
Andreia Marques Gomes

# ELUCIDATING BLOOD ORIGIN: FROM REPROGRAMMING TO ONTOGENY

Tese de Doutoramento em Biologia Experimental e Biomedicina, ramo Biotecnologia e Saúde,
orientada pelo Doutor Filipe Pereira e pelo Doutor Lino Ferreira, apresentada ao Instituto de Investigação Interdisciplinar da Universidade de Coimbra.

Abril 2018

· U C ·

UNIVERSIDADE DE COIMBRA

# Elucidating Blood Origin: From Reprogramming To Ontogeny

by Andreia Marques Gomes

Tese de Doutoramento em Biologia Experimental e Biomedicina, ramo Biotecnologia e Saúde, orientada pelo Doutor Filipe Pereira e pelo Doutor Lino Ferreira do Centro de Neurociências e Biologia Celular (CNC) da Universidade de Coimbra e apresentada ao Instituto de Investigação Interdisciplinar da Universidade de Coimbra. Este trabalho foi desenvolvido no UC-Biotech do CNC da Universidade de Coimbra, Portugal e no Icahn School of Medicine no Mount Sinai em Nova Iorque, Estados Unidos da América sob a supervisão do Professor Ihor R. Lemischka e Doutora Kateri Moore.

Doctoral thesis in Biomedicine and Experimental Biology in the field of Biotechnology and Health, supervised by Dr. Filipe Pereira and by Dr. Lino Ferreira from the Center of Neurosciences and Cell Biology (CNC) of the University of Coimbra, and presented to the Institute of Interdisciplinary Research of the University of Coimbra. This study was conducted at UC-Biotech, CNC, Portugal and at Icahn School of Medicine at Mount Sinai, New York, United States of America under the supervision of Professor Ihor R. Lemischka and Associate Professor Kateri Moore.

The study presented in this dissertation was supported by an individual PhD Fellowship from the Portuguese Foundation for Science and Technology (FCT) (SFRH/BD/51968/2012), attributed by the PhD Programme in Experimental Biology and Biomedicine (PDBEB) and by the research project (PTDC/BIM-MED/0075/2014).

"*omnis cellula e cellula*"

("every cell stems from another cell")

Rudolf Virchow

## Abstract

Hematopoiesis, the process by which all blood cell types are formed, rely on self-renewing and multipotent hematopoietic stem cells (HSCs) that arise from specialized endothelial cells by a process termed endothelial-to-hematopoietic transition (EHT). The generation of autologous long-term self-renewing HSCs *in vitro* as replacement therapy for blood disorders has been considered the holy grail of regenerative medicine. However, to achieve this, it is important to understand the complex genetic program driving HSC emergence during development.

The cellular reprogramming of mouse somatic cells into hematopoietic progenitors can be achieved by the enforced expression of four transcription factors – Gata2, Gfi1b, cFos, and Etv6 – and, it offers a powerful tool to better understand developmental processes and the underlying mechanisms. The induction of hematopoietic progenitors from fibroblasts progresses through hemogenic precursors that are $Prom^{+}Sca^{+}CD34^{+}CD45^{-}$ ($PS34CD45^{-}$). In this thesis, mouse placentas have shown to harbor a population with this phenotype that express endothelial and early hematopoietic markers and the global gene expression profile of $PS34CD45^{-}$ correlates with reprogrammed precursors. Upon stromal co-culture, $PS34CD45^{-}$ cells have shown to give rise to all blood cell lineages and engraft primary and secondary immunodeficient mice, establishing direct reprogramming as a model that can recapitulate developmental hematopoiesis.

Here, we also demonstrate that human fibroblasts can be reprogrammed into hemogenic cells by the same transcription factors. Induced cells display dynamic EHT transcriptional programs, generate hematopoietic progeny, express an HSC phenotype and repopulate immunodeficient mice. Mechanistically, GATA2 and GFI1B interact and co-occupy a cohort of targets. This cooperative binding is reflected by the engagement of open enhancers and promoters, initiating the silencing of fibroblast genes and activating the hemogenic program. However, GATA2 displays dominant and independent targeting activity during the early phases of reprogramming. These findings shed light on the processes controlling HSC specification and support the generation of reprogrammed HSCs for clinical applications.

## Resumo

O processo que gera todos os tipos de células do sangue, designa-se por hematopoiese e está dependente de células estamininais hematopoiéticas (CEH) multipotentes que possuem a capacidade de auto-renovação. Estas CEH emergem a partir de células endoteliais especializadas num processo designado por transição endotelial-hematopoiética. A geração de CEH autólogas com capacidade reconstitutiva a longo prazo *in vitro* para uso em terapia celular de doenças sanguíneas tem sido considerada o Santo Graal da medicina regenerativa, no entanto, para que tal aconteça é necessário primeiro desvendar o complexo programa genético que origina as CEH durante o desenvolvimento embrionário.

A reprogramação celular de células somáticas de ratinho em células hematopoiéticas progenitoras foi já estabelecido com a indução da expressão de quatro factores de transcrição – Gata2, Gfi1b, cFos e Etv6- que proporciona um método vantajoso para o estudo de processos biológicos e respectivos mecanismos durante o desenvolvimento embrionário. A indução de células hematopoiéticas progenitoras a partir de fibroblastos progride através de precursores hemogénicos que são $Prom^{+}Sca^{+}CD34^{+}CD45^{-}$ (PS34CD45$^{-}$). Neste trabalho, foi encontrada, em placentas de ratinho, uma população com este fenótipo que expressa marcadores endoteliais e hematopoiéticos e o perfil global de expressão genética das células PS34CD45$^{-}$ está associado com o de células percursoras reprogramadas. Após co-cultura com células do estroma, as PS34CD45$^{-}$ geraram toda a linhagem de células sanguíneas e formaram enxertos em ratinhos immunodeprimidos primários e secundários, estabelecendo a reprogramação directa como um modelo que consegue recapitular a hematopoiese durante o desenvolvimento.

Para além disso, esta tese demonstra que a reprogramação de fibroblastos humanos em células hemogénicas é feita pelos mesmos factores de transcrição e que as células induzidas apresentam um perfil transcricional dinâmico que mimetiza a transição endotelial-hematopoiética, geram descendência hematopoiética, expressam o fenótipo das CEH e enxertam ratinhos imunodeprimidos. Mecanisticamente, o GATA2 e o GFI1B interagem fisicamente e ocupam um elevado número de alvos genómicos. Esta cooperação reflete-se na ligação inicial a promotores e *enhancers* abertos que leva ao silenciamento de genes específicos de fibroblastos e à activação do programa hemogénico. Contudo, o GATA2 apresenta uma capacidade de ligação dominante e independente na fase inicial de reprogramação relativamente aos restantes factores de transcrição. Estas descobertas abrem novos caminhos para uma melhor percepção dos processos que controlam a especificação das CEH e suportam a geração de CEH reprogramadas que poderão ter uma aplicação clínica.

# Acknowledgements

This thesis also represents the commitment and hard work of many people, and here I would like to express my gratitude for your support.

Firstly, I would like to thank my supervisor, Dr. Filipe Pereira, who has been a great mentor and teacher. Thank you for your dedication and determination and for your guidance throughout the entire PhD. I also want to thank you for trusting me with responsibility and freedom, for the constant encouragement and for teaching how to see science beyond the experiments.

I would like to address a special thanks to my external supervisors, the late Professor Ihor Lemischka and Professor Kateri Moore, for the opportunity to develop this project at Mount Sinai, for the vast experience and steadfast support and for providing a positive working environment. Teri, thank you, for trusting me and for accepting me into your lab. Ihor, you will be very missed. I was very fortunate to work with both of you and, words cannot express fully my gratitude.

I would like to thank my collaborators, especially the late Dr. Dmitri Papatsenko for your invaluable bioinformatic input and for your guidance in this new world. Your expertise and feedback always brought a new level to the constantly mutating field of bioinformatics. Dmitri, you too will be very missed.

I would like to thank my internal supervisor, Dr. Lino Ferreira, for the opportunity and for the support in the beginning of this project.

I would like to acknowledge the Doctoral Programme in Biomedicine and Experimental Biology (BEB) for this opportunity and to thank the Portuguese Foundation for Science and Technology (FCT) for funding. A special thanks to my BEB fellows, with whom I have shared moments of big enthusiasm as well as anxiety. Thank you all for your help, support, and friendship.

I am indebted to all my colleagues and friends who have helped me during my journey at Mount Sinai, including Dung-Fang Lee, Avinash Waghray (Nash), Betty Chang, Huen Suk Kim (Crystal), Jeffrey Bernitz (Jeff), Michael Daniel (Mike), Yesai Fstkchyan, Ye Yaun (James), Ran Brosh, Christoph Schaniel, Timothy Blenkinsop (Tim), Marie Fernandes, Lauren Schiff, Megan Hogan, Ryan Wagner, Mingma Sherpa, John Wen, Ashish Jain, Julian Valdes, Miguel Fidalgo, Diana Guallar, Carlos Sanchez-Priego, Carolina Perdigoto, Alexandre Gaspar-Maia, Carolina Bigarella, Pauline Rimmele, Raymond Liang (Ray), Gaetan Barbet, Genis Camprecios and Sunita D'Souza. Thank you for all the discussions and special moments inside and outside the lab!

I would like to thank the Cell reprogramming lab members for welcoming me and for the helpful advice and discussions.

Diogo Bernardo, thank you for designing the best cover ever!

*Por último, o mais importante, agradecer à minha família. Aos meus pais, o meu sincero agradecimento pelo apoio incondicional, pela paciência, pelo exemplo de perseverança e acima de tudo pelos ensinamentos de uma vida. Aos meus eternos miúdos, Gonçalo e Catarina, continuem sempre a sonhar! O meu agradecimento especial à tia (madrinha) Eugénia que me recebeu nos Estados Unidos de braços abertos e me fez sentir em casa e aos primos Diogo e Fábio.*

## Contributions

The majority of FACS sorts were performed in the FACS core facility at Mount Sinai by the operators Jordi Ochando and Christopher Bare.

All high throughput sequencing from RNA-seq and ChIP-seq was performed in the Genomics core facility at Mount Sinai with Milind Mahajan and in Girihlet,inc. with Anitha Jayaprakash.

