chapter has been published in: Biochim Biophys Acta. 2012 Oct 12. doi:pii: S0304-4165(12)00284-X. 10.1016/j.bbagen.2012.10.004. [Epub ahead of print]
Review by Kaimakis P, Crisan M, Dzierzak E

1. General Introduction

1.1 Hematopoiesis and Hematopoietic stem cells (HSCs)

Hematopoiesis is a word originating from the two greek words αἷμα (haima) which means blood and the verb ποιεῖν (poien) which means to make/create. Hematopoiesis describes the process by which the organism creates and replenishes all the blood cell types that are required for the physiologic functions of the organism. The importance of the blood tissue can be highlighted by the many and discrete functions that it performs. These are accomplished through several different cell types forming the blood tissue (erythrocytes, platelets, macrophages, neutrophils, eosinophils, basophils, B-cells, T-cells, NK-cells). For example, the red blood cells or erythrocytes found in the circulating blood are mainly involved in the transport of O2 and CO2. Lymphocytes which are white blood cells are part of the immune system and actively participate in the defense of the organism against pathogens. In the adult organism hematopoietic cells are found not only in the blood but also in hematopoietic tissues such as the bone marrow, spleen, lymph nodes and thymus. Importantly, all mature hematopoietic cell types found in the blood tissues originate from rare hematopoietic stem cells (HSCs). These founder cells are quiescent, long-lived and are at the base of a well-studied cell differentiation hierarchy. HSCs robustly produce all the billions of mature blood cells that are required daily and throughout the entire life of the organism. HSCs are clinically relevant cells that have been used for over 50 years in transplantation and cell replacement therapies for leukemia and monogenic blood-related diseases.

The generation, growth and maintenance of HSCs are tightly controlled – massively reduced number of HSCs leads to anemia whereas unlimited growth could possibly result in leukemia. Hence, a variety of molecular mechanisms participates and controls HSC. These include extracellular signals that act in co-operation with cell surface receptors and intracellular signaling pathways, transcriptional regulation through pivotal transcription factors and transcription factor complexes and chromatin modifications that stabilize the genetic program of hematopoiesis. The harmonic cooperation of these mechanisms at the different developmental and adult stages of life, influence HSCs and allow for homeostasis of the blood tissue and rapid response following hematologic insult. Extensive investigations are being performed in this research field to provide insight and understanding into the precise events that control first the process of HSC generation and subsequently HSC expansion and maintenance.

In this Introduction we focus on the regulation of HSC development in the mouse embryo. By performing experiments exploring regulation of the HSC genetic program, we attempt to understand how the Gata2 transcription factor, a major

regulator of hematopoiesis, affects the fate of the cells that will give rise to HSCs and the fate of the HSCs themselves, as they differentiate to all different blood cell types.

1.2 Hematopoiesis in the mouse embryo

In the adult mouse the cohort of several hundred HSCs are located in the bone marrow (BM) niche in two anatomical locations, the endosteal niche (in proximity to the osteoblasts of the bones) ((Ellis, Grassinger et al. 2011)), and the vascular niche (in proximity to the vessels) ((Kiel and Morrison 2006; Sugiyama, Kohara et al. 2006)). Interactions with these niches allow for the HSCs to be maintained in a quiescent state until needed. At any point in time only a few of these HSCs are actively cycling. During cell division an HSC will produce a daughter cell that differentiates and produces progenitor cells at branchpoints along the hematopoietic hierarchy that will yield to specific mature blood cell types. Importantly, the other daughter cell retains HSC properties. The defining characteristic of the HSC is its self-renewal ability.

HSCs in the adult organism are not generated in bone marrow. They are generated much earlier during a short period of embryonic developmental time - at embryonic day (E) 10.5 to 12.5 in the mouse embryo and at 4-8 weeks gestation of the human embryo (Tavian, Robin et al. 2001; Tavian and Peault 2005). After their production, HSCs migrate and colonize several different hematopoietic sites during embryonic and fetal stages of development before reaching and colonizing the bone marrow (BM) niche. Much research is focused on determining what is the first site where the HSCs appear in the embryo, what tissues/cells generate these HSCs, at what timepoints they are made and what is the genetic program that directs their generation.

Hematopoiesis in the mouse embryo happens in three *de novo* waves of generation (reviewed by (Dzierzak and Speck 2008)) with each one being characterized by the production of different types of mature hematopoietic cells, hematopoietic progenitors and/or HSCs (Palis, Robertson et al. 1999). The initial two hematopoietic waves of generation provide the embryo with the reservoir of hematopoietic cells that can temporally and rapidly produce mature hematopoietic cells until the generation of HSCs in the third wave, that result in the establishment of the long-lived adult hematopoietic system. Several anatomical sites participate in hematopoietic cell generation in the mouse embryo. These are the yolk sac (YS), the aorta-gonads mesonephros (AGM) region, the vitelline and umbilical arteries (VA, UA), the placenta (PL) and the fetal liver (FL) (Figure 1A).

The first wave of blood generation produces short-lived primitive erythrocytes that are necessary to carry oxygen through the rapidly growing conceptus

and also primitive macrophages and megakaryocytes. Primitive erythrocytes are generated from aggregates of mesodermal precursors or "hemangioblasts", in the yolk sac blood islands. Described over 100 years ago, the overlapping ontogenic appearance of both erythroid and endothelial cells indicates a common mesodermal precursor with at least bi-lineage potential (Murray, Kilham et al. 1946)-(Sabin 1920). This is further supported by the overlap in genetic programs for the two lineages (i.e. expression of Flk-1 (KDR), Scl (Tal1) and CD34) and the lack of both lineages in embryos deficient for some of these genes (Shalaby, Rossant et al. 1995; Tavian, Hallais et al. 1999; Park, Ma et al. 2005). Hemangioblasts *in vivo* can first be detected not in the yolk sac but in the posterior primitive streak (Huber, Kouskoff et al. 2004). As they migrate to the yolk sac they begin their commitment to endothelial and hematopoietic progenitors, with several of these cells contributing to the formation of each blood island (Ueno and Weissman 2006). The first wave of primitive hematopoietic cell generation begins at E7.5 in the mouse conceptus and is highly conserved across vertebrate species, including man (at 16-20 days of gestation) (Tavian and Peault 2005).

In the mouse embryo the second wave of hematopoietic cell generation begins at E8/8.5, and overlaps with the first wave (Palis, Robertson et al. 1999; Medvinsky, Rybtsov et al. 2011). Definitive hematopoietic progenitors are *de novo* generated and some clusters of hematopoietic cells begin to appear in the major vasculature at E9.5. These hematopoietic progenitors are functionally more complex than primitive progenitors – they have multilineage potential (producing erythroid, myeloid and/or lymphoid cells), but they are not long-lived or self-renewing HSCs. *De novo* definitive progenitor generation occurs in the yolk sac, chorio-allantoic/placenta and the intraembryonic region around the aorta, as revealed by mouse embryo explant cultures and the *Ncx1-/-* mouse model (embryos lack circulation due to no heartbeat (Lux, Yoshimoto et al. 2008; Rhodes, Gekas et al. 2008) (reviewed in (Medvinsky, Rybtsov et al. 2011)). Thus, "definitive hematopoietic progenitors" constitute the second wave of hematopoietic specification.

The third wave of hematopoietic cell specification provides for the generation of adult type HSCs. Grafting studies in avian embryos provided unequivocal proof that the adult blood system is not derived from the yolk sac, but instead from an intraembryonic source of cells localizing to the dorsal aorta (reviewed in (Dzierzak and Speck 2008; Medvinsky, Rybtsov et al. 2011)). Clusters of hematopoietic cells consistently found on the ventral wall of the dorsal aorta and major arteries of the chick embryo, led to the proposition that HSCs of the adult hematopoietic system arise from vascular endothelial cells (Figure 1B). Work in the mouse embryo showed that the first adult-type HSCs are generated in the intrabody AGM region (Figure 1A and B). These transplantable HSCs (as potent as adult bone marrow

HSCs) are generated beginning at E10.5 (Medvinsky and Dzierzak 1996) and are thought to be contained within the vascular clusters within the aorta and vitelline and umbilical arteries (North, Gu et al. 1999; de Bruijn, Speck et al. 2000; de Bruijn, Ma et al. 2002). The real-time generation of hematopoietic cells from "hemogenic endothelial cells" lining the aorta (Figure 1C) has been demonstrated by vital confocal imaging in the mouse and zebrafish embryo (Bertrand, Chi et al. 2010; Boisset, van Cappellen et al. 2010; Kissa and Herbomel 2010). It is this third wave of hematopoietic cell generation that generates the long-lived self-renewing HSCs that migrate, colonize and reside in the bone marrow throughout adult life.

Hemogenic endothelium as a source of definitive hematopoietic progenitors and HSCs.

The generation of definitive hematopoietic progenitors (wave 2) and HSCs (wave 3) parallels the appearance of vascular hematopoietic clusters in the aorta, vitelline and umbilical arteries (Figure 1C). Histologic/immunostained sections through the midgestation embryo AGM region show that "hemogenic" endothelial cells express some hematopoietic markers, and some hematopoietic cluster cells express endothelial markers (Yokomizo, Yamada-Inagawa et al. 2012). Cluster numbers peak at E10.5, when HSCs first appear. However, not all hematopoietic cells in the clusters are HSCs and not all clusters contain HSCs. There are many more cluster cells in the aorta than there are HSCs between E9 and E11. Genetic studies using Cre-Lox recombination methods for deletion of pivotal intrinsic regulatory molecules (such as the Runx1 transcription factor) show that HSC generation occurs only during a short window of developmental time (Zovein, Hofmann et al. 2008; Chen, Yokomizo et al. 2009). It is unclear as yet whether all HSCs/cluster cells emerge from hemogenic endothelium, whether larger clusters form by proliferation of the emerging cell or through the recruitment of circulating cells. Recently, it was suggested that already hematopoietic committed cells (perhaps coming from the yolk sac, circulation (Yoder, Hiatt et al. 1997) or other areas (Rybtsov, Sobiesiak et al. 2011) could mature to HSCs as they integrate into the appropriate microenvironment (Rybtsov, Sobiesiak et al. 2011).

1.3 Molecular programming of HSC generation is distinct from programming of transient waves of hematopoietic cell generation.

The generation of HSCs in the embryo is special as compared to the earlier two waves of blood generation. Each wave produces differentiated hematopoietic cells that circulate and move through the embryo to perform specific functions that are not so different from the differentiated hematopoietic cells arising from adult bone marrow HSCs. However, it is only the third wave that generates HSCs. The unique production of HSCs has raised the following questions: Is it the vascular mi-

croenvironment or the intrinsic program of the hemogenic endothelial cells as they take on hematopoietic fate that directs the differences in hematopoietic cell/HSC potency, or is it a combination of both regulatory mechanisms? It is possible that extrinsic signals (growth factors and morphogens) produced by the microenvironment/niche and delivered to the HSCs or their cellular precursors initiate an intrinsic molecular program guiding hematopoietic fate by the activation or expression of pivotal transcription factors (Figure 2). Here we will describe some of the well known examples of intrinsic and extrinsic regulators of hematopoietic cell fate.

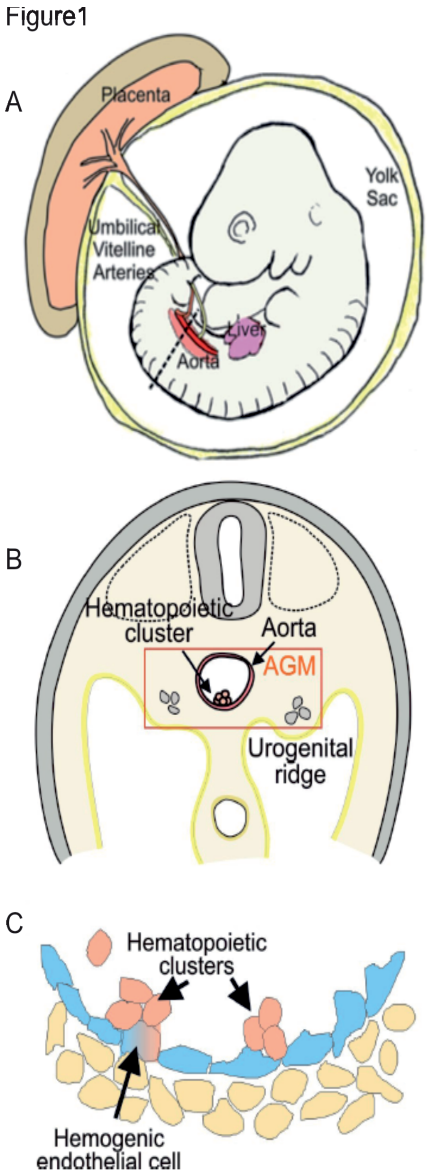

Figure 1

Hematopoietic stem cell development in the mouse embryo. A) Depiction of a mouse embryo at day 10.5 at the time when the first hematopoietic stem cells are generated in the aorta. Sites harboring (and/or generating) hematopoietic cells are shown: the extraembryonic yolk sac and placenta, the intraembryonic aorta and liver, and the umbilical and vitelline vessels that respectively connect the placenta and yolk sac to the aorta. The dotted line through the trunk of the embryo indicates the transverse section shown in panel B. B) Depiction of a transverse section through an E10.5 mouse embryo with the AGM (aorta-gonad-mesonephros/aorta and urogenital ridges) region in the red rectangle. The AGM is flanked on the dorsal side by the neural tube and the somites, and on the ventral side by the gut and peritoneum. A hematopoietic cluster is indicated on the ventral wall of the dorsal aorta. Hematopoietic stem cells are localized in the clusters. C) A close up of the ventral wall of the aorta showing cluster formation. A hemogenic endothelial cell is undergoing the transition from endothelial cell to a hematopoietic cell.

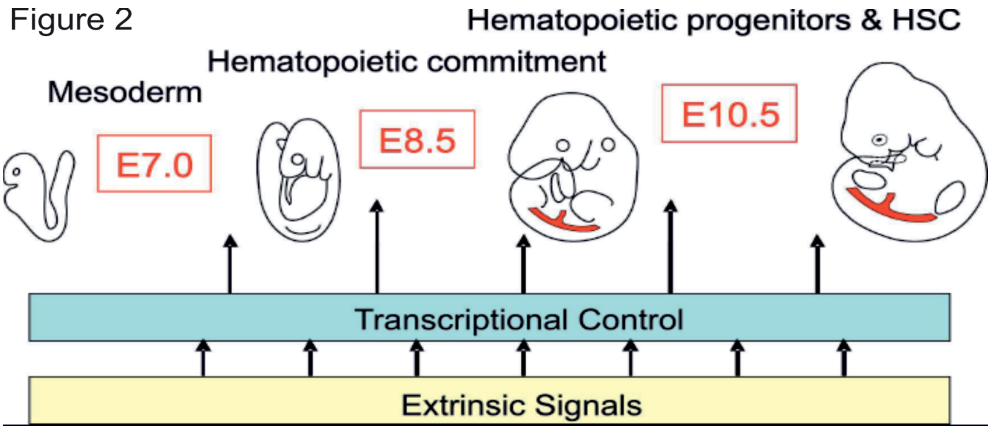

Figure 2

De novo generation of hematopoietic cells begins shortly after mesoderm formation and continues through mouse midgestation. Extrinsic factors that include morphogens such as FGF, Hh and BMPs produced in surrounding microenvironment affect mesodermal cells in hematopoietic fate choice and differentiation. These signaling pathways, as well as the Notch and VEGF pathways, impact directly or indirectly on the expression of several hematopoietic transcription factors in the presumptive hematopoietic cells in different hematopoietic sites and stages of development. The transcription factors directing hematopoietic fate and blood cell production include SCL, Gata2 and Runx1, amongst others. The specific temporal and spatial sequence of extrinsic signals, the combination and/or the levels of extrinsic signals play a role in the differential transcription factor expression and production of the distinct waves of hematopoietic cells in the developing embryo.

1.3.1 Intrinsic regulators of hematopoiesis

Intrinsic regulation of HSC generation and differentiation is performed through downstream effectors/genetic programs directed in their expression by pivotal hematopoietic transcription factors. Most transcription factors act specifically through the recognition of DNA sequences associated with the non-coding regions of target genes. Transcription factors act as single units or in complexes in combination with other proteins (other transcription factors, co-factors, chromatin remodelers, histone acetylases or deacetylases etc). Transcription factors act directly on promoter regions or indirectly through enhancer regions, leading to positive or negative effects on target gene transcription. The temporal and spatial expression patterns of transcription factors, especially during embryonic development, as well as their expression levels/dosage have been proven to play a significant role in regulation of the hematopoietic program affecting cellular identity, growth and physiologic function.

Several of the major transcriptional regulators of HSC specification and HSC generation have been identified and include the *Tal1* (Porcher, Swat et al. 1996;

Robb, Elwood et al. 1996), *Gata2* (Tsai, Keller et al. 1994) and *Runx1* (Okuda, van-Deursen et al. 1996; Wang, Stacy et al. 1996) transcription factors. Initially identified through chromosomal translocations and dysregulated expression in leukemic cells (*Runx1* and *Scl*), these hematopoietic transcription factors were found to play pivotal roles through genetic deletion approaches in the mouse. Germ line deletion of each of these genes, *Scl, Gata2* and *Runx1*, leads to midgestation embryonic lethality and profound anemia. Interestingly, these factors appear to work in complexes that bind to regulatory regions in a large overlapping set of hematopoietic specific genes (Wilson, Foster et al. 2010).

Below I discuss some characteristic examples of transcription factors with important roles in hematopoietic development, maintenance and differentiation. As major regulators of HSC generation, the Scl, Gata2 and Runx1 transcription factors are described in detail. Other factors such as Fli-1, PU.1, Gfi1 and Lmo2, that play specific roles in hematopoiesis, are briefly described.

1.3.1.a Scl

The Scl (Scl/TAL-1/TCL-5) transcription factor is characterized by a basic helix-loop-helix domain (bHLH). As a member of the bHLH protein family, it acts through the HLH domain in a heterodimerization complex with E-proteins that recognize specific DNA sequences called E-box motifs (G/ACANNTGG/A) (Lecuyer and Hoang 2004; Murre 2005). Whereas Scl is most commonly known for its activity in acute T-cell leukemias (reviewed by (Begley and Green 1999)), the importance of Scl in normal hematopoiesis is still under research, using approaches that include germline and conditional deletion of the transcription factor in embryonic stem (ES) cells and mice.

Scl germline deficient mice show early embryonic lethality at E9.5, with an absence of hematopoiesis. This extensive anemic phenotype is also thought to be the cause of death of these embryos. No primitive or definitive hematopoietic cells are produced, yolk sac vasculature is abnormal and no vitelline vessels are formed (Robb, Lyons et al. 1995; Shivdasani, Mayer et al. 1995; Robb, Elwood et al. 1996) . Scl is required for hematopoietic specification in ventral mesoderm or “hemangioblast“ generation (wave 1). Its requirement for the subsequent 2nd and 3rd waves of mouse hematopoietic development is as yet unknown. The contribution of *Scl-/-* ES cells to the formation of mouse chimeras has also been tested. *Scl-/-* ES cells do not contribute to the formation of hematopoietic tissues, whereas they contribute to other various tissues. Further studies in this chimeric context proved Scl to be essential for lymphoid development (Porcher, Swat et al. 1996). An *in vitro* approach has been followed to test the hematopoietic potential of *Scl-/-* ES cells. These cells show a complete lack of myeloid and erythroid hematopoiesis, an effect that can be rescued by retroviral introduction of *Scl* cDNA in these cells or by tamoxifen inducible Scl expression (Porcher, Swat et al. 1996; Robb, Elwood et al. 1996; Endoh,

Ogawa et al. 2002).

Scl is normally expressed in the developing embryo in hematopoietic, endothelial and neural tissues. Expression in endothelial cells and hematopoietic tissues is of particular interest because of the developmental connection of these two cell lineages (hemangioblast, hemogenic endothelium) (Kallianpur, Jordan et al. 1994; Elefanty, Begley et al. 1999). Several studies demonstrate the specific role of Scl in the proper development of both endothelial and hematopoietic cell lineages (reviewed by (Lecuyer and Hoang 2004) and suggest that it functions at the interface between vascular endothelial and hematopoietic lineages. Lancrin et al. found that *Scl-/-* Flk1+ ES-derived cells were unable to generate blast colonies (hemangioblasts) and also that there was a complete lack of cells expressing hematopoietic markers c-kit and CD41, indicating that Scl is critical for the generation of the hemogenic endothelial population. Further studies by this group places its role in hematopoietic hematopoietic specification before the Runx1 requirement (Lancrin, Sroczynska et al. 2009) (see below). Differentiation studies in Zebrafish embryos and in ES cells show that Scl is a regulator of mesodermal patterning (Ismailoglu, Yeamans et al. 2008; Peterkin, Gibson et al. 2009). Scl serves as an indirect negative regulator of cardiogenesis, and this function is independent of its later role in hematopoietic differentiation (Schoenebeck, Keegan et al. 2007; Van Handel, Montel-Hagen et al. 2012). Thus, Scl plays an important role in the presumptive hematopoietic progenitor/stem cell mesoderm during embryonic development (Hall, Curtis et al. 2003; Mikkola, Klintman et al. 2003; Van Handel, Montel-Hagen et al. 2012).

To define the effects of *Scl* deletion in adult mice, conditional knock out approaches have been taken, circumventing the embryonic lethality resulting from its germline deletion. When deleted in the adult, lineage-specific effects of Scl deficiency are found in erythroid cells and megakaryocytes, but not other myeloid lineage cells. Interestingly, Scl participates in a “core transcriptional network” (together with Gata1 and Klf1) to regulate erythroid lineage differentiation (Wontakal, Guo et al. 2012). It has been shown to be a direct regulator of a large group (>300) of erythroid specific genes, such as *Eklf1, Gata1, glycophorin A*, *c-kit* and *globin* (Kassouf, Hughes et al. 2010). Another study has revealed that Scl, through the transcriptional regulation of the cell cycle controller p21, regulates megakaryopoiesis. Importantly, although HSCs (and multipotent progenitors) express Scl, they do not rely on it for their ability to self-renew, differentiate and long-term reconstitute the blood system of irradiated adult recipients (Mikkola, Klintman et al. 2003). The apparent lack of effect of *Scl* deletion on adult HSCs is due to its redundancy with *Lyl1* (Souroullas, Salmon et al. 2009). Recent findings indicate that the levels of Scl expression control adult HSC quiescence by restricting the G1 entry of dormant HSCs (Lacombe, Herblot et al. 2010).

1.3.1.b Gata2

The Gata2 transcription factor which plays a role during the development of hematopoietic progenitor and stem cells is a member of the Gata family of factors. The 6 evolutionarily conserved proteins - Gata1 to Gata6 - recognize and bind the A/TGATAA/G DNA sequence from which they take their name, and have two highly conserved zinc finger (ZnF) domains (Ko and Engel 1993). Gata1, 2 and 3 are relevant to the hematopoietic system – Gata1 for the erythroid lineage, Gata2 for hematopoietic stem and progenitor cells, and Gata3 for T lymphocytes.

The expression pattern of *Gata2* in the mouse embryo provides some information on its importance in the hematopoietic development. It is highly expressed in hematopoietic cells of the fetal liver, placenta and in endothelial cells including the endothelium of the dorsal aorta from which HSCs emerge and intra-aortic hematopoietic cluster cells (Ng, George et al. 1994; Nardelli, Thiesson et al. 1999; Minegishi, Suzuki et al. 2003; Robert-Moreno, Espinosa et al. 2005). Moreover, high levels of *Gata2* mRNA are detected in cell populations enriched for HSCs (Orlic, Anderson et al. 1995). It is also expressed in the central nervous system (CNS) and fetal heart.

The importance of *Gata2* in the process of HSC generation was first highlighted by the creation of *Gatat2* deficient mice. *Gata2-/-* embryos exhibit embryonic lethality between E10-E10.5 and are characterized by severe anemia (Tsai, Keller et al. 1994). Primitive hematopoietic progenitor numbers are decreased. Definitive hematopoietic progenitors are most profoundly affected and no HSCs are produced (Ling, Ottersbach et al. 2004). The rescue of this lethal *Gata2-/-* phenotype by functional complementation was performed using a YAC clone bearing the murine *Gata2* genomic locus (Zhou, Lim et al. 1998). Further experiments performed with *Gata2-/-* ES cells differentiated into EBs under specific hematopoietic conditions, examined the hematopoietic progenitor compartment. The greatly reduced number of definitive hematopoietic progenitors and their poor expansion capacity lead to the conclusion that Gata2 is required for proliferation and survival of early hematopoietic cells (Tsai and Orkin 1997). Analysis of ES cell chimeric mice reveals a lack of contribution of *Gata2-/-* cells to any hematopoietic tissue (Tsai, Keller et al. 1994). Interestingly, *Gata2+/-* embryos are greatly reduced in AGM HSC numbers, as shown by transplantation experiments (Ling, Ottersbach et al. 2004). In these studies the number of Ly6A-GFP cells, in which the HSC population can be found, is decreased 10 fold. However, by the time such animals reach adulthood the number of HSCs is normal but they are qualitatively impaired, as shown in competitive transplantation assays. Thus, Gata2 plays several distinct roles during development - HSC production, expansion and potency.

Putative genome-wide target genes of Gata2 have been described recently

in the hematopoietic progenitor cell line HPC-7. In genes important for hematopoiesis, Gata2 binding sites overlap with those of other transcriptional factors (such as Scl, Runx1, Lmo2, Lyl1, Fli-1 and Erg), implying the collaboration of these factors in a "heptameric" transcription factor complex (Wilson, Foster et al. 2010), although the physical interaction of all these transcription factors still needs to be verified by mass spectrometry. Another study in bone marrow progenitor cell populations suggests that Gata2 and Scl function in combination with Ldb1, another transcriptional factor essential for hematopoiesis (Li, Jothi et al. 2011).

Although, the *Gata2-/-* mouse studies provide unequivocal proof that Gata2 is one of the key players in hematopoiesis, the precise function of this protein is still uncertain. Retroviral driven *Gata2* overexpression in adult bone marrow hematopoietic progenitor cells blocks their ability to contribute to hematopoiesis as well as their ability to expand (Persons, Allay et al. 1999). *Gata2/ER* enforced expression in FDCP cells (hematopoietic progenitor cell model) and in BA/F3 cells (IL-3 dependent hematopoietic progenitor cell line) triggers differentiation to monocyte and granulocyte cell fate, while it blocks self-renewal potential (Heyworth, Gale et al. 1999). These results reveal varying roles for Gata2 in multipotent hematopoietic progenitors and suggest that Gata2 levels make the difference. Particularly in embryonic development, the diploid dose of Gata2 is required to make the normal number of HSCs (Ling, Ottersbach et al. 2004). The ability to reach the threshold level of Gata2 expression is probabilistic in the haploid context, thus much fewer HSCs are generated. Recently, *GATA2* haploinsufficiency has been found to underlie some human immunodeficiency syndromes (Bresnick, Katsumura et al. 2012). The researchers showed in a mouse model that mutations in a specific *cis*- regulatory element of *Gata2* genomic locus disrupts vascular integrity *in vivo* and also results in a reduction of the number of HSCs and progenitors, while leaving unaffected the primitive hematopoiesis. This is thought to occur by preventing the positive autoregulation of Gata2 on its own locus due to the disruption of the Gata binding motif (Johnson, Hsu et al. 2012). Additionally, the effect of Gata2 in the growth suppression of hematopoietic progenitor cells may be explained by the formation of different complexes with other transcription factors that are expressed during later developmental stages (Ezoe, Matsumura et al. 2002). Despite progress in understanding the function of Gata2, its exact role in HSC generation still needs further investigation.

1.3.1.c Runx1

Runx1 (AML1, Cbfα) belongs to the family of the DNA binding proteins that also includes Runx2 and Runx3. These transcription factor genes have high homology with the Drosophila gene runt (reviewed in (Swiers, de Bruijn et al. 2010). Runx1 binds directly to the DNA sequence TGT/CGGT through its Runt homology domain (RHD), a 118 amino acid domain centrally located in the protein. Runx1 heterodi-

merizes with Cbfβ to form the core binding factor (CBF). This interaction increases its affinity for the DNA consensus sequence (Wang, Wang et al. 1993; Wang, Stacy et al. 1996).

The high frequency of chromosomal rearrangements in the Runx1 (AML) locus in acute myeloid leukemia (AML) cells was one of the reasons that made Runx1 an interesting subject for intense study in the blood system (reviewed by (Rabbitts 1994)). Runx1 binds to the regulatory elements (enhancer core motifs) and provides for tissue specific gene expression of molecules known to be important for hematopoiesis. Some of these genes encode GM-CSF, IL-3 and CSF1 receptor (Okuda, vanDeursen et al. 1996).

The importance of Runx1 in the development of the hematopoietic system was first revealed in *Runx1* deficient mice (Okuda, vanDeursen et al. 1996). Deletion of this transcription factor results in severe hematopoietic defects that lead to embryonicembryonic lethality at E12.5. In contrast the early normal production of the primitive hematopoietic system, the most obvious effect of *Runx1* deficiency is the complete absence of definitive hematopoiesis in the yolk sac (YS) and the fetal liver (FL) at midgestation. The AGM region was found to contain no HSCs (Cai, de Bruijn et al. 2000). Lack of contribution to the hematopoietic tissues by *Runx1-/-* cells was also observed in chimeric mice made with *Runx1-/-* ES cells. EB differentiation of *Runx1-/-* ES cells and analysis for the presence of hematopoietic progenitors confirmed the lack of definitive hematopoiesis (Okuda, vanDeursen et al. 1996; Wang, Stacy et al. 1996; Mukouyama, Chiba et al. 2000). Interestingly, deficiency of core binding factor β (Cbfβ), the partner of Runx1 in the core binding factor complex (CBF), results in the same phenotype as Runx1 deficiency (Sasaki, Yagi et al. 1996; Wang, Stacy et al. 1996; Wang, Stacy et al. 1996).

The temporal and spatial expression pattern of *Runx1* argues for its importance in the generation of the HSCs and progenitor cells. *Runx1* expression localizes to the endothelial cells of the yolk sac, the vitelline and umbilical arteries, the chorio/allantoic placenta and most importantly the ventral wall of the dorsal aorta (North, Gu et al. 1999; Ottersbach and Dzierzak 2005; Zeigler, Sugiyama et al. 2006; Rhodes, Gekas et al. 2008). It is expressed in the aorta just before the emergence of the HSCs and the progenitor cells, implicating *Runx1* in the hematopoietic fate process in hemogenic endothelial cells. Evidence to support this theory showed that *Runx1-/-* embryos lack hematopoietic clusters and *ex vivo* cultures of *Runx1-/-* endothelial cells possess no hematopoietic potential (North, Gu et al. 1999; Yokomizo, Ogawa et al. 2001). A more recent study (Chen, Yokomizo et al. 2009) confirmed the specific role of *Runx1* in the generation of hematopoietic cells from the hemogenic endothelial cells. Specific deletion of *Runx1* in VE-cadherin (VEC) expressing endothelial cells (VEC-Cre mediated *Runx1* excision) resulted in lack of HSC and progeni-

tor generation, phenocopying the hematopoietic defects found in the germline deleted embryos. Interestingly, the specific deletion of the gene in a hematopoietic cell context (Vav-cre mediated *Runx1* excision) showed that Runx1 is no longer required after HSCs are generated. Thus, Runx1 plays an important role in the endothelial to hematopoietic transition process by which definitive progenitors and HSCs are generated. However, although it is expressed in HSCs after they are generated, Runx1 function is no longer required.

Similarly to Gata2, haploinsufficiency of Runx1 has notable effects on HSC generation. A haploid dose of the transcriptional factor leads to only half the number of HSCs in the AGM region (Cai, de Bruijn et al. 2000; Robin, Ottersbach et al. 2006). It also results in the slightly earlier appearance of HSCs in the normal sites of hematopoiesis in the embryo (Cai, de Bruijn et al. 2000). Haploinsufficiency and trisomy of Runx1 (on chromosome 21) in man predisposes the individual and/or results results in hematopoietic dysfunction (leukemia and Downs Syndrome). Thus, proper regulated control of transcription factor levels in HSCs is an important but often less studied aspect. An interesting experiment to test whether Gata2 and Runx1 function in the same cells (precursors, hemogenic endothelial, hematopoietic cells) was performed. Since adult Runx1+/- and Gata2+/- mice are viable and present with a relatively normal hematopoietic blood profile, matings were established to generate double heterozygous mice. Interestingly, no Runx1+/-:Gata2+/- mice were born. However, Runx1+/-:Gata2+/- embryos could be harvested and fetal livers were found to contain fewer hematopoietic progenitors than single heterozygous mutants (Wilson, Foster et al. 2010) Chapter 4, this thesis). These findings, together with chromatin immunoprecipitation studies in hematopoietic cell lines, strongly suggest that Runx1 and Gata2 act in concert to control expression of hematopoietic genes involved in HSC and progenitor cell generation.

1.3.1.d Other Factors

Fli-1 is a member of the Ets transcription factor family. It has been found to be involved in human erythroid cell differentiation by affecting self renewal of erythroid progenitor cells. Reduced expression of Fli-1 drives differentiation of pluripotent hematopoietic progenitor cells and its overexpression results in inhibition of erythroid differentiation [(Howard, Yousefi et al. 1993; Athanasiou, Mavrothalassitis et al. 2000) and review by (Perry and Soreq 2002)]. Furthermore, Fli-1 deletion has been connected with dysmegakaryopoiesis and in combination with more evidence render it as a positive regulator of megakaryopoiesis [(Hart, Melet et al. 2000; Spyropoulos, Pharr et al. 2000; Kawada, Ito et al. 2001; Jackers, Szalai et al. 2004) and review by (Szalai, LaRue et al. 2006)].

PU.1 is another member of the Ets transcription factor family that participates in the control of hematopoietic differentiation. PU.1-/- mice show embryonic

lethality late in gestation with a fetal liver deficiency of B lymphocytes, granulocytes and macrophages (Scott, Simon et al. 1994; McKercher, Torbett et al. 1996; Lloberas, Soler et al. 1999; Back, Dierich et al. 2004). It has been suggested that expression levels of PU.1 in hematopoietic progenitor cells regulate their lineage commitment (Singh, DeKoter et al. 1999; DeKoter and Singh 2000). Interactions of PU.1 with different partners and formation of different protein complexes explain the ability to perform different functions [review in (Gupta, Gurudutta et al. 2009)].