# Table of Contents

# List of figures

# List of tables

# Common nomenclature abbreviations

ACE – Angiotensin Converting Enzyme
AGM – Aorta-Gonad-Mesonephros
AP-1 – Activator Protein-1
APC – Allophycocyanin
BJ – Neonatal foreskin fibroblasts
BM – Bone Marrow
BSA – Bovine Serum Albumin
bZIP –Basic leucine loop
CALC- Cobblestone Area Forming Cell
CD – Cluster of differentiation
cDNA – complementary DNA
CFC – Colony Forming Cell
CFU – Colony Forming Unit
CFU-S – Colony Forming Unit - Spleen
ChIP – chromatin immunoprecipitation
ChIP-seq – Chromatin ImmunoPrecipitation sequencing
CMP – common myeloid progenitor
CpG – Cytosine guanine dinucleotide
DAPI - 4',6-diamidino-2-phenylindole
DHS – DNase I hypersensitive sites
DMEM – Dulbecco's Modified Eagle Medium
DNA – Deoxyribonucleic acid
DNase I – Deoxyribonuclease I
EB – Elution Buffer
ECL – Enhanced Chemiluminescence Substrates
EHT – Endothelial-to-Hematopoietic Transition
EMT- Epithelial-to-mesenchymal transition
ESC – Embryonic Stem Cell
FACS – Fluorescence-Activated Cell Sorting
FBS – Fetal Bovine Serum
FITC – Fluorescein isothiocyanate
FT – Fetal liver
G-CSF – Granulocyte Colony-Stimulating Factor
GFI1B – Growth Factor Independent 1B
GGF – GATA2, GFI1B and FOS
GGFE – GATA2, GFI1B, FOS and ETV6
GMP – Granulocyte/macrophage progenitor
GO – Gene Ontology
GSEA – Gene Set Enrichment Analysis
GVHD – Graft Versus Host Disease
HDF – Human Dermal Fibroblasts
HE – Hemogenic Endothelium
HIF1B – Hypoxia Inducible Factor 1B
HLA – Human Leucocyte Antigen
HPC – Hematopoietic Progenitor Cell
HSC – Hematopoietic Stem Cell
HSCT – Hematopoietic Stem Cell Transplantation
HSPC – Hematopoietic Stem and Progenitor Cell
IAHC – Intra Aortic Hematopoietic Cluster
IF – Immunofluorescence
IP – Immunoprecipitation
iPSC – induced Pluripotent Stem Cell

ITGA6 – Integrin alpha 6
LTRC – Long-Term Repopulating Cell
MACS – Magnetic-Activated Cell Sorting
MGI – Mouse Genomics Informatics
mL – milliliter
mM – millimolar
MEP – megakaryocyte/erythroid progenitor
MET - Mesenchymal-to-epithelial transition
MPP – Multipotent Progenitor
MLP – MultiLymphoid Progenitor
NK – Natural Killer
NOG – NOD.Cg-Prkdc$^{scid}$Il2rγ$^{tm1Sug}$
NSG – NOD.Cg-Prkdc$^{scid}$ Il2rγ$^{tm1Wjl}$
OSKM – Oct4, Sox, Klf4, c-Myc cocktail
PBS – Phosphate Buffer Saline
PCA – Principal Component Analysis
PCR – Polymerase Chain Reaction
PE – PhycoErythrin
PFA – Paraformaldehyde
P-Sp – Para-aortic Splanchnopleura
RNA – Ribonucleic Acid
RNA-seq – RNA-sequencing
SEM – Standard Error Mean
TAD – Transactivation Domain
TBS – Tris-Buffered Saline
TBST – Tris-Buffered Saline with Tween
TF – Transcription Factor
TSS – Transcription Start Site
TTS – Transcription Terminal Site
UCB – Umbilical Cord Blood
WB – Western Blot
YS – Yolk Sac
ZF – Zinc Finger
µg – microgram
µL – microliter
µm – micrometer
µM – micromolar

## Publications emerging from this work

Carlos-Filipe Pereira*, Andreia M. Gomes*, Betty Chang, Ilia Kurochkin, Michael Daniel, Kenneth Law, Namita Satija, Alexander Lachmann, Zichen Wang, Lino Ferreira, Avi Ma'ayan, Benjamin Chen, Dmitri Papatsenko, Kateri A. Moore and Ihor R. Lemischka. Human Hemogenic Reprogramming is Mediated by Cooperative Transcription Factor Induction. *In revision.*

* Equal contribution

Carlos-Filipe Pereira, Betty Chang, Andreia Gomes, Jeffrey Bernitz, Dmitri Papatsenko, Xiaohong Niu, Gemma Swiers, Emanuele Azzoni, Marella F.R.T. de Bruijn, Christoph Schaniel, Ihor R. Lemischka, and Kateri A. Moore (2016). Hematopoietic Reprogramming *in vitro* informs *in vivo* identification of hemogenic precursors to definitive hematopoietic stem cells. Developmental Cell,*36*, 525–539. https://doi.org/10.1016/j.devcel.2016.02.011

# Chapter 1 – General introduction

## 1.1. Hematopoiesis

Every single-cell of a living organism needs a continuous intake of nutrients and gases as well as secretion of its metabolites. The development of any organism begins with the formation of a unicellular zygote and as it passes through consecutive mitotic divisions, the cell number increase which is followed by the organization of the specialized cells into distinct tissues. In vertebrates, the development of a blood system with a complex network of vessels fulfilled the needs of a multicellular organism independent of its size and complexity. Since the blood system is essential in sustaining the living organism, a key question is how the blood cells are generated.

The process of blood formation is known as hematopoiesis and is dependent on a rare cell type: hematopoietic stem cell (HSC). The idea of a unique cell generating the entire blood offspring was suggested in the late 19$^{th}$ century (reviewed by Ramalho-Santos and Willenbring, 2007), but this hypothesis was only experimentally verified seven decades later by the seminal work of McCulloch, Till and colleagues by demonstrating the presence of hematopoietic stem cells in a population of mouse bone marrow cells (Becker et al., 1963; Till and McCulloch, 1961). A series of retroviral marking HSCs studies provided important information supporting the existence of a common single stem cell for all lineages capable of long-term sequential clonality (Dick et al., 1985; Jordan and Lemischka, 1990; Lemischka et al., 1986).

HCSs are the most well studied adult stem cells and are at the top of the hematopoietic system (Figure 1.1). HSCs reside in the bone marrow and are responsible for maintaining continuous hematopoietic cell production throughout life. Similarly, to other stem cells, HSCs are capable of self-renewal, by maintaining the production of additional HSCs, and of differentiation into all blood lineages (Lemischka, 2005; Weissman, 2000). During cell division, HSCs give rise to multipotent progenitors (MPP) by losing self-renewal capacity and as a consequence, the differentiation events begin by sequential lineage commitment steps and the intermediate cells become more restricted within a lineage path (Figure 1.1) (Morrison et al., 1997).

The maintenance of HSC pool and lineage differentiation are balanced either by divisional asymmetry, in which specific cell fate factors are distributed to one of the two descendants cells; or by environmental asymmetry, in which one descendant cell leaves the niche that sustains HSC self-renewal and is then exposed to an

environment that stimulates lineage differentiation (Wilson and Trumpp, 2006). In the bone marrow, HSCs reside in specific niches that provide a specialized microenvironment for the regulation of these cells. At least, there are two well characterized niches, the osteoblastic or endosteal and the vascular niches, that exist to support HSCs survival, self-renewal and differentiation (Nakamura-Ishizu and Suda, 2013).

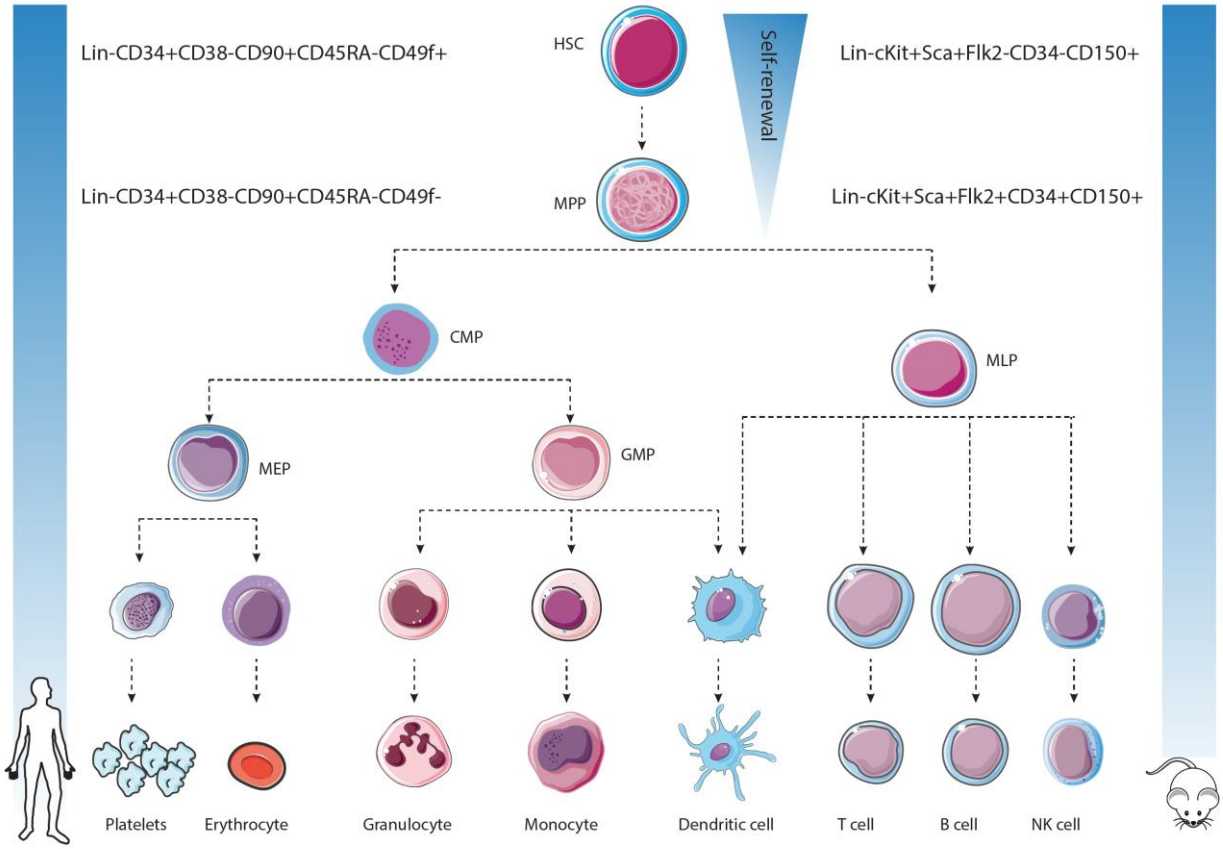



Figure 1.1- **The hematopoietic system**. Hierarchical simplified graphic of lineage fate during mouse and human hematopoiesis. The phenotypic cell surface marker for HSC and MPP population for mouse (right side) and human (left side) hematopoietic system is showed. Self-renewal capacity decreases progressively as HSCs start the differentiation trail. HSC, hematopoietic stem cell; MPP, multipotent progenitor; MLP, multi-lymphoid progenitor; CMP, common myeloid progenitor; MEP, megakaryocyte/erythroid progenitor; GMP, granulocyte/macrophage progenitor. HSC and MPP phenotype reviewed by Chotinantakul and Leeanansaksiri, 2012.

### 1.1.1. Clinical Relevance

HSC is also the functional unit of the hematopoietic stem cell transplantation (HSCT), a procedure that was originally developed to treat injury from irradiation in late 1950s and has now become standard for many disorders of the hematopoietic system including cancer (Copelan, 2006). According to the European Society for Blood and Marrow Transplantation 2014, the application of HSCT has increased over the last 20 years for both autologous and allogenic (Passweg et al., 2016). By definition, autologous relates to the collection of the patient's stem cells and later returned to the same patient. In allogenic transplantation, the stem cells transplanted are not genetically identical to the recipient patient, thus, increasing the risk of graft-versus-host disease (GVHD) (Copelan, 2006). 43% of all transplantations done in Europe are allogenic which requires Human Leucocyte Antigen (HLA) compatible units for transplantation (Kolb, 2008; Passweg et al., 2016). Currently, HSCs can be collected from three main sources: bone marrow, peripheral blood and umbilical cord blood, and nowadays, the more commonly used is mobilized cells from peripheral blood which is obtained with the cytokine granulocyte colony-stimulating factor (G-CSF) (Motabi and DiPersio, 2012; Nicolini et al., 2004). However, cord blood has a major advantage over the other sources due to a lower incidence of chronic GVHD (Gutman et al., 2016).

Recent efforts have been made to establish cord blood banks; however, there is a lack of suitable HLA-matched donors that can lead to an increase risk of graft failure, prolonged immunosuppression and delayed recovery (Norkin et al., 2013). As cultured HSCs *ex vivo* lose both their self-renewal capacity and their engraftment potential (Mikkola and Orkin, 2006), alternative sources of HSCs are required to fulfill the current needs.

### 1.1.2. Assessing functional HSCs

In both mouse and human, HSCs represent only 0.05% of the total cells in the bone marrow (Morrison et al., 1995) and yet the hematopoietic system is one of the most highly regenerative tissues with approximately one trillion cells emerging from the bone marrow (Doulatov et al., 2012). In the early 1960s, Till and McCulloch demonstrated the clonal capacity of marrow cells by analyzing their ability to produce colonies with myeloid and lymphoid cells in the spleen of the irradiated mice. In addition, they termed the cells with the ability to form colonies as “colony-forming units” or CFU (Till and McCulloch, 1961). These pivotal studies defined the concept of stem cell and established a functional method of hematopoiesis. After that, a number of methods have been created to measure the functional potential of HSCs in vitro and *in vivo* including Cobblestone Area Forming Cell (CAFC), colony forming cell (CFC) assay and long-term repopulating cells (LTRC) assay. CAFC assay allow the identification of primitive progenitor cells by plating the cells in a supportive stromal layer while CFC uses a viscous medium generally methylcellulose supplemented with cytokines to measure the proliferative capacity and the ability to differentiate in other cell types (Coulombel, 2004). An HSC is also characterized functionally by its ability to restore the hematopoietic system of an immunocompromised recipient (Wilson and Trumpp, 2006). The transplantation of hematopoietic cells into the murine system remains the gold standard to measure HSC activity *in vivo*.

The generation of humanized models in the late 1980s has provided invaluable information on the development of the immune system from HSCs (Willinger et al., 2011). For efficient engraftment, it is important to meet some pre-requisites such as 1) sufficient number of donor cells; 2) lack of immune rejection by the host; 3) appropriate physical space for the graft tissue; and 4) appropriate environment to support the graft (Rongvaux et al., 2013). Over the past 3 decades, gradual modifications have been made to develop more sophisticated humanized mouse (Willinger et al., 2011). The diversity of the strains reflects the technical challenges that have been faced throughout the years, despite the lack of a single model that mimics closely human hematopoiesis and immune system, the current models can address different biological questions (Shultz et al., 2007). Mice immunodeficiency was a major step forward to support long-term human engraftment. This advent was

achieved with mutations in the IL-2 receptor common γ chain causing the absent of natural killers (NK) cells and mutations in prdkc gene resulting in lack of T and B cells. Mice with deletion NOD.Cg-Prkdc$^{scid}$Il2rγ$^{tm1Sug}$ (NOG) (Cao et al., 1995) or truncation NOD.Cg-Prkdc$^{scid}$ Il2rγ$^{tm1Wjl}$ (NSG) (Ohbo et al., 1996) of common cytokine receptor gamma chain are highly immunodeficient (Ito, 2002). Comparison of HSC engraftment showed that NSG is more efficient than NOG mice up to 9-fold (McDermott et al., 2010). NSG mouse is the current most consensual strain for humanization and to improve the long-term multilineage reconstitution different strategies have been applied. Specifically, knock-in of cocktail of important cytokines such as human stem cell factor, granulocyte-macrophage colony stimulating factor and interleukin3 was the strategy used to generate NSG-SGM3 and reported a significant increase of T cells (Billerbeck et al., 2011). In addition, NSG strain allows quantitative assessment of limiting HSCs, as few as 10 HSCs, and secondary transplantation (Park et al., 2008), thus, providing an advantageous model to evaluate human HSC engraftment. Recently, mutation of the Kit receptor enabled robust multilineage engraftment and serial transplantation compare with NSG irradiated mice (Cosgun et al., 2014). Two studies showed that this mouse, NSGW41, also support erythropoiesis (Rahmig et al., 2016; Yurino et al., 2016).

### 1.2. Ontogeny of hematopoietic system

Several vertebrate animal models have been crucial for determining the current knowledge on hematopoietic development including zebrafish, amphibian, avian, and mouse (reviewed in Cumano and Godin, 2007). The external fertilization for zebrafish, amphibian and avian models has allowed an easier monitorization of the embryonic tissues. Furthermore, the transparency of the zebrafish embryos has enabled tracing the emergence, circulation and behavior of blood cells in real-time (Perlin et al., 2017). The murine system has been mostly used due to its mammalian background; in addition, the progress in generation of transgenic mice has strengthened its applicability (Schmitt et al., 2014).

In mammals, hematopoiesis occurs in sequential but overlapping waves to provide the oxygen and nutrient needs as the organism grow and complexity develops. Transient waves with distinct functions fulfill the rapid growth of the embryo and further the fetus,

before accomplish the final system that will sustain the organism throughout life. The multiple waves have been distinguished as primitive and definitive (reviewed in Frame et al., 2013).

Multiple anatomical locations participate in the production of nascent blood cells. This is a common process for other species including mouse (Mikkola and Orkin, 2006) and human (Tavian and Peault, 2005) (Figure 1.2). Seemingly, the separation of the blood cells from the tissues where they will function permits a more controlled process and while the separation seems to be an evolutionary conserved process, the organs for the HSCs production are variable within vertebrates. An example is the kidney as the main hematopoietic site in zebrafish compared to bone marrow for mammals (Ottersbach et al., 2010).

In addition, developmental stages in the mouse can be discriminated by morphological changes taking place in less than 24 hours by comparison, in the human embryo the developmental stages are extended (Ivanovs et al., 2017). Despite the chronology differences during development, similar events between mouse and human can be detected (Figure 1.2).

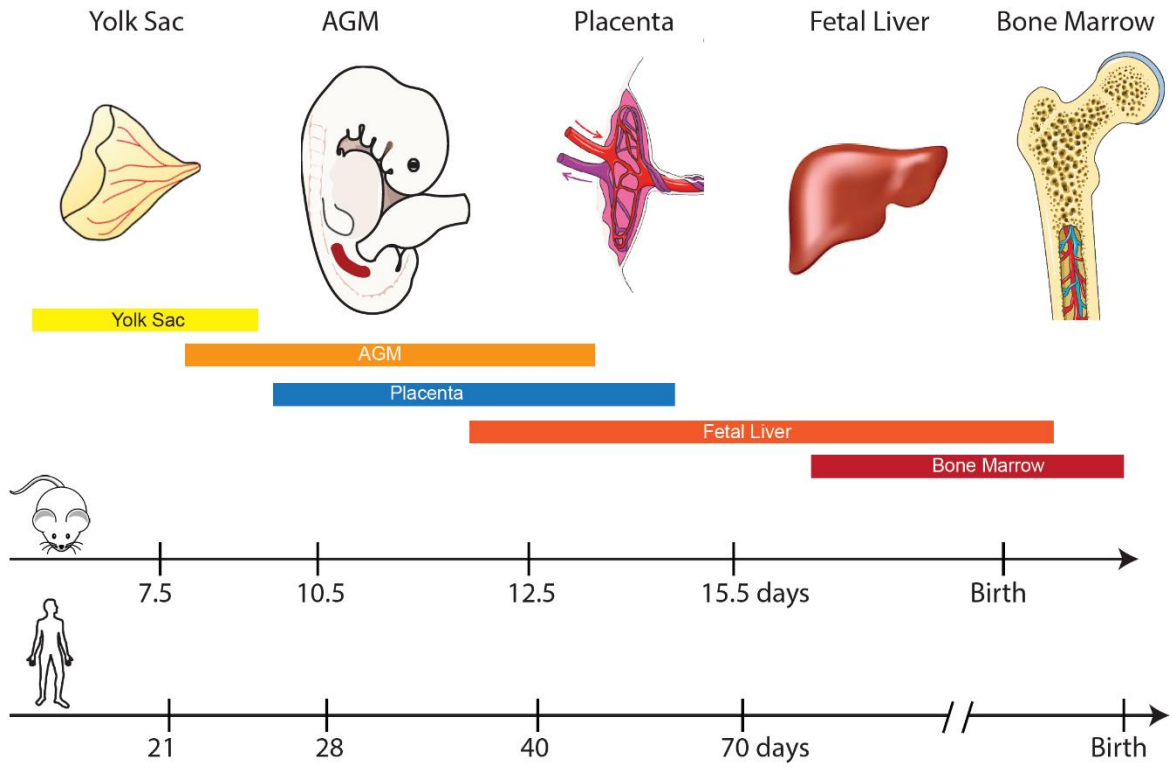



Figure 1.2 **- Chronology and anatomical sites of hematopoietic development**. Emergent cells from yolk sac, placenta and AGM region undergo self-renewal, expansion and differentiation in the hematopoietic niches, fetal liver and bone marrow. The chronology

shows the gestational ages in which the hematopoietic sites are active for both human and mouse. AGM- Aorta gonads-mesonephros. (Adapted from Mikkola and Orkin, 2006).

One of the earliest observations of blood cells was done by Maximov in 1909, who described the emergence of primitive blood cells from mesodermal clusters so called blood islands (Reviewed in Doulatov et al., 2012). Later, Florence Sabin postulated the hypothesis of a common precursor of blood and vascular lineages with a mesoderm origin. She named the precursor cell “angioblast” (Sabin, 1921). In 1932, P. Murray renamed the common precursor as “hemangioblast” by culturing sections of the primitive streak of chick embryos *ex vivo* supporting Florence’s discoveries (Murray, 1932). The existence of the hemangioblast with both vascular and hematopoietic potential was only demonstrated in vitro in the late 20$^{th}$ century in differentiating mouse ESCs (Choi et al., 1998; Kennedy et al., 1997). Hemangioblasts were also identified in zebrafish (Vogeli et al., 2006), in Drosophila (Mandal et al., 2004), and in differentiating human ESC (Kennedy et al., 2007). In addition, hemangioblast-like cells were detected in gastrulating mouse embryos (Huber et al., 2004). The concept of a common origin is reinforced by the expression of common genes between endothelial and hematopoietic lineages and by the lack of blood vessels and blood cells in gene knockout studies. The inactivation of *flk1*, an endothelial marker, disrupted both endothelial and hematopoietic lineages leading to embryonic death (Shalaby et al., 1997). However, it is still controversial in higher vertebrates if the hemangioblast generate both endothelial and blood cells (Lacaud and Kouskoff, 2017).

### 1.2.1. First hemogenic site: yolk sac

In the developing mouse embryo, the first wave of hematopoiesis generates a large number of short lived primitive cells including primitive erythrocytes, macrophages and megakaryocytes and it occurs at embryonic day 7.25 (E7.25) in the yolk sac (Figure 1.2) (Moore and Metcalf, 1970; Palis et al., 1999). Haar and Ackerman have shown the mesodermal origin of the first hematopoietic cells from structures called blood islands, that are capable of originating endothelium and blood cells (Haar and Ackerman, 1971). The classification as primitive was based initially in morphology as the primitive erythroid cells are larger with more cytoplasm than fetal-adult erythroid

cells and nucleated (Palis et al., 1999; Wong et al., 1986). Shortly after the primitive wave appear, as the heart starts to beat a definitive wave emerges in yolk sac at E8.25 constituted by erythroid, megakaryocyte, macrophages and myeloid cells (Bertrand, 2005; McGrath et al., 2015; Palis et al., 1999). Unlike primitive erythroid cells that express embryonic globin, definitive erythroid progenitors express β-globin and are enucleated (Frame et al., 2013). In addition, these definitive progenitors can be initially identified by the expression of the cell surface markers Cd41 and Kit and while maturation progresses the pan-hematopoietic marker Cd45 expression also increases (Mikkola and Orkin, 2006).

The emergence of primitive lineages with mostly myeloid cells and the lack of a colonized thymus at E10 led to the assumption that yolk sac could not generate lymphoid cells (Bertrand et al., 2007). However, this concept was challenged with findings showing that, regardless of the lack of long-term repopulating HSC in the yolk sac, derived cells can acquire this potential when transplanted into the fetal or neonatal liver (Yoder et al., 1997; Yoshimoto et al., 2011). With an inducible labeling model for early precursors, Samokhvalov and colleagues demonstrated that the yolk sac blood islands contain precursors that contribute to fetal lymphoid progenitors and to the adult HSC pool (Samokhvalov et al., 2007).