Growth factor independence 1 (Gfi1) gene is expressed in several types of blood cells. First of all it is expressed in HSCs and in subsets of T-cells, granulocytes, monocytes and macrophages (Karsunky, Zeng et al. 2002; Hock, Hamblen et al. 2004; Yucel, Kosan et al. 2004). Gfi-1 germline deficiency in mice leads to neutrophil ablation and to an HSC population with only short term hematopoietic activity ((Zeng, Yucel et al. 2004) and review by (Wilson, Calero-Nieto et al. 2011)). Recent data from the Lacaud lab suggest that Gfi1 is involved in directing the morphologic changes that hemogenic endothelial cells undergo in the endothelial to hematopoietic transition (ISEH abstract).

Lmo2 is an example of a transcriptional cofactor that works in collaboration with several important hematopoietic transcription factors (like Gata1, Gata2, Scl/Tal1). It has been shown that Lmo2 is essential for erythroid development and also in adult hematopoiesis. Germline deficiency leads to embryonic lethality with failure of YS erythropoiesis (Warren, Colledge et al. 1994; Yamada, Warren et al. 1998). It does not bind to the DNA directly but after the formation of protein complexes with other transcription factors like SCL/Tal1, Gata1 and Gata2 [review by (Wilson, Calero-Nieto et al. 2011)]. The formation of different transcription factor complexes explains the distinct roles and functions of the protein in different cell populations (review in (Wilson, Calero-Nieto et al. 2011) and (Loose, Swiers et al. 2007)).

1.3.2 Extrinsic regulation of HSC generation

Apart from transcription factors, extrinsic regulation, through extracellular signaling, plays an important role in the control of HSC generation. The study of the hematopoietic niches and the microenvironment of the hematopoietic sites has provided insight into the signals and signaling pathways participating in the control of HSC generation. Some of the identified signaling pathways are BMP, Notch, TGFβ, Hedgehog and Wnt.

Hedgehog (Hh) signaling has a role in hematopoiesis and its role might differ depending upon the developmental stage (embryo or adult), the cell type (mature hematopoietic cells and HSCs) and the conditions (physiologic homeostasis and stress hematopoiesis) (Lim and Matsui 2010). Through studies in mice and ES cells it has been suggested that Hedgehog (Hh) has a role in vasculogenesis and

primitive hematopoiesis (Dyer, Farrington et al. 2001; Byrd, Becker et al. 2002). In contrast, Gering et al. have shown in zebrafish that Hh signaling is involved in definitive but not primitive hematopoiesis and they suggest that it plays distinct roles in embryonic development depending on the downstream mediators of Hh signaling. More zebrafish studies show that Hh acts in a polarizing manner in the dorsal part of the dorsal aorta in maintaining the arterial program and in an opposing manner with Bmp signaling which acts on the ventral side of the dorsal aorta (Wilkinson, Pouget et al. 2009). Mouse embryo studies in our laboratory have shown that Hh expression from the gut tissue (ventrally located from the dorsal aorta) induces the HSC activity of E10 AGM. The same result has also been achieved by exogenous addition of Hh (Peeters, Ottersbach et al. 2009).

Another signaling molecule with a role in hematopoiesis is bone morphogenetic protein 4 (BMP4). In the mouse embryo, BMP4 is expressed in the mesenchyme ventrally of the DA and it has been suggested that it has a direct effect on the HSCs since BMP4 receptors (BMP4R) are expressed in an enriched HSC population (Durand, Robin et al. 2007). Furthermore, *in vitro* studies with hematopoietic differentiation of ES cells have revealed that BMP4 regulates differentiation of ES cells into cells with hematopoietic potential (Johansson and Wiles 1995; Chadwick, Wang et al. 2003). Dose-dependent effects on the role of BMP4 at the proliferation and differentiation of primitive human hematopoietic cells have also been described. Low levels of BMP-4 seem to induce proliferation and differentiation of this cell population while higher levels maintain the cells in a quiescent state (Bhatia, Bonnet et al. 1999). Experiments performed on Xenopus and zebrafish indicate that BMP signaling induces Fli1 expression which acts upstream of other specific hematopoietic transcription factors such as Gata2, Scl, Lmo2 and Etsrp.

Notch is another signaling pathway controlling various developmental processes. Different members of the Notch family seem to have different effects in embryonic development. Notch2, Notch3 and Notch4 have been shown to affect vasculogenesis while Notch1 has been implicated with hematopoiesis [(Joutel, Corpechot et al. 1996; Krebs, Xue et al. 2000; McCright, Gao et al. 2001; Kumano, Chiba et al. 2003) and review by (Bigas, Robert-Moreno et al. 2010)]. It has been shown by studies in *Notch1-/-* embryos that the deficiency for Notch1 effects only definitive but not primitive hematopoiesis since the number of hematopoietic progenitors in the YS of *Notch1-/-* embryos is similar to the number in WT embryos [(Kumano, Chiba et al. 2003) and review by (Bigas, Robert-Moreno et al. 2010)]. It was further proposed by this study is that Notch1 plays a role in the endothelial to hematopoietic transition and acquisition of hematopoietic fate from hemogenic endothelial cells. These findings are strengthened by RBPjκ/CSL mutant (Robert-Moreno, Espinosa et al. 2005) and Jagged deficient embryos (Robert-Moreno, Guiu et al. 2008).

It has been proposed that these effects of Notch1 signaling are mediated by the regulation of Gata2 expression. This suggestion is supported by the fact that the presumptive hematopoietic cells located ventrally in the aorta expressing Gata2 also highly express Notch1. In zebrafish studies it has also been shown that *Runx1* is a target of Notch1 signaling with an effect on hematopoiesis (Burns, Traver et al. 2005; Bugeon, Taylor et al. 2011). Apart from effects in the generation of HSCs Notch signaling pathways also participate in other hematopoietic specific processes like HSCs self renewal, lymphoid and myeloid differentiation as well as in the development of the vascular network [review by (Bigas, Robert-Moreno et al. 2010)].

1.4 Scope of the thesis

Blood generation during mouse embryogenesis is a very complex and tightly regulated process. It happens in three waves, that can be temporally, spatially as well as functionally distinguished from each other. The third wave of hematopoiesis is characterized by the generation of HSCs from hemogenic endothelial cells. The HSCs are the most important cells of the adult hematopoietic system since they can provide an organism with all the different cells of the blood tissue during its lifespan. The regulation of HSC generation is the result of the combination of several extrinsic and intrinsic signals. Several of these signals, including cytokines, growth factors and transcription factors, have already been described. What is still lacking is an understanding of the exact molecular events and their interplay in the control of HSC emergence, expansion and differentiation.

The focus of this thesis is on the transcription factor Gata2 and its role in HSC generation and growth. To date it has not been possible to detect and examine Gata2 expressing cells *in vivo* and also sort them and test their potential *in vitro*. Only in this way is it possible to understand the role of this pivotal HSC transcription factor in the process of HSC generation. To address this need and study Gata2 expression in live imaging experiment I generated a new Gata2 reporter ES cell line and mouse model that facilitates the identification and isolation of Gata2 expressing cells. In these studies I explore the potential of the Gata2 expressing cell populations so as to understand the role of the transcription factor in embryonic hematopoiesis.

In the second chapter we describe the generation of an ES cell *Gata2-Venus* reporter line. The ES cell differentiation system is used as a model of embryonic hematopoietic development in order to test the potential of the Gata2 expressing cells. We show that the Gata2 expressing cell population is highly enriched in hematopoietic progenitors and that Gata2 expressing cells also have endothelial potential. In the third chapter, we describe the generation of a *Gata2-Venus* reporter mouse model. The major advantage of our mouse model is that it does not interfere with Gata2 expression levels or function. Using these reporter mice we show that

Venus expression recapitulates *Gata2* expression in the embryo, rendering our reporter mouse a suitable model for the prospective isolation of Gata2 expressing cells and testing their potential in i*n vivo* and *in vitro* functional assays. We show that Gata2 expressing cell fraction is highly enriched in hematopoietic progenitors. We also identify a Gata2 independent progenitor population. Finally, we show that the HSC activity can be found only in the Gata2 expressing cell fraction. In chapter 4, the genome-wide binding profile of ten important transcriptional regulators of hematopoiesis is presented. Bioinformatics analysis of these data sets provides comprehensive genome-wide combinatorial interactions between these key regulators and suggests the existence of a heptamere of transcription factors associated with HSCs and progenitor cells. Some of these protein-protein interactions were previously unrecognized. *In vivo* evidence that two of the central hematopoietic regulators (Runx1 and Gata2) work in the same cells is also provided. This study gives a good a good example of how bioinformatics analysis of genome-wide binding profiling of several transcription factors can provide new insights into understanding how transcription factors cooperate in the regulation of specific cell processes.

1.5 References

Athanasiou, M., G. Mavrothalassitis, et al. (2000). "FLI-1 is a suppressor of erythroid differentiation in human hematopoietic cells." Leukemia 14(3): 439-445.

Back, J., A. Dierich, et al. (2004). "PU.1 determines the self-renewal capacity of erythroid progenitor cells." Blood 103(10): 3615-3623.

Begley, C. G. and A. R. Green (1999). "The SCL gene: from case report to critical hematopoietic regulator." Blood 93(9): 2760-2770.

Bertrand, J. Y., N. C. Chi, et al. (2010). "Haematopoietic stem cells derive directly from aortic endothelium during development." Nature 464(7285): 108-111.

Bhatia, M., D. Bonnet, et al. (1999). "Bone morphogenetic proteins regulate the developmental program of human hematopoietic stem cells." J Exp Med 189(7): 1139-1148.

Bigas, A., A. Robert-Moreno, et al. (2010). "The Notch pathway in the developing hematopoietic system." Int J Dev Biol 54(6-7): 1175-1188.

Boisset, J. C. and C. Robin (2012). "On the origin of hematopoietic stem cells: progress and controversy." Stem Cell Res 8(1): 1-13.

Bresnick, E. H., K. R. Katsumura, et al. (2012). "Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies." Nucleic Acids Res 40(13): 5819-5831.

Bugeon, L., H. B. Taylor, et al. (2011). "The NOTCH pathway contributes to cell fate decision in myelopoiesis." Haematologica 96(12): 1753-1760.

Burns, C. E., D. Traver, et al. (2005). "Hematopoietic stem cell fate is established by the Notch-Runx pathway." Genes Dev 19(19): 2331-2342.

Byrd, N., S. Becker, et al. (2002). "Hedgehog is required for murine yolk sac angiogenesis." Development 129(2): 361-372.

Cai, Z., M. de Bruijn, et al. (2000). "Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo." Immunity 13(4): 423-431.

Chadwick, K., L. S. Wang, et al. (2003). "Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells." Blood 102(3): 906-915.

Chen, M. J., T. Yokomizo, et al. (2009). "Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter." Nature 457(7231): 887-891.

de Bruijn, M. F., X. Ma, et al. (2002). "Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta." Immunity 16(5): 673-683.

de Bruijn, M. F., N. A. Speck, et al. (2000). "Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo." EMBO J 19(11): 2465-2474.
DeKoter, R. P. and H. Singh (2000). "Regulation of B lymphocyte and macrophage development by graded expression of PU.1." Science 288(5470): 1439-1441.
Durand, C., C. Robin, et al. (2007). "Embryonic stromal clones reveal developmental regulators of definitive hematopoietic stem cells." Proc Natl Acad Sci U S A 104(52): 20838-20843.
Dyer, M. A., S. M. Farrington, et al. (2001). "Indian hedgehog activates hematopoiesis and vasculogenesis and can respecify prospective neurectodermal cell fate in the mouse embryo." Development 128(10): 1717-1730.
Dzierzak, E. and N. A. Speck (2008). "Of lineage and legacy: the development of mammalian hematopoietic stem cells." Nat Immunol 9(2): 129-136.
Elefanty, A. G., C. G. Begley, et al. (1999). "SCL expression in the mouse embryo detected with a targeted lacZ reporter gene demonstrates its localization to hematopoietic, vascular, and neural tissues." Blood 94(11): 3754-3763.
Ellis, S. L., J. Grassinger, et al. (2011). "The relationship between bone, hemopoietic stem cells, and vasculature." Blood 118(6): 1516-1524.
Endoh, M., M. Ogawa, et al. (2002). "SCL/tal-1-dependent process determines a competence to select the definitive hematopoietic lineage prior to endothelial differentiation." EMBO J 21(24): 6700-6708.
Ezoe, S., I. Matsumura, et al. (2002). "GATA-2/estrogen receptor chimera regulates cytokine-dependent growth of hematopoietic cells through accumulation of p21(WAF1) and p27(Kip1) proteins." Blood 100(10): 3512-3520.
Gupta, P., G. U. Gurudutta, et al. (2009). "PU.1 and partners: regulation of haematopoietic stem cell fate in normal and malignant haematopoiesis." J Cell Mol Med 13(11-12): 4349-4363.
Hall, M. A., D. J. Curtis, et al. (2003). "The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12." Proc Natl Acad Sci U S A 100(3): 992-997.
Hart, A., F. Melet, et al. (2000). "Fli-1 is required for murine vascular and megakaryocytic development and is hemizygously deleted in patients with thrombocytopenia." Immunity 13(2): 167-177.
Heyworth, C., K. Gale, et al. (1999). "A GATA-2/estrogen receptor chimera functions as a ligand-dependent negative regulator of self-renewal." Genes Dev 13(14): 1847-1860.
Hock, H., M. J. Hamblen, et al. (2004). "Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells." Nature 431(7011): 1002-1007.
Howard, J. C., S. Yousefi, et al. (1993). "Temporal order and functional analysis of mutations within the Fli-1 and p53 genes during the erythroleukemias induced by F-MuLV." Oncogene 8(10): 2721-2729.
Huber, T. L., V. Kouskoff, et al. (2004). "Haemangioblast commitment is initiated in the primitive streak of the mouse embryo." Nature 432(7017): 625-630.
Ismailoglu, I., G. Yeamans, et al. (2008). "Mesodermal patterning activity of SCL." Exp Hematol 36(12): 1593-1603.
Jackers, P., G. Szalai, et al. (2004). "Ets-dependent regulation of target gene expression during megakaryopoiesis." J Biol Chem 279(50): 52183-52190.
Johansson, B. M. and M. V. Wiles (1995). "Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development." Mol Cell Biol 15(1): 141-151.
Johnson, K. D., A. P. Hsu, et al. (2012). "Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity." J Clin Invest.
Joutel, A., C. Corpechot, et al. (1996). "Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia." Nature 383(6602): 707-710.
Kallianpur, A. R., J. E. Jordan, et al. (1994). "The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis." Blood 83(5): 1200-1208.
Karsunky, H., H. Zeng, et al. (2002). "Inflammatory reactions and severe neutropenia in mice lacking the transcriptional repressor Gfi1." Nat Genet 30(3): 295-300.
Kassouf, M. T., J. R. Hughes, et al. (2010). "Genome-wide identification of TAL1's functional

targets: insights into its mechanisms of action in primary erythroid cells." Genome Res 20(8): 1064-1083.
Kawada, H., T. Ito, et al. (2001). "Defective megakaryopoiesis and abnormal erythroid development in Fli-1 gene-targeted mice." Int J Hematol 73(4): 463-468.
Kiel, M. J. and S. J. Morrison (2006). "Maintaining hematopoietic stem cells in the vascular niche." Immunity 25(6): 862-864.
Kissa, K. and P. Herbomel (2010). "Blood stem cells emerge from aortic endothelium by a novel type of cell transition." Nature 464(7285): 112-115.
Ko, L. J. and J. D. Engel (1993). "DNA-binding specificities of the GATA transcription factor family." Mol Cell Biol 13(7): 4011-4022.
Krebs, L. T., Y. Xue, et al. (2000). "Notch signaling is essential for vascular morphogenesis in mice." Genes Dev 14(11): 1343-1352.
Kumano, K., S. Chiba, et al. (2003). "Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells." Immunity 18(5): 699-711.
Lacombe, J., S. Herblot, et al. (2010). "Scl regulates the quiescence and the long-term competence of hematopoietic stem cells." Blood 115(4): 792-803.
Lancrin, C., P. Sroczynska, et al. (2009). "The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage." Nature 457(7231): 892-895.
Lecuyer, E. and T. Hoang (2004). "SCL: from the origin of hematopoiesis to stem cells and leukemia." Exp Hematol 32(1): 11-24.
Li, L., R. Jothi, et al. (2011). "Nuclear adaptor Ldb1 regulates a transcriptional program essential for the maintenance of hematopoietic stem cells." Nat Immunol 12(2): 129-136.
Lim, Y. and W. Matsui (2010). "Hedgehog signaling in hematopoiesis." Crit Rev Eukaryot Gene Expr 20(2): 129-139.
Ling, K. W., K. Ottersbach, et al. (2004). "GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells." J Exp Med 200(7): 871-882.
Lloberas, J., C. Soler, et al. (1999). "The key role of PU.1/SPI-1 in B cells, myeloid cells and macrophages." Immunol Today 20(4): 184-189.
Loose, M., G. Swiers, et al. (2007). "Transcriptional networks regulating hematopoietic cell fate decisions." Curr Opin Hematol 14(4): 307-314.
Lux, C. T., M. Yoshimoto, et al. (2008). "All primitive and definitive hematopoietic progenitor cells emerging before E10 in the mouse embryo are products of the yolk sac." Blood 111(7): 3435-3438.
McCright, B., X. Gao, et al. (2001). "Defects in development of the kidney, heart and eye vasculature in mice homozygous for a hypomorphic Notch2 mutation." Development 128(4): 491-502.
McKercher, S. R., B. E. Torbett, et al. (1996). "Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities." EMBO J 15(20): 5647-5658.
Medvinsky, A. and E. Dzierzak (1996). "Definitive hematopoiesis is autonomously initiated by the AGM region." Cell 86(6): 897-906.
Medvinsky, A., S. Rybtsov, et al. (2011). "Embryonic origin of the adult hematopoietic system: advances and questions." Development 138(6): 1017-1031.
Mikkola, H. K., J. Klintman, et al. (2003). "Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene." Nature 421(6922): 547-551.
Minegishi, N., N. Suzuki, et al. (2003). "Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos." Blood 102(3): 896-905.
Mukouyama, Y., N. Chiba, et al. (2000). "The AML1 transcription factor functions to develop and maintain hematogenic precursor cells in the embryonic aorta-gonad-mesonephros region." Dev Biol 220(1): 27-36.
Murray, R., L. Kilham, et al. (1946). "Development of streptomycin resistances of gram-negative bacilli in vitro and during treatment." Proc Soc Exp Biol Med 63(2): 470-474.
Murre, C. (2005). "Helix-loop-helix proteins and lymphocyte development." Nat Immunol 6(11): 1079-1086.
Nardelli, J., D. Thiesson, et al. (1999). "Expression and genetic interaction of transcription factors GATA-2 and GATA-3 during development of the mouse central nervous system." Dev Biol 210(2): 305-321.
Ng, Y. K., K. M. George, et al. (1994). "GATA factor activity is required for the trophoblast-specific transcriptional regulation of the mouse

placental lactogen I gene." Development 120(11): 3257-3266.
North, T., T. L. Gu, et al. (1999). "Cbfa2 is required for the formation of intra-aortic hematopoietic clusters." Development 126(11): 2563-2575.
Okuda, T., J. vanDeursen, et al. (1996). "AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis." Cell 84(2): 321-330.
Orlic, D., S. Anderson, et al. (1995). "Pluripotent hematopoietic stem cells contain high levels of mRNA for c-kit, GATA-2, p45 NF-E2, and c-myb and low levels or no mRNA for c-fms and the receptors for granulocyte colony-stimulating factor and interleukins 5 and 7." Proc Natl Acad Sci U S A 92(10): 4601-4605.
Ottersbach, K. and E. Dzierzak (2005). "The murine placenta contains hematopoietic stem cells within the vascular labyrinth region." Dev Cell 8(3): 377-387.
Palis, J., S. Robertson, et al. (1999). "Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse." Development 126(22): 5073-5084.
Park, C., Y. D. Ma, et al. (2005). "Evidence for the hemangioblast." Exp Hematol 33(9): 965-970.
Peeters, M., K. Ottersbach, et al. (2009). "Ventral embryonic tissues and Hedgehog proteins induce early AGM hematopoietic stem cell development." Development 136(15): 2613-2621.
Perry, C. and H. Soreq (2002). "Transcriptional regulation of erythropoiesis. Fine tuning of combinatorial multi-domain elements." Eur J Biochem 269(15): 3607-3618.
Persons, D. A., J. A. Allay, et al. (1999). "Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis." Blood 93(2): 488-499.
Peterkin, T., A. Gibson, et al. (2009). "Common genetic control of haemangioblast and cardiac development in zebrafish." Development 136(9): 1465-1474.
Porcher, C., W. Swat, et al. (1996). "The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages." Cell 86(1): 47-57.
Rabbitts, T. H. (1994). "Chromosomal translocations in human cancer." Nature 372(6502): 143-149.
Rhodes, K. E., C. Gekas, et al. (2008). "The emergence of hematopoietic stem cells is initiated in the placental vasculature in the absence of circulation." Cell Stem Cell 2(3): 252-263.
Robb, L., N. J. Elwood, et al. (1996). "The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse." EMBO J 15(16): 4123-4129.
Robb, L., I. Lyons, et al. (1995). "Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene." Proc Natl Acad Sci U S A 92(15): 7075-7079.
Robert-Moreno, A., L. Espinosa, et al. (2005). "RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells." Development 132(5): 1117-1126.
Robert-Moreno, A., J. Guiu, et al. (2008). "Impaired embryonic haematopoiesis yet normal arterial development in the absence of the Notch ligand Jagged1." EMBO J 27(13): 1886-1895.
Robin, C., K. Ottersbach, et al. (2006). "An unexpected role for IL-3 in the embryonic development of hematopoietic stem cells." Dev Cell 11(2): 171-180.
Rybtsov, S., M. Sobiesiak, et al. (2011). "Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region." J Exp Med 208(6): 1305-1315.
Sasaki, K., H. Yagi, et al. (1996). "Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor beta." Proc Natl Acad Sci U S A 93(22): 12359-12363.
Schoenebeck, J. J., B. R. Keegan, et al. (2007). "Vessel and blood specification override cardiac potential in anterior mesoderm." Dev Cell 13(2): 254-267.
Scott, E. W., M. C. Simon, et al. (1994). "Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages." Science 265(5178): 1573-1577.
Shalaby, F., J. Rossant, et al. (1995). "Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice." Nature 376(6535): 62-66.
Shivdasani, R. A., E. L. Mayer, et al. (1995). "Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL." Nature 373(6513): 432-434.

Singh, H., R. P. DeKoter, et al. (1999). "PU.1, a shared transcriptional regulator of lymphoid and myeloid cell fates." Cold Spring Harb Symp Quant Biol 64: 13-20.
Souroullas, G. P., J. M. Salmon, et al. (2009). "Adult hematopoietic stem and progenitor cells require either Lyl1 or Scl for survival." Cell Stem Cell 4(2): 180-186.
Spyropoulos, D. D., P. N. Pharr, et al. (2000). "Hemorrhage, impaired hematopoiesis, and lethality in mouse embryos carrying a targeted disruption of the Fli1 transcription factor." Mol Cell Biol 20(15): 5643-5652.
Sugiyama, T., H. Kohara, et al. (2006). "Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches." Immunity 25(6): 977-988.
Swiers, G., M. de Bruijn, et al. (2010). "Hematopoietic stem cell emergence in the conceptus and the role of Runx1." Int J Dev Biol 54(6-7): 1151-1163.
Szalai, G., A. C. LaRue, et al. (2006). "Molecular mechanisms of megakaryopoiesis." Cell Mol Life Sci 63(21): 2460-2476.
Tavian, M., M. F. Hallais, et al. (1999). "Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo." Development 126(4): 793-803.
Tavian, M. and B. Peault (2005). "Embryonic development of the human hematopoietic system." Int J Dev Biol 49(2-3): 243-250.
Tsai, F. Y., G. Keller, et al. (1994). "An early haematopoietic defect in mice lacking the transcription factor GATA-2." Nature 371(6494): 221-226.
Tsai, F. Y. and S. H. Orkin (1997). "Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation." Blood 89(10): 3636-3643.
Ueno, H. and I. L. Weissman (2006). "Clonal analysis of mouse development reveals a polyclonal origin for yolk sac blood islands." Dev Cell 11(4): 519-533.
Van Handel, B., A. Montel-Hagen, et al. (2012). "Scl represses cardiomyogenesis in prospective hemogenic endothelium and endocardium." Cell 150(3): 590-605.
Wang, Q., T. Stacy, et al. (1996). "Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis." Proc Natl Acad Sci U S A 93(8): 3444-3449.
Wang, Q., T. Stacy, et al. (1996). "The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo." Cell 87(4): 697-708.
Wang, S., Q. Wang, et al. (1993). "Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor." Mol Cell Biol 13(6): 3324-3339.
Warren, A. J., W. H. Colledge, et al. (1994). "The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development." Cell 78(1): 45-57.
Wilkinson, R. N., C. Pouget, et al. (2009). "Hedgehog and Bmp polarize hematopoietic stem cell emergence in the zebrafish dorsal aorta." Dev Cell 16(6): 909-916.
Wilson, N. K., F. J. Calero-Nieto, et al. (2011). "Transcriptional regulation of haematopoietic transcription factors." Stem Cell Res Ther 2(1): 6.
Wilson, N. K., S. D. Foster, et al. (2010). "Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators." Cell Stem Cell 7(4): 532-544.
Wontakal, S. N., X. Y. Guo, et al. (2012). "A core erythroid transcriptional network is repressed by a master regulator of myelo-lymphoid differentiation." Proc Natl Acad Sci U S A 109(10): 3832-3837.
Yamada, Y., A. J. Warren, et al. (1998). "The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis." Proc Natl Acad Sci U S A 95(7): 3890-3895.
Yoder, M. C., K. Hiatt, et al. (1997). "Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac." Immunity 7(3): 335-344.
Yokomizo, T., M. Ogawa, et al. (2001). "Requirement of Runx1/AML1/PEBP2alphaB for the generation of haematopoietic cells from endothelial cells." Genes Cells 6(1): 13-23.
Yokomizo, T., T. Yamada-Inagawa, et al. (2012). "Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos." Nat Protoc 7(3): 421-431.
Yucel, R., C. Kosan, et al. (2004). "Gfi1:green fluorescent protein knock-in mutant reveals differential expression and autoregulation of the growth factor independence 1 (Gfi1) gene

during lymphocyte development." J Biol Chem 279(39): 40906-40917.
Zeigler, B. M., D. Sugiyama, et al. (2006). "The allantois and chorion, when isolated before circulation or chorio-allantoic fusion, have hematopoietic potential." Development 133(21): 4183-4192.
Zeng, H., R. Yucel, et al. (2004). "Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells." EMBO J 23(20): 4116-4125.
Zhou, Y., K. C. Lim, et al. (1998). "Rescue of the embryonic lethal hematopoietic defect reveals a critical role for GATA-2 in urogenital development." EMBO J 17(22): 6689-6700.
Zovein, A. C., J. J. Hofmann, et al. (2008). "Fate tracing reveals the endothelial origin of hematopoietic stem cells." Cell Stem Cell 3(6): 625-636.

Chapter 2

Gata2 expressing cells in embryoid bodies (EB) show a hemangioblast phenotype resembling that of early embryonic hematopoiesis

Polynikis Kaimakis[1], Mihaela Crisan[1], Siska Driegen[2], Martine Jeagle[2], Dies Meijer[2] and Elaine Dzierzak[1]

1. ESI, Erasmus Stem cell Institute, Department of Cell Biology, Erasmus MC, Rotterdam, the Netherlands
2. Department of Cell Biology and Genetics, Erasmus MC, Rotterdam, the Netherlands

(Work in progress)

Gata2 expressing cells in embryoid bodies (EB) show a hemangioblast phenotype resembling that of early embryonic hematopoiesis

Polynikis Kaimakis, Mihaela Crisan, Siska Driegen, Martine Jeagle, Dies Meijer and Elaine Dzierzak

Abstract

The process of embryonic hematopoiesis is tightly controlled by a combination of extracellular signals as well as intracellular regulators such as transcription factors. An understanding of this developmental process is of high importance for the field of hematopoiesis. Detailed descriptions of the molecular events regulating the generation of hematopoietic cells could provide us with new options for the successful generation and manipulation of hematopoietic stem cells (HSCs) *in vitro*. It has already been demonstrated that the hematopoietic cells are derived from a type of progenitor with both endothelial and hematopoietic potential (hemangioblast) and from specialized endothelial cells with hematopoietic potential (hemogenic endothelium). The transition between these developmental stages is under intense investigation.

Several transcription factors have been described having specific and important effects in the generation of HSCs. Gata2 is considered to be one of the master regulators of hematopoiesis, since its deletion leads to embryonic lethality due to severe hematopoietic defects. The specific role of Gata2 in this process has not been defined yet.

The embryonic stem cell (ES) differentiation system, in which embryoid bodies (EB) are formed, is a suitable *in vitro* model that recapitulates the early embryonic development and has already been used to study of the role of some transcription factors in hematopoietic development. In this study we report the generation and characterization of a novel ES cell reporter line (*Gata2-Venus*). We find that Venus is expressed coordinately with Gata2. For the first time Gata2 expressing cells can be directly sorted and tested for lineage potential in functional assays. We show that the Gata2 expressing cell population is highly enriched in hematopoietic potential and also possesses endothelial potential.

Introduction

The *in vitro* differentiation of embryonic stem (ES) cells into embryoid bodies (EB) has been widely used as a model to study embryonic hematopoiesis. Specific culturing conditions have been successfully described for the differentiation of EB derived cells into the different and distinct types of hematopoietic progenitors. The developmental kinetics of the different progenitor types parallels that of the

early embryo, particularly yolk sac hematopoiesis (Keller, Webb et al. 2002; Choi, Chung et al. 2005). Blast colony forming cells (BL-CFCs), a cell type with hemangioblast characteristics, appear first (Kennedy, Firpo et al. 1997; Choi, Kennedy et al. 1998), followed by the development of primitive erythroid and later on definitive erythroid and myeloid progenitors (Wiles and Keller 1991; Nakano, Kodama et al. 1996; Palis, Robertson et al. 1999). Lymphoid differentiation from EB derived cells has also been successfully described, with the differentiation to B, NK and T-lineage cells (Nakano, Kodama et al. 1994; Tabatabaei-Zavareh, Vlasova et al. 2007). Whereas most distinct lineages of differentiated blood cells can be produced in this *in vitro* model, it has not yet been possible to obtain adult-type transplantable hematopoietic stem cells (HSCs) without genetic manipulation.

The developmental similarities between the early stage conceptus and the EB differentiation culture system offer some significant practical advantages. This system provides access to various distinct cell types representing different stages of hematopoietic development. These cell types can be easily identified, characterized and isolated in sufficient numbers for genetic and epigenetic analysis, something not feasible in an in vivo embryo system because of cell number limitations. Moreover, ES cells can be easily manipulated to study the effects of specific genes in the process of embryonic hematopoiesis. ES cell gene targeting has been used to produce deletions and mutations in genes regulating hematopoietic progenitor/stem cells, such as *SCL*, *Gata2* and *Runx1* (Tsai, Keller et al. 1994; Porcher, Swat et al. 1996), as well as lineage-specific regulators function like *Gata1* (Weiss, Keller et al. 1994). Studies with *SCL*$^{-/-}$ ES cells have shown that this transcription factor is required for the specification of hemangioblasts during EB differentiation.

In this study we investigated the role of Gata2 in hematopoietic development. The importance of the Gata2 transcription factor in hematopoietic development has been highlighted in several *in vitro* and *in vivo* models. *Gata2*$^{-/-}$ mice die during development at embryonic day (E)10 – E10.5 of gestation. These mice show severe anaemia and complete absence of definitive hematopoietic progenitors. Furthermore, chimeric studies with *Gata2*$^{-/-}$ ES cells showed no contribution of these cells to any hematopoietic tissue (Tsai, Keller et al. 1994). Apart from these functional experiments, *Gata2* localization studies have shown expression of *Gata2* in hematopoietic tissues at most developmental timepoints. A characteristic example is the expression of Gata2 in the endothelial cells forming the wall of the dorsal aorta, the first site of autonomous generation of the HSCs in the mouse embryo (Kornhauser, Leonard et al. 1994; Ng, George et al. 1994; Minegishi, Ohta et al. 1998; Nardelli, Thiesson et al. 1999; Robert-Moreno, Espinosa et al. 2005). Another characteristic of pivotal transcription factors in hematopoietic development is the connection between expression levels of the protein and its function. Mice hetero-

zygous for Gata2 have reduced Gata2 expression and show a greatly reduced number of hematopoietic progenitors (Ling, Ottersbach et al. 2004), demonstrating that precise levels of this transcription factor in generation or expansion of the HSC pool. What is lacking in these models is the ability to directly isolate Gata2 expressing cells.

To prospectively study the role of Gata2 during hematopoietic development, we generated a *Gata2-Venus* fluorochrome ES cell reporter cell line. We report here the generation of a *3'UTR Gata2* targeting construct that allows expression of an *IRES Venus* gene concomitantly with normal levels of *Gata2*. This approach avoids Gata2 haploinsufficiency. Our results show proper, coordinately regulated expression of Gata2 and Venus. We show that almost all EB-derived hematopoietic progenitors are Gata2 expressing, whereas endothelial cells are both Gata2-expressing and non-expressing. These ES cells represent a unique resource for further genetic and epigenetic studies of hematopoietic development.

Results

Gata2-Venus ES cell generation

To better understand the role of the Gata2 transcription factor in the embryonic generation of definitive progenitors and stem cells, we established a mouse ES cell model that would allow the tracking of Gata2 expression and the prospective isolation of Gata2 expressing cells during hematopoietic cell development. A knock in approach was followed, whereby the coding sequence for the Venus fluorochrome preceded by an *IRES* sequence was recombined into the *3' UTR* of the *Gata2* genomic locus (fig 1A). A *LoxP* flanked *puromycin* gene was included in the construct for positive selection. Using this strategy, the *Gata2* coding sequence is preserved, avoiding *Gata2* haploinsufficiency and negative effects on hematopoietic development. Moreover, the expression of the fluorochrome will not interfere with Gata2 protein function, as may be the case with the use of a Gata2-Venus fusion protein.

Gata2-Venus targeting was performed in ES cells, followed by puromycin selection. 460 ES clones were expanded. DNA was isolated from each clone and PCR for the targeting construct was performed. Six clones were found to contain the targeting construct and these were further screened by Southern blotting analysis (fig 1B), to ensure the correct integration into the *3'UTR* of the *Gata2* genomic locus. Two positive clones showed the proper integration of the construct and these were karyotyped. One of them had the normal number of chromosomes and was used in the rest of the studies presented here.