In the human embryo, the yolk sac is the initial hematopoietic site up to 6 weeks until the fetal liver surpasses it as the primary site (Palis and Yoder, 2001) (Figure 1.2). The first blood cells can be detected around day 16 of development reaching the highest frequency at day 19 (Tavian and Peault, 2005). The presence of erythroid and granulo-macrophage progenitors in the human yolk sac was first detected at 4.5 weeks (Migliaccio et al., 1986). Similarly to the mouse, primitive nucleated erythrocytes are abundant in the yolk sac between week 4 and 5. However, a gradual increase of the frequency of fetal and adult hemoglobin molecules occurs from week 5 and further and by week 12 an overall decline of primitive erythroblasts is observed in peripheral blood (reviewed in Palis and Yoder, 2001). In contrast to the mouse yolk sac, in human this structure develops in a balloon-like shape and starts degenerating around 10 weeks of gestation (Ivanovs et al., 2017; Palis and Yoder, 2001).

### 1.2.2. AGM is the first intra-embryonic site of definitive hematopoiesis

Until the 1970s, the origin of hematopoiesis was believed to be in the blood islands of the yolk sac (Moore and Owen, 1967), when in 1975 Françoise Dieterlen-Lièvre altered the perception of yolk sac as the origin site for definitive HSCs for the first time by grafting the quail's embryo onto the yolk sac of the chick and by observing that all definitive hematopoietic cells were of quail origin (Dieterlen-Lièvre, 1975). These experiments demonstrated the transient contribution of the yolk sac-derived cells and the intra-embryonic origin of the definitive avian hematopoiesis; thus, creating a paradigm shift in hematopoiesis (Dieterlen-Lièvre, 1975; Martin et al., 1978). Subsequently, several functional studies documented that the definitive hematopoietic system had its origin in the avian aortic region (review in Jaffredo et al., 2005). In mid 1990s, Medvinsky and colleagues demonstrated in mice for the first time that the intra-embryonic aorta-gonad-mesonephros (AGM) region hold a higher frequency of CFU-S (colony forming units- splenic) compared with embryonic yolk sac (Medvinsky et al., 1993). In addition, lymphoid progenitors were found to be present in the mouse para-aortic splanchnopleura (P-Sp) prior their appearance in the yolk sac (Godin et al., 1993). Studies claimed that the AGM/P-Sp region has demonstrated to be a powerful source of definitive HSCs at E10.5 in the mouse (Medvinsky and Dzierzak, 1996; Müller et al., 1994). In addition, it was shown that before the onset of circulation, both myeloid and lymphocyte progenitors arise at E7.5 within the P-Sp and only the explants from this region, and not yolk sac, demonstrate long-term multilineage hematopoietic reconstitution (Cumano et al., 1996, 2001). Close examination of the AGM region allowed the identification of intra-aortic hematopoietic clusters (IAHC) emerging from endothelial cells (de Bruijn et al., 2002; Jaffredo et al., 1998; Yokomizo and Dzierzak, 2010). These cells bud off predominantly from the ventral floor of the dorsal aorta and co-express both endothelial and hematopoietic markers including c-Kit, CD31, Sca-1, CD34, MAC1, VE-cadherin, flk1, tie2 and CD45 (Boisset et al., 2010; Medvinsky et al., 2011; Rybtsov et al., 2014; Taoudi and Medvinsky, 2007). Cells located closer to the endothelial layer express CD31, c-Kit, Flk-1 and low CD41 but they do not show CD45, on the other hand hematopoietic cells closer to the aortic clusters express both CD45 and Flk-1 (Yokomizo and Dzierzak, 2010). IAHC have been identified to occur around E9.5 in the AGM, and between

E10.5 and E11.5 the HSCs reach their higher emergence (de Bruijn et al., 2000; Medvinsky and Dzierzak, 1996; Müller et al., 1994). In the human embryo, the appearance of the first definitive hematopoietic progenitor cells occurs at day 27 of gestation that rapidly proliferate in clusters and disappear around day 40, and similarly to the mouse, the localization of the clusters was confirmed at the ventral wall of the dorsal aorta (Tavian et al., 1996, 1999). These nascent cells were recently shown to have long-term engraftment capacity when transplanted *in vivo* (Ivanovs et al., 2011). Although there is no established phenotype that discriminates emergent human HSCs from their precursors or progenitors, some molecules have been identified that are present in developing HSCs. Angiotensin-converting enzyme (ACE) marks fetal liver HSCs (Jokubaitis et al., 2008) and ACE+CD34- cells beneath the human dorsal aorta (Sinka et al., 2012). ACE+CD34- cells may represent HSC precursors that give rise to ACE+CD34+ cells contained in aortic clusters. During mouse ontogeny, the identification and isolation of these cells in the AGM region has been widely based on the expression of CD41, which is a marker expressed only in nascent HSCs (Ferkowicz et al., 2003), Aa4.1 (Petrenko et al., 1999), c-Kit and CD34 (Mikkola and Orkin, 2006). Vascular endothelial cadherin, which is typically thought of as an endothelial marker, is transiently expressed during the emergence and maturation of HSCs (Kim et al., 2005).

### 1.2.3. Additional anatomical hemogenic sites

The placenta is an important interface between maternal and fetal environments and is indispensable for fetal development (Rossant and Cross, 2001), and even though it has not been considered a traditional hematopoietic site, early studies indicated the presence of cells with hematopoietic capacity (Dancis et al., 1977). In fact, the placenta was found to be an early reservoir for precursors of B cells prior to the fetal liver (Melchers, 1979). A comprehensive study for the hematopoietic potential within the placenta tissue was performed using a GFP reporter transgenic mice, showing that it harbors a multipotent pool of hematopoietic cells of fetal origin and they can be found as early as E9.0. Between E15 and E17, placenta showed the highest CFU counts compared to the yolk sac, fetal liver and caudal half of embryo (Alvarez-Silva et al., 2003). However, this study did not show the full contribution of the placenta to

adult hematopoiesis. A further step was taken by demonstrating the existence of definitive HSCs in the placenta at E11 in two independent studies (Gekas et al., 2005; Ottersbach and Dzierzak, 2005). These cells express CD34, Sca1 and cKit, a similar phenotype to fetal liver HSCs, and show multilineage engraftment capacity for longer than 4 months. In addition, the onset of HSC appearance in the placenta matches with the AGM region and, the HSC pool peaks by E12.5 outnumbering the HSCs found in AGM or the yolk sac and by E15.5 the number of HSCs present in the placenta declines while in the fetal liver the HSC pool increases (Gekas et al., 2005; Ottersbach and Dzierzak, 2005), suggesting that the HSCs re-renter in circulation to colonize the fetal liver. Moreover, the presence in the placenta labyrinth of endothelial cells expressing hematopoietic markers highlight the close relationship between the two cell types and supports the idea that HSCs derived from hemogenic cells similar to AGM hematopoiesis (Ottersbach and Dzierzak, 2005). The independent generation of definitive HSCs in the placenta was acclaimed in placentas of Ncx1 knockout mouse models lacking heartbeat and therefore with an inactive blood circulation. Hematopoietic cells formed clusters and the early marker CD41 were detected in nascent Ncx1-/- placenta cells, and it was shown that these cells give rise to T, B lymphoid and myeloid cells *in vitro* (Rhodes et al., 2008). However, this study failed to show generation of functional repopulating cells, as consequence the question of whether the placenta generates HSCs during ontogeny or only allow their maturation remains controversial. The human placenta was also shown to harbor as early as week 6 in gestation hematopoietic progenitors and HSCs, more specifically CD34+CD45+ were identified in placental villi stroma and CD45+ were emerging from the vasculature (Robin et al., 2009). Additionally, the human placental cells could repopulate immunocompromised mice by xenotransplantation similarly to UCB cells (Robin et al., 2009). A recent study showed that a population of hematopoietic progenitor cells could be detected in the human chorion in close association of the blood vessels at all gestational ages, but the cells with multilineage and long-term capacity were only present from 15 to 24 weeks of gestation (Muench et al., 2017). Mouse and human placentas show some anatomical resemblances and have genes with analogue identity (Cox et al., 2009; Georgiades et al., 2002). The extraembryonic placenta is formed by multiple tissues such as trophectoderm, allantois, mesodermal and chorionic mesoderm (Rossant and Cross, 2001) and in during mouse ontogeny, the chorionic mesoderm and allantois fuse to form the placenta with the distal

mesoderm invading the trophoblast layer to form the vascular compartment of the placental labyrinth while in the proximal site becomes the umbilical cord linking the embryo and the placenta (Gekas et al., 2010).
In recent years, many other locations have been correlated with hematopoietic sites namely umbilical and vitelline arteries (de Bruijn et al., 2000; Gordon-Keylock et al., 2013; Müller et al., 1994; Zovein et al., 2010). Functional repopulating HSCs were detected in the dorsal aorta, vitelline artery and umbilical artery by E10.5 onwards (de Bruijn et al., 2000), confirmed by a three-dimensional imaging using endothelial and hematopoietic markers to distinguish the hematopoietic clusters within the arteries (Yokomizo and Dzierzak, 2010). Another study showed that both arteries contain pre-HSCs that in the presence of IL3 are capable to generate matured HSCs (Gordon-Keylock et al., 2013). The main transcriptional regulator of adult hematopoiesis, Runx1, was shown to be expressed in vitelline and umbilical arteries (North et al., 1999). Recently, the endothelium and the head were identified as new sites of hematopoietic cell emergence from endothelium (Li et al., 2012; Nakano et al., 2013). Nevertheless, the head remains a controversial site (Iizuka et al., 2016).

It has been shown that fetal liver does not produce de novo HSCs (Johnson and Moore, 1975) and with the onset of heart beat at E8.5, the progenitor cells from the yolk sac are found in the blood stream and further in embryo proper. Shortly after, at E9.5, the first myeloerythroid progenitors seed in the fetal liver to generate definitive erythroid cells. From E12.5, it becomes the primary site of hematopoiesis when it is colonized by circulating hematopoietic progenitor cells to provide the niche for expansion and maturation of the progenitors that will further be part of the HSC pool (Ema and Nakauchi, 2000; Gekas et al., 2005). As gestation progresses, the hematopoietic differentiation also changes and the purpose of the fetal liver lasts until 2 weeks after birth with the bone marrow gradually taking over as the main site of residing HSCs (Mikkola and Orkin, 2006). Despite the little knowledge about the niche that sustain HSCs, attempts have been made to establish key components that support HSCs expansion and as a result fetal liver stromal cell lines led to a considerable increase in mouse long-term HSC activity (Moore et al., 1997).

Bone marrow serves as the final destination of hematopoietic stem and progenitor cells, where they will reside throughout the lifetime of the individual. How the

hematopoietic cells switch from fetal liver to the bone marrow remains unclear. The unique microenvironment within the bone marrow, consisting of osteoblasts and stromal cells, attracts circulating HSCs that start seeding in this niche at E15 of the mouse embryo (Medvinsky et al., 2011). In addition, functional HSC are detected since E17.5 (Christensen et al., 2004) and depending on the stimulus from different cytokines and growth factors, HSPCs either differentiate and form committed hematopoietic lineages or keep the quiescent state creating an HSC reservoir (Mikkola and Orkin, 2006).

HSCs from the AGM region or bone marrow differ in many aspects, regarding cell-surface markers, developmental potential and cell-cycle status. As development progresses, mature mouse HSCs loose CD41 and CD34. Concomitantly, Sca-1, a marker present in adult HSCs, is also upregulated (Mikkola and Orkin, 2006). Functionally, HSC are defined by their capacity to provide long-term hematopoiesis repopulation of all blood-cell lineages upon transplantation to wild-type irradiated adult recipients (Cumano and Godin, 2007). Although all functional mouse HSCs are found in the Lineage−Sca1+c-Kit+ (LSK) population, less than 1 out of 10 of these LSK-cells has repopulation capacity, suggesting substantial functional heterogeneity (Wilson et al., 2007). Kiel *et al*. identified a combination of signaling lymphocyte activation molecule (SLAM) family of cell surface receptors (Cd150+Cd244+Cd48-) that showed to be differentially expressed between stem and progenitor populations in the bone marrow (Kiel et al., 2005). Isolation of functional human HSCs are based on the expression of CD34 (Krause et al., 1996), CD45RA (Mayani et al., 1993) and CD38 (Bhatia et al., 1997; Hao et al., 1995), studies have shown that Lin-CD34+CD38-CD90-CD45RA- population is enriched for MMPs with an incomplete self-renewal capacity while Lin-CD34+CD38-CD90+CD45RA- is enriched for HSCs (Majeti et al., 2007). However, Notta and colleagues have demonstrated that both Lin-CD34+CD38-CD90-CD45RA- and Lin-CD34+CD38-CD90+CD45RA- display long-term repopulating capacity in secondary immunocompromised mice and, they also identified a specific human HSC marker, CD49f, within the population enriched for Lin-CD34+CD38-CD45RA-. Loss of CD49f is linked to absence of long-term engraftments while single CD49f+ cells were shown to be associated with highly efficient long-term multilineage engraftments (Notta et al., 2011) (Figure 1.1).

### 1.2.4. Mechanisms of HSC emergence

As described in the previous section, despite the doubt about the existence of the hemangioblast, this concept seems to apply to the primitive hematopoiesis emerging in the yolk sac. In the AGM region, evidence supporting a common precursor is favored by the expression of common markers indicating a developmental link between both endothelial and hematopoietic lineages. Expression of an important HSC marker, CD34, is detected at the vascular lining and hematopoietic cells of E9.0 mouse yolk sac (Young et al., 1995) and, in human embryos, CD34 is also expressed in intra-aortic hematopoietic clusters and endothelial cells of the dorsal aorta at week 5 of gestation (Tavian et al., 1996). The transcription factor Tal1 was also shown to be expressed in both types of cells in multiple species, such as mouse (Kallianpur et al., 1994), chicken (Drake et al., 1997) and in human (Hwang et al., 1993). The initial generation of definitive hematopoietic stem cells have been intensively scrutinized in the past few decades in the AGM region, in which there is strong evidence that multipotent HSPCs emerge directly from a specialized vascular endothelial cell named hemogenic endothelium in a transient manner and in a process called endothelial-to-hematopoietic transition (EHT) (Kissa and Herbomel, 2010) (Figure 1.3).

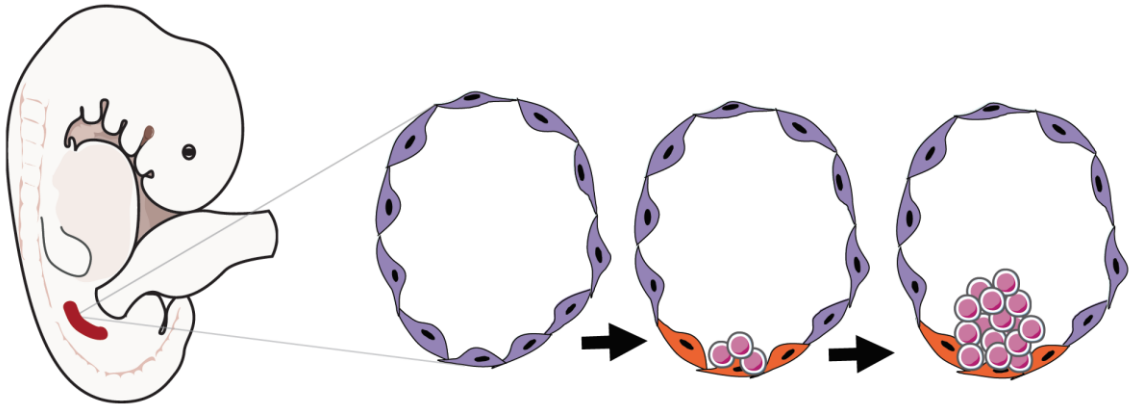


Figure 1.3 - **Schematic representation of the cellular origin of definitive HSCs in the AGM region**. Endothelial layer of vascular cells forms tight adhesion within each other (grey) and harbors vascular endothelial cells with hemogenic potential (orange) and as ontogeny progresses, hemogenic cells lose their attachments with the neighboring cells and endothelial-to-hematopoietic transition ends with the emergence of round hematopoietic clusters and multilineage hematopoietic stem and progenitor cells (pink) that will enter in the vessel space.

The initial observations of endothelial cells originating blood cells were done by the group of Dieterlen-Lièvre in chick embryos, in which endothelial cells were marked with Kdr and hematopoietic cells were marked with CD45 to classify the relation of both cell types. Double staining confirmed the endothelial cells as Kdr+ CD45- and the intra-aortic hematopoietic clusters as Kdr-CD45+. Moreover, the group identified endothelial cells based on their acetylated-low density lipoprotein (Ac-LDL) uptake (Voyta et al., 1984). Using this Ac-LDL as a tracing method for endothelial cells, the same group later observed some hematopoietic cells marked with Ac-LDL, demonstrating a connection between endothelial and hematopoietic transitioning in dorsal aorta of the chick embryos (Jaffredo et al., 1998). During mouse embryogenesis, parallel results were observed using Ac-LDL tracing method (Sugiyama et al., 2003). Furthermore, two studies have shown that mouse VE-cadherin+ endothelial cells can give rise to blood cells including of lymphoid origin *in vitro* (Nishikawa et al., 1998) and generate HSC with multilineage hematopoietic differentiation, using a genetic tracing strategy *in vivo* (Zovein et al., 2008). In fact, live cell fate tracking and time-lapse imaging elucidated the complete process of EHT. The transparency of zebrafish facilitates the visualization of EHT by live microscopy and traces the origin of HSCs back to the ventral wall of the dorsal aorta (Bertrand et al., 2010; Kissa and Herbomel, 2010; Lam et al., 2010). Using *in vivo* time-lapse imaging, Kissa and Harbomel captured the emergence of definitive HSCs not inside intra-aortic clusters but budding from the aortic wall of the zebrafish embryo (Kissa and Herbomel, 2010). Real-time imaging has also been performed in mice, demonstrating the emergence of multipotent HSC, phenotypically defined as Sca+, CD41+, cKit+, arising directly from the ventral wall of the dorsal aorta (Boisset et al., 2010). In the human embryo, sorted endothelial cells CD34+CD45- from the AGM region were co-cultured in a stromal cell layer that support multilineage hematopoiesis were capable to differentiate into CD45+ hematopoietic cells with multilineage progeny (Oberlin et al., 2002). In addition, the frequency in the AGM region of hemogenic endothelial cells was closely correlated with hematopoietic activity of this tissue, calculated around 1/120 at day 27, supporting the idea of hemogenic endothelium in the human intra-aortic region (Oberlin et al., 2002).

Studies on EHT *in vivo* are useful to visualize the correct localization and to time the chronology within the living embryo; however, *in vivo* techniques still need

improvements in some models like mouse. *In vitro* mouse ES cell differentiation seem to overcome some limitations of *in vivo* tracking systems and two studies have presented invaluable information on the stages of EHT and the identification of transient cells. Embryonic stem cell-derived cultures were used to generate putative hematopoietic cells via an intermediate hemangioblast equivalent expressing Flk1, nevertheless no hematopoietic cells ($CD41^{+}Tie2^{hi}cKit^{+}$) were detected. However, by using a Runx1 deficient line they found a $Tie2^{hi}cKit^{+}$ population still lacking CD41+ cells, that could be rescued by the re-expression of Runx1 (Lancrin et al., 2009). Another study used a similar approach to the one aforementioned by Lancrin and colleagues, in which hemangioblast-like cells expressing Flk1 were isolated and cultured on a stromal layer to allow the formation of hemogenic endothelium. Extended time-lapse permitted the observation of the formation of endothelial colonies and later CD41+CD45+ emergent hematopoietic progenitors were then traced at the single cell level (Eilken et al., 2009) demonstrating that EHT can be recapitulated *in vitro*.

Overall, these studies provide *in vivo* and *in vitro* evidence of EHT by showing that the ventral side of the dorsal aorta harbor specialized endothelial cells that undergo several changes including transcriptional with loss of endothelial markers and acquisition of hematopoietic markers and, morphological changes in which adherent cells bud off the endothelial lining turning into round shape cells that enter circulation (Figure 1.3). In addition, the emergence of the first hematopoietic cells can be observed among multiple species indicating that EHT is a well-conserved mechanism responsible for blood formation in vertebrates.

### 1.2.5. Important regulators of HSC emergence and EHT

Endothelial-to-hematopoietic transition is a highly regulated process both spatially and temporally and can be defined by the morphological changes as well as transcriptional alterations. This transition has been reported in numerous vertebrate species indicating a conserved process strictly controlled by both extrinsic and intrinsic molecular and cellular mechanisms. To further characterize these mechanisms, a clear phenotypic definition of hemogenic cells would allow their thorough classification; however, isolation of these specific cells proved to be challenging due to the low

representation -less than 1% of yolk sac and less than 5% of AGM endothelial cells - and the transient nature of these cells within hematopoietic tissues (Goldie et al., 2008; Marcelo et al., 2013).

Nonetheless, up to date, the identification of a single marker that distinguishes non-hemogenic endothelial cells from hemogenic remains to be identified, and as a consequence a larger set of regulators have been implicated in the emergence of hematopoietic stem cells (Table 1.1). In agreement, several signaling pathways have also been identified in the mechanism activating an endothelial cell to undergo EHT (Table 1.1). In the next section, I will highlight some important regulators of EHT and HSCs emergence.

Runx1

Runx1, runt-related transcription factor 1, is one of the key genes of EHT during definitive hematopoiesis. Consistent with this role, mutant Runx1 mice shows normal yolk sac erythropoiesis but lacks adult hematopoiesis and died around E12.5 (Okuda et al., 1996). It was demonstrated that Runx1 is required for the hemogenic endothelial cells to HSC transition (Chen et al., 2009) and, hematopoietic clusters are not formed in Runx-/- mutant embryos (North et al., 1999). Expression of Runx1 is detected in all definitive hematopoietic sites including placenta, AGM, and major arteries in all vertebrate species studied so far and, it is expressed also in endothelial tissues that are in contact with nascent blood cells (North et al., 1999, 2002; Swiers et al., 2010). Interestingly, its expression is not essential during primitive hematopoiesis nor after EHT (Chen et al., 2009; Lacaud et al., 2002; Lancrin et al., 2009). Deletion of a single allele of Runx in the mouse embryo give rise to fewer hematopoietic stem cells suggesting that Runx1 expression is dose dependent (Cai et al., 2000; Mukouyama et al., 2000). Runx1+/- embryos change the temporal expression pattern and show a temporal shift in HSCs emergence characterized by earlier detection at E10 in the yolk sac and AGM and the generation of HSCs ends prematurely in these embryos at E11 in the AGM. Interestingly, the E11 AGM Runx1+/- embryos show fewer CFU-S when compared to wild type tissue, suggesting that haploinsufficiency of Runx1 reduces HSC emergence, maintenance and/or proliferation (Cai et al., 2000).

The expression pattern of Runx1 during ontogeny highlights its relevant role in EHT and recently studies have been focusing on downstream targets of this regulator to potentially broad the understanding about the mechanism of EHT. By studying the transcriptome of Runx1 -/- in ES cell lines, Lancrin and colleagues identified Gfi1 and Gfi1b as direct targets of Runx1 (Lancrin et al., 2012).