To minimize the potential effects of the inserted construct on proper regu-

lation of Gata2 expression, we excised the *puromycin* selection cassette. A diffusible Cre protein was added directly into ES cell cultures (Peitz, Pfannkuche et al. 2002; Peitz, Jager et al. 2007) for 18 hours. Thereafter, ES cells were plated as single cells and allowed to expand in culture. Individual clones were analyzed for recombination excision events by DNA PCR designed to detect the presence of the *LoxP* site (fig 1C). Two clones (G2V5 and G2V6) that had excised the puromycin selection cassette were found. G2V5 was used for the experiments described in further studies.

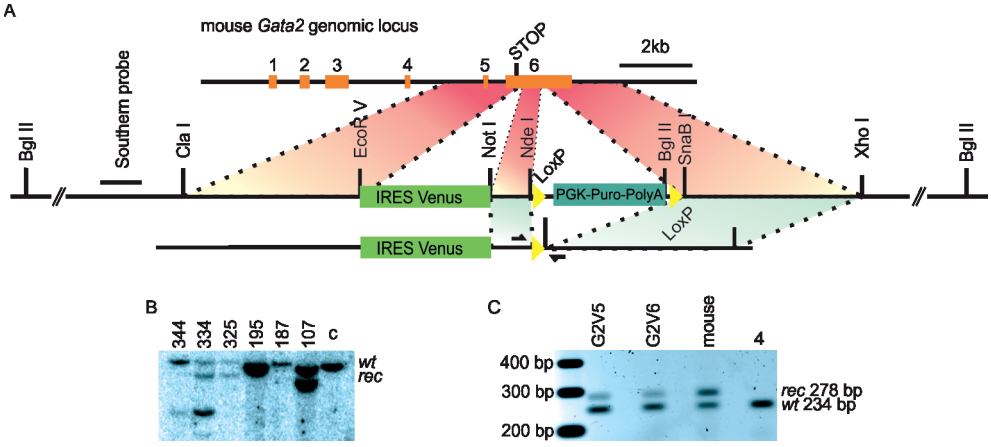

Fig 1 Generation of *Gata2-Venus* reporter ES cell line

The *Gata2-Venus* targeting construct is designed so that a full length Gata2 protein and Venus fluorochrome are produced by a bicistronic mRNA transcript. (A) Mouse *Gata2* genomic locus (top). Exons 1-6 depicted as orange boxes and STOP codon position is indicated. Predicted structure and restriction map of *Gata2-Venus* targeting construct (middle). Homology borders between wild type *Gata2* genomic locus and *Gata2* targeting construct are highlighted. Triangles represent *LoxP* sites. The positions of the restriction sites used in the cloning strategy are indicated. Corresponding boxes for the *IRES-Venus* and the *PGK-Puro* selection cassette are noted in the scheme. Predicted structure of the *Gata2-Venus* targeted allele after the excision of the selection cassette by Cre recombination (bottom). The set of *LoxP* flanking primers used for the detection of recombined clones are shown as small arrows (actual primer size not proportional with the length of the arrow). (B) Southern blot analysis of ES cell clones. Genomic DNA was isolated and digested with BglII. Detection probe was a genomic Gata2 DNA sequence (600 bp) localized upstream of the left recombination arm (not depicted). Expected fragment sizes (wt *Gata2* 12 Kbp; *Gata2-Venus* 11 Kbp). Lanes 1, 2, 3, 4, 5, 6, c represent ES cell clones 344, 334, 325, 195, 187, 107 and wt DNA control respectively. Clones 325 and 107 show the expected pattern of correct integration. (C) PCR reaction for the detection of the recombined clones that have excised the *Puromycin* selection cassette. In lane 1 and 2 the PCR results of clones G2V5 and G2V6 respectively are depicted. The upper band (278 bp) detects the recombined allele and the lower band (234 bp) the *wt* one. Lane 3 shows the result of a control of a recombined mouse DNA (described in next chapter) and lane 4 is *wt* control DNA.

Venus expression parallels that of Gata2 during EB hematopoietic differentiation

To check whether G2V5 ES cells express the Venus fluorochrome during hematopoietic development, EB differentiation cultures were established. At day (d)5 of differentiation, G2V5 and wild type (WT) control EBs were observed under the fluorescent microscope (fig 2A). G2V5 EBs showed Venus expressing cells that specifically localize in the core of the EB. To quantitate Venus expressing cells in such cultures, d6 EBs were harvested, made into a single cell suspension and tested by FACs analysis for Venus fluorochrome expression (fig 2B). Venus expression was found in approximated 3 - 4% of the d6 G2V5 EB cell population, confirming the Venus expression observed in whole EBs.

To test whether the Venus expressing cells also express *Gata2*, we took a qRT-PCR approach. Venus positive and Venus negative cell fractions were flowcytometrically sorted from d6 EBs and examined for *Venus* and *Gata2* gene expression. *Venus* and *Gata2* mRNAs were detected only in the Venus positive sorted cell fraction (fig 2C) demonstrating, as expected from the *IRES* containing construct, that *Venus* gene expression is coordinated with *Gata2* gene expression. Venus and Gata2 protein expression in the sorted cell fractions was also examined. In accordance with the mRNA data, Western blot analysis showed the presence of Gata2 and Venus protein only in the Venus-positive sorted EB cell fraction (fig 2D). Thus, the *3'UTR Gata2 IRES Venus* knock-in approach leads to faithful and exclusive expression of the Venus fluorochrome in Gata2 expressing cells.

To further examine the expression of *Gata2* and *Venus*, we performed a time course analysis during ES cell differentiation. EBs were harvested at different timepoints (d0, d2, d3, etc.) and the relative levels of *Gata2* and *Venus* mRNA were tested by qRT-PCR. As shown in fig 2E there is a coordinated up- and down-regulation of expression of *Venus* and *Gata2* mRNA, indicating that both genes are subject to correct developmental regulation.

These results demonstrate that Venus fluorochrome expression is an excellent surrogate marker for Gata2 expression in the ES cell differentiation culture system, with G2V ES cells as a suitable in vitro model for studying the role of Gata2 in hematopoietic development.

Venus expressing EB cells possess endothelial and hematopoietic potential

In the mouse embryo, *Gata2* is known to be expressed both in endothelial and hematopoietic cell lineages. We analyzed G2V5 cells for expression of Gata2 specifically in endothelial and hematopoietic cells derived from d6 EB differentiation cultures. Flow cytometric analysis for endothelial marker Flk1, endothelial and hematopoietic marker CD31, and hematopoietic markers c-kit and CD45 was performed (fig 3A, fig 3B). A small percentage of EB-derived endothelial and hemato-

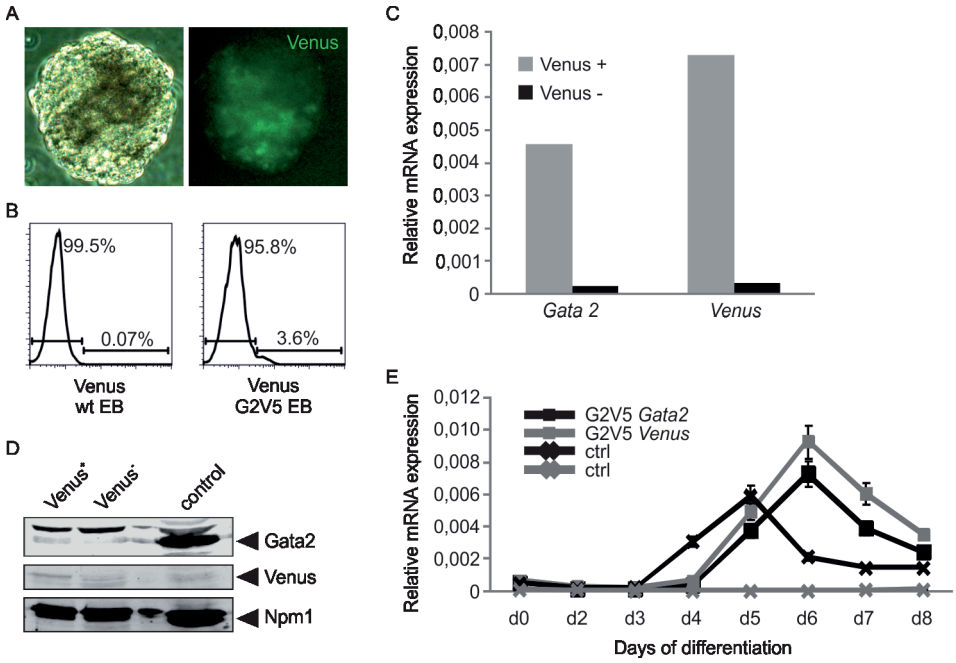

Fig 2 Venus expression during EB differentiation

(A) Venus fluorochrome expression in a single EB after 5 days (d5) of differentiation. Brightfield and fluorescence images are depicted in left and right panel respectively. (B) Flow cytometric analyses of Venus expression in single cell suspensions of d6 EBs. Wild type (wt) control EB cells (left panel) and G2V5 EB cells (right panel). Percentages (%) of positive and negative cell fractions are shown. (C) qRT-PCR analysis of *Gata2* and *Venus* expression by Venus positive and Venus negative sorted cell populations from G2V5 d6 EBs. *Gata2* and *Venus* mRNA was detected only in the Venus positive sorted cell fraction. (D) Western blot analysis for Gata2 and Venus protein expression in d6 EBs. Anti-Gata2 antibody was used for Gata2 detection and anti-GFP antibody for Venus detection in Venus positive (+) and Venus negative (-) sorted cell fractions. BaF3, a Gata2 overexpressing cell line was used as a positive control for Gata2 expression. Npm1 (Nucleophosmin 1) was used as a protein loading control. Only the Venus positive sorted cell extracts show Gata2 and Venus protein expression. (E) qRT-PCR analysis of *Gata2* and *Venus* mRNA expression between d0 and d8 of EB differentiation. Relative mRNA quantity was normalized to *beta actin* mRNA. On x axis the day of differentiation is shown. Wild type ES cells were used as negative control (ctrl) for *Venus* expression and as a positive control for *Gata2* expression. *Venus* and *Gata2* expression is coordinately regulated during G2V5 differentiation.

poietic cells were Gata2 Venus expressing. Of the 2.5% of EB cells that were Flk1 positive, 20% of these also expressed Gata2 Venus. CD31 positive cells represented 14% of the total EB cell population and 10% of CD31 positive cells expressed Gata2 Venus. No CD45 positive EB cells were found. However, of the 35% of EB cells expressing c-kit, only 0.3% expressed Gata2 Venus. The fact that only a small proportion of the EB-derived hematopoietic and endothelial cells are Gata2 expressing suggests that Gata2 expression delineates the hematopoietic progenitors and possibly hemogenic endothelial cells.

We next examined d6 G2V5 EBs for endothelial and hematopoietic function. To test the endothelial potential of Gata2 expressing cells, Venus positive and negative cells were sorted and plated in Matrigel under endothelial growth conditions. Cells of both fractions showed endothelial differentiation potential. Tubules with matrix and junction formation typical of endothelial cells were found (fig 3C). Thus, some EB-derived Gata2 expressing cells are functional endothelial cells.

The hematopoietic function of d6 G2V EB cells was tested in methylcellulose progenitor assays. Highly enriched colony-forming unit-culture (CFU-C) activity was found in the Gata2 expressing sorted cell population. The frequency of progenitors in the Gata2 expressing cell fraction was 43-fold increased as compared to the Gata2 negative fraction. The vast majority of the colonies were CFU-Es, which are primitive erythroid lineage restricted colonies. This data matches what has been previously reported concerning the types of progenitors present at this stage of EB differentiation (Keller, Webb et al. 2002; Mikkola, Fujiwara et al. 2003; Choi, Chung et al. 2005), the types of progenitors lacking in *Gata2*$^{-/-}$ differentiated ES cells and also parallels the primitive erythroid colony formation observed with yolk sac-derived cells. Other colony types indicative of granulocyte and macrophage progenitors were also found, thus suggesting a role for Gata2 in cells with primitive and definitive hematopoietic potential (fig 3D).

Discussion

We generated a novel *Gata2 Venus* reporter ES cell line, that expresses Venus in a temporally and lineage correct manner. The targeted insertion of the *IRES Venus* reporter into the *Gata2 3'UTR* rather than the coding region prevents a Gata2 haploinsufficiency state. This is important since Gata2 haploinsufficiency results in a decrease in the number of the HSCs in the AGM as well as in qualitative defects in the adult HSCs population (Ling, Ottersbach et al. 2004). A similar targeting approach was used by others for generation of a *PU.1* fluorescent reporter. The direct detection of PU.1 expression levels was useful in identifying functionally distinct common myeloid progenitors. Moreover, it was found that PU.1 downregulation is

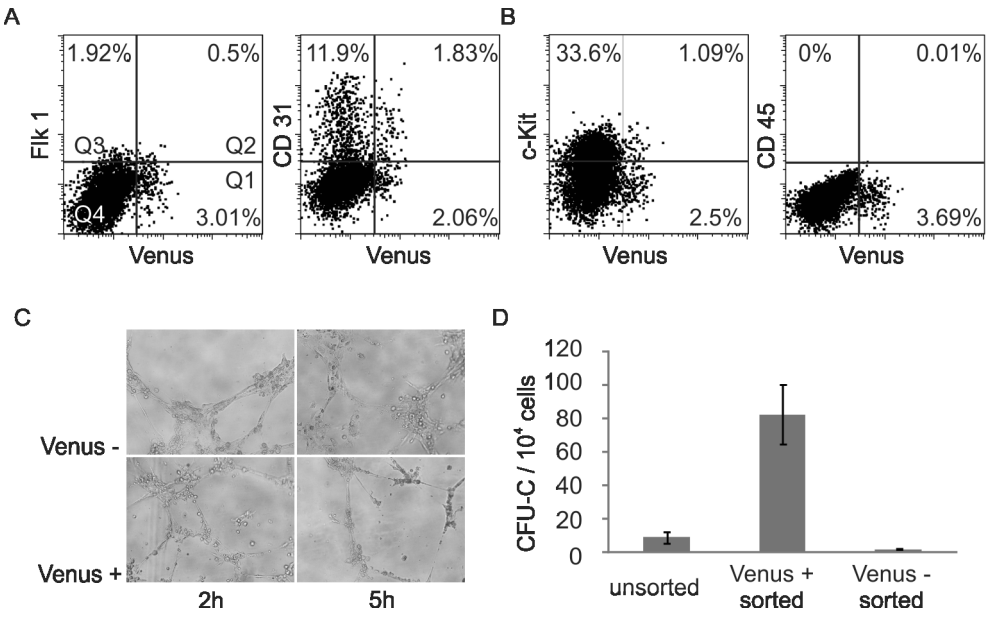

Fig 3 Endothelial and hematopoietic potential of Gata2 expressing cells

Flow cytometric analysis of d6 G2V5 EB cells. (A) Cells stained for endothelial markers Flk1 and CD31. Flk1 is also expressed on early mesodermal and hemangioblast cells. CD31 is also expressed on hematopoietic cells. (B) Cells stained for hematopoietic markers c-Kit and CD45. % of cells in each of the quadrants are shown. 20-50x103 number of events were analyzed. (C) Endothelial differentiation of EB derived cells. Venus + and Venus – cells were plated on matrigel coated dishes. Images show tubule and junction formation which was observed from both Venus + and Venus – cells at 2 hours (2h) and 5h after plating. (D) Methylcellulose colony forming unit-culture (CFU-C) progenitor assay. The frequency of progenitors (CFU-C/104 cells) is presented in Venus +, Venus – and unsorted populations. The Venus + population shows a 43-fold enrichment of progenitors as compared to the Venus– population.

associated with restriction to megakaryocyte and erythroid lineage differentiation (Nutt, Metcalf et al. 2005). Whereas *3'UTR* sequences represents accessible sites for gene targeting, specific sequences within the *3'UTR* of many genes are target sites for *miRNAs*. Disruption of such sequences may result in aberrant gene regulation. Indeed, the *Gata2 3'UTR* contains a *miRNA* target sequence. The *miR-451* has been shown in zebrafish to regulate *Gata2* expression during erythroid differentiation (Pase, Layton et al. 2009) and the Gata2 gene has 2 target sequences for this miR. Data mining of the mouse genome predicts 3 target sites in the *3'UTR* of the mouse *Gata2* gene. However, our construct was designed so that the insertion of the *IRES Venus* reporter does not interfere with the miRNA target sequences.

Our *Gata2-Venus* ES cell reporter line allows the prospective isolation of

cells expressing the *Gata2* transcription factor. Previously, Gata2 expression could be assessed only retrospectively by transcription analysis of hematopoietic progenitor and stem cell populations obtained by flow cytometrically sorting for specific cell surface markers that characterize these cells types. *Gata2* was found to be expressed by hematopoietic progenitors and stem cells but not by mature blood cells. Here we show that directly sorted Gata2 expressing d6 EB cell populations are 43-fold enriched in hematopoietic progenitors. Our results support a role for this transcription factor in EB hematopoietic progenitor potential. Additionally, we found that Gata2 expressing cells (and also Gata2 non-expressing cells) have endothelial potential. The fact that both endothelial cells and hematopoietic progenitors express Gata2 suggests that Gata2 may function in the precursors to these two lineages, i.e. hemangioblasts and/or hemogenic endothelial cells. We are attempting to establish whether cells with these characteristics can be enriched from G2V5 EBs. Single cell studies and live imaging of single Gata2 expressing EB cells will be performed to define more accurately their lineage potentials and differentiation properties. Prospectively isolated Gata2 expressing cells will allow further characterization, such as cell cycle status, signaling pathway activation and gene expression profiling and provide insight into the possible roles of Gata2 in hematopoietic progenitor and stem cell development.

A more detailed characterization of the cell population expressing Gata2 is currently underway. In combination with Gata2, panels of cell surface markers are being used to identify cells as they begin to take on hematopoietic lineage fate. Expression of a marker such as CD41, which is one of the earliest hematopoietic markers, would indicate commitment to the hematopoietic lineage (Mikkola, Fujiwara et al. 2003; Robin, Ottersbach et al. 2011). Expression of intermediate levels of CD41 would be indicative of HSC development, as has been found in the AGM. Endothelial cells expressing Gata2 could then be followed in real-time to observe the acquisition of hematopoietic stem cell potential. Another marker that is expressed on the hematopoietic stem cell population is Ly-6A (Sca-1). Ly-6A GFP ES cells have been produced and are an excellent resource to discover appropriate culture conditions allowing the in vitro generation of HSCs. Onset of GFP expression should mark the endothelial to HSC transition, as has been observed in real-time In vivo. Double fluorescent marker ES cells would facilitate a more efficient enrichment of EB cells with HSC potential. The use of endothelial markers such as endoglin or Tie2 could provide further information about possible presence of distinct endothelial cell types, including hemogenic endothelium, and different stages of the endothelial to hematopoietic transition in ES differentiation cultures, and provide candidates for further HSC enrichment procedures.

In summary, in this study we describe the generation and characterization

of a novel *Gata2-Venus* ES cell reporter line. This new system gives us the opportunity to directly isolate Gata2 expressing cells during EB differentiation, study their lineage potential during distinct stages of differentiation and to illuminate the function of Gata2 during hematopoietic development *in vitro* and *in vivo*.

Materials and methods

Cloning strategy

The *Gata2* homology arms were amplified from either IB10 derived genomic DNA or from BACs clone containing the last exon and the *3' UTR* of the Gata2 genomic locus. The amplification was done using Phusion™ taq polymerase (Finnzymes F-530L). The left recombination arm was amplified with the use of 5' AATCGATGCCGAGGGAGTTCAGTGCTAG 3' (forward primer) and 5' AGATATCACAGTAATGGCGGCACAAGGC 3' (reverse primer). The right recombination arm was amplified using 5' ATACGTACAGGAAGGAAACATTCTCTGG 3' (forward primer) and 5' ACTCGAGGTCTCAGGCAAGACTATGG 3'(reverse primer)and the middle recombination arm with 5' CATATGGCCTGGGAATCTGCGCAGGAC 3' and 5' GCGGCCGCAATATTTCTAACTGGGCTGC 3'. The amplified were cloned into TOPO vector (Zero Blunt® TOPO® Invitrogen) and the vectors were used for bacterial transformation. Each fragment was then isolated from bacterial clones and cut out by restriction enzymes reactions. Left arm was cut with ClaI and EcoRV, middle arm with NdeI and NotI and the right arm with SnaBI and XhoI. All the fragments were cloned in Psp72 vector containing the LoxP-PGK-Puro-LoxP fragment.

ES cells were transfected with 20 µg of linearized Psp72 vector caring the final construct, and cultured in ES cell medium supplemented with puromycin for the positive selection.Individual ES cell colonies that were grown in a Puromycin selection environment were picked and expanded individually and DNA was isolated from each one of them. The detection of the integrated construct was done by PCR using the following primers in the Venus sequence TATAGACAAACGCACACCGGC (RV) and in the Gata2 genomic sequence CAGCTTCAGCCTGCTTACTCA (FRW). The expected length of the amplified fragment was 2292 kb.

The Southern probe used for the verification of the correct integration was a 382 bp fragment located upstream of the left homology arm The fragment was amplified using Phusion™ taq polymerase (Finnzymes F-530L) and the primers used were 5' CTGAAGGATT CCCAGAGCCC A 3'and 5' AGCAAGAGGCAGGACTGAGTC 3'.

ES cell and EB culturing

ES cells were cultured on a monolayer of irradiated mouse embryonic fibroblast (MEF) feeder cells in ES cell culture medium (DMEM, FBS 15%, Glutamax 1%,

Non-Essential Amino Acids 1%, sodium pyruvate 1%, PS 1%, 50μM β mercaptoethanol) with 1000U/ml LIF (Leukemia Inhibitor Factor) at 37°C and 5% CO_2. ES cells were split 2-3 times per week.

EB formation was performed as described by Keller (Keller, Webb et al. 2002). ES cell cultures were harvested and single ES were seeded in suspension in ES cell culture medium in the absence of LIF at 37°C and 5% CO_2. ES cell cultures were left to differentiate into EBs on a shaker with a slow shaking motion of 50 rpm for the desired number of days. For d6-d8 EBs the medium was refreshed at d3-d4. EBs were made into single cell suspensions by trypsinization. Pelleted EBs were incubated at 37°C for 3-5 min. Trypsin was deactivated by addition of 1ml IMDM (50%FCS). The EBs dissociation was perfomed by passing the cells through a 20-gauge needle twice.

Recombination of selection cassette and selection of recombined clones

The excision of the selection cassette was performed by culturing targeted ES cells with diffusible HTN-Cre protein in the culturing medium. HTN-Cre production and Cre protein transduction of ES cells were performed as described by Peitz et al. (2007). Briefly, single cell suspensions of ES cells were plated, transduced 6 hours later and cultured in normal ES cell medium containing diluted HTN-Cre glycerol stock at a concentration of 5μM. The ES cells were incubated with the HTN-Cre protein for 18h, washed with PBS and cultured until single ES colonies could be picked and expanded . 55-60 colonies were picked and expanded, and DNA was isolated from each one of them. A PCR reaction designed to detect the presence of one LoxP site in the *Gata2* genomic locus was performed to detect recombined clones. 2 recombined clones were detected G2V5 and G2V6. The primers used were: (FRW) GACTTAGAGGTCCTCAGCCT (RV) TAGAGCACAGGCTGCAGCTC. The expected length of the wt product is 244 bp and of the recombined one 278 bp.

qRT-PCR

The sequences of the primers used for the qRT-PCR are presented here: ATCTTCTTCAAGGACGACGG Venus Forward (FRW) primer, GGCTGTTGTAGTTGTACTCC Venus Reverse (RV) primer, CACCCCTAAGCAGAGAAGCAA Gata2 FRW, TGGCACCACAGTTGACACACT Gata2 RV, β-actin. The PCR programme of the reaction was: 95oC 3’ min, followed by 40 cycles of 95°C 10’’ sec, 60°C 30’’ sec, 72°C 30’’ sec.

Protein extraction and Western blotting

Venus positive and Venus negative sorted cell suspensions were washed two times with Phosphate- buffered saline(PBS) and pelleted cells were dissolved in RIPA buffer and incubated for 20’-30’ on ice. Samples were centrifuged at 16000 g for 15’min and the supernatant was collected and snap frozen in liquid nitrogen. The samples were stored in -80°C. For western blotting the equivalent of the same

number of cells for each sample was boiled (95°C for 5'min) in SDS sample buffer. Samples were loaded on SDS-polyacrylamide gel, electrophoresed and then blotted for 1h 100 volts – 240 mA on a nitrocellulose membrane. The membrane was blocked with 1% BSA immunostained with anti Gata2 (Santa Cruz, sc9008x), anti GFP (MBL 598) and anti nucleophosmin 1 (Npm1) (B23 (FC-8791) SC-32236) antibodies. The staining was done in TBS buffer and the membrane was washed in TBS-T. Secondary antibodies used were IRDye 800CW Goat Anti mouse (LI-926-32210) and Goat Anti rabbit (LI-926-32211) were used and membranes were analyzed by using the Odyssey Infrared Imaging system (Li-Cor Biosciences)

Flow cytometric cell sorting and analysis

For analyzing cell surface markers expression of EBs cells, EBs were dissociated (as described before) and made into single cell suspensions and stained with PE anti-Flk1 (555308 BD Pharmigen), PE anti-CD31 (5553371 BD Pharmigen), PE anti-c-Kit (5553355 BD Pharmigen), PE anti-CD45 (553081 BD Pharmigen). Dead cells were excluded by Hoechst33258 (Invitrogen) staining and analyzed or sorted on a FACS aria II SORP. Results analysis was performed by FlowJo software (Tree star).

Matrigel differentiation assay

96 well dishes were coated with 50 µl of Matrigel™ Basement Mambrane Matrix (BD Biosciences cat no 354234). The thin gel method was used to coat the plates as described by the manufacturer. Plates were kept at 37°C until use. All other materials were pre-cooled materials and kept on ice. Cell populations to be used in the matrigel differentiation assay were pre-plated in a high glucose containing medium (DMEM, 20% FCS, 1% PS) on gelatin coated dishes. After 1-2 days, the cells were trypsinized, harvested, suspended in EGM-2 medium (EGM-2 Bulletkit, Lonza, cat num CC-3162) and plated on the matrigel coated wells. Plates were incubated at 37°C and 5% CO_2. Tubule and junction formation was observed under a IX70 Olympus fluorescent inverted microscope every 2-3 hours.

Methylcellulose hematopoietic progenitor assay

Cells were plated in methylcellulose (stem cell technologies) at a concentration of 2x104 cells/ml. In most cases triplicate plating (300 ul) was performed. After 10-12 days of incubation at 37°C, CO_2 the number of progenitor colonies was scored by microscopic observation.

References

Choi, K., Y. S. Chung, et al. (2005). "Hematopoietic and endothelial development of mouse embryonic stem cells in culture." Methods Mol Med 105: 359-368.

Choi, K., M. Kennedy, et al. (1998). "A common precursor for hematopoietic and endothelial cells." Development 125(4): 725-732.

Keller, G. M., S. Webb, et al. (2002). "Hematopoietic Development of ES Cells in Culture." Methods Mol Med 63: 209-230.

Kennedy, M., M. Firpo, et al. (1997). "A common precursor for primitive erythropoiesis and definitive haematopoiesis." Nature 386(6624): 488-493.

Kornhauser, J. M., M. W. Leonard, et al. (1994). "Temporal and spatial changes in GATA transcription factor expression are coincident with development of the chicken optic tectum." Brain Res Mol Brain Res 23(1-2): 100-110.

Ling, K. W., K. Ottersbach, et al. (2004). "GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells." J Exp Med 200(7): 871-882.

Mikkola, H. K., Y. Fujiwara, et al. (2003). "Expression of CD41 marks the initiation of definitive hematopoiesis in the mouse embryo." Blood 101(2): 508-516.

Minegishi, N., J. Ohta, et al. (1998). "Alternative promoters regulate transcription of the mouse GATA-2 gene." J Biol Chem 273(6): 3625-3634.

Nakano, T., H. Kodama, et al. (1994). "Generation of lymphohematopoietic cells from embryonic stem cells in culture." Science 265(5175): 1098-1101.

Nakano, T., H. Kodama, et al. (1996). "In vitro development of primitive and definitive erythrocytes from different precursors." Science 272(5262): 722-724.

Nardelli, J., D. Thiesson, et al. (1999). "Expression and genetic interaction of transcription factors GATA-2 and GATA-3 during development of the mouse central nervous system." Dev Biol 210(2): 305-321.

Ng, Y. K., K. M. George, et al. (1994). "GATA factor activity is required for the trophoblast-specific transcriptional regulation of the mouse placental lactogen I gene." Development 120(11): 3257-3266.

Nutt, S. L., D. Metcalf, et al. (2005). "Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors." J Exp Med 201(2): 221-231.

Palis, J., S. Robertson, et al. (1999). "Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse." Development 126(22): 5073-5084.

Pase, L., J. E. Layton, et al. (2009). "miR-451 regulates zebrafish erythroid maturation in vivo via its target gata2." Blood 113(8): 1794-1804.

Peitz, M., R. Jager, et al. (2007). "Enhanced purification of cell-permeant Cre and germline transmission after transduction into mouse embryonic stem cells." Genesis 45(8): 508-517.

Peitz, M., K. Pfannkuche, et al. (2002). "Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes." Proc Natl Acad Sci U S A 99(7): 4489-4494.

Porcher, C., W. Swat, et al. (1996). "The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages." Cell 86(1): 47-57.

Robert-Moreno, A., L. Espinosa, et al. (2005). "RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells." Development 132(5): 1117-1126.

Robin, C., K. Ottersbach, et al. (2011). "CD41 is developmentally regulated and differentially expressed on mouse hematopoietic stem cells." Blood 117(19): 5088-5091.

Tabatabaei-Zavareh, N., A. Vlasova, et al. (2007). "Characterization of developmental pathway of natural killer cells from embryonic stem cells in vitro." PLoS One 2(2): e232.

Tsai, F. Y., G. Keller, et al. (1994). "An early haematopoietic defect in mice lacking the transcription factor GATA-2." Nature 371(6494): 221-226.

Weiss, M. J., G. Keller, et al. (1994). "Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells." Genes Dev 8(10): 1184-1197.

Wiles, M. V. and G. Keller (1991). "Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture." Development 111(2): 259-267.

Chapter 3

The dynamics of Gata2 expression actively affect hematopoietic development

Polynikis Kaimakis, Emma de Pater, Undine Hill, Reinier van der Linden, Jean Charles Boisset, Catherine Robin, Tomomasa Yokomizo, Elaine Dzierzak

Erasmus Stem cell Institute, Department of Cell Biology, Erasmus MC, Rotterdam, the Netherlands

(Work in progress)

The dynamics of Gata2 expression actively affect hematopoietic development

Polynikis Kaimakis, Emma de Pater, Undine Hill, Reinier van der Linden, Jean Charles Boisset, Catherine Robin, Tomomasa Yokomizo, Elaine Dzierzak

Abstract

The Gata2 transcription factor is one of the master regulators of hematopoietic stem cell (HSC) generation in the mouse embryo. *Gata2* germline deletion leads to embryonic lethality, severe anemia and an absence of HSCs. Localization studies show that it is expressed in some cells in the embryonic hematopoietic sites during the different waves of hematopoietic cell generation. ES cell differentiation studies and overexpression studies in cell lines implicate a role for Gata2 in the expansion, differentiation and/or self-renewal ability of hematopoietic progenitors. However, the precise role of Gata2 in the development of HSCs is not fully understood. Previous studies have shown that cell populations sorted for hematopoietic markers such as c-kit and Sca-1 express *Gata2*, but prospective specific isolation of viable *Gata2* expressing cells for examination of their functional attributes has not been feasible.

In this study we describe the generation of a *Gata2-Venus* reporter mouse model which allows us to directly isolate, characterize and test the *in vivo* and *in vitro* functional potential of Gata2 expressing cells. We show that Gata2 expressing cell fractions in the aorta, vitelline and umbilical arteries and fetal liver are highly enriched in hematopoietic progenitors, and identify a novel subset of progenitors that are Gata2 independent. Importantly, we show that all the HSCs are Gata2 expressing. Live imaging studies demonstrate real-time Gata2 expression in intraaortic cluster cells and aortic endothelial cells at midgestation. These results highlight for the first time the dynamic expression of Gata2 in HSC development and suggest that changing levels in single cells may influence functional output.

Introduction

Hematopoiesis is the process by which blood cells are generated. During mouse embryonic development several distinct waves of hematopoiesis have been described. The first one is characterized by the generation of the blood islands in the yolk sac (YS). These are primitive erythroid cells originating from a cell with both hematopoietic and endothelial potential, the hemangioblast [review by (Dzierzak and Speck 2008)]. In the second wave erythro-myeloid hematopoietic progenitors arise (Palis, Robertson et al. 1999). Later on, at E9, cells with lymphoid potential can also be detected (Godin, Dieterlen-Lievre et al. 1995; Cumano, Dieterlen-Lievre et al. 1996). Finally, in the third wave, the hematopoietic stem cells (HSCs) are generated at midgestation (E10.5) in the mouse embryo. These are the cells with the

potential to differentiate into all the distinct cell types of the blood during the whole life span of an individual. They are characterized by their unique ability to self renew. In the adult they reside in the bone marrow (BM) and their development is under current intense investigation.

In the mouse embryo it has been demonstrated that the first HSCs are autonomously generated in the aorta gonad mesonephros (AGM) region at E10.5 (Muller, Medvinsky et al. 1994; Medvinsky and Dzierzak 1996). Apart from the AGM, other embryonic tissues have been shown to contain cells with HSC activity. These are the vitelline and umbilical arteries (VA and UA) beginning at E10.5 (de Bruijn, Speck et al. 2000), the yolk sac (YS) at E11 (Muller, Medvinsky et al. 1994; Medvinsky and Dzierzak 1996; Kumaravelu, Hook et al. 2002), the placenta (PL) at E11 (Gekas, Dieterlen-Lievre et al. 2005; Ottersbach and Dzierzak 2005) and the fetal liver (FL) where the HSCs migrate and expand at E11 - E11.5 (Muller, Medvinsky et al. 1994; Medvinsky and Dzierzak 1996; Kumaravelu, Hook et al. 2002; Gekas, Dieterlen-Lievre et al. 2005). The study and understanding of the processes governing the generation of the HSCs are of major importance for regenerative medicine.

Cell intrinsic regulation of HSC generation is provided mainly by transcription factors which play an important role in directing the hematopoietic gene expression program. Some of the major transcriptional regulators in HSC generation are Gata2, SCL and Runx1 (review by (Dzierzak and Speck 2008; Wilson, Calero-Nieto et al. 2011)). Recently it has been shown that these transcription factors work in a combinatorial manner by forming transcription factor complexes (Wilson, Foster et al. 2010).