The Runx1 +23 hematopoietic intronic enhancer serves as a cis-regulatory element for various transcription factors with key roles in *de novo* generation of HSCs and contains conserved GATA, SCL, ETS and RUNX motifs (Ng et al., 2010; Nottingham et al., 2007). Mutation of Runx1 showed almost no effect on Runx1 expression during ontogeny, while mutation of the GATA and ETS motifs in the Runx1 +23 enhancer completely abrogated its activity in the dorsal aorta suggesting that GATA and ETS motifs control the Runx1 +23 enhancer (Nottingham et al., 2007). Furthermore, studies have shown an overlap between the expression of Gata2 and in the lateral plate mesoderm and all endothelial cells and later in hematopoietic clusters where Runx1 expression is detected (Minegishi et al., 2003; North et al., 1999).

Gata2

The transcription factor GATA2 is a member of a conserved family of proteins, together with five other members (Gata1-6) that play a key role in differentiation of many cell types. The origin of the family name GATA is related with the ability that these proteins show to specifically recognize the DNA sequence (A/T)GATA(A/G) (Patient and McGhee, 2002) via two conserved zing fingers domains located at the terminals of the protein, both zinc fingers can mediate binding to the DNA and protein-protein interactions (Bates et al., 2008). Three Gata family members are related with hematopoiesis, Gata1 with erythropoiesis, Gata2 with HSCs and HPCs, and Gata3 with T lymphopoiesis (Gao et al., 2015; Kitajima et al., 2006; Ko and Engel, 1993). Gata2 and Gata3 are also required for differentiation of the central nervous system. Gata4, -5 and -6 are associated with cardiovascular embryogenesis in addition to other sites (Lentjes et al., 2016).

The expression of Gata2 is observed in AGM region of the embryo in early HSCs (Minegishi et al., 2003). During ontogeny, Gata2 was shown to play an important role for the expansion of HSCs in both the AGM region and in the bone marrow in which

also controls the balance between cell quiescence and self-renew (Ling et al., 2004; Wilson et al., 2010). Besides the AGM region and bone marrow, Gata2 is highly expressed in megakaryocytes (Visvader and Adams, 1993) and in mast cells (Tsai and Orkin, 1997). However, the importance of Gata2 in HSC generation was only highlighted by knockout mice model. Gata2 deficient mice suffer from severely impaired primitive erythropoiesis and a complete lack of other committed progenitors and HSCs and die at E10.5 (corresponding to the time of the first HSC emergence) (Tsai et al., 1994). Due to this lethally, most studies on loss of function for GATA2 expression are focused on the inactivation of one allele, this haploinsufficiency affects the generation and expansion of HSCs especially in the AGM region (Ling et al., 2004; Rodrigues, 2005; de Pater et al., 2013). In the mouse AGM, conditional deletion of +9.5 regulatory element results in decreased expression of the transcription factors Runx and Scl and abrogated the hemogenic endothelium capacity to give rise to HSCs. This deletion ultimately leads to embryonic lethally at E13-14 (Gao et al., 2013). In contrast, gain of function studies show that enforced Gata2 constrain HSCs function by blocking differentiation and inhibiting cell cycle (Minegishi et al., 2003; Persons et al., 1999; Tipping et al., 2009). Additionally, Gata2 is under the control of bone morphogenesis protein (BMP4), a factor involved in blood specification, which was also detected in the AGM region (Bhatia et al., 1999) (Table 1.1).

Gfi1b

The growth factor independence 1B (Gfi1B), a transcriptional repressor harboring six zinc-finger motifs and a Snail/Gfi1 (SNAG) domain, is a member of the GFI1 family that recognize the DNA sequence 5'-TAAATCACTG-3' and is differentially expressed during hematopoiesis (Grimes et al., 1996; Tong et al., 1998; Zweidler-McKay et al., 1996). Gfi1b is most predominant in the earliest HSCs, erythroid precursors and megakaryocytes and evidence suggests that deletion of Gfi1b does not affect myeloid development. Gfi1b-/- embryos die by E15 from a failure of fetal liver erythropoiesis (Hock et al., 2004; Saleque et al., 2002). Deletion of Gfi1b in HSCs led to a significant expansion of these cells in the bone marrow and peripheral blood and showed a decreased expression of adhesion molecules necessary to retain HSCs within their niches (Khandanpour et al., 2010). Conditional knockout mice allowed the study of the

role of Gfi1b in adult erythropoiesis and two recent studies showed that deletion of Gfi1b revealed to be critical in megakaryocytic cell lineage and erythropoiesis with reduced number of erythroid precursors, delayed erythroid maturation and anemia (Foudi et al., 2014; Vassen et al., 2014) whereas enforced Gfi1b expression in more differentiated erythroid precursors arrested proliferation and induced erythroid differentiation (Garcon, 2005). In addition, it was reported as having a specific role during the endothelial-to-hematopoietic transition. In the absence of Runx1, Gfi1 and Gfi1b were capable to silence the endothelial identity of hemogenic endothelial cells (Lancrin et al., 2012). Moreover, mutations in the GFI1B gene have been recently identified in 2 families with autosomal dominant platelet disorder (Monteferrario et al., 2014; Stevenson et al., 2013). Thus, Gfi1b seems to play an important role during normal HSC emergence and erythroid-megakaryocyte lineage and specification as well as platelet formation in human diseases.

Pathways

Several pathways have been described as crucial for HSCs emergence such as Notch, Wnt and Hedgehog (Table 1.1). Notch signaling has been shown to be important for HCS generation namely in the AGM region. By comparing the analysis of deficient mice for Notch1 and Notch2, Kumano and colleagues revealed that Notch1 is required for definitive hematopoietic progenitor and HSC generation (Kumano et al., 2003). In the dorsal aorta, it was revealed that Gata2 is one of the direct targets of Notch1 (Robert-Moreno et al., 2005). In zebrafish, a recent study identified an ortholog of Gata2- gata2b – as being exclusively restricted to HE and needed to initiate runx1 expression in a Notch-dependent manner (Butko et al., 2015). Interestingly, the hematopoietic defect that occurs in the absence of Notch signaling can be rescued by the artificial induction of Runx1 but not Gata2 (Nakagawa et al 2006) suggesting a link between Notch activation and the generation of HSCs and the interplay of Notch, Gata2 and Runx1. Hes2, a transcription factor that acts upstream of the Notch signaling, was shown to be crucial in HSCs generation (Rowlinson and Gering, 2010).

Even before the emergence of definitive HSCs, blood flow starts in the mouse at E8.5 while in the human around day 22. Surprisingly, biomechanical forces and blood flow were shown to be implicated in EHT. Inactivation of the gene Ncx1 results in

embryonic lethal phenotype with a complete vasculature and normal endothelial cells but lacking heart beating (Koushik et al., 2001). Ncx mutant mouse is devoided of hematopoietic stem cell in the embryo due to a lack of biomechanical stimulation (Adamo et al., 2009; Lux et al., 2008). The fluid shear stress was also shown to increase Runx1 expression and mediate hematopoietic emergence via nitric oxide signaling (Adamo et al., 2009).

A recent discovery in the field was the contribution of the inflammatory response pathway in HSC emergence. One study in zebrafish has shown that activation of the expression of Notch ligand Jagged1 and NF-κB in aortic endothelial cells by Tnfα/Tnfr2 signaling promotes HSC specification and emergence through interaction with Notch1 (Espín-Palazón et al., 2014). In mouse and zebrafish, it was also demonstrated that HSPC emergence requires TLR4-MyD88-NF-κB to promote Notch signaling, and knockdown of gcsfr and tnfr2 decreased Notch target gene expression and HSPC numbers (He et al., 2015). Likewise, mouse embryos lacking IFNα or IFNγ exhibit few HSPC in the AGM (Li et al., 2014). Analyses in zebrafish and fetal human HSPCs also show that IFNγ plays a role in promoting HSC generation, downstream of Notch, using a STAT3-mediated pathway to drive HSC specification (Sawamiphak et al., 2014).

Table 1.1 - Regulators involved in HSCs emergence.

| Regulators | Organism | References |
|---|---|---|
| Gata2 | Mouse and human | (Labastie et al., 1998; Ling et al., 2004; de Pater et al., 2013; Tsai et al., 1994) |
| GATA3 | Human | (Labastie et al., 1998) |
| Runx1 | Mouse and zebrafish | (Chen et al., 2009; Kissa and Herbomel, 2010) |
| Gfi1 | Mouse | (Lancrin et al., 2012; Thambyrajah et al., 2016) |
| Gfi1b | Mouse | (Lancrin et al., 2012; Thambyrajah et al., 2016) |
| Gpr56 | Mouse and human | (Kartalaei et al., 2015) |
| Gpr183 | Zebrafish | (Zhang et al., 2015) |
| Vegf | Zebrafish | (Rowlinson and Gering, 2010; Burns et al., 2009) |
| cKit | Human and mouse | (Labastie et al., 1998; Marcelo et al., 2013) |
| Fgf | Zebrafish | (Lee et al., 2014) |
| Scl/Tal1 | Mouse, human and zebrafish | (Labastie et al., 1998; Porcher et al., 1996; Zhen et al., 2013) |
| Sox17 | Mouse | (Nobuhisa et al., 2014) |
| FLK1 | Human | (Labastie et al., 1998) |
| cMYB | Human | (Labastie et al., 1998) |
| Hif1α | Zebrafish and mouse | (Harris et al., 2012; Imanirad et al., 2014) |
| Hif2α | Zebrafish | (Gerri et al., 2018) |
| Igf2 | Mouse | (Mascarenhas et al., 2009) |
| Hdac1 | Zebrafish | (Burns et al., 2009) |
| Evi1 | Zebrafish | (Konantz et al., 2016) |
| Etv2 | Mouse | (Lee et al., 2008; Sumanas et al., 2008) |
| Etv6 | Mouse and xenopus | (Ciau-Uitz et al., 2010; Wang et al., 1997) |
| Lmo2 | Mouse | (Yamada et al., 1998) |
| Eto2 | Xenopus | (Leung et al., 2013) |
| Fli1 | Xenopus, zebrafish and mouse | (Hart et al., 2000; Liu et al., 2008) |
| Fos | Mouse | (Goode et al., 2016; McKinney-Freeman et al., 2012) |
| F2r | Mouse | (Yue et al., 2012) |
| Hes1 | Mouse | (Guiu et al., 2013) |
| Hes 5 | Mouse | (Guiu et al., 2013) |
| RBP-jκ | Mouse | (Oka et al., 1995; Robert-Moreno et al., 2005) |
| Hey2 | Zebrafish | (Rowlinson and Gering, 2010) |
| Raldh2 | Mouse | (Chanda et al., 2013) |

| | | |
|---|---|---|
| Rac 1 | Mouse | (Ghiaur et al., 2008) |
| Cdca7 | Mouse | (Guiu et al., 2014) |
| Foxc2 | Mouse | (Jang et al., 2015) |
| Cdkn1c | Mouse | (Mascarenhas et al., 2009) |
| Tbx16 | Zebrafish | (Burns et al., 2009) |
| VegfA | Mouse | (Carmeliet et al., 1996; Ferrara et al., 1996) |
| Jagged1 | Mouse and zebrafish | (Espín-Palazón et al., 2014; Robert-Moreno et al., 2008; Xue et al., 1999) |
| Mib | Zebrafish | (Burns et al., 2005; Gering and Patient, 2005) |
| Ncx1 | Mouse | (Adamo et al., 2009; Lux et al., 2008; Rhodes et al., 2008) |
| IFNγ | Zebrafish | (Li et al., 2014; Sawamiphak et al., 2014) |
| TNFα | Zebrafish | (Espín-Palazón et al., 2014) |
| IFNα | Mouse | (Li et al., 2014) |
| Notch1 | Mouse and zebrafish | (Hadland, 2004; Kim et al., 2014; Kumano et al., 2003) |
| Jam1a/Jam2a | Zebrafish | (Kobayashi et al., 2014) |
| CBFβ | Mouse | (Sasaki et al., 1996; Wang et al., 1996) |
| IL-3 | Mouse | (Robin et al., 2006) |
| IL-6 | Mouse | (Xu et al., 1998) |
| Adenosine | Zebrafish | (Jing et al., 2015) |
| BMP4 | Zebrafish and mouse | (Crisan et al., 2015; Wilkinson et al., 2009) |
| β-catenin | Mouse | (Ruiz-Herguido et al., 2012) |
| Hedgehog | Zebrafish | (Wilkinson et al., 2009) |
| Wnt9a | Zebrafish | (Grainger et al., 2016) |
| Retinoic acid | Mouse | (Chanda et al., 2013; Goldie et al., 2008) |

In summary, intensive analysis have revealed the contribution of specific transcription factors, important signaling pathways as well as physiological pathways including inflammation and blood flow in the emergence of HSCs. It has also been discovered the interplay and overlap between regulators. EHT is sustained by a network of transcription factors that are responsible to regulate highly-tuned temporal and spatial events and to trigger signaling cascades, how all these regulators integrate in a HE gene regulatory core is yet to be determined.

## 1.3. Reprogramming

### 1.3.1. Nuclear reprogramming

Following fertilization, the resulting zygote has the highest developmental capacity known as totipotency, to generate all types of cells in the organism. In the course of embryonic development, cells of the early embryo gradually constrain their potential by committing to one of the three germ layers – endoderm, mesoderm, and ectoderm. As differentiation occurs this developmental capacity is lost, and the intermediate populations give rise to cells with specialized functions that will contribute to complex tissues in a unidirectional and frequently stable process (Figure 1.4). However, somatic cells in particular conditions can change their cell fate (ex. cancer). In addition, several studies questioned the malleability of cell fate by showing that the differentiated state may be reversible experimentally. This process is defined as nuclear reprogramming and can be achieved *in vitro* by different approaches.

*Somatic Cell Nuclear Transfer (SCNT)*

Not a long time ago, biologists believed that cell fate was irreversible and permanent. This concept was challenged by the experiments of Briggs and King in frogs, in which they conducted a nuclear transplantation of embryonic cell nuclei from early phases of development into enucleated eggs of *Rana pipiens*. They observed normal development in animals when the nuclei were taken from the blastula stage; however, the transfer of nuclei from a later stage, such as gastrula stage, resulted in significant abnormal development. These observations led the researchers to conclude that throughout cell differentiation, the cell faces irreversible changes (Briggs and King, 1952). The same methodology was extended to *Xenopus laevis* but with a different interpretation. In these studies, Gurdon and colleagues transplanted nuclei from differentiated intestinal cells of tadpoles into enucleated oocytes and successfully produce tadpoles with normal development (Gurdon, 1962; Gurdon et al., 1958). This result demonstrated that even fully differentiated, the nucleus provided the genetic information required to generate all cell types of an entire organism and that the differentiation is a reversible process without altering the gene content. In mammals, SCNT was only accomplished four decades later with the generation of the sheep

“Dolly” (Campbell et al., 1996; Wilmut et al., 1997). In this breakthrough experiment, Campbell and colleagues transplanted the nuclei of culture mammary glands into enucleated sheep oocytes and the resulting blastocyst was placed in a surrogate female sheep to continued embryonic development until birth (Campbell et al., 1996). Afterwards, nuclear transfer was performed in many other species including rabbit, cow, pig, and goat (reviewed in Gurdon and Byrne, 2003). In the mouse, the successful generation of mice from lymphocytes was achieved with ES cells derived from cloned embryos (Hochedlinger and Jaenisch, 2002). Despite cloning in diverse animals and many attempts in human cells, the generation of human SCNT embryos was only demonstrated recently using ESC as intermediates (Tachibana et al., 2013). The nuclear transplantation method described in these experiments was later named cloning or SCNT (Figure 1.4).

*Cell fusion*

Somatic cell hybrids are produced through cell fusion involving two or more cells. This single identity has more than one nucleus and until the nuclei fuse, the transient cell is designated heterokaryons if the fusion results from different tissues or species, or homokaryons resulting from the same cell type or species and fused share the cytoplasm. Despite being an inefficient process, the discovery of hybridization made important contributions by revealing the molecular mechanisms of somatic cells and PSCs (Yamanaka and Blau, 2010). Tada and colleagues fused embryonic stem cells with adult thymocytes and showed the acquisition of pluripotency by the fused tetraploid cells through the activation of GFP reporter driven by the promoter of Oct4 (Tada et al., 2001). Another study showed that the efficiency of somatic cell-ESC cell fusion can be improved if either the somatic cells or ESCs used overexpress Nanog (Silva et al., 2006). This study led to the discovery of the first reprogramming factor, which together with the ESC machinery can enhance reprogramming efficiency. However, the overexpression of Nanog alone in somatic cells is insufficient to drive their reprogramming to a pluripotent state (Silva et al., 2006). Fusion of B cells with ESC allowed the identification of Oct4 as another key regulator of reprogramming towards pluripotency (Pereira et al., 2008) and revealed the impact of the epigenetic modulator PRC2 in this process (Pereira et al., 2010).

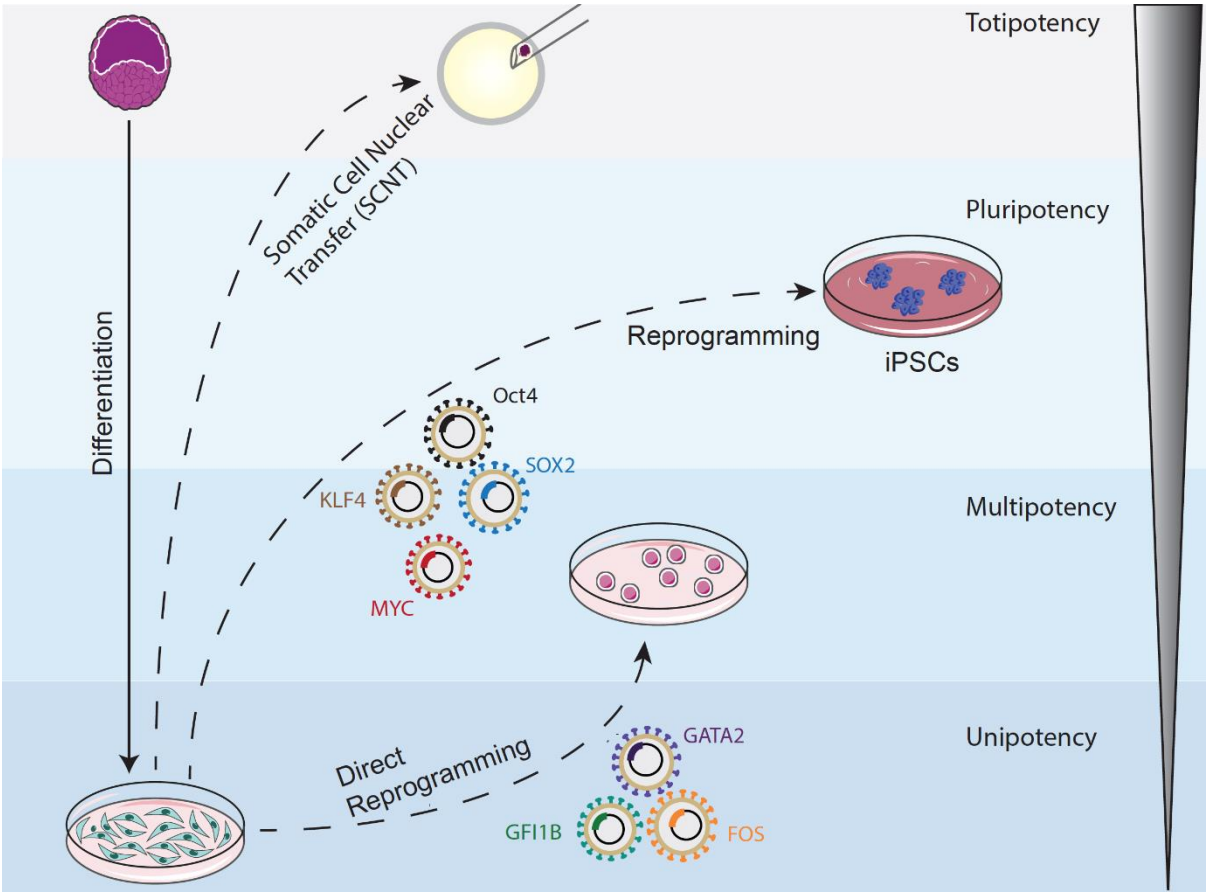



Figure 1.4 - **Schematic illustration of the alternative cell fate changes and differentiation**. During embryonic development, pluripotent stem cells can be differentiated into multiple somatic lineages represented in the upper left side. The illustration on the right side shows the 3 main approaches of nuclear reprogramming. *Somatic cell nuclear transfer*, removal of the nucleus from an oocyte and injection of the somatic nucleus into the enucleated oocyte. *Transcription-factor mediated reprogramming* Induction of pluripotent stem cells by the specific factors Oct4, Sox2, Klf4 and Myc. Direct reprogramming, induction of somatic cell using tissue-specific factors (ex. GATA2, GFI1B and FOS) into lineage committed cells.

### *TF-mediated reprogramming*

The conversion of cell fate by the enforced expression of tissue-specific transcription factors is representative of another nuclear reprogramming (Figure 1.4). The concept that a single transcription factor can unleash a cascade of events, altering the gene expression profile that eventually result in phenotypic modification was first described in the fruit fly and later in mammals. The ectopic expression of transcription factors like Antp gene in *D.melanogaster* larvae induced antenna to leg modifications (Schneuwly et al., 1987) while induction of the master regulator of eye Pax-6 was capable to introduce ectopically eye structures on the wings, antennae and legs (Halder et al., 1995). In mammals, the gene MyoD was identified as a master regulator

of myogenesis and ectopic expression of this factor induced the conversion of somatic cells into myoblasts (Davis et al., 1987). These studies showing the programming of one cell type into another, sidestepping a pluripotent state intermediate, pioneered further experiments of what is now termed transdifferentiation, lineage conversion or direct reprogramming (Graf and Enver, 2009). More recent work has demonstrated the induction of somatic cell into several cell types mediated by tissue-specific transcription factors, namely, neurons (Vierbuchen et al., 2010) hepatocytes (Sekiya and Suzuki, 2011), and cardiomyocytes (Ieda et al., 2010). This process will be discussed in the next section with particular focus on hematopoietic fate.

The missing evidence supporting the idea that cell fate is reversible and stable without the need of oocytes arrived with the Takahashi and Yamanaka findings in 2006. In this pioneering study, 24 genes that are expressed in ESCs were cloned and delivered into mouse fibroblasts using retrovirus. The pluripotency induction was screened by examining the endogenous reactivation of a specific ESC marker in transduced cells. Clones emerged two weeks later at a 0.01–0.1% efficiency, showing similar morphology, growth rate and gene expression compared to ESCs. To confirm the acquisition of pluripotency, transduced cells were injected into immunodeficient mice forming teratomas containing all three germ layers (Takahashi and Yamanaka, 2006). To selectively eliminate factors that were not critical for this process, by withdrawal of each candidate from the initial pool, Takahashi and Yamanaka reached a minimal combination of four factors that is sufficient to induce pluripotency from fibroblasts: Oct4, Sox2, Klf4 and c-Myc, As these cells were converted to pluripotency, they were named induced pluripotent stem cells (iPSCs) and the combination of TFs are known as OSKM or the “Yamanaka factors” (Takahashi and Yamanaka, 2006). This remarkable finding paved the way for the demonstration of cell fate plasticity and for providing mechanistic understandings of the reprogramming process. Consequently, human iPSCs were generated from human fibroblasts using the same combination of transcription factors OSKM (Park et al., 2008; Takahashi et al., 2007), or using a similar one - OCT4, SOX2, LIN28, and NANOG (Yu et al., 2007). In addition, since the first published report, interest in the application of iPSCs increased and has enabled the generation of iPSCs from disease-specific patients that allow either drug screening or identification of molecular mechanisms of the pathology (Kim, 2015).