Gata2 is one of the major regulators of HSC generation. Germline deficiency for *Gata2* results in embryonic lethality between E10-E10.5 and an anemic phenotype, with decreased number of primitive hematopoietic progenitors and a massive decrease in the number of definitive hematopoietic progenitors (Tsai, Keller et al. 1994). Profound effects of *Gata2* deficiency are also detected in chimeras with *Gata2*$^{-/-}$ ES cells, where no contribution to the hematopoietic tissues by deficient cells was found. Further studies on hematopoietic differentiation of *Gata2*$^{-/-}$ ES cells imply that this transcription factor plays a role in the survival or proliferation of the multipotent hematopoietic progenitor cells (Tsai and Orkin 1997). The embryonic lethality of *Gata2*$^{-/-}$ embryos at E10.5 precludes the study of AGM HSC generation. Instead, studies in *Gata2*$^{+/-}$ embryos show a greatly reduced number of phenotypic and functional HSCs in the AGM region, suggesting an important role for Gata2 in HSC generation (Ling, Ottersbach et al. 2004). *In situ* hybridization studies show that *Gata2* is expressed in the hematopoietic tissues during all the waves of embryonic development. More specifically, *Gata2* can be found in endothelial cells and in the intraaortic hematopoietic cluster cells. It is also highly expressed in the PL and FL

(Nardelli, Thiesson et al. 1999; Minegishi, Suzuki et al. 2003; Robert-Moreno, Espinosa et al. 2005; Pimanda, Ottersbach et al. 2007). High levels of *Gata2 mRNA* expression can be detected in cell populations enriched in HSCs (Orlic, Anderson et al. 1995). Overexpression studies and studies in haploinsufficient animals reveal that quantitative levels of Gata2 expression play an important role in its function. (Heyworth, Gale et al. 1999; Persons, Allay et al. 1999).

To date, the analysis of specifically enriched *Gata2* expressing cells has been limited to a reporter mouse model that results in Gata2 haploinsufficiency state (Minegishi, Suzuki et al. 2003). In this study we generate and characterize a new *Gata2-Venus* mouse model in which the *Venus* fluorochrome gene expressed in tandem with *Gata2* does not affect the normal level or function of Gata2. In the embryo, we show that Venus expression recapitulates *Gata2* expression and that hematopoietic progenitors are enriched in the Venus positive sorted AGM fraction. Interestingly, a Gata2 independent hematopoietic progenitor population is identified in the Venus negative AGM fraction and verified in *Gata2*$^{-/-}$ embryos. The HSC population on the other hand is exclusively Gata2 Venus expressing. Using *in vivo* live imaging we show that Gata2 expression in the aorta is dynamic during the period of the onset of HSC generation, thus suggesting that the active onset and changing levels of expression of Gata2 drives hematopoietic development.

Results

Generation and characterization of *Gata2-Venus* mice

To directly isolate and examine the potential of *Gata2* expressing cells a knockin reporter mouse model was generated. Briefly, the coding sequence of the Venus fluorochrome gene following an internal ribosome entry site (IRES) sequence was recombined into the *3'UTR* of the *Gata2* genomic locus (fig 1A) in ES cells. A puromycin cassette flanked by lox recombination sites was included in the targeting construct for selection purposes. The G2V construct and recombination strategy are described in detail in chapter 2. This approach allows for the normal expression of Gata2 from the regulatory elements within its genomic locus without affecting the levels of expression or protein function. These characteristics are particularly important since Gata2 haploinsufficiency greatly reduces the number of HSCs generated during development (Ling, Ottersbach et al. 2004). After blastocyst injections and identification of germline transmitting founders, the first generation offspring carrying the G2V knockin were crossed with CAG-Cre transgenic mice to excise the puromycin selection cassette. Offspring were backcrossed onto the C57BL/6 background. A *Gata2*$^{Venus/Venus}$ mouse colony was established. These mice breed normally and show no growth defects. They also show no hematopoietic defects since

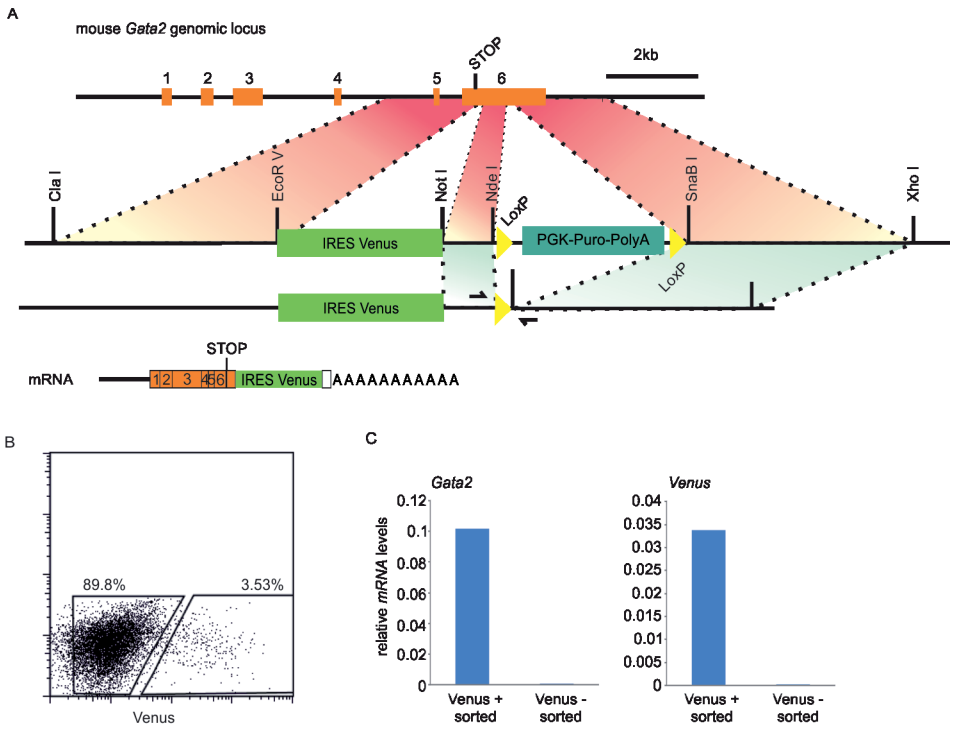

Figure 1 *Gata2-Venus* model can be used for the prospective isolation of Gata2 expressing cells

(A) The targeting strategy for the generation of the *Gata2-Venus* mouse model is presented. In the upper panel *Gata2* genomic locus is depicted. Exons are indicated by orange boxes and the position of the STOP codon is shown. In the middle panel the targeting construct is depicted with the *IRES-Venus* coding sequence and the puromycin selection cassette coding sequence are indicated by rectangular boxes. The *LoxP* sites are indicated by triangles and the homology regions highlighted. In the bottom panel the predicted structure of the targeted allele is shown after the excision of the selection cassette. (B) FACS analysis of adult BM. Percentage of Venus expression is indicated in the Venus gate. (C) Relative *mRNA* expression of *Venus* and *Gata2* in Venus positive sorted and Venus negative sorted cell populations of adult BM samples. Both *Venus* and *Gata2* can only be found in the Venus positive sorted cell fraction.

they have the same lineage subset distribution with wild type (WT) animals (supplementary table 1), indicating that Gata2 function is normal.

To determine whether Venus expression parallels that of Gata2 in hematopoietic cells, bone marrow was isolated from adult *Gata2*$^{Venus/Venus}$ mice. Cells were flow cytometrically sorted into Venus positive and negative fractions (figure 1B) and RNA was isolated. qRT-PCR was performed for *Gata2* and *Venus* sequences. As

shown in figure 1C, only sorted Venus positive cells express *Venus* and *Gata2 mRNA*. These results indicate that *Venus* expression from the targeted allele correctly reflects *Gata2* expression and that our *Gata2-Venus* mouse model is suitable for the isolation and study of Gata2-expressing cells.

Gata2-Venus is expressed in the midgestation aorta and other embryonic hematopoietic tissues

We next examined whether the *Venus* expression pattern in *Gata2-Venus* embryos is similar to that described by *Gata2 in situ* hybridization studies (Nardelli, Thiesson et al. 1999; Robert-Moreno, Espinosa et al. 2005; Pimanda, Ottersbach et al. 2007). Whole mount embryo immunostainings at E10.5 revealed high level Venus expression in the AGM and FL, the olfactory bulb (OB) and along the length of the neural tube (NT) (fig 2A,B,C). These findings correspond to the previously described *in situ* pattern of *Gata2* in these tissues (Nardelli, Thiesson et al. 1999). Confocal imaging of thick transverse sections of E10 *Gata2-Venus* embryos show that Venus is expressed in the aorta in the endothelial cell lining of the aorta and in intraaortic cluster cells (fig 2D,E). Venus expression is also found in the urogenital region of the AGM (fig 2D). A more detailed analysis of Venus expression in the major vasculature of the embryo - the aorta, VA and UA - and other hematopoietic tissues was performed by immunostaining. Transverse sections of Gata2-Venus embryos at different embryonic timepoints E9.5, E10.5 and E11.5) (fig 2E-I) were stained with anti-CD31 antibody to detect endothelial and intraaortic hematopoietic cluster cells. Confocal imaging revealed that Venus is expressed in the endothelial cells of the dorsal aorta and the VA/UA at E 9.5 (fig 2E,F) and similarly at E10.5 and E11.5 (fig 2G,H). At E10.5 high Venus expression in the hematopoietic cluster cells is observed (fig 2H). Apart from the aorta, Venus expression is found in other embryonic hematopoietic sites such as the FL (fig 2I), yolk sac and placenta (not shown). In the FL, the punctuate pattern of Venus expressing cells is similar to what has been previously documented for *Gata2 in situ* analysis (Pimanda, Ottersbach et al. 2007). Thus, Venus expression in *Gata2-Venus* embryos recapitulates normal Gata2 expression.

Gata2 expression levels are dynamic in the midgestation aortic endothelium and clusters

Previous studies have used the Ly6A GFP embryo model (which marks HSCs) (de Bruijn, Ma et al. 2002) to track the generation of hematopoietic cells in real-time (Boisset, van Cappellen et al. 2010). To track the onset of Gata2 expression and its relationship to the endothelial to hematopoietic cell transition we performed real-time imaging on thick sections of *Gata2*$^{Venus/Venus}$ embryos. Live imaging of *Gata2-Venus* embryo sections show that at E10.5 Gata2 is expressed in the aorta

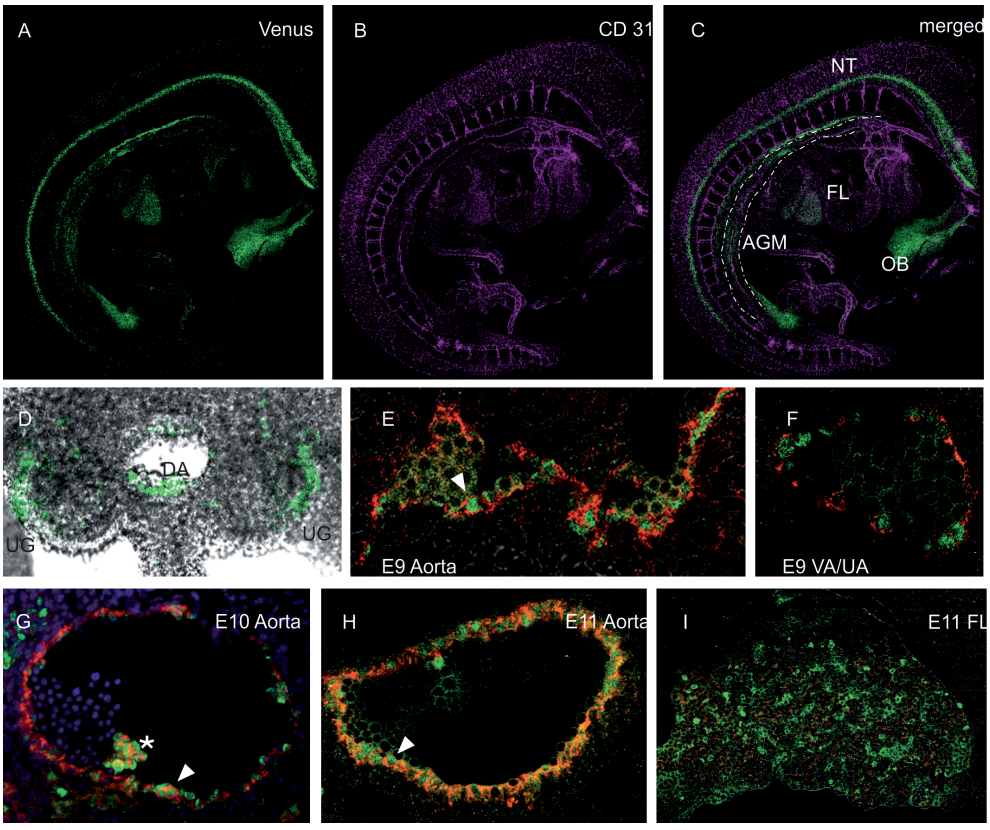

Figure 2 Venus expression in hematopoietic tissues during mouse embryogenesis

(A-C) Confocal imaging of E10.5 whole mount *Gata2-Venus* embryos immunostained with CD31. Venus expression can be detected in the hematopoietic sites of the aorta (indicated in C with a dotted line) and the FL and in the olfactory bulb (indicated in C with the initials OB) and alongside the neural tube (indicated in C with the initials NT). (D) Confocal imaging of thick transverse *Gata2-Venus* embryo sections of E10 embryos. Brightfield merged with fluorescent image is shown. The tissues are imaged fresh and Venus fluorochrome expression without any staining is detected. The ventral side of the embryo is positioned at the bottom of the image. DA stands for Dorsal Aorta and UG for UoroGenital regions. (F-J) Fluorescent microscopic images of transverse sections of E9 (F), E10 (H) and E11 aortas (I), E9 VA/UA (G) and E11 FL (J). Venus expression co-localizes with CD31 expression in cells of the aortic endothelium and in the endothelium of the VA and UA and is expressed in a characteristic punctuate pattern in the FL. The arrowheads in 2E,2G,2H indicate a Venus expressing endothelial cell. The asterisk in 2G indicates a hematopoietic cluster.

in a very dynamic manner. During 12 hours, we observe emergence of Gata2 expressing aortic hematopoietic cells and the active movement of hematopoietic cluster cells (video 1 and fig 3). Both the onset and extinction of Venus expression can be observed, revealing the highly active regulation of Gata2 transcription in single

hematopoietic and/or endothelial cells. The differential expression, as well as the fluctuation/oscillation of Venus expression in single hematopoietic cluster/endothelial cells in this small window of time, suggests that the dynamic regulation of Gata2 levels is important in the function of the individual cells.

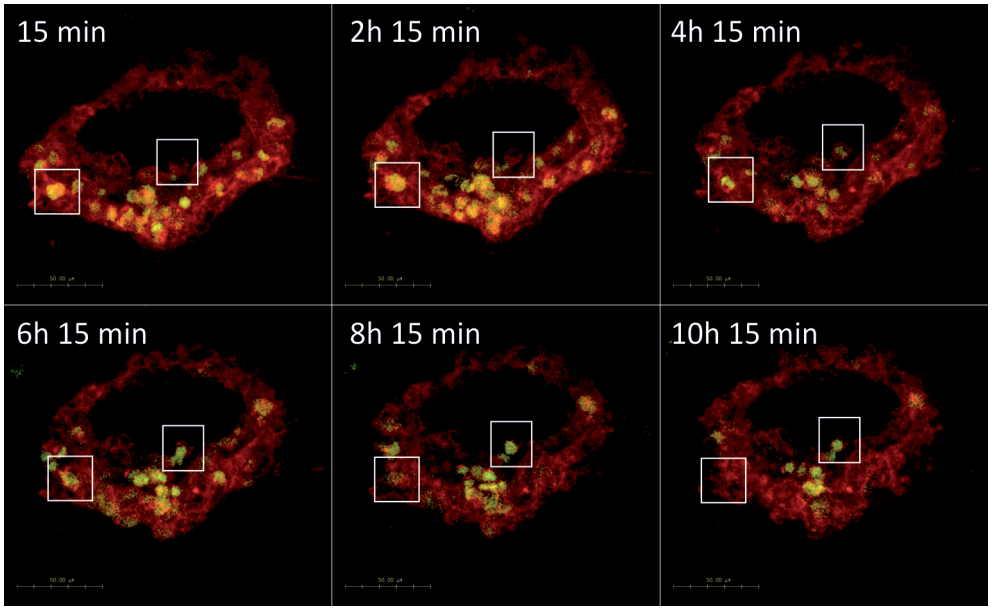

Figure 3 Gata2 expression during E10 AGMs is very dynamic

Time lapse imaging of E10 *Gata2-Venus* embryo slices. X20 lens with the ventral side in the bottom. Three dimensional image stacks of a thick section of the dorsal aorta are shown for different timepoints during a 12 hour confocal imaging session. In all of the sections the ventral side of the dorsal aorta is in the bottom of the image. Each of the images has a two hours time difference. Box on the right indicates the presence of a cell turning on Venus expression (at 4h 15min) and box on the left indicates a cell which turns off Venus expression (at 10h 15 min) during 10-12h of imaging. Massive morphological changes and cell movements in the intraaortic cell cluster are observed. The intraaortic cluster is located in the area between the two boxes. In addition, many cell movements in the endothelial cell layer can be observed in the video.

Gata2 expressing cells increase in number during development in hematopoietic tissues

To study the role of Gata2 directly in the cells that express it during embryonic stages, flow cytometric analysis was performed at different time points on *Gata2-Venus* hematopoietic tissues (fig 4A). At E9 Venus expression is most prominent in the YS, with 5.8% ±0.6 of viable cells expressing Venus and in the AGM, 1.7% ±0.1 of cells are Venus expressing. At E10 - E10.5, the time when the first HSCs are

generated and the number of intraaortic hematopoietic clusters reaches a maximum, 1.8% ±0.3 of AGM cells are Venus positive and at E11.5 it increases to 6% ±0.3. In the FL, the percentages of Gata2-Venus expressing cells increase 100-fold from E10 to E11.5, with 0.13% ±0.09 and 12.3% ±0.1 Venus+ cells at these respective time points. The total number of Gata2-expressing cells in these tissues is shown in table 1 and fig 4B.

Table 1

		Gata2-Venus expressing cells/ee		
Tissue	**Cell fraction**	**E9.5**	**E10.5**	**E11.5**
AGM	**total**	1428 ±85	2186 ±328	16648 ±1154
	CD31 hi c-Kit+	ND	373	175
	CD31 hi c-Kit-	ND	355	322
	CD31 int	ND	206	1728
	CD31 -	ND	1988	12345
YS	**total**	7749 ±715	ND	ND
FL	**total**	ND	66	29520 ±240

Table 1 Number of Venus+ cells / embryonic tissue

Absolute number of Venus positive cells in each embryo equivalent (ee) is presented. Dissected embryonic tissues (YS, AGM, FL) were made into single cell suspensions and FACS analysis was performed to define the percentage of cells expressing Venus. For E10.5, the AGM tissue (186x103 cells/ee) was used but for E11.5 AGM, we included the embryo body part of the VA and UA (270x103cells/ee). Our data for the total number of cells in each tissue correlated with published data for YS and E9.5 AGM (Dzierzak and de Bruijn 2002) and for FL (Gekas, Dieterlen-Lievre et al. 2005).

Staining of *Gata2-Venus* embryo hematopoietic tissues with c-Kit and CD31 antibodies can provide a more in depth analysis of the Gata2 expressing cell population. The triple positive ($CD31^{hi}$c-$Kit^{+}Venus^{+}$) cell fraction represents hematopoietic cells present at the intraaortic clusters. At E10.5 the absolute number of these triple positive cells in one embryo equivalent (ee) is 373 and at E11.5 this number drops to 175 (table 1). This reduction is comparable to the reduction of hematopoietic intraaortic clusters and cells that has been previously reported (Yokomizo and Dzierzak 2010) from E10.5 to E11.5. At the same time points the $CD31^{int}$c-$Kit^{-}Venus^{+}$ cell fraction represents endothelial cells expressing Gata2. It has already been show by the immunostainings that endothelial cells express Gata2. Part of this cell fraction might represent hemogenic endothelial cells that have not yet acquired hematopoietic fate. The remainder of the $Venus^{+}$ cells are likely to be in the gonads region of the AGM (fig 2D). The absolute number of Gata2-Venus expressing cells in the hematopoietic tissues at E9.5, E10.5 and E11.5 is presented in detail in table 1.

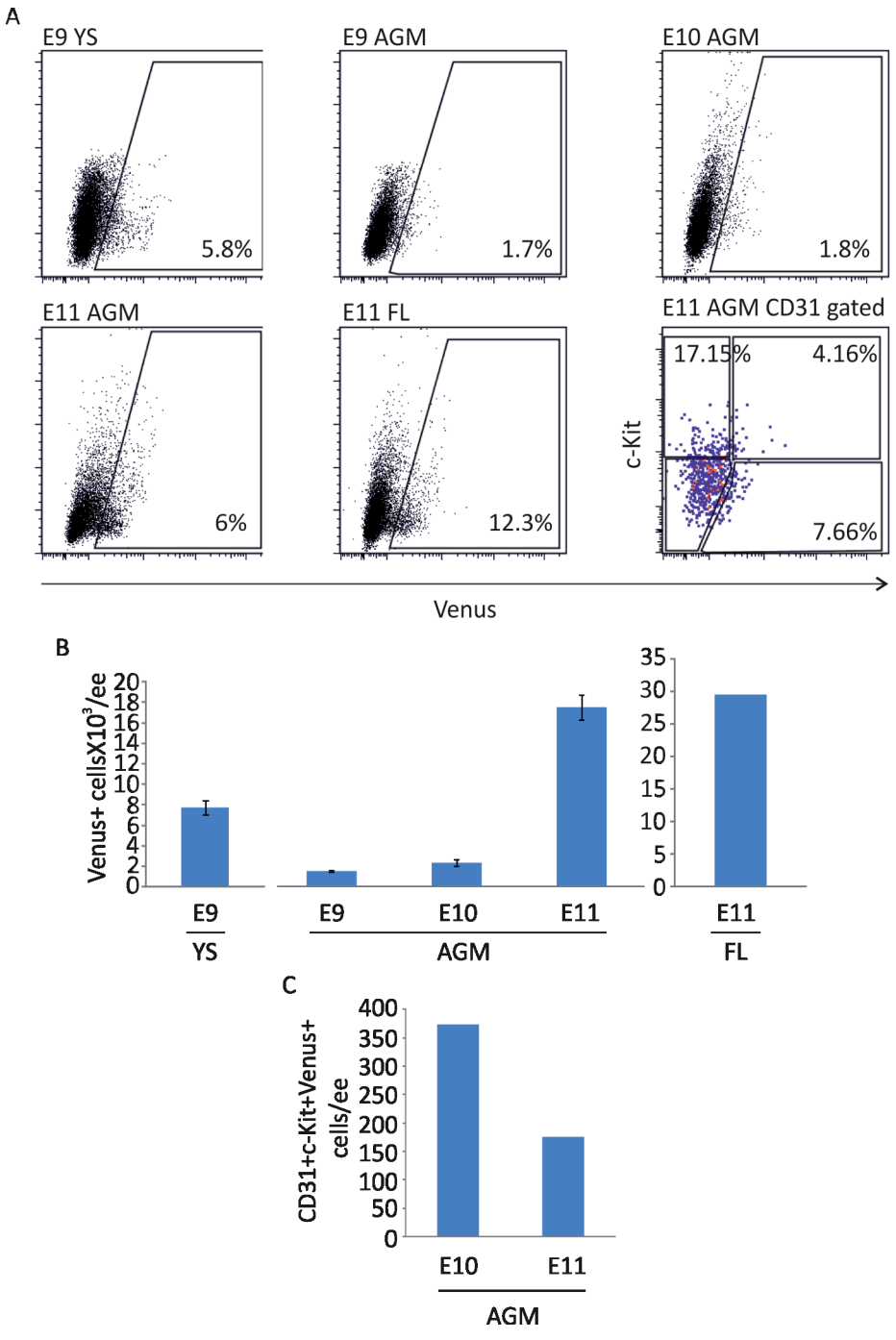
A
E9 YS
5.8%
E9 AGM
1.7%
E10 AGM
1.8%
E11 AGM
6%
E11 FL
12.3%
E11 AGM CD31 gated
17.15%
4.16%
7.66%
c-Kit
Venus
B
Venus+ cellsX10³/ee
E9
YS
E9
E10
E11
AGM
E11
FL
C
CD31+c-Kit+Venus+ cells/ee
E10
E11
AGM

Figure 4 Gata2 is expressed in all the hematopoietic sites during mouse embryonic development

(A) Representative FACS plots demonstrating % of cells expressing Venus in: E9 YS, E9 AGM, E10 AGM, E11 AGM and E11 FL. Gates indicate the Venus positive cell fraction. Same gatings were used for the sorting of the cell populations in functional experiments. FACS analysis of Gata2-Venus E11.5 AGM stained against CD31 and c-Kit. In the plot the gated CD31 positive cell population is shown for Venus and c-Kit expression. 4.16% of CD31+ are c-Kit+ Venus+ positive hematopoietic cluster cells. The C31+ c-Kit- Venus+ cell fraction represents endothelial cells and the C31+ c-Kit+ Venus- cell fraction possibly represents Gata2 independent hematopoietic progenitors. (B) Absolute numbers of Venus positive cells in individual embryonic tissues (ee=embryo equivalent). (C) Absolute number of triple positive cell population (CD31hic-Kit+Venus+) at E10.5 and E11.5 embryos. This cell population represents hematopoietic cells of the intraaortic clusters. The reduction of hematopoietic intraaortic cluster cells between E10.5 and E11.5 has already been reported in the literature (Yokomizo and Dzierzak 2010).

Gata2 expression in the major embryonic vasculature defines most hematopoietic progenitors and reveals a Gata2 independent subset

To test the hematopoietic function of Gata2 expressing cells, AGM cells were sorted into Venus positive and negative fractions and plated in the methylcellulose progenitor assay. High enrichment of hematopoietic progenitors was found in the Venus positive cell fraction. At E9 the Venus expressing cell fraction yielded 51.61 ±9.58 colony forming units (CFU) per embryo, 2.2-fold more than in the Venus negative fraction (22.95 ±3.40) (fig 5A). At E10, CFU numbers increased: 3.5-fold more CFU were found in the Gata2 expressing AGM cell fraction (93.03 ±15.42) as compared to the Gata2 negative fraction (26.36 ±11.3) (fig 5B). Similarly at E11 there were 3.8-fold more CFU in the Venus positive fraction as compared to the Venus negative fraction (167.30 ±29.94 and 43.65 ±7.79 respectively: fig 5C). Total CFU numbers in both sorted fractions increased with developmental time and all types of progenitors (uni, bi and multipotential) were found. Most notably, multipotent progenitors forming mixed colonies of granulocyte, erythroid, megakaryocyte, macrophage cells (CFU-GEMM) were found almost exclusively in the Venus expressing cell fraction. The Venus expressing fraction at E9, E10 and E11 showed 7.9-, 8.1- and 21.4- fold more CFU-GEMM than the Venus negative fraction. Detailed results of methylcellulose progenitor assays are shown in table 2. These data clearly demonstrate that the Gata2 expressing cell fraction is highly enriched in hematopoietic progenitors, especially the most immature multipotential progenitors.

Surprisingly a significant number of non-Gata2 expressing hematopoietic progenitors were found. These progenitors may have once expressed *Gata2* and then extinguished its expression. Alternatively, a Gata2 independent progenitor subset may exist. To test this we performed methylcellulose assays on AGM cells

Table 2

		hematopoietic progenitors / ee					
		BFU-E	CFU-G	CFU-M	CFU-GM	CFU-GEMM	Total
E9 AGM	Venus +	2,91 ± 2,16	10,94 ± 2,88	17,87 ± 4,26	13,85 ± 2,23	6,03 ± 3,28	51,61 ± 9,58
	venus -	2,21 ± 0,29	5,81 ± 1,33	7,52 ± 1,76	6,64 ± 2,31	0,76 ± 0,34	22,95 ± 3,4
	total	5,12	16,75	25,39	20,50	6,80	74,56
E10 AGM	Venus+	16,03 ± 2,95	5,4 ± 1,29	37,04 ± 7,26	19,41 ± 4,96	15,15 ± 2,23	93,03 ± 15,42
	Venus-	2,90 ± 0,78	0,81 ± 0,45	17,55 ± 11,86	3,22 ± 1,05	1,87 ± 0,51	26,36 ± 11,3
	Total	18,93	6,22	54,59	22,63	17,02	119,39
E11 AGM	Venus+	40,74 ± 14,05	3,67 ± 1,14	76,47 ± 19,31	33,35 ± 7,53	13,07 ± 3,72	167,30 ± 29,94
	Venus-	5,27 ± 1,92	1,67 ± 1,07	28,08 ± 7,71	8 ± 2,42	0,61 ± 0,61	43,65 ± 7,79
	Total	46,01	5,37	104,55	41,35	13,69	210,96

Table 2 Hematopoietic progenitor numbers from Venus positive and Venus negative sorted cells of E9, E10 and E11 AGMs

AGM tissues were dissected and single cell suspensions were sorted based on Venus expression. The two different cell populations were plated on methylcellulose progenitor assays. 1 or 0.3 embryo equivalents were plated in one plate or in triplicate regarding the number of expected progenitors in each tissue. Number of colonies was scored 12 days after culture. E9 embryos: 24-29 sp (n=2). E10 embryos: 30-35 sp (n=2). E11 embryos: 45-48 sp (n=2). Bar graphs representative of these data are shown in fig 5 A, B and C.Gata2-/- embryos.

from *Gata2*$^{-/-}$ embryos. Early E9 *Gata2*$^{-/-}$ AGMs contained 16.7 ±5.7 CFU, an almost identical number of CFUs as found in the Venus negative fraction of *Gata2-Venus* E9 AGMs (comparison between fig 5A and 5D, and tables 2 and 3). Whereas most colony types were found, few *Gata2*$^{-/-}$ CFU-GM and no CFU-GEMM were observed. Thus, Gata2 independent progenitors exist and for the first time this distinct subset of cells can be isolated from progenitors that express and require Gata2.

Further evidence concerning the existence of hematopoietic progenitors in the *Gata2*$^{-/-}$ embryos comes from whole mount immunostaining with the hematopoietic marker c-Kit. c-Kit^{+} hematopoietic cells can be observed in *Gata2*$^{-/-}$ embryos in VA and UA. These clusters are reduced in size and number as compared to WT and *Gata2*$^{+/-}$ embryo (fig 5E and F).

Further hematopoietic progenitor enrichment and subset discrimination in the *Gata2-Venus:Ly6A-GFP* double reporter mouse model

Mouse models similar to our *Gata2-Venus* have been used as a means of defining cell populations enriched in HSC activity. Studies with the *Ly6A-GFP* reporter mouse model (de Bruijn, Ma et al. 2002; Ma, Robin et al. 2002) have been particularly informative concerning HSC generation. Interestingly, the *Ly6A-GFP* and

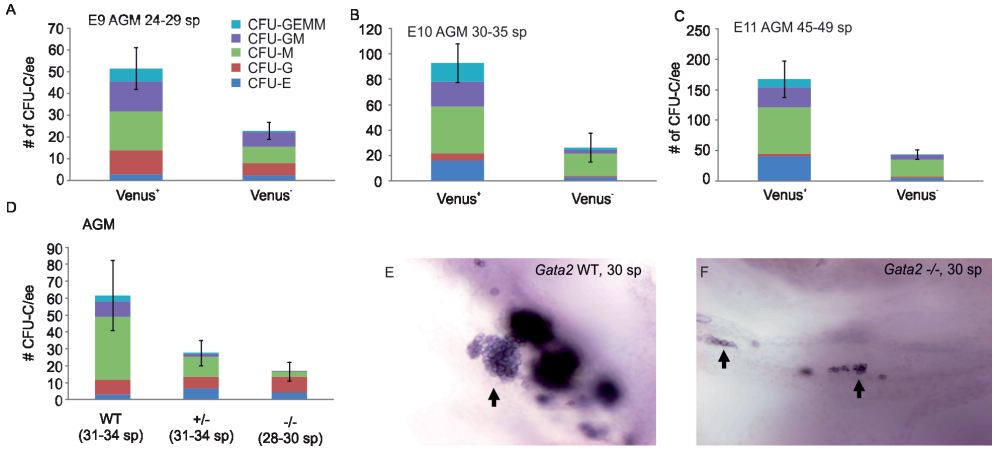

Figure 5 Gata2 expressing cell fraction defines most hematopoietic progenitors

(A-C) Methylcellulose progenitor assays of (A) E9, (B) E10 and (C) E11 AGMs of Venus positive and Venus negative cell fractions. In the graphs the total number of progenitors in 1 ee of Venus positive and the Venus negative sorted cell fractions are presented. Single cell suspensions of E9, E10 and E11 AGMs were sorted based on Venus expression and 1-0,3 embryo equivalents of cells were plated in a single dish or in a triplicate of methylcellulose respectively. The number of colonies was scored after 12 days in culture. (D) Wild type (WT), heterozygote (*Gata2*$^{+/-}$) and mutant (*Gata2*$^{-/-}$) AGMs derived from crossings between *Gata2*$^{+/-}$ mice were tested for the number and type of myeloid progenitors present. In the early E10 AGM tissues of WT embryos 62.4 ±20.8 progenitors were found, whereas *Gata2*$^{+/-}$ and *Gata2*$^{-/-}$ AGMs showed a great reduction in the number of progenitors (27.6 ±10.1 and 16.7 ±5.7 respectively) which represents a 3.7 fold decrease between the WT and the *Gata2*$^{-/-}$ animals ($p<0.05$). Even greater differences were observed in the number of multipotent progenitors (CFU-GEMM, CFU-GM). The number of CFU-GEMMs was 4.7 fold decreased in *Gata2*$^{+/-}$ AGMSs ($p<0.05$) while completely absent in the *Gata2*$^{-/-}$ AGMs and the number of CFU-GMs was 5.2 fold decreased in *Gata2*$^{+/-}$ AGMs (decrease not significant) and 75.4 fold decreased in *Gata2*$^{-/-}$ AGMs ($p<0.1$) (table 3). Detailed description of the number and different types of colonies of each cell population in the different developmental time points is presented in table 2. (E,F) Whole mount immunostaining against c-Kit. Hematopoietic clusters and cells are observed in WT VA/UA (C) and *Gata2*$^{-/-}$ animals. The size and the number of the clusters and the number of the c-Kit^{+} cells are greatly reduced in the *Gata2*$^{-/-}$ embryos.

Gata2-Venus expression patterns in the E10 aortic endothelium and cluster cells are very similar. To examine whether the Ly6A-GFP expressing cell population completely overlaps with the Venus expressing population we crossed our *Gata2-Venus* mice with *Ly6A-GFP* mice. Double reporter embryos were generated and AGM cells were sorted into four distinct populations (fig 6A). A small percentage (1%) of AGM cells expresses both markers. The results of methylcellulose assays show that all

Table 3

tissue		Hematopoietic progenitors					
		BFU-E	CFU-G	CFU-M	CFU-GM	CFU-GEMM	Total
AGM	+/+ (31-34 sp)	3,02 ± 1,39	8,92 ± 2,04	37,17 ± 13,66	9,05 ± 5,43	3,49 ± 1,42	62,38 ± 20,80
	+/- (31-34 sp)	6,67 ± 5,16	7,04 ± 2,79	11,48 ± 2,78	1,73 ± 0,53	0,74 ± 0,25	27,65 ± 10,11
	-/- (28-30 sp)	4,07 ± 2,71	9,38 ± 2,61	3,09 ± 0,92	0,12 ± 0,10	0	16,67 ± 5,69
YS	+/+ (31-34 sp)	105,71 ± 18,63	305,71 ± 70,60	312,86 ± 40,69	120,00 ± 43,00	74,29 ± 36,51	918,57 ± 147,90
	+/- (31-34 sp)	78,75 ± 14,19	177,50 ± 30,51	213,75 ± 20,90	123,75 ± 18,12	48,75 ± 15,62	642,50 ± 51,32
	-/- (28-30 sp)	2,30 ± 0,90	10,81 ± 1,70	8,34 ± 2,31	1,81 ± 1,01	2,49 ± 1,61	25,75 ± 5,27

Table 3 Hematopoietic progenitors number in late E9 - early E10 *Gata2*$^{+/+}$, *Gata2*$^{+/-}$ and *Gata2*$^{-/-}$ AGM and YS

AGM and YS tissues were dissected from the embryos and single cell suspensions were plated on methylcellulose progenitor assays. Three experiments were performed (n=3) and the number of *wt,+/-* and *-/-* embryos analyzed was 7, 10 and 9 respectively. The numbers of the different types of progenitors present in one embryo equivalent (1ee) are presented in the table. CFU-E: erythroid colonies, CFU-G: granulocyte colonies, CFU-M: macrophage colonies, CFU-GM: granulocyte, macrophage colonies, CFU-GEMM: granulocyte, erythrocyte, monocyte, megakaryocyte colonies. The differences between the progenitor numbers in +/+ and -/- animals are significant ($p<0.05$ for the total AGM, CFU-GEMM AGM, CFU-M AGM, total YS, CFU-GM YS, CFU-M YS, CFU-G YS, CFU-E YS and $p<0.1$ for CFU-GM AGM, CFU-G, CFU-E, CFU-GEMM.)

four sorted populations contain hematopietic progenitors. The double positive cell population (Venus+GFP+) is the most highly enriched in progenitor activity (53.8 ±22.9 CFU/ee) and it contains the highest proportion of multipotent progenitors, CFU-GEMM (fig 6B). Although the single positive and double negative fractions contain more cells, 2-3 times fewer CFU/ee are found. CFU-G appear to be enriched in the Venus- fractions, whereas the Venus+ fractions are enriched in CFU-M, suggesting that Gata2 influences lineage fate. Detailed results of the methylcellulose progenitor assay are presented in table 4.

All long-term repopulating AGM HSCs are Gata2 Venus expressing and phenotypic analysis indicate that they are Gata2 intermediate expressing

Despite previous data showing that HSC enriched populations transcribe *Gata2*, it has been unclear whether all HSCs express Gata2 and to what level. To test this, the Venus positive and negative E11 AGM cell fractions were examined for HSC activity by in vivo transplantation assays. The two sorted cell populations were injected into irradiated adult animals and engraftment was analyzed at four months post transplantation. Long term, high level engraftment (>10%) by Venus+ donor cells was found only in the Venus expressing cell population (table 5). No engraftment was found in any of the 11 recipient mice injected with up to 8 embryo equivalents of Venus- AGM cells. Thus, all adult repopulating HSCs express Gata2. At E14 and E16, when the HSCs are located in the FL, flow cytometric analysis for SLAM

Table 4

E11,5-E12 embryos 46-53 sp	CFU-C/AGM+UA+VA					
	BFU-E	CFU-G	CFU-M	CFU-GM	CFU-GEMM	Total
GFP-/venus-	0,87	15,46	3,48	1,75	0,44	22,00
Venus+	3,13	4,80	16,31	3,45	0,44	28,13
GFP+	0,00	12,13	12,96	4,77	1,30	31,15
GFP+/Venus+	3,16	3,80	32,22	9,26	5,36	53,79
Total	7,16	36,19	64,97	19,22	7,53	135,07

Table 4 Ly6A-GFP$^+$:Gata2-Venus$^+$ cell fraction is highly enriched in hematopoietic progenitors

Methylcellulose progenitor assay of the four cell populations isolated from *Ly6A-GFP:Gata2-Venus* double positive animals. In the table the number of progenitors present in each cell fraction in 1 ee is presented. The double positive cell fraction is highly enriched in hematopoietic progenitors and especially in GEMMs.

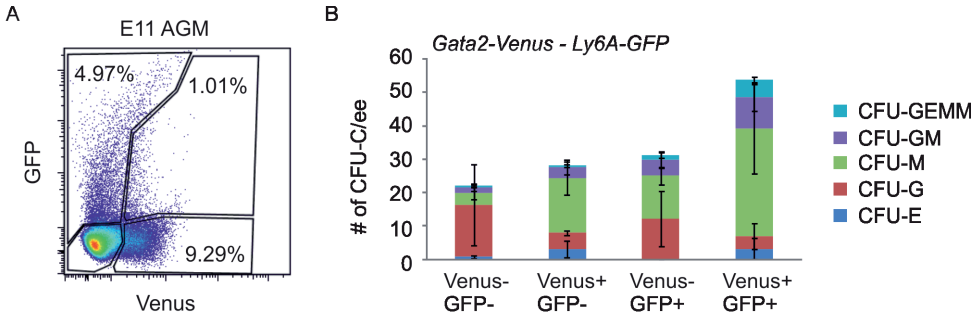

Figure 6 Venus+GFP+ cell fraction is highly enriched in hematopoietic progenitors

(A) FACS analysis of E11 AGM of a *Ly6A-GFP:Gata2-Venus* embryo. The four distinct cell populations were sorted and plated on methylcellulose progenitor assays. (B) Number of progenitors present in each cell fraction in 1 embryo equivalent of tissue. The double positive cell fraction is highly enriched in hematopoietic progenitors and especially in CFU-GEMM.

markers strongly indicates that the HSCs are located in the Venus^{int} cell fraction. More specifically, at E14 the CD150$^+$CD48$^-$ cell population, which includes the HSCs (Kim, He et al. 2006), cannot be found in the Venus^{hi} cell fraction, while 8% of Venus^{int} expressing cells is CD150$^+$CD48$^-$ (data not shown). At E16, the Venus^{hi} only a 1.45% of Venus^{hi} cells express CD150 and only 29.5% of them express c-Kit while in the Venus^{int} cell fraction 95.7% of the cells are c-Kit$^+$ and 41.5% are Cd150$^+$, again indicating that the HSCs are located in the Venus^{int} cell fraction(data not shown).

Table 5

HSC activity of Venus+ and Venus- cell populations		
E11 AGM cell fraction	Venus +	Venus -
time after injection	4 months	4 months
number reconst/numbers injected	5/11	0/11

Table 5 All of the long term HSCs are found in the Venus^{+} cell fraction

HSC activity in Venus positive and Venus negative cell populations of E11 AGMs (41-49 sp fluctuation). 2-5 ee of each cell fraction was injected into adult lethally irradiated mice. Results of 5 experiments are presented for the long term (4 months) reconstitution capacity. In the table the number of animals showing reconstitution out of the number of total injected animals after 4 months. The reconstitution efficiency was tested by PCR designed to detect the *Venus* allele in DNA isolated by peripheral blood.

Discussion

In this study we describe the generation of a new *Gata2-Venus* reporter mouse model in which Venus fluorochrome is used as a marker of the Gata2 expressing cells. For the first time it is possible to show that all HSCs in the developing mouse embryo express Gata2. Moreover, we show that most but not all hematopoietic progenitors are Gata2 dependent. We followed a strategy in which *Venus* coding sequence is recombined into the *3'UTR* of *Gata2* genomic locus. This strategy minimizes the possibility of altering in the expression levels and function of the Gata2 transcription factor. A similar strategy has previously been followed for the generation of a PU.1 transcription factor fluorescent reporter mouse. This mouse model (Nutt, Metcalf et al. 2005) revealed that different levels of PU.1 expression define distinct progenitor populations and correlates the downregulation of PU.1 expression with loss of myeloid linege potential and restriction to megakaryocyte and erythroid (MegE) lineages.

Analyses of our reporter model show that *Gata2*$^{Venus/Venus}$ mice have no profound hematopoietic defects and that they breed normally. We found in sorted cell fractions that *Venus* and *Gata2* are only expressed in the Venus positive cells. Localization analyses of Venus expression in *Gata2*$^{Venus/Venus}$ embryos show a similar expression pattern with published *Gata2 in situ* hybridization results (Nardelli, Thiesson et al. 1999; Robert-Moreno, Espinosa et al. 2005; Pimanda, Ottersbach et al. 2007). These data verify that our new *Gata2-Venus* reporter mouse model is suitable for prospective isolation and analysis of Gata2 expressing cells directly from the normal *in vivo* physiologic setting without manipulation.

Imaging and FACS analyses of *Gata2-Venus* embryos confirm that Gata2 is expressed in all the hematopoietic sites during mouse embryogenesis. Interestingly,

the numbers of Gata2 expressing cells change in accordance with developmental and temporal hematopoietic changes occurring in these sites. At E9 Venus expressing cells are found predominantly in the YS, which at this time produces the highest numbers of the hematopoietic progenitors in the conceptus. Slightly later as hematopoiesis begins in the AGM and then the FL, the numbers of Gata2 expressing cells also increase. At E10.5, the highest numbers of hematopoietic $CD31^{+}ckit^{+}$ cluster cells are found in the aorta, vitelline and umbilical arteries as quantitatively shown by whole mount embryo imaging (Yokomizo and Dzierzak 2010). Similarly, $CD31^{+}c\text{-}Kit^{+}Venus^{+}$ cell numbers are also highest in the vasculature at E10.5 and are decreased at E11.5. These quantitative data indicate that Gata2 is expressed in hematopoietic cluster cells but that not all cluster cells express Gata2 perhaps indicating that as Gata2 expression is downregulated hematopoietic cells mature and or differentiate in the clusters. Importantly, Gata2 is also expressed in the endothelial cells of the dorsal aorta. Already at E8.5, Gata2 is expressed in the endothelial cells lining the paired dorsal aorta and continues to be expressed in the E10.5 aorta when HSCs are generated, thus highlighting a possible role for Gata2 in directing the hemogenic program of endothelial cells.

Live imaging of *Gata2-Venus* embryos shows that Venus expression is highly dynamic in the aorta at the time of HSC generation. The turning on and turning off of *Gata2* expression, as well as the active movements of *Gata2* expressing cells in intraaortic clusters, suggests that Gata2 may also affect other processes at this developmental timepoint. For example, changes in the levels of Gata2 expression at the single cell level may influence hematopoietic function. The half life (1/2 life) of Gata2 protein still needs to be precisely calculated and compared with that of Venus protein. For this sensitive western blotting on the possible maximum number of Venus positive sorted cells will be attempted.

It is well known that Gata2 haploinsufficiency affects the quantitative ap pearance of HSCs in the AGM (Ling, Ottersbach et al. 2004). The dynamic expression of Gata2 in the aorta brings to mind recent findings of transcription factor oscillation in pluripotent stem cells (Glauche, Herberg et al. 2010; Wee, Yio et al. 2012). Fluctuations in transcription factor expression levels are thought to influence cell fate (Nutt, Metcalf et al. 2005). Further live imaging, the analysis of more cell surface markers (greater enrichment) and functional analyses are necessary to confirm a role for Gata2 and levels of Gata2 expression in the transition from hemogenic endothelial to a hematopoietic cell fate.

The advantage of the *Gata2-Venus* mouse model is in the direct isolation of Gata2 expressing cells. The functional characterization of the Venus positive cell fraction showed that it is enriched in hematopoietic progenitors and contains all long term repopulating HSCs. Interestingly, the Venus negative cell fraction always

contained a limited number of hematopoietic progenitors – a Gata2 independent subset of hematopoietic progenitors. We showed that the number of hematopoietic progenitors present in the Venus negative cell fraction of the E9 embryos is comparable to progenitor numbers in the *Gata2*$^{-/-}$ embryos, thus verifying the existence of Gata2 independent progenitor subset (fig 5 and tables 2 and 3). These progenitors are restricted in their differentiation potential predominantly to the macrophage lineage and may represent the source of primitive macrophages. In addition, *Gata2*$^{-/-}$ embryos show only a few hematopoietic c-kit^{+} cluster cells in the vitelline and umbilical arteries indicating that Gata2 might have an important role in the hemogenic endothelial cell compartment before or during the generation/emergence of hematopoietic cells. Gata2 independent progenitors may arise differently, perhaps directly from hemangioblasts. It is also possible that Gata2 is required for the proliferation and/or maintenance of hemogenic endothelial cells.

Finally, a more precise characterization of the Gata2 expressing cell fraction by multiple cell surface markers and combination of this reporter with other reporter mouse models will lead to a better understanding of the role Gata2 in the development and function of multipotential hematopoietic progenitors and HSCs. Preliminary results of immunostaining experiments of *Gata2-Venus* embryonic cells with hematopoietic and endothelial markers have identified a population highly enriched in multipotent hematopoietic progenitor activity (data not shown). Combination of *Gata2-Venus* mice with *Ly6A-GFP* reporter mice allowed us to identify a double positive cell population highly enriched in hematopoietic progenitors and especially multipotent progenitors (CFU-GEMM). This double positive cell fraction is a promising candidate for being highly enriched in HSC activity. Purification of HSCs and their direct precursors, the hemogenic endothelium, together with transcriptome and genetic analyses, will allow for an understanding of the role of Gata2 in embryonic hematopoiesis and the generation of HSCs.

Materials and Methods

Mice and embryos

For the generation of *Gata2-Venus* mouse model IB10 G2V ES cells (129SV background) targeted with *Gata2-Venus* targeting construct, showing correct construct integration and karyotype, were used for blastocyst injections into C57BL/6 mice. The produced chimeric mice were crossed again with c57bl6 mice and the ones with germline transmition were used for the used for the production of the f1 of the *Gata2-Venus* mice. The positive mice for the *Gata2-Venus* construct were backcrossed with CAG-Cre mice for the excision of the puromycin selection cassette. Correct selection cassette excision was tested by PCR detecting the presence of a *LoxP* site in the *Gata2* genomic locus (details in chapter 2).

Gata2$^{-/-}$ and *Ly6A-GFP* mice are kept in a C57BL/6 background. Timed matings were performed with *Gata2-Venus* males and *Ly6A-GFP* or WT (c57bl6) females and *Gata2*$^{+/-}$ males with *Gata2*$^{+/-}$ females. The day of the plug detection was scored as E0. Mice are housed according to institutional guidelines with free access to food and water.

qRT-PCR

The sequences of the primers used for the qRT-PCR are presented here: ATCTTCTTCAAGGACGACGG *Venus* Forward (FRW) primer, GGCTGTTGTAGTTGTACTCC *Venus* Reverse (RV) primer, CACCCCTAAGCAGAGAAGCAA *Gata2* FRW, TGGCACCACAGTTGACACACT *Gata2* RV, *β-actin*. The PCR program of the reaction was: 95^{o}C 3' min, followed by 40 cycles of 95^{o}C 10'' sec, 60^{o}C 30'' sec, 72^{o}C 30'' sec. The relative mRNA expression was normalized according to β-actin expression.

Whole mount- immunostainings

Preparation and immunostaining of mouse embryos was performed as described by (Yokomizo and Dzierzak 2010; Yokomizo, Yamada-Inagawa et al. 2012). Rat anti mouse primary antibodies for c-Kit (2B8) and biotinylated anti–CD31 (MEC13.3) from BD Biosciences. Anti GFP (MBL 598) from MBL. Secondary antibodies used: goat anti-rat IgG antibody-Alexa Flour 647, anti-rabbit IgG antibody Alexa Flour 488 from Invitrogen and streptavidin Cy3 (Jackson Immunoresearch).

In vivo live imaging

E10 embryos were dissected in in phosphate-buffered saline (PBS)/10% fetal calf serum (FCS)/1% penicillin/streptomycin solution. The somite pairs of the E10 embryos were ranging between 30-34 somite pairs when the embryos were taken out of the uterus. Embryos were dissected, immunostained, cut in to slices, imaged by confocal microscope [Leica SP5 AOBS DMI600 inverse microscope (Leica Microsystems)] and the collected videos were processed as described by (Boisset,

van Cappellen et al. 2010; Boisset, Andrieu-Soler et al. 2011).

Immunohistochemical staining

Gata2-Venus embryos were fixed in 2% paraformaldehyde/PBS for 30' in room temperature (RT), equilibrated with 20% sucrose overnight at 4°C and quick frozen in Tissue Tek. Frozen samples were kept at -80°C. Cryosections of transverse sections were treated in 100% acetone for 10'min, washed in PBS, blocked in 5% goat serum in PBS, washed (3x 5'min) in PBS and stained with anti-GPF antibody (Rb pAb to GFP ab290-50 ab-cam) at RT for 1-2 hours (h), washed in PBS and stained with goat anti-rabbit Alexa-488 (Goat a Rabbit IgG Alexa Fluor 488 Invitrogen). Subsequent washing in PBS and staining against CD34-biotinilated (cl RAM34 BD-Biosciences PharMigen) for 1h RT was followed by PBS washing step and streptavidin-Cy5 (cat#016-170-084 Jackson Immunoresearch), PBS washings and mounting with vactashield with DAPI (Vector Labs). The sections were observed either in a confocal microscope (Leica SP5 AOBS DMI600 inverse microscope (Leica Microsystems)) or in a fluorescent microscope (Zeiss AxioImager Z1 upright microscope).

Confocal microscopy

Whole mount immunostained embryos were imaged in a confocal microscope [Leica SP5 AOBS DMI600 inverse microscope (Leica Microsystems)]. Lasers used: 488-, 561-, 633- nm wavelengths. Images were obtained with 20x objection lense, pinhole was set at 1 Airy unit and the z steps were 0.8-2.3 μm as previously described by (Yokomizo and Dzierzak 2010; Yokomizo, Yamada-Inagawa et al. 2012).

Cell preparation for flow cytometry

Embryo tissues from E9, E10 and E11 were dissected in phosphate-buffered saline (PBS)/10% fetal calf serum (FCS)/1% penicillin/streptomycin solution and made into single cell suspensions by collagenase (0.125% in PBS/10% FCS/1% penicillin/streptomycin) treatment (45'min 37°C). The antibodies used for the FACS analysis were CD31 PE (MEC13.3) and c-Kit APC (2B8) both from BD Pharmigen for better combination with Venus fluorochrome. Viability selection was performed with Hoechst 33258 (Invitrogen) staining. FACS analysis was performed in a FACS Aria II SORP and data analyzed by FlowJo (Tri Star).

Methycellulose hematopoietic progenitor assay

In vitro clonogenic analysis was performed by plating sorted cell populations in methylcellulose medium (M3434, StemCell Technologies) and incubate in 37°C, 5% CO_2. The number and type of colonies were scored in an inverted microscope 10-12 days after plating.

Acknowledgements

We thank Alex Maas and John Kong a San for performing blastocyst injections, Henk Dronk and Experimental Dieren Centrum (Erasmus MC) for mouse care. We thank Dr Dies Meijer, Martine Jaegle and Siska Driegen for help and input in molecular biology work. This work was supported by ZonMW (Dutch Medical Research Council (911-09-036), FES NIRM (Dutch Innovation Grant) and NIH (RO37 DK54077).

References

Boisset, J. C., C. Andrieu-Soler, et al. (2011). "Ex vivo time-lapse confocal imaging of the mouse embryo aorta." Nat Protoc 6(11): 1792-1805.

Boisset, J. C., W. van Cappellen, et al. (2010). "In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium." Nature 464(7285): 116-120.

Cumano, A., F. Dieterlen-Lievre, et al. (1996). "Lymphoid potential, probed before circulation in mouse, is restricted to caudal intraembryonic splanchnopleura." Cell 86(6): 907-916.

de Bruijn, M. F., X. Ma, et al. (2002). "Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta." Immunity 16(5): 673-683.

de Bruijn, M. F., N. A. Speck, et al. (2000). "Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo." EMBO J 19(11): 2465-2474.

Dzierzak, E. and M. de Bruijn (2002). "Isolation and analysis of hematopoietic stem cells from mouse embryos." Methods Mol Med 63: 1-14.

Dzierzak, E. and N. A. Speck (2008). "Of lineage and legacy: the development of mammalian hematopoietic stem cells." Nat Immunol 9(2): 129-136.

Gekas, C., F. Dieterlen-Lievre, et al. (2005). "The placenta is a niche for hematopoietic stem cells." Dev Cell 8(3): 365-375.

Glauche, I., M. Herberg, et al. (2010). "Nanog variability and pluripotency regulation of embryonic stem cells--insights from a mathematical model analysis." PLoS One 5(6): e11238.

Godin, I., F. Dieterlen-Lievre, et al. (1995). "Emergence of multipotent hemopoietic cells in the yolk sac and paraaortic splanchnopleura in mouse embryos, beginning at 8.5 days postcoitus." Proc Natl Acad Sci U S A 92(3): 773-777.

Heyworth, C., K. Gale, et al. (1999). "A GATA-2/estrogen receptor chimera functions as a ligand-dependent negative regulator of self-renewal." Genes Dev 13(14): 1847-1860.

Kim, I., S. He, et al. (2006). "Enhanced purification of fetal liver hematopoietic stem cells using SLAM family receptors." Blood 108(2): 737-744.

Kumaravelu, P., L. Hook, et al. (2002). "Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver." Development 129(21): 4891-4899.

Ling, K. W., K. Ottersbach, et al. (2004). "GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells." J Exp Med 200(7): 871-882.

Ma, X., C. Robin, et al. (2002). "The Ly-6A (Sca-1) GFP transgene is expressed in all adult mouse hematopoietic stem cells." Stem Cells 20(6): 514-521.

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

Minegishi, N., N. Suzuki, et al. (2003). "Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos." Blood 102(3): 896-905.

Muller, A. M., A. Medvinsky, et al. (1994). "Development of hematopoietic stem cell activity in the mouse embryo." Immunity 1(4): 291-301.

Nardelli, J., D. Thiesson, et al. (1999). "Expression and genetic interaction of transcription factors GATA-2 and GATA-3 during development of the mouse central nervous system." Dev Biol 210(2): 305-321.

Nutt, S. L., D. Metcalf, et al. (2005). "Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors." J Exp Med 201(2): 221-231.

References

Boisset, J. C., C. Andrieu-Soler, et al. (2011). "Ex vivo time-lapse confocal imaging of the mouse embryo aorta." Nat Protoc 6(11): 1792-1805.

Boisset, J. C., W. van Cappellen, et al. (2010). "In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium." Nature 464(7285): 116-120.

Cumano, A., F. Dieterlen-Lievre, et al. (1996). "Lymphoid potential, probed before circulation in mouse, is restricted to caudal intraembryonic splanchnopleura." Cell 86(6): 907-916.

de Bruijn, M. F., X. Ma, et al. (2002). "Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta." Immunity 16(5): 673-683.

de Bruijn, M. F., N. A. Speck, et al. (2000). "Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo." EMBO J 19(11): 2465-2474.

Dzierzak, E. and M. de Bruijn (2002). "Isolation and analysis of hematopoietic stem cells from mouse embryos." Methods Mol Med 63: 1-14.

Dzierzak, E. and N. A. Speck (2008). "Of lineage and legacy: the development of mammalian hematopoietic stem cells." Nat Immunol 9(2): 129-136.

Gekas, C., F. Dieterlen-Lievre, et al. (2005). "The placenta is a niche for hematopoietic stem cells." Dev Cell 8(3): 365-375.

Glauche, I., M. Herberg, et al. (2010). "Nanog variability and pluripotency regulation of embryonic stem cells--insights from a mathematical model analysis." PLoS One 5(6): e11238.

Godin, I., F. Dieterlen-Lievre, et al. (1995). "Emergence of multipotent hemopoietic cells in the yolk sac and paraaortic splanchnopleura in mouse embryos, beginning at 8.5 days postcoitus." Proc Natl Acad Sci U S A 92(3): 773-777.

Heyworth, C., K. Gale, et al. (1999). "A GATA-2/estrogen receptor chimera functions as a ligand-dependent negative regulator of self-renewal." Genes Dev 13(14): 1847-1860.

Kim, I., S. He, et al. (2006). "Enhanced purification of fetal liver hematopoietic stem cells using SLAM family receptors." Blood 108(2): 737-744.

Kumaravelu, P., L. Hook, et al. (2002). "Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver." Development 129(21): 4891-4899.

Ling, K. W., K. Ottersbach, et al. (2004). "GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells." J Exp Med 200(7): 871-882.

Ma, X., C. Robin, et al. (2002). "The Ly-6A (Sca-1) GFP transgene is expressed in all adult mouse hematopoietic stem cells." Stem Cells 20(6): 514-521.

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

Minegishi, N., N. Suzuki, et al. (2003). "Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos." Blood 102(3): 896-905.

Muller, A. M., A. Medvinsky, et al. (1994). "Development of hematopoietic stem cell activity in the mouse embryo." Immunity 1(4): 291-301.

Nardelli, J., D. Thiesson, et al. (1999). "Expression and genetic interaction of transcription factors GATA-2 and GATA-3 during development of the mouse central nervous system." Dev Biol 210(2): 305-321.

Nutt, S. L., D. Metcalf, et al. (2005). "Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors." J Exp Med 201(2): 221-231.

Orlic, D., S. Anderson, et al. (1995). "Pluripotent hematopoietic stem cells contain high levels of mRNA for c-kit, GATA-2, p45 NF-E2, and c-myb and low levels or no mRNA for c-fms and the receptors for granulocyte colony-stimulating factor and interleukins 5 and 7." Proc Natl Acad Sci U S A 92(10): 4601-4605.

Ottersbach, K. and E. Dzierzak (2005). "The murine placenta contains hematopoietic stem cells within the vascular labyrinth region." Dev Cell 8(3): 377-387.

Palis, J., S. Robertson, et al. (1999). "Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse." Development 126(22): 5073-5084.

Persons, D. A., J. A. Allay, et al. (1999). "Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis." Blood 93(2): 488-499.

Pimanda, J. E., K. Ottersbach, et al. (2007).

"Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development." Proc Natl Acad Sci U S A 104(45): 17692-17697.
Robert-Moreno, A., L. Espinosa, et al. (2005). "RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells." Development 132(5): 1117-1126.
Tsai, F. Y., G. Keller, et al. (1994). "An early haematopoietic defect in mice lacking the transcription factor GATA-2." Nature 371(6494): 221-226.
Tsai, F. Y. and S. H. Orkin (1997). "Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation." Blood 89(10): 3636-3643.
Wee, K. B., W. K. Yio, et al. (2012). "Transcription factor oscillations induce differential gene expressions." Biophys J 102(11): 2413-2423.
Wilson, N. K., F. J. Calero-Nieto, et al. (2011). "Transcriptional regulation of haematopoietic transcription factors." Stem Cell Res Ther 2(1): 6.
Wilson, N. K., S. D. Foster, et al. (2010). "Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators." Cell Stem Cell 7(4): 532-544.
Yokomizo, T. and E. Dzierzak (2010). "Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos." Development 137(21): 3651-3661.
Yokomizo, T., T. Yamada-Inagawa, et al. (2012). "Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos." Nat Protoc 7(3): 421-431.

Chapter 4

Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators

Nicola K. Wilson[1], Samuel D. Foster[1], Xiaonan Wang[1], Kathy Knezevic[2], Judith Schütte[1], Polynikis Kaimakis[3], Paulina M. Chilarska[1], Sarah Kinston[1], Willem H. Ouwehand[1], Elaine Dzierzak[3], John E. Pimanda[2], Marella F.T.R. de Bruijn[4], and Berthold Göttgens[1]*

1University of Cambridge Department of Haematology, Cambridge Institute for Medical Research, Hills Road, Cambridge, CB2 0XY, UK
2The Lowy Cancer Research Centre and the Prince of Wales Clinical School, University of New South Wales, Sydney, NSW 2052, Australia
3Erasmus Medical Center, Department of Cell Biology and Erasmus Stem Cell Institute, 3000 CA Rotterdam, the Netherlands
4MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, OX3 9DS, UK
*Correspondence: bg200@cam.ac.uk
DOI 10.1016/j.stem.2010.07.016

Cell Stem Cell

Cell Stem Cell
Resource

Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators

Nicola K. Wilson,[1] Samuel D. Foster,[1] Xiaonan Wang,[1] Kathy Knezevic,[2] Judith Schütte,[1] Polynikis Kaimakis,[3] Paulina M. Chilarska,[1] Sarah Kinston,[1] Willem H. Ouwehand,[1] Elaine Dzierzak,[3] John E. Pimanda,[2] Marella F.T.R. de Bruijn,[4] and Berthold Göttgens[1,*]

[1]University of Cambridge Department of Haematology, Cambridge Institute for Medical Research, Hills Road, Cambridge, CB2 0XY, UK
[2]The Lowy Cancer Research Centre and the Prince of Wales Clinical School, University of New South Wales, Sydney, NSW 2052, Australia
[3]Erasmus Medical Center, Department of Cell Biology and Erasmus Stem Cell Institute, 3000 CA Rotterdam, the Netherlands
[4]MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, OX3 9DS, UK
*Correspondence: bg200@cam.ac.uk
DOI 10.1016/j.stem.2010.07.016

SUMMARY

Combinatorial transcription factor (TF) interactions control cellular phenotypes and, therefore, underpin stem cell formation, maintenance, and differentiation. Here, we report the genome-wide binding patterns and combinatorial interactions for ten key regulators of blood stem/progenitor cells (SCL/TAL1, LYL1, LMO2, GATA2, RUNX1, MEIS1, PU.1, ERG, FLI-1, and GFI1B), thus providing the most comprehensive TF data set for any adult stem/progenitor cell type to date. Genome-wide computational analysis of complex binding patterns, followed by functional validation, revealed the following: first, a previously unrecognized combinatorial interaction between a heptad of TFs (SCL, LYL1, LMO2, GATA2, RUNX1, ERG, and FLI-1). Second, we implicate direct protein-protein interactions between four key regulators (RUNX1, GATA2, SCL, and ERG) in stabilizing complex binding to DNA. Third, *Runx1*$^{+/-}$::*Gata2*$^{+/-}$ compound heterozygous mice are not viable with severe hematopoietic defects at midgestation. Taken together, this study demonstrates the power of genome-wide analysis in generating novel functional insights into the transcriptional control of stem and progenitor cells.

INTRODUCTION

Combinatorial interactions of transcription factors are key determinants of tissue identity. Early recognition of this came from overexpression experiments where, for example, the myogenic family of transcription factors (TF) could induce muscle phenotype in non-muscle cells (Davis et al., 1987) or combinations of hematopoietic TFs reprogram nonhemogenic mesoderm into blood during development (Gering et al., 2003). The particular power of combinatorial TF interactions has been highlighted more recently by the derivation of induced pluripotent cells through expression of pluripotency-associated TFs in differentiated cells (Takahashi and Yamanaka, 2006), as well as the transdifferentiation of pancreatic exocrine cells into insulin-secreting β-cells (Zhou et al., 2008) and the direct conversion of fibroblasts to functional neurons (Vierbuchen et al., 2010) following exogenous expression of specific trios of transcriptional regulators.

The requirement for specific TF combinations is likely to reflect combinatorial interactions of those TFs when binding to regulatory elements of their target genes. In line with this hypothesis, it has been shown that the pluripotency factors OCT4, SOX2, and NANOG co-occupy regulatory elements of many gene loci in embryonic stem cells (ESCs) (Boyer et al., 2005). Systematic analysis of combinatorial transcriptional control mechanisms will require genome-wide information on combinatorial transcription factor binding. Significant progress in this regard has been made to further our understanding of ESCs, where genome-wide occupancy patterns have been established for more than a dozen TFs (Chen et al., 2008). Recent sophisticated computational analysis of these comprehensive data sets is beginning to reveal mechanistic insights into the transcriptional programs operating in ESCs (Gao et al., 2009). However, with the exception of ESCs, no comprehensive TF binding data sets have been reported for any other mammalian stem/progenitor cell types.

Hematopoiesis has long served as a model system for studying transcriptional control of cell-fate choice and subsequent differentiation. Single or pairs of transcription factors can specify differentiation of progenitor cells into distinct mature lineages (Laslo et al., 2006; McNagny et al., 1998; Nerlov and Graf, 1998; Nerlov et al., 1998; Spooner et al., 2009; Ye and Graf, 2007), and detailed functional analysis of individual regulatory elements controlled by these factors has led to the identification of specific regulatory codes such as the Gata/E-box motif in red blood cells or the Ets/Ets/Gata motif in stem and progenitor cells (Donaldson et al., 2005; Pimanda et al., 2007b; Wadman et al., 1997). However, because identification of these regulatory codes was largely based on detailed functional studies of specific regulatory elements, it is not clear how widely relevant they are across the whole genome nor how they would interface with other transcription factor complexes.

The advent of ChIP-Seq technology coupling chromatin immunoprecipitation with high-throughput sequencing has

made the generation of genome-wide maps of transcription factor binding a realistic undertaking. Especially when coupled with extensive validation and sophisticated bioinformatic analysis, these new approaches can provide novel hypotheses on transcriptional control mechanisms. For example, our recent ChIP-Seq analysis of the SCL/TAL1 transcription factor in the multipotent hematopoietic progenitor cell line 7 (HPC-7) followed by extensive validation in transgenic mice established a suitable model system for global analysis of transcriptional programs in early progenitor cells by providing in vivo validated links between SCL and 15 other hematopoietic TFs. Similarly, a recent analysis of a model cell line for terminal erythroid differentiation provided a genome-wide view of SCL interplay with the transcriptional regulators LDB1, GATA1, and ETO2/MTGR1 in a mature hematopoietic lineage (Soler et al., 2010).

Here, we report genome-wide binding profiles for ten key hematopoietic transcriptional regulators (SCL/TAL1, LYL1, LMO2, GATA2, RUNX1, MEIS1, PU.1, ERG, FLI-1, and GFI1B) in hematopoietic progenitor cells. Integrated analysis of these data sets provides a comprehensive genome-wide view of combinatorial interactions between these ten key regulators. In particular, multiple independent lines of evidence demonstrate close collaboration between seven hematopoietic stem/progenitor cell (HSPC)-associated TFs, including functional links between RUNX1 and other key HSPC regulators. Subsequent in vitro and in vivo validation confirmed a previously unrecognized synergism between the three central regulators of blood stem cell development, SCL, GATA2, and RUNX1, thus demonstrating how comprehensive analysis of genome-wide data sets can provide novel insights into transcriptional control of stem/progenitor cells.

RESULTS

Key Hematopoietic Transcription Factors Differ in the Number of Their Genomic Targets

We recently reported genomic targets of the SCL transcription factor in the multipotential model hematopoietic progenitor cell line HPC-7 (Wilson et al., 2009). Extensive in vivo validation of SCL-bound enhancers in transgenic mice validated HPC-7 as an excellent model system for early hematopoietic stem/progenitor cells. However, because of the absence of a negative control ChIP-Seq data set, we had limited peak identification to those with the lowest false discovery rate, which resulted in 228 high confidence peaks. To address this limitation, we have now generated a new Scl ChIP-Seq data set (15,350,104 uniquely mappable reads) together with a control IgG ChIP sample (9,012,777 uniquely mappable reads), which allowed us to analyze our Scl ChIP-Seq data using software tools specifically designed to distinguish artifactual peaks found in both test and control samples from true binding events (see Experimental Procedures). This resulted in the identification of 7096 statistically significant peaks (see Experimental Procedures). Importantly, analysis using the new negative control sample not only retained all except ten of the original peaks, but also allowed us to identify more than 6,800 new genomic locations bound by SCL in hematopoietic cells (Table S1).

We had previously shown that Gata, Ets, Runx, Gfi, and Meis consensus binding motifs were highly overrepresented in SCL-bound regions (Wilson et al., 2009), suggesting cooperation between Scl and other major hematopoietic regulators. To directly investigate potential combinatorial interactions at a genome-wide scale, we have now performed ChIP-Seq assays for an additional nine transcriptional regulators of HSCs that would be expected to bind to these motifs. LYL1, LMO2, and GATA2 would be expected to bind to GATA-E-box sites; RUNX1 to the Runx motifs; MEIS1 to the Meis sites; PU.1, ERG, and FLI-1 to the Ets sites; and GFI1b to the Gfi consensus (the latter was chosen because even though both GFI1 and GFI1b are expressed in HSPC, GFI1b shows higher expression levels in HPC-7 cells). Peaks of transcription factor binding were identified using the negative control IgG sample as before yielding 4350 LYL1, 9604 LMO2, 9234 GATA2, 5258 RUNX1, 8386 MEIS1, 22720 PU.1, 19601 FLI-1, 36137 ERG, and 8853 GFI1B (for peak coordinates, see Table S1). To characterize the distribution of binding events across genomic features, we determined the number of peaks within ± 1 kb from the transcription start sites (promoter peaks) and partitioned the remaining peaks into those within and between genes (intragenic and intergenic, respectively; see Figure 1). For LYL1, LMO2, GATA2, and

Factor	Peaks	Distribution of Peaks			
		Promoter	Intragenic	Intergenic	
Scl/Tal1	7096	2021	2497	2578	
Lyl1	4350	69	1941	2340	
Lmo2	9604	263	4374	4967	
Gata2	9234	271	4191	4772	
Runx1	5258	273	2354	2642	
Meis1	8386	969	3701	3731	
Pu.1	22720	5165	9139	8439	
Erg	36167	8646	14049	13472	
Fli-1	19601	2019	8567	9015	
Gfi1b	8853	2401	3296	3156	

Figure 1. Key Hematopoietic Transcription Factors Differ in the Number of Their Genomic Targets

The number of peaks for each transcription factor was determined as outlined in Experimental Procedures. Each peak was then allocated to be either within a promoter, intragenic, or intergenic region. The pie chart visually shows the distribution of the peaks across those three categories (blue, green, and red, respectively). For lists of peaks bound by 10 TFs and genes associated with those peaks, see Tables S1 and S2.

RUNX1, very few peaks were observed over promoter regions, with the bulk of binding events more or less evenly distributed between intra- and intergenic sites. SCL, MEIS1, PU.1, ERG, FLI-1, and GFI1B showed binding to a larger number of promoter regions, and again, nonpromoter peaks were evenly distributed between intra- and intergenic regions. To facilitate use of this very large data set for the wider community, we generated comprehensive supplemental data, all of which can be downloaded from our website http://hscl.cimr.cam.ac.uk/ChIP-Seq_Data/ChIP-Seq.html. These include the following files: (1) custom track (.wig) files to display ChIP-Seq profiles for each of the 10 factors in genome browsers; (2) a table listing the peak positions for all ten TFs (Table S1); and (3) a table listing all UCSC genes indicating which (if any) of the ten factors bind to their respective gene loci (within the body of the gene and up to 50 kb 5′ and 3′ intergenic flanking sequences, Table S2). This new ten factor ChIP-Seq data set, therefore, provides a uniquely powerful resource for the future characterization of target genes for these ten major hematopoietic regulators as well as the characterization of regulatory networks in HSPCs. For the remainder of this manuscript, we illustrate how this new data set can provide insights into combinatorial transcriptional control mechanisms in HSPCs.

Genome-wide Analysis Identifies Preferential Transcription Factor Binding Pairs

Having generated genome-wide maps for ten important regulators of HSPCs, we reasoned that comprehensive bioinformatic analysis of this unique data set might allow us to obtain new insights into combinatorial transcription factor interactions. Visual inspection of individual gene loci demonstrated a wide variety of combinatorial binding patterns (Figure 2). For example, the +1 HSPC enhancer of the Hhex gene (Donaldson et al., 2005) was bound by all ten TFs (Figure 2A), whereas multiple peaks bound by different combinations of TFs were seen in the *Zfpm1*/*Fog1* gene locus (Figure 2B).

Given the large scale of genome-wide data sets, inspection of specific gene loci is completely impractical for discovering globally important combinatorial TF patterns. In order to identify such control mechanisms, we first asked whether the frequency of colocalization of transcription factor pairs on the same target regions was greater than would be expected by chance, thus indicating preferential coregulatory activities. To this end, we determined the number of overlapping peaks for all 45 possible pair-wise combinations (Figure 3). This analysis identified overlapping binding events for all pair-wise combinations but did not address whether the observed overlaps were statistically significant. To get an estimate of the total available space for transcription factor binding in HPC-7 cells, we performed a ChIP-Seq experiment using an antibody against acetylated histone H3 (Histone H3 lysine 9 and 14), which allowed us to define 46598 regions of enriched histone acetylation. We then determined the total number of regions that were acetylated and/or bound by at least one of the ten factors (e.g., the union of acetylation and transcription factor peaks), which resulted in a total of ~80,000 unique peaks. This number is similar to the ~100,000 DNase I sites reported in human CD4 T cells and lymphoblastoid cells (Boyle et al., 2008; McDaniell et al., 2010). 100,000 was, therefore, used as an estimate of the number of potentially available regions for binding for the ten TFs. Using the observed numbers of peaks for each of the ten factors, we performed simulations (1000 iterations) to determine how many regions would be expected to be co-occupied by chance for any pair of factors. This bootstrapping exercise demonstrated that some overlaps were much more significant than others and that all pairs except those involving PU.1 occur more often than expected by chance. Many of the most significant overlaps were observed for pairs involving known partners, such as SCL/LMO2/LYL1/GATA2, or closely related family members, such as Erg/Fli-1. Interestingly, pairs of factors not previously thought to collaborate showed equally high significance, particularly those involving RUNX1 with either SCL, LYL1, LMO2, or GATA2 (Figure 3 and Figure S2).

SCL/LYL1/LMO2/GATA2 are known to bind as components of multiprotein complexes to regulatory elements (Anguita et al., 2004; Wadman et al., 1997), but our analysis suggests that there may also be close collaboration between factors not previously known to operate together, such as SCL/RUNX1, GATA2/RUNX1, LYL1/RUNX1, and also RUNX1/LMO2. To further investigate the likelihood of cooperative TF binding, we determined the relative distances for all pair-wise binding events taking advantage of the fact that ChIP-Seq peak calling algorithms are highly accurate at identifying peak summits (Zhang et al., 2008). This analysis demonstrated that the median distance for all 45 pairs was variable and ranged from 22 bp (LYL1 ↔ LMO2) to 85 bp (PU.1 ↔ GFI1b). As different members of the same TF family (e.g., PU.1/FLI-1/ERG and SCL/LYL1) would be expected to often bind to the same site within a given population of cells, we grouped all ''within family'' comparisons and then ranked all ''across family'' comparisons based on their median distance (Figure 3B). This analysis demonstrated that median distances between peak centers for pair wise combinations that involved PU.1, Gfi1b, and Meis1 were larger than those for ''within family'' comparisons, whereas all other combinations were within the same range (Figure 3B). Reassuringly, known partners were identified within this analysis, such as SCL/LMO2, GATA2/LMO2, and GATA2/ERG or GATA2/FLI-1. However, previously unknown potentially collaborative binding was also identified between SCL/RUNX1 or LYL1/RUNX1, which together highlighted a potential new HSPC regulatory complex composed of SCL, LYL1, LMO2, GATA2, RUNX1, ERG, and FLI-1.

Characterization of Wider Combinatorial Interactions among HSPC Transcription Factors

Having identified potential interplay between seven factors from pairwise analysis, we next interrogated whether an examination of wider combinatorial interactions would corroborate these initial observations. To analyze combinatorial patterns in an unbiased fashion, we performed principal component analysis (PCA) and hierarchical clustering (see Experimental Procedures) of all regions bound by five or more factors (Figure 4). Analysis of the ten factor multivariate data set by PCA revealed close correlation between GATA2/SCL/LYL1 and ERG/FLI-1 occupancy patterns with RUNX1 and LMO2 situated near these two clusters. PU.1, GFI1B, and MEIS1 binding patterns, on the other hand, were all found to be much more separated from the other seven factors.

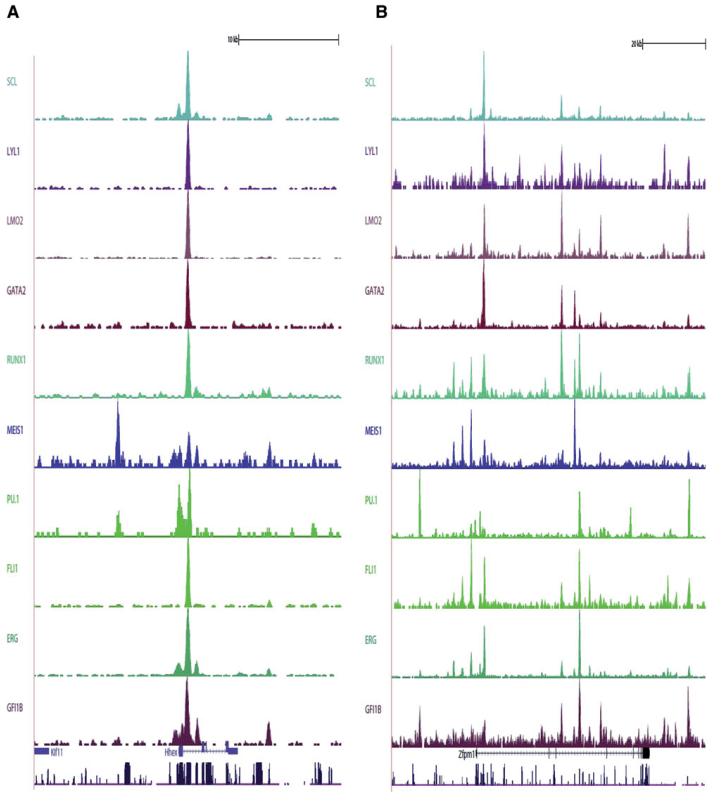

Figure 2. Specific Gene Loci Exhibit a Wide Variety of Combinatorial Binding Patterns

Raw ChIP-Seq read data were transformed into a density plot for each TF and loaded into the UCSC genome browser as custom tracks above the UCSC tracks for gene structure and mammalian homology. Visual inspection of gene loci showed many different combinations of binding for the ten different transcription factors. Displayed here are two loci that have been tested previously in transgenic mice (Donaldson et al., 2005; Wilson et al., 2009). (A) Hhex. (B) *Zfpm1/Fog1*. Screenshots for ten additional gene loci are provided as Figure S2. Real-time PCR validation of ChIP material in HPC-7 and E11.5 FL can be found in Figure S2.

To corroborate these findings using an independent mathematical approach, we also analyzed occupancy patterns using hierarchical clustering. Interestingly, this analysis also separated the ten factors into two major clusters separating the TFs in the same manner as analysis of peak summit distances from the pair wise comparisons and PCA (e.g., MEIS1/GFI1B/PU.1 versus LYL1/GATA2/RUNX1/SCL/LMO2/FLI-1/ERG). Taken together, therefore, our comprehensive bioinformatic analysis suggests that close combinatorial interactions among seven regulators (SCL, LYL1, LMO2, GATA2, RUNX1, FLI-1, and ERG) play an important role in HSPC transcriptional control. Although all of these TFs are known to be important individually for HSPCs, our unbiased bioinformatic analysis implicates all seven factors together as being extensively involved in cooperative transcriptional control.

Genome-wide Analysis of Coordinated Binding of LYL1, SCL, LMO2, GATA2, RUNX1, FLI1, and ERG

Statistically significant pairwise co-occupancy, close proximity of binding, and combinatorial clustering all highlighted a heptad of transcription factors important for HSPCs (SCL, LYL1, LMO2, ERG, FLI-1, GATA2, and RUNX1). We, therefore, investigated co-occupancy involving these seven factors further. There were 1015 regions where all seven factors were bound within 200 bp or less. De novo motif discovery on this set of sequences recovered the consensus binding motifs for all seven factors, as well as one additional overrepresented motif. We were unable to identify a cognate binding factor for this latter motif, but a similar motif has previously been recovered from a purely computational genome-wide analysis of conserved noncoding sequences (Xie et al., 2007) (Figures 5A and B). Recovery of all

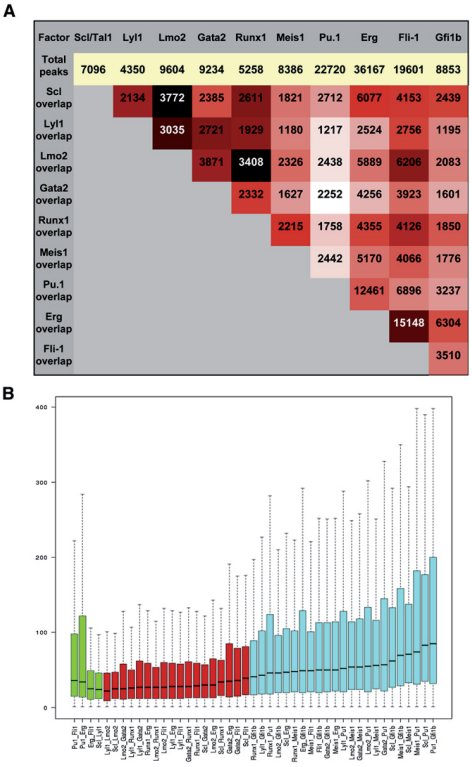

Figure 3. Genome-wide Analysis Identifies Preferential Transcription Factor Binding Pairs

(A) Shown at the top is the total number of peaks for each factor followed by all 45 possible pairwise combinations (number of overlapping peaks). To obtain a measure for the significance of pairwise overlaps, bootstrapping was performed (see Experimental Procedures) to calculate Z scores with increased color intensity of shading indicating increasingly higher Z scores (a full color coded legend and actual Z scores can be found in Figure S2).

(B) Box plots showing the distances (in bp on the y axis) between peak summits for all 45 lists of pairwise overlaps from (A). Pairwise combinations of the 10 TFs were separated into three groups: within TF families (shaded in green), across families with median distance equal to or below that of "within family" pairs (shaded in red), and across families with median distance greater than for the "within family" comparisons (shaded in blue).

expected consensus motifs underlines the high quality of the data sets produced in the current study. Moreover, identification of only a single additional motif suggests that the set of factors analyzed in this study provides most of the regulatory context of regions bound by the SCL/LYL1/GATA2/RUNX1/LMO2/FLI-1/ERG heptad.

While de novo motif discovery identifies statistically overrepresented motifs within a sequence collection, it does not directly address the question of motif presence/absence or the presence of motif combinations in individual sequences. To address this issue, we mapped Ets, GATA, E-box, and Runx consensus sequences across all regions bound by the seven factors and performed hierarchical clustering to recover potential combinatorial patterns. This analysis demonstrated that the vast majority of heptad-bound peaks contained GATA and Ets sites (Figure 5C). By contrast, only 76% contained an E-box and even less (39%) contained a Runx consensus motif, indicating potentially indirect binding of E-Box and RUNX transcription factors to a significant proportion of regulatory elements. Of note, each region contained at least one of the four motifs, and over 99% contained at least two. Taken together, this analysis demonstrates that expected motifs are present in heptad-bound regions but also suggests that RUNX1 and SCL/LYL1 may be recruited indirectly to many elements.

The SCL/LYL1/GATA2/RUNX1/LMO2/FLI-1/ERG Heptad Displays Previously Unrecognized Protein-Protein Interactions

Comprehensive analysis of our ten factor ChIP-Seq data highlighted a preference for SCL/LYL1/GATA2/LMO2/ERG/FLI-1/RUNX1 to bind in close proximity to each other to specific DNA sequences. This observation is suggestive of cooperative binding, especially in the light of the absence of bona fide Runx1 and Scl/Lyl1 binding motifs in a substantial number of regions. Such cooperative binding would likely be facilitated by extensive protein-protein interactions. Previous studies have reported interactions between GATA2/LMO2/LYL1/SCL and also GATA2/FLI-1/ERG, yet no interactions involving RUNX1 and other heptad members have been reported in HSPCs. To investigate protein interactions between RUNX1 and other complex constituents, we performed coimmunoprecipitation (Co-IP) experiments for RUNX1, SCL, ERG, and GATA2 (Figure 6A). Following transfection of myc-tagged Runx1 and flag-tagged Scl, lysates were immunoprecipitated using an anti-Flag antibody, and immune complexes were analyzed using an anti-Myc antibody to detect the presence of coimmunoprecipitated RUNX1. Protein complex formation between SCL and RUNX1 can clearly be seen in Figure 6Ai, which only shows the presence of RUNX1 when both SCL and RUNX1 have been expressed together. To further test whether other transcription factors bound within the heptad could act as further docking factors for RUNX1, additional Co-IP experiments were performed. A combination of myc-tagged RUNX1 and ERG or myc-tagged RUNX1 and GATA2 were transfected and immunoprecipitated. Immune complexes were then analyzed using either an anti-myc or an anti-ERG antibody (Figures 6Aii and 6Aiii, respectively). Prominent interactions between RUNX1 and GATA2/ERG were seen when the relevant TFs were coexpressed, consistent with the notion that protein-protein interactions facilitate binding of RUNX1 to those heptad-bound regions that lack Runx1 consensus binding motifs (Figure 6B).

GATA2/RUNX1 Compound Heterozygous Mice Are Not Viable with Severe Hematopoietic Defects at Midgestation

Constituents of the new HSPC heptad, such as RUNX1 or GATA2, are individually essential for the development of

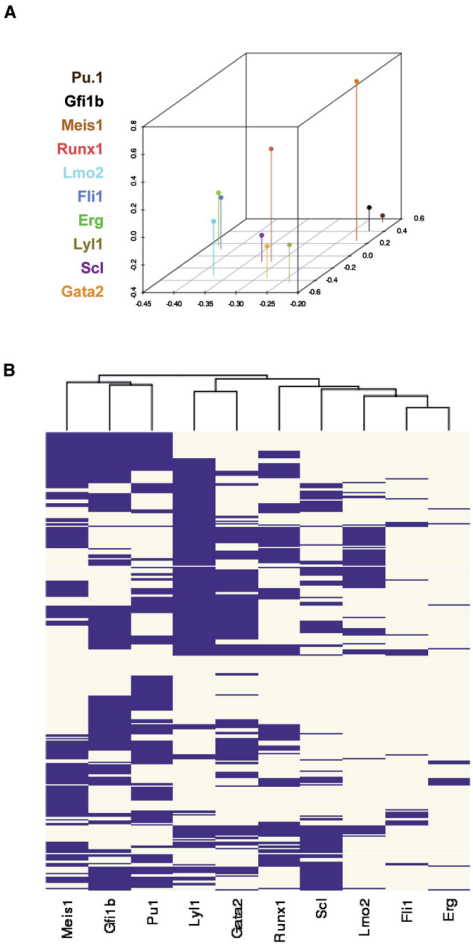

Figure 4. Characterization of Wider Combinatorial Interactions among HSPC Transcription Factors

(A) Principle component analysis (PCA) identifies preferential combinatorial TF associations. All genomic regions bound by five or more TFs were listed in a ten-column table with 0/1 indicating absence/presence of each of the 10 TFs. PCA was performed and displayed using the R software environment (http://www.r-project.org/). The plot shown displays the first three principle components on the x, y, and z axis, with each colored dot corresponding to the relative location of the TF indicated on the left.

(B) Hierarchical clustering of all genomic regions bound by 5 or more TFs based on factor occupancy patterns. The major partition of columns separates the 10 TFs into two main branches, with MEIS1/GFI1B/PU.1 to the left and SCL/LYL1/GATA2/RUNX1/LMO2/FLI-1/ERG to the right. Each line corresponds to an individual genomic region where blue/white coloring indicates the presence/absence of the given TF. Peak regions for individual TFs were set at a standard width of 400 bp, and after merging overlapping regions, the average size was 460 bp ± 260 bp. PCA and hierarchical clustering identify overlapping specific multifactor associations, such as the FLI1/ERG/LMO2 triplet or the LYL1/GATA2/RUNX1/SCL/LMO2/FLI-1/ERG heptad. (See Table S3 for list of heptad peaks).

definitive hematopoietic cells (Cai et al., 2000; Okuda et al., 1996; Tsai et al., 1994; Wang et al., 1996). However, heterozygous *Gata2*$^{+/-}$ and *Runx1*$^{+/-}$ mice are viable with only minor hematopoietic phenotypes (Cai et al., 2000; Ling et al., 2004; Rodrigues et al., 2005; Sun and Downing, 2004). Cooperative control of important common target genes would suggest that *Gata2*$^{+/-}$::*Runx1*$^{+/-}$ compound heterozygous mice might display a much more severe phenotype than animals heterozygous for a single gene only. To investigate potential consequences of *Gata2*/*Runx1* compound heterozygosity, *Gata2*$^{+/-}$ and *Runx1*$^{+/-}$ mice were intercrossed (Figure 6C). Interestingly, after numerous matings, only offspring which were heterozygous for either *Runx1* or *Gata2* were viable, which occurred at a normal expected Mendelian ratio. By contrast, no *Gata2*$^{+/-}$::*Runx1*$^{+/-}$ animals were found, even though one quarter of all offspring would be expected to be compound heterozygous ($p = 0.004$). These results, therefore, demonstrate that *Runx1*$^{+/-}$ and *Gata2*$^{+/-}$ function as classical synthetic lethal alleles, which are widely recognized as genetic evidence of buffering relationships between genes that function in common pathways.

To examine a possible hematopoietic phenotype of *Gata2*$^{+/-}$::*Runx1*$^{+/-}$ compound heterozygous mice, colony assays were performed on fetal livers (FL) from embryonic (E) day 12.5 to determine the total number of clonogenic progenitor cells (colony forming unit-culture [CFU-C]) per FL (Figure 6D). As reported previously, CFU-Cs were compromised in *Runx1*$^{+/-}$ and *Gata2*$^{+/-}$ fetal liver when compared with wild-type (data not shown). Interestingly, the number of CFU-Cs was reduced even further in fetal livers from compound heterozygous embryos (Figure 6D), which was paralleled by a significant reduction in overall cellularity (see Figure S3). Furthermore, gene expression analysis demonstrated the expected reductions in Gata2 and Runx1 levels for heterozygous samples and expression changes specific for the *Gata2*$^{+/-}$::*Runx1*$^{+/-}$ compound heterozygous samples for some GATA2/RUNX1 targets, such as *Meis1*, *Lyl1*, *Zfpm1/Fog1*, *Scl/Tal1*, *Gfi1b*, and *Tet2* (see Figure S3). Whether any of these expression changes are responsible for the phenotype of *Gata2*$^{+/-}$::*Runx1*$^{+/-}$ compound heterozygous embryos will be the subject of future investigations. Taken together, genetic analysis allowed us to identify interplay between two of the most powerful regulators of HSPC function and, therefore, demonstrate the power of comprehensive bioinformatic analysis of large genome-scale data sets for generating new biological insight.

Integrated Analysis of the Ten Factor ChIP-Seq Resource with External Data Generates New Hypotheses

The utility of a resource, such as the ten-factor ChIP-Seq data set described here, ultimately depends on how readily it can be accessed by other scientists to analyze their own data and gain novel insights. To provide a couple of vignettes of the types of analyses made possible, we present below integrated analysis

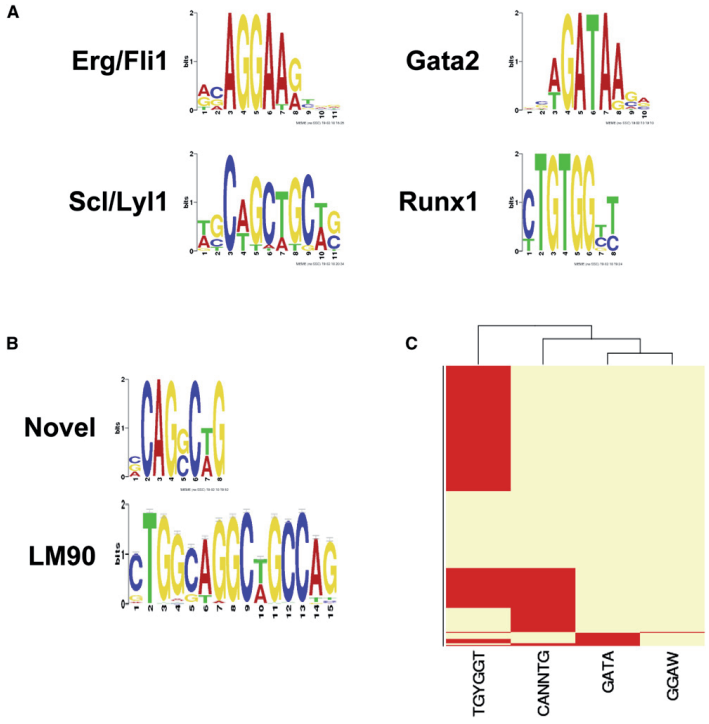

Figure 5. Motif Analysis Highlights the Core Regulatory Context of the Regions Bound by the SCL/LYL1/GATA2/RUNX1/LMO2/FLI-1/ERG Heptad

(A) De novo motif discovery performed on the set of regions bound by the SCL/LYL1/GATA2/RUNX1/LMO2/FLI-1/ERG heptad identified Ets, GATA, E-box, and Runx motifs precisely corresponding to the four motifs that would be expected for these seven factors.

(B) De novo motif discovery also highlighted one additional novel motif with no match to known TF binding motifs but significant similarity to a motif of unknown function (LM90) identified through genome-wide computational analysis of noncoding conserved sequences (Xie et al., 2007).

(C) Hierarchical clustering of all regions bound by the SCL/LYL1/GATA2/RUNX1/LMO2/FLI-1/ERG heptad based on the presence (white) or absence (red) of expected consensus DNA binding motifs. Consensus binding sites were mapped using TFBSSearch (Chapman et al., 2004). A significant proportion of regions did not contain Runx (TGYGGT) and/or E-box (CANNTG) consensus binding motifs, whereas most regions contain GATA and Ets (GGAW) sites. (See Table S4 for list of peaks and associated genes which do not contain either an E-box or a Runx motif).

of the ten-factor ChIP-Seq data with both external expression and single factor ChIP-Seq data. Given the wealth of high quality gene expression profiles generated by other labs, we first asked whether the 927 genes next to peaks bound by the heptad showed positive enrichment for HSC-specific expression. To this end, we performed gene set enrichment analysis (GSEA) by loading microarray expression data from the hematopoietic fingerprints database (Chambers et al., 2007) and the GNF Gene Atlas (Wu et al., 2009) into the standalone GSEA program (http://www.broadinstitute.org/gsea/) and then used the 927 genes next to heptad-bound peaks as the query gene set. The heptad gene set showed highly significant enrichment for HSC specific expression with both data sets (See Figure 7A), which was not seen with three randomly chosen control gene sets (Figure S4). Moreover, Gene Ontology analysis demonstrated that a substantial proportion of the 927 heptad targets are involved in transcriptional control, signaling, cell death, and cell cycle (Figure 7B). These results, therefore, suggest that the heptad target gene set (see http://hscl.cimr.cam.ac.uk/ChIP-Seq_Data/ChIP-Seq.html) represents a rich source of potentially new regulators of HSCs.

We next explored the value of integrating single-factor ChIP-Seq data with our ten-factor resource. Given that SCL/GATA complexes were originally discovered between SCL and the erythroid factor GATA1 (Osada et al., 1995), we asked whether there are any potential correlations between regions bound by GATA1 in erythroid cells and those bound by SCL and its ortholog LYL1, as well as GATA2, in HPC-7 cells (Figure 7C). A recent

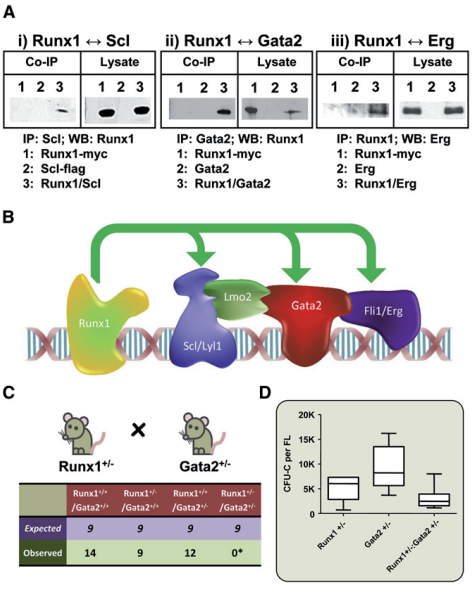

Figure 6. The SCL/LYL1/GATA2/RUNX1/LMO2/FLI-1/ERG Heptad Displays Previously Unrecognized Protein-Protein as Well as Genetic Interactions

(A) Identification of protein-protein interactions between RUNX1-SCL, RUNX1-GATA2 and RUNX1-ERG. Expression constructs were transfected into 293T cells and putative protein interactions assayed by coimmunoprecipitation/western blot analysis. (Ai) αFLAG immunoprecipitation (SCL), followed by western blot using αMYC-tag antibody (RUNX1), demonstrates coimmunoprecipitation of Runx1 by the αSCL (Flag) antibody. A western blot of lysates on the right is shown as expression control. (Aii) αGATA2 immunoprecipitation, followed by western blot using αRUNX1 (myc), demonstrates coimmunoprecipitation of Runx1 by the αGATA2 antibody. (Aiii) αMYC immunoprecitation (Runx1), followed by western blot using α-ERG, demonstrates coimmunoprecipitation of ERG by the αRUNX1 (myc) antibody.

(B) Schematic of heptad proteins bound to DNA. Previously known protein complexes are shown bound to DNA, and the newly identified links between RUNX1 and SCL/GATA2/ERG are indicated by green arrows. The order of proteins shown is for illustrative purposes rather than reflecting a particularly common arrangement of binding sites.

(C) Expected and actual numbers of progeny from the *Runx1*$^{+/-}$ × *Gata2*$^{+/-}$ crosses. The failure to obtain live *Runx1*$^{+/-}$::*Gata2*$^{+/-}$ compound heterozygote animals is highly significant ($p = 0.004$, χ^2 test).

(D) Colony assays on fetal livers from *Runx1*$^{+/-}$, *Gata2*$^{+/-}$ and *Runx1*$^{+/-}$::*Gata2*$^{+/-}$ E12.5 embryos. The data shown correspond to three, four, and six livers, respectively, each analyzed in duplicate. Fetal livers from compound heterozygous embryos give rise to fewer colonies than those heterozygous for just a single gene. (See Figure S3 for differences in cellularity of E12.5 fetal livers and qRT-PCR expression analysis of WT, *Runx1*$^{+/-}$, *Gata2*$^{+/-}$, and *Runx1*$^{+/-}$::*Gata2*$^{+/-}$ fetal liver samples).

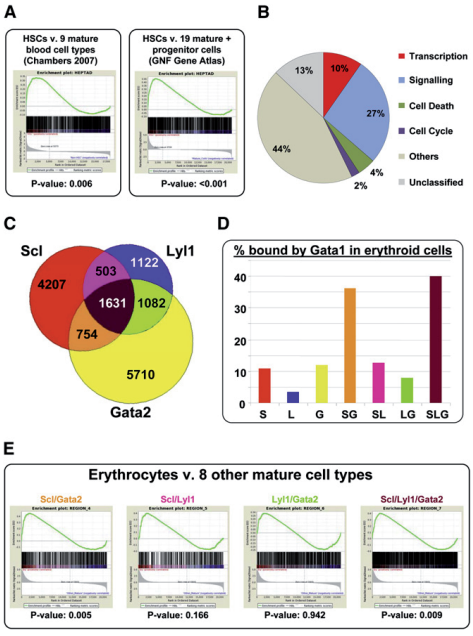

Figure 7. Integration of the Ten-Factor ChIP-Seq Resource with External Data

(A) Gene set enrichment analysis (GSEA) demonstrates highly significant HSC specific expression for the 927 candidate target genes of the SCL/LYL1/GATA2/RUNX1/LMO2/FLI-1/ERG heptad. GSEA was performed using standard settings with two independent expression profiling compendia (Chambers et al., 2007; Wu et al., 2009). The graphical outputs show enrichment (green curve) of the heptad gene set along ranked lists of all genes ordered based on their HSC-specificity of expression in the two data sets.

(B) Analysis of putative gene function for 927 heptad target genes reveals an abundance of genes involved in transcriptional control and signaling.

(C) SCL, LYL1, and GATA2 binding combinations in HPC-7 cells (Venn diagram showing the numbers of peaks bound by combinations of SCL, LYL1, and GATA2).

(D) Peaks bound by both SCL and GATA2 in HPC-7 cells are commonly bound by GATA1 in erythroid cells. The bar chart shows the proportion of all regions shown in (C) that are bound by GATA1 in erythroid cells (GATA1 peak data taken from Cheng et al. [2009]); S, SCL; L, LYL1; G, GATA2.

(E) Gene set enrichment analysis (GSEA) identifies subsets of genes next to peaks bound by both SCL and GATA2 with erythroid-specific expression. Neighboring genes were determined for the SG, SL, LG, and SLG peak categories from (D), and GSEA was performed using standard settings. Erythroid specificity of gene expression was determined by comparing erythroid expression to the eight other mature cell types present in the hematopoietic fingerprints database (Chambers et al., 2007). (See Figure S4 for GSEA analysis with control gene sets).

genome-wide ChIP-Seq analysis of Gata1 in the erythroid model cell line G1E-ER4 identified 14,351 peaks bound by GATA1 in erythroid cells (Cheng et al., 2009). We used this data set to ask what proportion of the regions bound by combinations of SCL, LYL1, and GATA2 in HPC-7 would, later on in differentiation, be bound by GATA1. As shown in Figure 7D, those regions bound by both GATA2 and SCL in HPC-7 are much more likely to be bound by GATA1 than any of the other subsets. In contrast to

LYL1, SCL is a key regulator of erythropoiesis, where it commonly functions in SCL/GATA1 containing multiprotein complexes. Our data, therefore, suggest that these functional differences between SCL and LYL1 are already apparent in the set of target regions bound differentially by SCL and LYL1 in progenitors. Moreover, our results are consistent with the notion that a significant proportion of those regions bound later on by SCL/GATA1 are already "primed" by SCL/GATA2 binding in progenitors.

To further corroborate potential priming of erythroid loci by SCL/GATA2, we determined the genes next to peaks bound by SCL/GATA2, SCL/LYL1, LYL1/GATA2, and SCL/LYL1/GATA2 and performed gene set enrichment analysis. Comparison of the various peak classes demonstrated that only those genes next to peaks bound by SCL and GATA2 (and, therefore, more likely to be bound by GATA1 later on) showed a modest but nevertheless significant enrichment for erythroid specific expression (Figure 7E). This analysis illustrates, therefore, how integrated analysis of ChIP-Seq peak regions might be utilized to predict gene function (e.g., that SCL has a major erythroid phenotype, whereas LYL1 does not). Moreover, by linking differentially-bound ChIP-Seq peaks to neighboring genes, this analysis also provides possible candidate target genes responsible for the different functions of Scl and Lyl1. Taken together, therefore, the above examples demonstrate that the ten-factor ChIP-Seq resource reported here can be integrated readily with external data, such as expression or ChIP-Seq profiling, to provide a wealth of novel hypotheses.

DISCUSSION

Identification of the "rules" that govern gene expression in stem cells is of fundamental scientific and clinical significance. Here, we report genome-wide binding patterns for ten key regulators of HSPCs. Comprehensive bioinformatic analysis of combinatorial binding profiles suggested new direct functional links between some of these vital regulators of HSC function, which were validated using both cell biological/biochemical as well as knockout mouse in vivo approaches.

Extracting Biological Insights from ChIP-Seq Analysis

Single-protein ChIP-Seq studies reporting long lists of binding peaks often provide only limited scientific insights because there are far too many putative candidate target genes to provide specific new biological functions for the upstream regulator under investigation. The challenge, therefore, remains to develop widely applicable strategies for extracting biological meaning from genome-wide ChIP-Seq data. Studies in lower model organisms, such as yeast and *Drosophila*, have demonstrated that the true power of genome-wide TF binding studies may only be revealed by studying multiple factors together. For example, determination of genome-wide binding patterns for a near-full complement of yeast transcription factors has served as an important cornerstone for the development of regulatory network models of unprecedented scale and substantial predictive power (Lee et al., 2002). In addition, a genome-wide survey of five TFs over a time series of 15 developmental stages of *Drosophila* mesoderm development demonstrated that spatiotemporal gene expression patterns could be predicted with accuracy from clustered TF binding patterns (Zinzen et al., 2009). Similarly, a computational framework to predict absolute as well as differential gene expression from comprehensive mouse ESC ChIP-Seq data collections has recently been proposed (Ouyang et al., 2009), and a comprehensive resource of ESCs modified to express affinity-tagged transcription factor proteins has recently been reported (Nishiyama et al., 2009). The study reported here demonstrates that comprehensive bioinformatic analysis of multifactor ChIP-Seq data has the power to reveal new insights into combinatorial transcriptional control mechanisms in complex mammalian settings.

Apart from the clear limitations of single-factor surveys, another important consideration for ChIP-Seq studies is the homogeneity of the starting population, which presents particular challenges with rare and heterogeneous adult stem/progenitor cells. The success of downstream bioinformatic analysis heavily depends on sample homogeneity, and therefore, recent studies using even much less challenging experimental systems, such as mature erythroid or T cells, have opted for the use cell line systems (Cheng et al., 2009; Fujiwara et al., 2009; Soler et al., 2010; Yu et al., 2009). Importantly, we had previously validated Scl-bound regions in the multipotential progenitor cell line HPC-7 as bona fide early hematopoietic enhancers using extensive transgenic mouse studies (Wilson et al., 2009). These demonstrated that 10 out of 11 regions bound by Scl had fetal liver activity. Four of these ten (situated in the *Erg*, *Gfi1b*, *Klf2*, and *Nfe2* loci, respectively) are bound by all seven factors, and all four not only have fetal liver activity, but also display specific staining within the midgestation dorsal aorta region (data not shown) where HSCs emerge at this time point during mouse embryogenesis (de Bruijn et al., 2000). Further evidence of the validity of our experimental system comes from our analysis of the genes next to peaks bound by the newly identified heptad of SCL, LYL1, LMO2, GATA2, RUNX1, ERG, and FLI-1. Gene set enrichment analysis (Subramanian et al., 2005) demonstrated highly significant enrichment of genes specifically expressed in HSCs, as well as signaling and transcriptional regulatory proteins and was, therefore, consistent with the notion that genes controlled by this heptad are critical regulators of HSPC function. Taken together, therefore, the data presented here suggest that integrated analysis of ChIP-Seq results for multiple factors has the potential to identify the key downstream mediators of cellular phenotypes in complex mammalian systems.

Identification of Transcriptional Control Mechanisms in HSPCs

Initial insights into gene regulatory mechanisms in HSPCs came from studies of regulatory elements specifically active in HSPCs, such as the *Scl* +19 or RUNX1 +23 enhancers (Landry et al., 2008; Nottingham et al., 2007; Sánchez et al., 1999). Detailed functional analysis demonstrated synergistic activation by protein complexes containing the GATA factor GATA2 and Ets factors FLI-1 or ERG (Göttgens et al., 2002). Subsequent computational analysis based on evolutionarily conserved TF binding motifs identified three structurally related elements with similar in vivo activity leading to the hypotheses that an Ets/Ets/Gata regulatory code plays a significant role in HSPC transcriptional control (Donaldson et al., 2005; Pimanda et al., 2007a). However, because of their short length and degeneracy, TF consensus binding sequences occur many more times in mammalian

genomes than there are genuine regulatory sites. Consequently, genome-wide computational characterization of transcriptional target sites has remained an elusive proposition. The data presented here show that multi-TF ChIP-Seq provides a much more promising route toward gaining a comprehensive view of transcriptional control mechanisms. Indeed, the most compelling reason why computational screens will struggle lies in the absence of recognized consensus binding sites in many regions that are nevertheless bound by the relevant TFs in vivo, as observed here for SCL/LYL1 and RUNX1.

Absence of conventional binding sites suggests tethering by other factors. This is precisely what was shown previously through a series of elegant studies where GATA1 was implicated into recruiting SCL to DNA in erythroid cells because a mutant Scl protein devoid of its own DNA binding domain could be detected on regions also bound by GATA1 (Kassouf et al., 2008). Given that Runx consensus binding motifs were absent in an even greater proportion of heptad-bound regions than E-boxes, we postulated that RUNX1 might also be recruited to DNA by interacting with other TFs. However, it was not known whether components of SCL/GATA2 complexes can interact with RUNX1 and, thus, contribute to its tethering to DNA. In this study, we provide evidence not only for biochemical but also genetic in vivo interaction between RUNX1 and components of SCL/GATA2 complexes. Bioinformatically informed functional validation, therefore, provided previously unrecognized links between some of the most powerful regulators of HSC fate.

Combinatorial Control Mechanisms in HSPCs

The potential for different combinatorial interactions within complex transcriptional regulatory networks is truly immense. Even with only ten factors such as the set reported in this study, there are 2^{10} (1,024) different possible binding patterns based on simple present/absent calls for each of the ten factors. In reality, the situation is going to be much more complex still; not only because additional factors will play a role but also because quantitative effects are likely to be important (e.g., regions bound strongly by both SCL and PU.1 may be functionally distinct from those bound more strongly by SCL than PU.1). For the current study, we focused our analysis on pairwise binding as well as those regions bound by five or more factors. However, the data set readily permits analysis of any other combinatorial pattern that may be of interest to the wider community, including any one of the 120 possible three-way and 210 possible four-way combinations. The data presented here, therefore, represent a platform for very substantive future investigations into the function of ten critical regulators of HSPC function.

When faced with the prospect of having to explore potentially huge numbers of combinatorial patterns such as the 1024 possible occupancy patterns for the 10 factors studied here, it is salient to bear in mind that rigorous statistical analysis of the experimental data can reduce significantly the number of patterns when used as a means of prioritization. For example, of the 1024 patterns possible with our ten factors, 221 are found much more often than would be expected by chance ($p < 0.01$), and importantly, many of these ''overrepresented'' patterns are closely related. Integrated analysis taking into account for example expression patterns or biological functions of genes associated with specific categories of peaks should serve to further prioritize hypotheses for subsequent experimental testing.

With respect to the particular combinatorial interaction between RUNX1 and GATA2/SCL complexes explored here, it is noteworthy that despite their apparent functional differences during early hematopoietic development (Lancrin et al., 2009), recent studies are consistent with overlapping functions once the hematopoietic program is established. Deletion of RUNX1 in adult bone marrow causes expansion of immature progenitors with a concomitant reduction of long-term HSC activity (Growney et al., 2005), and *Runx1*$^{+/-}$ mice have increased numbers of progenitors yet reduced numbers of LT-HSCs when compared with wild-type mice (Sun and Downing, 2004). At the same time, increased levels of GATA2 induce quiescence of HSCs (Tipping et al., 2009) and SCL has also recently been reported to be a positive regulator of HSC quiescence (Lacombe et al., 2010). Taken together, therefore, these observations are consistent with the notion that RUNX1, GATA2, and SCL collaborate by jointly controlling a set of genes critical for setting the balance between quiescence and proliferation of HSCs. Given that loss of function mutations of *Runx1* have been found in several types of leukemia (Osato, 2004), identifying the molecular mechanisms responsible for controlling the expression of such critical target genes in normal cells will have important implications for our understanding of the molecular consequences of aberrant transcriptional control in hematopoietic malignancies.

The fundamental importance of combinatorial transcriptional control of development is particularly well documented through studies in model organisms such as sea urchin and *Drosophila*. Of note, the *Drosophila* Gata and Runx factors Serpent and Lozenge control key stages of blood development (Lebestky et al., 2000) and a series of elegant experiments demonstrated both genetic as well as direct protein-protein interactions between Serpent and Lozenge (Waltzer et al., 2003). Moreover, interactions between RUNX1 and GATA1 have been identified as a potentially important transcriptional mechanism in megakaryocytic cells (Elagib et al., 2003; Wu et al., 2009; Xu et al., 2006). These observations, therefore, not only emphasize the deep evolutionary conservation of interactions between Gata and Runx proteins but also highlight the potential for combinatorial interactions between different members of these two transcription factor families in different cellular contexts. It will be interesting to investigate to what extent larger combinatorial interactions such as the SCL/LYL1/GATA2/RUNX1/LMO2/FLI-1/ERG heptad identified here may be involved in controlling transcription in these other lineages.

Concluding Remarks

Here, we report the genome-wide binding profiles for ten key transcriptional regulators of HSPCs which will be of major interest to researchers studying any one of these ten factors. Moreover, generation of a unified data set from a single cell type represents an important prerequisite to performing sophisticated integrated analysis of combinatorial TF interactions. The experimental data provided in this study serve as an illustration of the potential power of such future investigations. Integrated bioinformatic analysis prompted us to investigate interactions

between RUNX1 and GATA2/SCL. Functional validation of bioinformatic predictions not only validated the overall approach but also provided a significant advance of our understanding of blood stem cells by establishing previously unrecognized and functionally important interactions between some of the most powerful regulators of blood stem cells.

EXPERIMENTAL PROCEDURES

Chromatin Immunoprecipitation

HPC-7 cells (Pinto do O et al., 1998) were grown in Stem Cell Factor, and ChIP assays were performed as previously described using polyclonal antibodies against acetylated histone H3 (Milipore 06-599), SCL (Santa Cruz, sc12984x), LYL1 (Abcam, ab30334-200), GATA2 (Santa Cruz, sc9008x), LMO2 (R&D, AF2726), RUNX1 (Abcam, ab23980-100), MEIS1 (Santa Cruz, sc-10599x), PU.1 (Santa Cruz, sc-352x), ERG (Santa Cruz, sc354x), FLI-1 (Abcam, ab15289-500), GFI1B (Santa Cruz, sc8559x), and control nonspecific rabbit IgG (Sigma I5006); see Supplemental Experimental Procedures for more details. Each sample was amplified and sequenced using the Illumina 2G Genome Analyzer, following manufacturer's instructions. Sequencing reads were mapped to the mouse reference genome using MACS (Zhang et al., 2008), converted to a density plot, and displayed as UCSC genome browser custom tracks. Raw sequence data, custom track (.wig) files, and peak/gene tables can be downloaded from the NCBI Gene Expression Omnibus portal (www.ncbi.nlm.gov/geo; accession number GSE22178).

Peak Calling

For all ten transcription factors, three peak finding programs were used with the same parameters used for each factor. Parameters for programs were as follows: Findpeaks—default setting for ChIP-Seq analysis, peaks with FDR < 0.05 called; Peakseq—default setting; MACS—MACS was used with the following command line parameters: mfold = 16, t size = 35, bw = 100, pvalue = 1e−9, g size = 2,200,000,000, no lambda. The overlap from the three programs was then taken (1 bp minimum). MACS peak coordinates were used as reference for further analysis. Regions that lie at the end of chromosomes that were high in repeat regions were subtracted from the high-confidence peak coordinates. Finally, all peaks were standarized to 399 bp wide based on MACS summits and average length of ChIP DNA fragments. Distances between overlapping peaks from pairwise TF comparisons were calculated and plotted as box plots using R (http://www.r-project.org/).

Analysis of Complex Binding Patterns

Principal component analysis (PCA) was used to reduce the dimensionality of the data matrix. Ten TFs were plotted in a 3D scatter plot, corresponding to the first three principle components. In R, the svd function was used for PCA and scatterplot3d for the 3D scatter plot. Hierarchical Clustering was performed on the data matrix in R using the hclust function with default parameters. Euclidean distance was used for distance matrix calculations, and the "complete method" was used for clustering.

Motif Analysis

For motif analysis, the middle 200 bp was taken for all peak regions, and peaks were excluded from analysis if they had more than 60% of repetitve sequences. For de novo motif discovery, the MEME program (Bailey and Elkan, 1994) was used with standard settings. Matches to consensus sequences were determined as described using TFBSSearch (Chapman et al., 2004). To simplify the result matrix, 0 was used when there were no sites, and 1 was used if 1 or more sites were found. The resulting matrix was clustered and displayed using the heatmap function in R.

Coimmunoprecipitation Analysis

293T cells were transiently transfected with expression plasmids using Protransfection Mammlian Transfection System (Promega) and incubated 48 hr before analysis. Cells were lysed and cleared of debris and DNA. Supernatants were precleared with IgG and protein G-agarose beads. Then, anti-FLAG antibody (Sigma M2, F3165), anti-GATA2 antibody (SC-9008x), or MYC-agarose (sc-90 AC) were added to all corresponding samples. The immune complexes were washed, boiled for 10 min in sample buffer, and analyzed by western blot. Antibodies used for western blot are as follows: ERG (Santa Cruz, sc354x) or MYC (Santa Cruz, sc-90).

Colony Assays

Clonogenic progenitor cells within the FL cell populations were assayed in the colony-forming unit culture (CFU-C) assay. Cells were plated in duplicate in 35 mm culture dishes in MethoCult M3434 (StemCell Technologies, London, United Kingdom) according to the manufacturer's instructions. Cultures were grown at 37°C, 5% CO_2, and colonies were counted after 7 days (Nottingham et al., 2007).

ACCESSION NUMBERS

New microarray data have been deposited into the NCBI Gene Expression Omnibus portal under the accession number GSE22178.

SUPPLEMENTAL INFORMATION

Supplemental information includes four figures, Supplemental Experimental Procedures, and four tables and can be found online at doi:10.1016/j.stem.2010.07.016.

ACKNOWLEDGMENTS

We would like to thank A. Rendon and O. Burren for help with mapping ChIP-Seq reads, S. Anand for advice on immunoprecipitation assays, D. Hebenstreit for help with generating density profiles, and S. Teichmann and L. Wernisch for helpful discussions. Work in the authors' laboratories is supported by Leukaemia and Lymphoma Research, the Leukaemia and Lymphoma Foundation, the UK Medical Research Council, the National Institute for Health Research Cambridge Biomedical Research Centre, and the National Health and Medical Research Council of Australia and the Australian Research Council (K.K. and J.P.).

Received: March 13, 2010
Revised: June 10, 2010
Accepted: July 21, 2010
Published: October 7, 2010

REFERENCES

Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., and Higgs, D.R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. EMBO J. *23*, 2841–2852.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. *2*, 28–36.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Cai, Z., de Bruijn, M., Ma, X., Dortland, B., Luteijn, T., Downing, R.J., and Dzierzak, E. (2000). Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. Immunity *13*, 423–431.

Chambers, S.M., Boles, N.C., Lin, K.Y., Tierney, M.P., Bowman, T.V., Bradfute, S.B., Chen, A.J., Merchant, A.A., Sirin, O., Weksberg, D.C., et al. (2007). Hematopoietic fingerprints: an expression database of stem cells and their progeny. Cell Stem Cell *1*, 578–591.

Chapman, M.A., Donaldson, I.J., Gilbert, J., Grafham, D., Rogers, J., Green, A.R., and Göttgens, B. (2004). Analysis of multiple genomic sequence

alignments: a web resource, online tools, and lessons learned from analysis of mammalian SCL loci. Genome Res. *14*, 313–318.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Cheng, Y., Wu, W., Kumar, S.A., Yu, D., Deng, W., Tripic, T., King, D.C., Chen, K.B., Zhang, Y., Drautz, D., et al. (2009). Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. Genome Res. *19*, 2172–2184.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

de Bruijn, M.F., Speck, N.A., Peeters, M.C., and Dzierzak, E. (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. EMBO J. *19*, 2465–2474.

Donaldson, I.J., Chapman, M., Kinston, S., Landry, J.R., Knezevic, K., Piltz, S., Buckley, N., Green, A.R., and Göttgens, B. (2005). Genome-wide identification of cis-regulatory sequences controlling blood and endothelial development. Hum. Mol. Genet. *14*, 595–601.

Elagib, K.E., Racke, F.K., Mogass, M., Khetawat, R., Delehanty, L.L., and Goldfarb, A.N. (2003). RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. Blood *101*, 4333–4341.

Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A.K., Kang, Y.A., Choi, K., Farnham, P.J., and Bresnick, E.H. (2009). Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. Mol. Cell *36*, 667–681.

Gao, J.W., Zheng, R.T., Ohtani, H., Zhu, D.S., and Chen, G. (2009). Experimental investigation of heat conduction mechanisms in nanofluids. Clue on clustering. Nano Lett. *9*, 4128–4132.

Gering, M., Yamada, Y., Rabbitts, T.H., and Patient, R.K. (2003). Lmo2 and Scl/Tal1 convert non-axial mesoderm into haemangioblasts which differentiate into endothelial cells in the absence of Gata1. Development *130*, 6187–6199.

Göttgens, B., Nastos, A., Kinston, S., Piltz, S., Delabesse, E.C., Stanley, M., Sanchez, M.J., Ciau-Uitz, A., Patient, R., and Green, A.R. (2002). Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. EMBO J. *21*, 3039–3050.

Growney, J.D., Shigematsu, H., Li, Z., Lee, B.H., Adelsperger, J., Rowan, R., Curley, D.P., Kutok, J.L., Akashi, K., Williams, I.R., et al. (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. Blood *106*, 494–504.

Kassouf, M.T., Chagraoui, H., Vyas, P., and Porcher, C. (2008). Differential use of SCL/TAL-1 DNA-binding domain in developmental hematopoiesis. Blood *112*, 1056–1067.

Lacombe, J., Herblot, S., Rojas-Sutterlin, S., Haman, A., Barakat, S., Iscove, N.N., Sauvageau, G., and Hoang, T. (2010). Scl regulates the quiescence and the long-term competence of hematopoietic stem cells. Blood *115*, 792–803.

Lancrin, C., Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V., and Lacaud, G. (2009). The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. Nature *457*, 892–895.

Landry, J.R., Kinston, S., Knezevic, K., de Bruijn, M.F., Wilson, N., Nottingham, W.T., Peitz, M., Edenhofer, F., Pimanda, J.E., Ottersbach, K., and Göttgens, B. (2008). Runx genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. Blood *111*, 3005–3014.

Laslo, P., Spooner, C.J., Warmflash, A., Lancki, D.W., Lee, H.J., Sciammas, R., Gantner, B.N., Dinner, A.R., and Singh, H. (2006). Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. Cell *126*, 755–766.

Lebestky, T., Chang, T., Hartenstein, V., and Banerjee, U. (2000). Specification of Drosophila hematopoietic lineage by conserved transcription factors. Science *288*, 146–149.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799–804.

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

McDaniell, R., Lee, B.K., Song, L., Liu, Z., Boyle, A.P., Erdos, M.R., Scott, L.J., Morken, M.A., Kucera, K.S., Battenhouse, A., et al. (2010). Heritable individual-specific and allele-specific chromatin signatures in humans. Science *328*, 235–239.

McNagny, K.M., Sieweke, M.H., Döderlein, G., Graf, T., and Nerlov, C. (1998). Regulation of eosinophil-specific gene expression by a C/EBP-Ets complex and GATA-1. EMBO J. *17*, 3669–3680.

Nerlov, C., and Graf, T. (1998). PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. Genes Dev. *12*, 2403–2412.

Nerlov, C., McNagny, K.M., Döderlein, G., Kowenz-Leutz, E., and Graf, T. (1998). Distinct C/EBP functions are required for eosinophil lineage commitment and maturation. Genes Dev. *12*, 2413–2423.

Nishiyama, A., Xin, L., Sharov, A.A., Thomas, M., Mowrer, G., Meyers, E., Piao, Y., Mehta, S., Yee, S., Nakatake, Y., et al. (2009). Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. Cell Stem Cell *5*, 420–433.

Nottingham, W.T., Jarratt, A., Burgess, M., Speck, C.L., Cheng, J.F., Prabhakar, S., Rubin, E.M., Li, P.S., Sloane-Stanley, J., Kong-A-San, J., and de Bruijn, M.F. (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. Blood *110*, 4188–4197.

Okuda, T., van Deursen, J., Hiebert, S.W., Grosveld, G., and Downing, J.R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell *84*, 321–330.

Osada, H., Grutz, G., Axelson, H., Forster, A., and Rabbitts, T.H. (1995). Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. Proc. Natl. Acad. Sci. USA *92*, 9585–9589.

Osato, M. (2004). Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia. Oncogene *23*, 4284–4296.

Ouyang, Z., Zhou, Q., and Wong, W.H. (2009). ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. Proc. Natl. Acad. Sci. USA *106*, 21521–21526.

Pimanda, J.E., Donaldson, I.J., de Bruijn, M.F., Kinston, S., Knezevic, K., Huckle, L., Piltz, S., Landry, J.R., Green, A.R., Tannahill, D., and Göttgens, B. (2007a). The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. Proc. Natl. Acad. Sci. USA *104*, 840–845.

Pimanda, J.E., Ottersbach, K., Knezevic, K., Kinston, S., Chan, W.Y., Wilson, N.K., Landry, J.R., Wood, A.D., Kolb-Kokocinski, A., Green, A.R., et al. (2007b). Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. Proc. Natl. Acad. Sci. USA *104*, 17692–17697.

Pinto do O, P., Kolterud, A., and Carlsson, L. (1998). Expression of the LIM-homeobox gene LH2 generates immortalized steel factor-dependent multipotent hematopoietic precursors. EMBO J. *17*, 5744–5756.

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

Sánchez, M., Göttgens, B., Sinclair, A.M., Stanley, M., Begley, C.G., Hunter, S., and Green, A.R. (1999). An SCL 3′ enhancer targets developing endothelium together with embryonic and adult haematopoietic progenitors. Development *126*, 3891–3904.

Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.J., Stevens, M., Kockx, C., van Ijcken, W., et al. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes Dev. *24*, 277–289.

Spooner, C.J., Cheng, J.X., Pujadas, E., Laslo, P., and Singh, H. (2009). A recurrent network involving the transcription factors PU.1 and Gfi1 orchestrates innate and adaptive immune cell fates. Immunity *31*, 576–586.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Sun, W., and Downing, J.R. (2004). Haploinsufficiency of AML1 results in a decrease in the number of LTR-HSCs while simultaneously inducing an increase in more mature progenitors. Blood *104*, 3565–3572.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Tipping, A.J., Pina, C., Castor, A., Hong, D., Rodrigues, N.P., Lazzari, L., May, G.E., Jacobsen, S.E., and Enver, T. (2009). High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. Blood *113*, 2661–2672.

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

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035–1041.

Wadman, I.A., Osada, H., Grütz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. EMBO J. *16*, 3145–3157.

Waltzer, L., Ferjoux, G., Bataillé, L., and Haenlin, M. (2003). Cooperation between the GATA and RUNX factors Serpent and Lozenge during Drosophila hematopoiesis. EMBO J. *22*, 6516–6525.

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A.H., and Speck, N.A. (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proc. Natl. Acad. Sci. USA *93*, 3444–3449.

Wilson, N.K., Miranda-Saavedra, D., Kinston, S., Bonadies, N., Foster, S.D., Calero-Nieto, F., Dawson, M.A., Donaldson, I.J., Dumon, S., Frampton, J., et al. (2009). The transcriptional program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development. Blood *113*, 5456–5465.

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd, and Su, A.I. (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. *10*, R130.

Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc. Natl. Acad. Sci. USA *104*, 7145–7150.

Xu, G., Kanezaki, R., Toki, T., Watanabe, S., Takahashi, Y., Terui, K., Kitabayashi, I., and Ito, E. (2006). Physical association of the patient-specific GATA1 mutants with RUNX1 in acute megakaryoblastic leukemia accompanying Down syndrome. Leukemia *20*, 1002–1008.

Ye, M., and Graf, T. (2007). Early decisions in lymphoid development. Curr. Opin. Immunol. *19*, 123–128.

Yu, M., Riva, L., Xie, H., Schindler, Y., Moran, T.B., Cheng, Y., Yu, D., Hardison, R., Weiss, M.J., Orkin, S.H., et al. (2009). Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. Mol. Cell *36*, 682–695.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627–632.

Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., and Furlong, E.E. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature *462*, 65–70.

Chapter 5

General Discussion

General discussion and future experiments

HSCs are the cell type responsible for the generation and establishment of the blood tissue in an organism, through their two main characteristics: self renewal capacity and differentiation ability in to all the different blood cell types. These two characteristics establish them as the founders of the adult hematopoietic tissue. The understanding of the molecular events controlling the processes of differentiation and self renewal, as well as the generation of the HSC population, are of major importance in the field of regenerative medicine, since they could suggest strategies to manipulate different cell types (non-hematopoietic and/or hematopoietic) to change lineage fate and allow the acquisition of HSC properties.

Transcriptional control is one of the most important intracellular regulatory mechanisms for establishment of cell fate, identity and function. The transcription factor Gata2 is a major regulator of embryonic hematopoiesis. The exact role of Gata2 in the generation, differentiation and self-renewal of HSCs is still largely unknown. In this thesis we provide insight into the role of *Gata2* in HSCs during mouse development by testing the hematopoietic potential of Gata2 expressing cells. The direct analysis of viable Gata2 expressing cells is possible only through the use of a *Gata2* reporter system. The previously existing *Gata2-GFP* reporter mouse model (Minegishi, Suzuki et al. 2003) resulted in detectable GFP expression, however the genetic insertion of the GFP gene yielded a *Gata2* haploinsufficient state. It has already been shown by Ling et al. (Ling, Ottersbach et al. 2004) that *Gata2* haploinsufficient embryos have a greatly reduced number of HSCs and these HSCs are also qualitatively defective, clearly proving that the level of expression of the transcription factor is important for its physiologic function. This fact renders the *Gata2-GFP* reporter mouse model unsuitable for testing the potential of the Gata2 expressing cells under physiologic conditions. Thus the generation of a new tool was necessary - a new reporter mouse model in which the level of expression and the function of the transcription factor would be unaffected, allowing the prospective isolation of Gata2 expressing cells. For this reason we generated the *Gata2-Venus ES* cell and mouse model (described in chapters 2 and 3 of this thesis).

Gata2: A role in hematopoietic cell generation?

A few studies have provided functional data that help hypothesize about the role of *Gata2* during embryonic hematopoiesis. The phenotypes of both *Gata2*$^{-/-}$ and *Gata2*$^{+/-}$ embryos, showing significantly reduced numbers of HSCs and hematopoietic progenitor cells, imply that Gata2 plays a role in the generation of hematopoietic cells. Whether Gata2 controls the acquisition of hematopoietic fate from the hemogenic endothelial cells and/or some other processes in the endothelial to hematopoietic transition is not known. Nevertheless, *Gata2* is expressed in the lateral mesoderm and endothelial cells of the paired dorsal aorta and also the en-

endothelial cells of the dorsal aorta and hematopoietic cluster cells at the time-points of hematopoiesis in the mouse embryo, indicating that Gata2 has a functional role in these specific cell types. This fact, in combination with FACS analysis data presented in this thesis (chapter 3) which show that *Gata2* is expressed in hemogenic endothelial cells, further supports the idea of *Gata2* playing a role in the generation of hematopoietic cells from the hemogenic endothelial cell population. If that suggestion is true, it is expected that *Gata2* expressing cell populations would be enriched in hematopoietic progenitors and HSCs. Gene expression profiling studies of various hematopoietic cell populations have already shown that *Gata2* is expressed in high levels in cell populations enriched in HSC activity and in more primitive hematopoietic cell populations. Lower levels of *Gata2* expression were observed in more mature cell subsets (Orlic, Anderson et al. 1995). In this thesis (chapter 3) we prove that *Gata2* expressing cell population is highly enriched in hematopoietic progenitor cells and that all the HSCs are found in the *Gata2* expressing cell fraction. The specific role of Gata2 in the HSC population is at present unknown.

Evidence coming from imaging experiments further supports the suggestion that Gata2 is involved in the generation of hematopoietic cells. Whole mount immunostaining results demonstrate that the *Gata2*$^{-/-}$ embryos have no intra-aortic clusters in the lumen of the dorsal aorta (DA), indicating that Gata2 is somehow involved in the process of hematopoietic cell generation. On the other hand c-Kit^{+} hematopoietic cells in the form of small hematopoietic clusters are present in the vitelline and the umbilical arteries. In addition, small numbers of hematopoietic progenitors can be detected by methylcellulose assays in the YS and the major vasculature of the *Gata2*$^{-/-}$ embryos. The presence of hematopoietic progenitors in the *Gata2*$^{-/-}$ embryos is further supported by the finding that hematopoietic progenitors are always present in the Venus negative cell fractions of the hematopoietic tissues (this thesis, chapter 3). The above mentioned findings clearly demonstrate that Gata2 independent progenitors exist in the embryos. This fact seems somewhat contradictory with the general idea about the importance of Gata2 function in hematopoietic cell generation. One possible explanation about the existence of these Gata2 independent hematopoietic progenitors in the *Gata2*$^{-/-}$ mice could be that Gata2 function is partially compensated by the binding of other Gata factors, since it is already shown from the literature that the same Gata binding sequences can be recognized by more than one Gata transcription factor (Takahashi, Shimizu et al. 2000; Grass, Boyer et al. 2003; Jing, Vakoc et al. 2008). Experiments examining whether other Gata factors are expressed in the relevant cells in embryo are necessary to confirm this. Another hypothesis could be that these *Gata2* independent progenitors which are also present in physiologic conditions (Venus negative cells)

represent hematopoietic progenitors of the first waves of progenitor generation before the emergence of the HSCs. Transcriptional profiling of these cells should provide some insight into this.

To specifically identify the cell types in which Gata2 plays an important role, a conditional knock out approach can be followed. A similar approach has already been taken to examine in which cells the function of Runx1 - another major regulator of hematopoiesis - is required. Chen and colleagues (Chen, Yokomizo et al. 2009) proved that the role of Runx1 is localized in the endothelial to hematopoietic transition in the midgestation embryo. Conditional knock out of the transcription factor in vascular-endothelial-cadherin positive (VE-cad) cells resulted in loss of HSCs, hematopoietic progenitors and intraaortic cluster formation. The findings described in the Runx1 study raise a similar question concerning the role of Gata2, and its localized function in specific cell types and during a specific developmental window of time. At present, studies in our lab on the conditional knock out of *Gata2* in endothelial (driven by *VEcad-Cre*) or hematopoietic cells (driven by *Vav-Cre*) suggest that *Gata2* may have different and distinct roles during the different stages of hematopoietic development in the mouse embryo (manuscript in preparation by E. de Pater, data not shown). How Gata2 manages to perform the different tasks and different functions is still unknown. More studies need to be done towards the precise identification of Gata2 target genes and the protein partners of Gata2 in regulating gene expression in every different cell population. Technical limitations, with very small cell number in each population being the most important of them, still need to be surpassed in order to answer all these important questions.

Different levels of Gata2 expression might define cell subsets with distinct potential or role

As it has been mentioned previously, reduced levels of *Gata2* affects hematopoietic cell generation/maintenance in the embryo as well as hematopoietic function in the adult (Ling, Ottersbach et al. 2004; Rodrigues, Janzen et al. 2005). Similar to Gata2, Runx1 also functions in a dose dependent manner, as it has been shown in experiments with *Runx1*$^{+/-}$ embryos (Robin, Ottersbach et al. 2006). The fact that both of these major regulators of hematopoiesis depend on the expression levels for their physiologic function suggest that there is a direct association between transcription factor levels and their role, and this seems to be an important characteristic of the intrinsic regulation of hematopoiesis. This aspect raises the question of whether it is possible to use this attribute to define and isolate separate cell populations with distinct functions and potential based on the levels of transcription factor of expression. Studies with the PU.1 transcription factor in a reporter mouse model have already shown that varying levels of PU.1 expression can define distinct populations with different hematopoietic potential (Nutt, Met-

calf et al. 2005). More specifically, high levels of PU.1 expression in hematopoietic progenitors are associated with the maintenance of the myeloid potential, whereas PU.1 downregulation leads to restriction in megakaryocyte and erythroid differentiation. The test of this hypothesis for the Gata2 expressing cell population in the embryo and adult would be of particular interest because of its important role as a major regulator of hematopoiesis. Isolation of the Gata2 expressing cell fraction into subsets based on its expression levels could identify distinct cell populations with differential cell fate potential and function. Cell population discrimination and characterization should help the sub-categorization of the hematopoietic cells into a developmental and/or functional hierarchy. This is of particular importance since no single marker currently allows the discrimination of long and short term HSCs and hematopoietic progenitors in early embryonic hematopoietic development. Our preliminary FACS analysis results indicate that such discrimination is possible for the Gata2 expressing cells and that the HSC activity in the E14 and E16 FL can be localized in the Venus intermediate expressing cell population (chapter 3 of this thesis).

The next step towards a more accurate characterization of a HSC cell population would be the combination of our *Gata2-Venus* reporter mouse with other major regulators of hematopoiesis reporter mouse models like Runx1 and SCL. As part of my early thesis work, I initiated the construction of a *Runx1* reporter with *mKate* fluorochrome gene, similar to the *Gata2-Venus* construct. This *Runx1-mKate* construct has been successfully recombined in ES cells and the creation of a mouse model will soon be underway. The combination of the *Gata2-Venus* and *Runx1-mKate* models will allow the isolation of cells expressing both of the transcription factors.

The combination of both the multiple reporters with the differential gene expression discrimination and the addition of cell surface markers discrimination could be a major step towards the goal of isolation of a pure embryonic HSCs, hematopoietic progenitor populations or pre-HSCs. Gene expression profiling of pure cell populations and the comparison between them and between the cells they derive from, like hemogenic endothelial cells, would be a significant advancement towards the understanding of the molecular mechanisms controlling processes like HSC generation, proliferation and differentiation into more restricted lineages.

Gata2 expression is very dynamic

All the above mentioned attributes of Gata2 have clearly emphasized that the timing and the level of expression are of major importance for the physiologic function of the transcription factor. The onset of *Gata2-Venus* expression in the hematopoietic sites and the following of the behavior of the individual cells expressing *Gata2* can further provide us with information about the role and function of

A novel technique of time lapse imaging has been developed in order to demonstrate specific cell behaviors like the endothelial to hematopoietic transition (Boisset, van Cappellen et al. 2010). Time lapse imaging of thick sections of the aorta at the important hematopoietic developmental stages, using *Gata2-Venus* embryos, will help in addressing of some interesting questions about the *Gata2* expressing cells. Are they proliferating? Do they undergo endothelial to hematopoietic transitions? Is the onset of *Gata2* expression associated with specific functions like proliferation or differentiation? More time lapse imaging videos need to be produced from different developmental timepoints to yield a better understanding of the behavior of the *Gata2* expressing cells. What is already clear from the existing data is that *Gata2* expression is very dynamic. The turning on and off as well as fluctuations in the Venus expression have been observed in individual cells. Whether these fluctuations can be functionally associated with the fluctuations in other regulatory molecules like transcription factors or signaling molecules needs further investigation.

The combination of the ¬Gata2-Venus reporter with the ¬Runx1-mkate in the future will provide an even further advantageous tool in addressing important questions like the onset of each transcription factor in specific cells and whether the expression of one affects the expression of the other. It will be interesting to observe whether endothelial to hematopoietic transitions can be observed from double expressing cells. Most importantly, the double reporter models will provide us with the opportunity to study the movements, behavior and the hematopoietic function of the cells simultaneously, singularly or sequentially expressing both of these pivotal transcription factors.

Gata2 works in a combinatorial manner with other transcription factors.

One of the questions raised about function of Gata2 is its different and distinct roles, as suggested by our results in germline *Gata2*$^{+/-}$ and *Gata2* conditional knock out embryos and adult mice. In the study by Wilson and colleagues presented in this thesis (chapter 4) the genome wide binding sites of ten key transcription regulators of hematopoiesis was investigated. Bioinformatics analysis of the data sets show that a heptad of transcription factors is very often bound in genomic sequences controlling hematopoiesis-related genes. These data also show that transcription factors work in a combinatorial manner and it can be hypothesized that different combinations of transcription factors in a transcription factor “complex” might have different effects and serve different functions in hematopoietic cells. This mechanism can possibly explain the differential roles of Gata2 during development in lateral mesoderm, endothelial and hemogenic endothelial cells, hematopoietic progenitors and HSCs. The formation of different transcription factor complexes, based on the set of transcription factors expressed in a specific cell type at a

certain developmental timepoint, could lead to the transcriptional regulation of a distinct set of genes, thus controlling in this way various developmental processes. This combinatorial manner of gene regulation involving Gata2 together with the quantitatively different levels of Gata2 expression levels is likely to be the underlying mechanism by which Gata2 exhibits distinct effects at different developmental timepoints.

In conclusion, in this thesis we show that all HSCs in the embryo express Gata2. The Gata2 expressing cell population of the embryo and embryoid bodies is also enriched in hematopoietic progenitor activity. Thus, Gata2 is relevant to generation and maintenance of the adult hematopoietic system. We provide novel data proving the existence of *Gata2* independent progenitors that may be helpful in understanding how the early wave of hematogenesis differs from the permanent adult wave. For the first time we show by time lapse imaging, that *Gata2-Venus* expression in the aortic endothelium and hematopoietic cluster cells is very dynamic at E10.5, suggesting the stochastic activation of target genes. Finally, in the Wilson et al. study we show that Gata2 works in combinatorial manner along with other transcription factors that are major players in hematopoietic regulation. As a major transcriptional regulator in HSCs and progenitors during embryonic development, the relevance of Gata2 and its regulation and function in human health has been recently demonstrated (Johnson, Hsu et al. 2012). In this study the researchers describe the case of a patient with monocytopenia and mycobacterial infection syndrome (MonoMAC). MonoMAC can be caused by missense mutations or intragenic deletions of *GATA2*. In this case the causative mutation was the disruption of a conserved *E-box-GATA* composite element located in an intron (+9.5). The study revealed that disruption of this cis regulatory element in mice ($+9.5^{-/-}$ and $+9.5^{+/-}$) leads to impaired vascular integrity and also in hematopoietic defects different from those seen in $Gata2^{-/-}$ embryos. It seems that this deletion affects a specific cell subset of Gata2 expressing cells (definitive but not primitive). This study clearly demonstrates the importance of Gata2 expression in the normal physiologic function of human hematopoietic cells and highlights the need for further understanding *Gata2* regulation and function in adult hematopoiesis.

References

Boisset, J. C., W. van Cappellen, et al. (2010). "In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium." Nature 464(7285): 116-120.

Chen, M. J., T. Yokomizo, et al. (2009). "Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter." Nature 457(7231): 887-891.

Grass, J. A., M. E. Boyer, et al. (2003). "GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling." Proc Natl

Acad Sci U S A 100(15): 8811-8816.

Jing, H., C. R. Vakoc, et al. (2008). “Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus.” Mol Cell 29(2): 232-242.

Johnson, K. D., A. P. Hsu, et al. (2012). “Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity.” J Clin Invest.

Ling, K. W., K. Ottersbach, et al. (2004). “GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells.” J Exp Med 200(7): 871-882.

Minegishi, N., N. Suzuki, et al. (2003). “Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos.” Blood 102(3): 896-905.

Nutt, S. L., D. Metcalf, et al. (2005). “Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors.” J Exp Med 201(2): 221-231.

Orlic, D., S. Anderson, et al. (1995). “Pluripotent hematopoietic stem cells contain high levels of mRNA for c-kit, GATA-2, p45 NF-E2, and c-myb and low levels or no mRNA for c-fms and the receptors for granulocyte colony-stimulating factor and interleukins 5 and 7.” Proc Natl Acad Sci U S A 92(10): 4601-4605.

Robin, C., K. Ottersbach, et al. (2006). “An unexpected role for IL-3 in the embryonic development of hematopoietic stem cells.” Dev Cell 11(2): 171-180.

Rodrigues, N. P., V. Janzen, et al. (2005). “Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis.” Blood 106(2): 477-484.

Takahashi, S., R. Shimizu, et al. (2000). “GATA factor transgenes under GATA-1 locus control rescue germline GATA-1 mutant deficiencies.” Blood 96(3): 910-916.

Summary
Samenvatting

Summary

Hematopoietic stem cells (HSCs) are the most important cell type of the adult hematopoietic system. They produce the billions of new blood cells needed each day and are characterized by the abilities of self renewal and differentiation. Self renewal capacity is responsible for maintaining an intact pool of HSCs during the life span of an organism. Through differentiation, HSCs provide the organism with all the distinct hematopoietic cell types of the blood tissue. During embryonic development, the first adult type HSCs are generated in the aorta gonads mesonephros (AGM) region – embryonic day 10.5 in the mouse. The generation of the HSCs and their differentiation into more restricted hematopoietic progenitors is tightly controlled. The molecular mechanisms regulating these processes involve extrinsic and intrinsic regulators. Transcriptional factors are the most important intrinsic regulators of the HSCs generation and differentiation. Gata2 is a key transcription factor controlling HSCs generation.

To understand the role of Gata2 during embryonic hematopoiesis it is essential to study the characteristics of the cells expressing Gata2. Until now it has not been feasible to isolate Gata2 expressing cells and test their potential in *in vivo* and *in vitro* functional assays or to follow them *in vivo* with live time lapse imaging. To this end, I generated in my thesis research a new *Gata2* reporter embryonic stem (ES) cell line and a mouse model in which Gata2 expression is followed by the expression of Venus fluorochrome. In this way the observation of Gata2 expressing cells is facilitated and their isolation has become feasible.

In **Chapter 2** I describe the generation of this new ES cell reporter line and the embryoid body (EB) differentiation culture system used to explore the potential of *Gata2* expressing cells. I show that *Venus* and *Gata2* are coordinately expressed in our reporter line (G2V5) and further demonstrate that during hematopoietic differentiation of EBs Gata2 expressing cells also express endothelial and hematopoietic cell surface markers (CD31 and c-Kit, respectively). Prospective isolation of Gata2 expressing cells and functional assays show that they possess endothelial potential in matrigel cultures, and also highly enriched in hematopoietic differentiation potential.

In **Chapter 3** the generation of the *Gata2-Venus* reporter mouse model is described. I demonstrate that *Gata2-Venus* adult mice are hematopoietically normal and that *Venus* and *Gata2* are coordinately expressed. Confocal microscopy of the mouse embryo shows that Venus fluorochrome expression is localized to the same anatomical regions and cells that have been shown to express *Gata2* mRNA. Flow cytometric analyses of the hematopoietic anatomical sites show that Gata2 expressing cells increase in numbers in the AGM and fetal liver (FL) during embryonic development. Furthermore, the number of $CD31^{+}c\text{-}Kit^{+}Venus^{+}$ intraaortic cluster

cells also decrease between E10 – E11 in agreement with previously published data on cluster cells. Functional analyses of Venus sorted cell populations show that the Gata2 expressing cell fraction is highly enriched in hematopoietic progenitors. Hematopoietic progenitors were also observed in the Venus negative cell fraction. Hematopoietic progenitor assays performed on *Gata2*$^{-/-}$ embryonic tissues confirm that Gata2 independent hematopoietic progenitors exist. These results have been further validated by the identification of c-kit^{+} hematopoietic clusters in the vitelline artery (VA) of *Gata2*$^{-/-}$ embryos. Hematopoietic progenitor enrichment can be further enhanced by combining our *Gata2-Venus* mouse model with the *Ly6A-GFP* reporter mouse. The identification of a small double positive cell population (GFP^{+}Venus^{+}) very highly enriched in immature hematopoietic progenitors. Importantly, HSC activity was found exclusively in the Gata2 expressing cell fraction.

In **Chapter 4** I present the genome-wide binding profile of ten transcription factors that have been shown to play an important role in regulating hematopoiesis. Bioinformatics analyses of these data sets suggest genome-wide combinatorial interactions between these key regulators of hematopoiesis and also further suggest the existence of a "heptamere" of transcription factors associated with hematopoietic cells. These data suggest the existence of previously unidentified transcription factors interactions. Finally, functional experiments in *Gata2*$^{+/-}$*Runx1*$^{+/-}$ embryos show these two major hematopoietic regulators have a combined effect in their function, suggesting that the bioinformatics analysis data can be a useful tool in understanding transcription factors cooperation in regulating specific processes like hematopoiesis.

In brief, in this thesis I describe the generation of a new *Gata2-Venus* reporter mouse and ES cell line model that allows the identification of Gata2 expressing cells *in vivo* and their isolation for use in functional *in vivo* and *in vitro* assays. I show that all the HSC activity in the mouse embryo is found in the Gata2 expressing cell fraction and that this cell population is highly enriched in hematopoietic progenitors. Gata2 is found to work in a combinatorial manner along with other major regulators of hematopoiesis. Most importantly, I show for the first time, by time-lapse imaging, that Gata2 expression at E10 embryonic aorta undergoes dynamic changes. These studies have expanded our knowledge of Gata2 and the regulation of hematopoietic stem and progenitor cell development.

Samenvatting

Hematopoietische stam cellen (HSCs) zijn de belangrijkste celtypes van het volwassen hematopoietische systeem. Zij produceren de miljarden nieuwe bloedcellen per dag en worden gekenmerkt door het vermogen van zelfvernieuwing en differentiatie. Zelfvernieuwing capaciteit is verantwoordelijk voor een intact poel van HSCs tijdens de levensduur van een organisme. Door differentiatie, verzorgen de HSCs het organisme met alle verschillende hematopoietische celtypes van het bloed weefsel. Tijdens de embryonale ontwikkeling, worden de eerste volwassen type HSCs gegenereerd in de aorta gonaden mesonephros (AGM) regio – op embryonale dag 10,5 in de muis. Het genereren van de HSCs en hun differentiatie in meer beperkte hematopoëtische voorloper cellen wordt streng gecontroleerd. De moleculaire mechanismen dat deze processen regulieren worden aangestuurd door extrinsieke en intrinsieke regulatoren. Transcriptiefactoren zijn de belangrijkste intrinsieke regulatoren van de HSC generatie en differentiatie. Gata2 is een essentiële transcriptiefactor dat HSCs generatie aanstuurt.

Om de rol van Gata2 tijdens de embryonale hematopoiese te begrijpen is het essentieel om de kenmerken van de cellen waar Gata2 actief is te bestuderen. Tot nu toe was het niet mogelijk om de cellen dat Gata2 tot expressie brengen te isoleren en te analyseren in *in vivo* en *in vitro* functionele assays of te volgen *in vivo* met live time lapse imaging. Te dien einde, heb ik in mijn proefschrift onderzoek een nieuwe Gata2 verslaggever embryonale stamcel (ES) cellijn en een muismodel gegenereerd waarin Gata2 expressie wordt gevolgd door de expressie van Venus fluorochroom. Op deze manier is de waarneming van Gata2 expressie brengende cellen en de isolatie ervan mogelijk gemaakt.

In **Hoofdstuk 2** beschrijf ik het ontstaan van deze nieuwe ES-cel reporter lijn en de embryoid body (EB) differentiatie kweek systeem dat wordt gebruikt om het potentieel van deze Gata2 expressie brengende cellen te verkennen. Ik laat zien dat *Venus* en *Gata2* gecoördineerd to expressie komen in onze reporter lijn (G2V5) en toon bovendien aan dat gedurende hematopoietische differentiatie van EBs Gata2 expressie brengende cellen ook endotheliale en hemopoëtische celoppervlakte eiwitten (CD31 en c-Kit, respectievelijk) bezitten. Prospectieve isolatie van Gata2 expressie brengen cellen en functionele toetsen laten zien dat deze cellen endotheliale vermogen in Matrigel kweek systeem bezitten, en ook zeer verrijkt zijn in hematopoietische differentiatie vermogen.

In **Hoofdstuk 3** wordt het genereren van de *Gata2-Venus* reporter muismodel beschreven. Ik laat zien dat *Gata2-Venus* volwassen muizen zijn hematopoietically normaal en dat *Venus* en *Gata2* worden gecoördineerd uitgedrukt. Uit confocale microscopie van de muis embryo blijkt dat Venus fluorochroom expressie is gelokaliseerd aan dezelfde anatomische regio en cellen waarvan aangetoond is dat

deze *Gata2 mRNA* expressie hebben. Flow cytometrische analyses van de hematopoietische lokaties laat zien dat Gata2 expressie brengende cellen in aantallen toenemen in de AGM en foetale lever (FL) tijdens de embryonale ontwikkeling. Bovendien is het aantal CD31^{+}c-Kit^{+}Venus^{+} intraaortic cluster cellen nemen ook af tussen E10 - E11 in overeenstemming met eerder gepubliceerde gegevens over cluster cellen. Functionele analyse van Venus gesorteerd celpopulaties laten zien dat de Gata2 expressie brengende cel fractie is sterk verrijkt in hematopoietische voorlopercellen. Hematopoëtische voorlopers werden ook waargenomen in de Venus negatieve cel fractie. Hematopoietische voorlopercel assays uitgevoerd op *Gata2*$^{-/-}$ embryonale weefsels bevestigen dat Gata2 onafhankelijke hematopoietische voorlopercellen bestaan. Deze resultaten werden verder bevestigd door de identificatie van c-Kit^{+} hematopoietische clusters in de vitelline slagader (VA) van *Gata2*$^{-/-}$ embryo's. Hematopoietische voorlopercel verrijking kan verder worden verbeterd door het combineren van onze *Gata2-Venus* muismodel met *Ly6A-GFP* reporter muis. Met als gevolg, de identificatie van een kleine dubbele positieve celpopulatie (GFP^{+} Venus^{+}) zeer hoog verrijkt met onvolwassen hematopoietische voorlopercellen. Belangrijk is dat HSC activiteit is uitsluitend geobserveerd in de Gata2 expressie brengende cel fractie.

In **Hoofdstuk 4** presenteer ik de genoomwijde bindingsprofiel van tien transcriptiefactoren waarvan is aangetoond dat ze een belangrijke rol spelen bij het reguleren van hematopoiese. Bioinformatische analyses van deze datasets suggereren genoomwijde combinatorische interacties tussen deze belangrijke regulatoren van hematopoiese en suggereren verder het bestaan van een "heptameer" van transcriptiefactoren geassocieerd met hematopoïetische cellen. Deze gegevens suggereren het bestaan van eerder niet geïdentificeerde transcriptiefactoren interacties. Tot slot, functionele experimenten in *Gata2*$^{+/-}$ *Runx1*$^{+/-}$ embryo's tonen aan dat deze twee belangrijke hematopoietische regulatoren hebben een samenwerkend effect in hun functie, wat suggereert dat de gegevens uit de bioinformatische analyse een nuttig hulpmiddel kan zijn bij het begrijpen van transcriptiefactoren samenwerking in het reguleren van specifieke processen zoals hematopoiese.

Kortom, in dit proefschrift beschrijf ik het genereren van een nieuwe Gata2-Venus reporter muis ES cellijn model dat de identificatie van Gata2 expressie brengende cellen in vivo en hun isolatie mogelijk maakt voor de functionele in vivo en in vitro assays. Ik toon aan dat alle HSC activiteit in de muis embryo wordt gevonden in de Gata2 expressie brengende celfractie en dat deze celpopulatie is hoog verrijkt in hematopoëtische voorlopercellen. Gata2 blijkt te werken in een combinatorische wijze samen met andere belangrijke regulatoren van hematopoiese. Het belangrijkste is dat ik voor het eerste keer, door time-lapse imaging, laat zien dat Gata2 expressie op E10 embryonale slagader dynamische veranderingen ondergaat. Deze

studies hebben onze kennis van Gata2 en de regulering en ontwikkeling van hematopoietische stamcellen uitgebreid.

Abbreviations

AGM	Aorta-Gonads Mesonephros region
AML	Acute Myeloid Leukemia
BAC	Bacterial Artificial Chromosome
BL6	Black 6 (mouse strain)
BL-CFCs	Blast Colony Forming Cells
BM	Bone Marrow
BMP	Bone Morphogenetic Protein
bp	base pairs
CAG	Chicken β-actin promoter with CMV enhancer
CBF	Core Binding Factor
CD	Cluster of Differentiation
CFU	Colony Forming Unit
CFU-C	Colony Forming Unit - Culture
CFU-E	Colony Forming Unit - Erythrocyte
CFU-G	Colony Forming Unit - Granulocyte
CFU-GEMM	Colony Forming Unit – Granulocyte, Erythrocyte, Macrophage, Mega-karyocyte
CFU-GM Colony	Forming Unit – Granulocyte, Macrophage
CFU-M	Colony Forming Unit – Macrophage
CMV	Cytomegalovirus
CNS	Central Nervous System
c-Kit	Cellular Kit
DA	Dorsal Aorta
DNA	Deoxyribonucleic acid
E	Embryonic day
ee	embryo equivalent
EB	Embryoid Body
EGM-2	Endothelial Growth Medium-2
ER	Estrogen Receptor
ES cells	Embryonic Stem cells
FACS	Fluorescent-Activated Cell Sorting
FCS	Fetal Calf Serum
FDCP	Factor Dependent Cell Progenitors
FGF	Fibroblast Growth Factor
FL	Fetal Liver
Fli 1	Friend leukemia integration 1 (transcription factor)
Flk-1	Fetal liver kinase - 1
Gfi1	Growth factor independece 1 (gene)

GFP	Green Fluorescent Protein
Hh	Hedgehog
HSC	Hematopoietic Stem Cell
HPC-7	Hematopoietic Precursor Cell-7
HTN-Cre	His tag, Tat peptide, NLS (Nuclear Localization Signal) - Cre
IRES	Internal Ribosome Entry Site
Lmo2	LIM domain only 2
Ly-6A	Lymphocyte antigen 6A
MEF	Mouse Embryonic Fibroblast
MegE	Megakaryocyte, Erythrocyte
MonoMac	Monocytopenia and mycobacterial infection
NK	Natural Killer (cells)
Npm1	Nucleophosmin 1
NT	Neural Tube
OB	Olfactory Bulb
PE	Phycoerythrin
PGK	Phosphoglycerate kinase (promoter)
PL	Placenta
qRT-PCR	quantitative Reverse Transcriptase – Polymerase Chain Reaction
RHD	Runt Homology Domain
RNA	Ribonucleic Acid
Runx1	Runt-related transcription factor 1
Sca-1	Stem cell antigen-1
Sp	somite pairs
UA	Umbilical Artery
UG	Urogenital Regions
UTR	Untranslated Region
VA	Vitelline Artery
VEC	Vascular Endothelial Cadherin
VEGF	Vascular Endothelial Growth Factor
WT	Wild Type
YAC	Yeast Artificial Chromosome
YS	Yolk Sac
ZnF	Zink Finger

Curriculum Vitae

Personal details

Name:	Polynikis Kaimakis
Birth date:	27-09-1984
Birth place:	Thessaloniki Greece

Education and research

2008-2013	ESI Stem Cell Institute Department of Cell Biology Erasmus MC, Rotterdam, the Netherlands (Prof dr E.A. Dzierzak) PhD student: Genetic Gata2 in embryonic hematopoiesis
2002-2007	Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupoli, Greece BSc in Molecular Biology and Genetics
03/2006 – 08/2006	Internship Department of Cell Biology, Erasmus MC, Rotterdam, The Netherlands Undergraduate senior year diploma thesis: “Elucidation of the molecular basis of atypical β-thalassemia patients with high fetal hemoglobin levels” (dr G. Patrinos)
1999-2002	Graduate student of 1st Senior High School of Kalamaria in Thessaloniki. Greece High school education

PhD Portfolio	
Summary of PhD training and teaching	
Name PhD student: Polynikis Kaimakis	PhD period: May 2008 – Dec 2012
Erasmus MC Department: ESI Erasmus Stem cell Institute, Cell Biology	Promotor: Prof. Dr Elaine Dzierzak
Research School: MGC graduate school	
1. PhD training	Year
General courses	
- Laboratory animal science course(Article 9), 2008	2008
- Safety working in the laboratory, 2008	2008
- Radioactivity course, 2008	2008
- Molecular and cell biology course, 2009	2009
Specific courses	
- Winterschool Kleinwalsertal (oral presentation contribution)	2009-2010
- ESI retreat (oral presentation contribution)	2011
Seminars and workshops	
- MGC PhD student workshop	2009
- 19th MGC-symposium, Rotterdam, the Netherlands	2009
- Erasmus lectures in cell biology and development	2008-2012
- Erasmus lectures on stem cell and regenerative medicine	2008-2012
- Monday morning meetings	2008-2012
- ESI site visit	2010
Presentations	
- Monday Morning Meetings	2008-2012
- Work discussion (lab meetings)	2008-2012
Internatinoal Conferences	
- 2nd International Symposium, Stem Cells, Development and Regulation, Amsterdam 2008	2008
- DSSCR 2nd Dutch Stem Cell meeting, Rotterdam	2009
- ISEH 38th Annual Scientific meeting, Athens (poster presentation)	2009

- 3rd International Symposium, Stem Cells, Development and Regulation, Amsterdam	2009
- DSSCR 3rd Dutch Stem Cell meeting, Utrecht	2010
- 4th International Symposium, Stem Cells, Development and Regulation, Amsterdam (poster presentation	2010
- ISEH 39th Annual Scientific meeting, Vancuver, Canada (oral presentation)	2011
- ISEH 40th Annual Scientific meeting, Amsterdam, the Netherlands (oral presentation, poster presentation)	2012
2. Teaching	
Supervising practicals and excursions, Tutoring	
- High School students (occasionally)	2009-2012
- Master thesis student	2012

Publications

Kaimakis P, Crisan M, Dzierzak E.(2012) The biochemistry of hematopoietic stem cell development Biochim Biophys Acta

Wilson NK, Foster SD, Wang X, Knezevic K, Schütte J, Kaimakis P, Chilarska PM, Kinston S, Ouwehand WH, Dzierzak E, Pimanda JE, de Bruijn MF, Göttgens B. (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell

Robin C, Bollerot K, Mendes S, Haak E, Crisan M, Cerisoli F, Lauw I, Kaimakis P, Jorna R, Vermeulen M, Kayser M, van der Linden R, Imanirad P, Verstegen M, Nawaz-Yousaf H, Papazian N, Steegers E, Cupedo T, Dzierzak E. (2009) Human placenta is a potent hematopoietic niche containing hematopoietic stem and progenitor cells throughout development. Cell Stem Cell

Papachatzopoulou A, Kaimakis P, Pourfarzad F, Menounos PG, Evangelakou P, Kollia P, Grosveld FG, Patrinos GP. (2007) Increased gamma-globin gene expression in beta-thalassemia intermedia patients correlates with a mutation in 3'HS1. Am J Hematol.

Giardine B, van Baal S, Kaimakis P, Riemer C, Miller W, Samara M, Kollia P, Anagnou NP, Chui DH, Wajcman H, Hardison RC, Patrinos GP. (2007) HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. Hum Mutat

van Baal S, Kaimakis P, Phommarinh M, Koumbi D, Cuppens H, Riccardino F, Macek M Jr, Scriver CR, Patrinos GP. (2007) FINDbase: a relational database recording frequencies of genetic defects leading to inherited disorders worldwide. Nucleic Acids Res.

Acknowledgements

This is a part of the thesis I never thought I would have problem writing but somehow now I find it difficult, and that is mostly because I cannot thank enough all the people who helped me surpass all of the various problems that someone will cross during the process of a PhD.

First of all, I would like to thank Prof.dr E.A. Dzierzak. Thank you Elaine for giving me the opportunity to explore this so interesting and promising project and for opening for me the door to this fascinating world of hematopoiesis and stem cells! Thank you for being so patient with me and for doing all you could do to provide me with the scientific environment (through wonderful lectures, meetings and congresses) that is so necessary for a PhD student. Finally, thank you for being so kind and taking care of my needs a PhD student and as a human being.

Dies, thank you for reading, correcting and commending on my thesis. I always enjoyed all the scientific conversations we had because they were always up to the point, precise and were giving solutions or directions. I would also like to thank you for the great cooking and the interesting conversations we had during the nice dinners in your house or garden.

Sjaak, thank you for agreeing to be in my small committee and for reading and commending on my thesis and above all thank you for organizing the Kleinwalsertal meeting, a wonderful environment for learning about other people's work, communicating, socializing and sports! I had great time there as everybody else!

Raymond, thank you for agreeing to be in my small committee, reading my thesis and providing your corrections and your useful suggestions. Also, thank you for organizing this very interesting series of talks with Dutch scientists.

Special thanks to George Patrinos, my first supervisor in Erasmus MC during my internship for my bachelor thesis! George I will always be grateful to you for your help and support in my first steps in the Netherlands!

And now I will begin with the old school first! Catherine, thanks a lot for all of your advices and for being always friendly, kind and willing to help! Also, thanks for sharing the time lapse imaging technique! Esther, I consider you to be my first teacher (and friend) in the lab practice! Your help in the beginning was invaluable. Unfortunately, I do not think that you would be very proud of your student since I can hardly reach any of your organization and planning skills even after four and a half years in the lab! Francesco, thanks for being always there to give me a hand and an advice when I needed it and for being a good company in the lab! Yoko, thank you for your literature advices and of course for the football discussions we had! I miss them both! Karine (the nicest blue eyes ever!) I always enjoyed your company!

Emma, I would like to honestly thank you for all of your really important

help in experiments, discussions, writing, analyzing and everything else I cannot think of at the moment. Your presence in the lab made my life easier!

Mihaela, you were always an excellent example in the lab with your hard working attitude, your organization skills and your high spirit! Life is more beautiful now!

Tomoko, thanks a lot for the help in the whole mount imaging, the confocal and the scientific discussions. Also thanks for the company and the discussions in the difficult nights late in the lab! It was important support!

Dorota, I think you are a remarkable person with a positive aura surrounding you!!! You also compete with Tomoko for the kindness award!

Derk and Gerben, thank you both for any help, suggestion and advice in the lab meetings discussions the 4,5 years! Your experience was always more than welcome! And Derk also thanks for the Wnt and ES cell culturing advices!!!

PhD students (master students and technicians)!!! The most important people of my everyday life in the lab!!! Shared with you everything! Frustration, anger, success, happiness and any other emotion that you can possibly experience during a PhD period.

JC! I am proud to have worked with you! Thanks for music sharing and good company in Athens and other conferences! Parisa! It has been a fun four years next to you in the office and the lab! Keep going and enjoy US! Fanny! Always useful in multiple ways! I had a great time in any occasion inside and outside of the lab! Thomas! Thank you for every fun moment in the lab (and they are many with you!) Nesrin! I really enjoyed your company! Wish you to always be cheerful! Melody, you little sweet creature, thank you for every smile in the corridors! Wish you the best of luck and success in US! Keep going! Chris! Thanks for all the help and your music taste sharing! Reinier, thanks for your help and patience! Alvin! Great guy! Always nice and helpful! Was a pleasure working in the same environment with you! Alex! You are a positive presence in the lab!

Manny thanks to Siska and Martine for all the technical advices and help in the cloning and ES cell culturing! I should not forget to mention Alex Maas and John Kong a San for blastocyst injections and animal work and also every person of the EDC for taking care of our precious mice!

Special thanks to my friends Pavlos and Parham! Παυλάρα! Ήσουν πάντα παράδειγμα και φίλος! Ότι και να πω θα είναι λίγο! Ευχαριστώ για όλα! Parham! I cannot express how important you were in the lab for me! Thank you for all the discussions, for the support and for being a good friend and letting me share with you all of my problems, thoughts and happiness! Your presence was invaluable for me!

Every single person I met the past 4,5 yours in the 10th ,7th and 9th floor!!! Thank you from the bottom of my heart! Even being there was important for me! I would forget a lot of people for sure if I start listing them and that would be unfair so I will skip this point! I just have to thank Petros and Farzin because they were the first persons who took care of me during my internship in 2006! Thank “mastores”!

Σοφούλα! Τζούτζουμπρούτζου και ευχαριστώ για την παρέα και όλο το γέλιο που έχουμε ρίξει μαζί! Αριστέα! Thanks για την κουβέντούλα και την παρέα! Μαρία και Πέτρο ευχαριστώ για την παρέα και την παρουσία σας! Πάντα είναι σημαντικό να μπορείς να επικοινωνήσεις με κάποιον που μπορεί να σε καταλάβει πιο εύκολα!

Finally, I would honestly like to apologize to everyone in the lab and in the department that I annoyed or caused problems to! I never had any intention of doing so!

Last but not least! Ανεστάκη και Κατερίνα! Τα πολύτιμα κουμπαράκια μου! Χριστάκη! Χωρίς εσάς δε ξέρω αν θα τα κατάφερνα στην «κακούργα ξενιτιά»! Πάντα ήσασταν και είστε τα πρόσωπα που θα με ξελασπώσουν από την κούραση και τα ψυχολογικά αδιέξοδα!

Ένα μεγάλο ευχαριστώ σε όλους τους φίλους στην Ολλανδία! Και μόνο το να σας δω την Κυριακή το πρωί ήταν ένα κίνητρο για μένα για να βγει κάθε δύσκολη βδομάδα! Ευχαριστώ επίσης όλους τους φίλους σε ελλάδα και εξωτερικό! Και μόνο ότι δε μας ξεχνάτε είναι πολύ σημαντικό! Ευχαριστώ τον πνευματικό μου και όλους τους πατέρες στο μοναστήρι για τις ευχές τους και για την παρουσία τους στη ζωή μου!

Τέλως, θα ήθελα να ευχαριστήσω την οικογένειά μου για όλες τις δυσκολίες που τους προξένησα, για όλη τη στήριξη και την αγάπη και για κάθε λόγο και βοήθεια που μου έδιναν απλόχερα όλο αυτόν τον καιρό!

THANK YOU ALL!!!