Almost six decades have passed since the discovery that somatic nucleus can be reprogrammed to a pluripotent state, and that reprogrammed cells are capable to re-enter embryo development and create an entirely new animal (Gurdon, 1962). Now it is possible to convert somatic cells directly into lineage committed cells by the enforced expression of transcription factors, without the use of oocytes, ESCs, or embryos. From the previous section, the prevailing information of *in vivo* development of definitive hematopoiesis indicates that emergent HSCs follows commitment through hematoendothelial lineage, and later reaches hematopoietic specification and maturation. In the next section, I will address the approaches that have successfully derived engraftable HSCs from iPSCs differentiation and from direct lineage reprogramming as well as the approaches that describe an endothelial-to-hematopoietic transition.

### 1.3.2. Reprogramming as a hemogenic platform to generate functional HSCs

#### 1.3.2.1 Derivation of hematopoietic cells from ESCs and iPSCs

The initial attempts to generate engraftable HSCs in vitro were made through differentiation of pluripotent stem cells using either embryoid body formation supplemented with hematopoietic cytokines (Bigas et al., 1995; Hole et al., 1996; Keller et al., 1993) or co-culture with supportive stromal cell lines such as OP9 (Choi et al., 2009; Takayama et al., 2008; Timmermans et al., 2009); however, with limited long-term reconstitution potential.

The first study showing a promising result overexpressed HoxB4 with an inducible system combined with OP9 co-culture in murine ESCs achieving an adult HSC-like phenotype that yielded multilineage engraftment potential in immunocompromised primary and secondary mice (Kyba et al., 2002). Even though promising, the transient expression of HoxB4 seems to have prevented lymphoid lineage reconstitution and upon withdrawal of doxycycline, no quantification was performed of transgene HoxB4 expression remaining. Contrary to the previous study, two recent reports showed increase hematopoietic cells but no engraftment capacity was described when HoxB4

overexpression was effectively transient (Jackson et al., 2012; Tashiro et al., 2012). Nevertheless, the long-term reconstitution was achieved when using constitutive HoxB4 overexpression (Bonde et al., 2008; Chan et al., 2008; Matsumoto et al., 2009). Hematopoietic differentiation from human ESCs was induced by constitutive overexpression of HOXB4 but did not give rise to engraftment potential (Bowles et al., 2006; Chadwick, 2003; Jackson et al., 2016; Unger et al., 2008; Wang et al., 2005). The heterogeneity between studies does not allow to infer clearly the contribution of HoxB4, however other studies used different approaches to derived hematopoietic stem cells from pluripotent cells. Several studies have shown that differentiation of iPSC contain both hematopoietic and endothelial potential (Choi et al., 2009; Vereide et al., 2014). In addition, Elcheva and colleagues selected two combinations of TFs (GATA2-ETV2 and GATA2-TAL1) to confer activation of pan-myeloid or erythro-megakaryocytic programmes, respectively, highlighting the importance of GATA2 in hematopoiesis (Elcheva et al., 2014). Although these strategies failed to generate long-term repopulating cells, short-term engraftment of erythroid and myeloid lineages was accomplished with overexpression of HOXA9, ERG, RORA, SOX4 and MYB of $CD34^{+}CD45^{+}$ hPSC-derived cells (Doulatov et al., 2013).

To more closely mimic developmental environment, co-culturing the PSCs with cell lines and using teratoma formation was another strategy used to generate transplantable HSCs. Teratomas hold the potential to form cells from all three germ layers and can develop ectopically in a host recipient. In this context, hPSCs were co-transplanted with OP9 into immunodeficient mice and the resulting teratomas generated human CD45+ cells that migrate to bone marrow, are detectable in the host peripheral blood and colonize lymphoid tissues (Amabile et al., 2013; Suzuki et al., 2013). Transplantation of human teratoma-derived CD34+CD45+ was serially transplanted from primary into secondary recipients showing reconstitution of all human blood cell lineages including lymphoid lineage (Amabile et al., 2013; Suzuki et al., 2013). However, the percentage of chimerism obtained were generally low and both studies did not show a robust reproducibility of the system. Despite the promising engraftment reconstitution results, the putative clinical use of teratoma formation is still limited.

In a recent publication, the generation of engraftable HSCs was described in a two-step protocol. First, hemogenic endothelium was generated from human PSCs using

growth factors and in a second step, hemogenic endothelial hPSC-derived cells were retrovirally transduced with seven transcription factors (ERG, HOXA5, HOXA9, HOXA10, LCOR, RUNX1, and SPI1). These hemogenic endothelial-derived cells were transplanted directly into the bone marrow of immunodeficient mice, showing multilineage reconstitution in primary and secondary recipients (Sugimura et al., 2017). Another recent report used a single transcription factor, MLL-AF4, to generate iPSC-derived HSCs and showed robust long-term engraftment reconstitution with multilineage potential. Nevertheless, the long-term post-transplanted mice developed leukemia (Tan et al., 2018). These studies represent an encouraging step closer to generate efficiently HSCs using PSC sources; however, further clinical application would require some considerations.

#### 1.3.2.2 Hemogenic somatic cell reprogramming

The advent of iPSCs has unlocked unprecedented new perspectives in regenerative medicine and the iPSCs-derived HSCs are no exception, but the tumorigenic risk associated with pluripotency, efforts have been made to bypass this state. This section will present a summary of studies that have used combinations of TFs to convert directly somatic cells into hematopoietic committed cells.

The earlier experiments by Davis and colleagues, showing the induction of a myoblast fate with ectopic expression of MyoD in fibroblasts, demonstrated proof-of-concept for transcripton factor-mediated transdifferentiation (Davis et al., 1987). This study had opened the way to the generation of other cellular conversions from cells within the same potency, or the same developmental system. However, the long-standing challenge of reprogramming in hematopoiesis is the successful conversion of a somatic cell into a highly functional multipotent HSC. the first evidence of cellular reprogramming in the hematopoietic system was described by Tomas Graf group. In 1995, his work pioneered the field by showing that GATA1 induce erythroblasts fate when expressed in myelomonocytic cells (Kulessa et al., 1995) providing evidence of the high plasticity of hematopoietic cells. Later, B lymphocytes were converted into functional macrophages using the enforced expression of C/EBPα or C/EBPβ (Xie et al., 2004) and with a similar approach his group demonstrated the conversion of

developmental distant cell types, disrupting the established idea that reprogramming was possible only between closely related cells. The combination of C/EBPα/β and PU.1 was sufficient to induce a macrophage fate in fibroblasts (Feng et al., 2008) (Figure 1.5).

The first attempt to generate a multipotent HSC was shown with ectopic overexpression of the pluripotent marker OCT4 with specific cytokines. The resulting CD45+ cells showed developmental plasticity by generating a variety of hematopoietic progeny including erythroid, granulocytic, monocytic, and megakaryocytic lineages although with limited engraftment capacity (Mitchell et al., 2014a; Szabo et al., 2010). In addition, many lines of evidence during reprogramming have established OCT4 involvement at the onset of cell fate commitment in the generation of neural stem cell (Mitchell et al., 2014b) underscoring the importance of this TF in reprogramming. However, as OCT4 induces pluripotency, the downstream potential of generating blood or neural cells with tumorigenic capacity remains unclear.

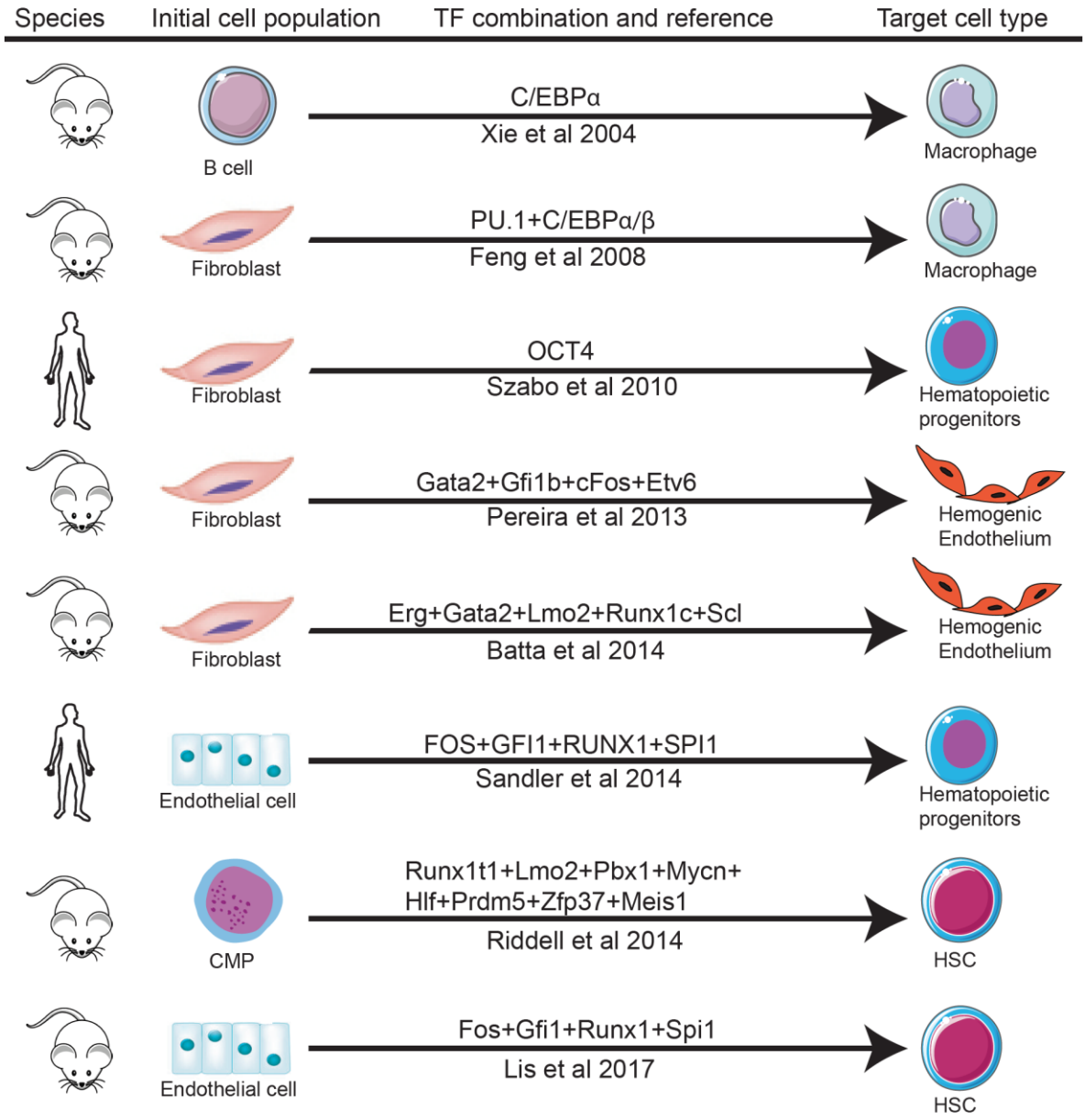



Figure 1.5 - **Strategies for somatic cell fate reprogramming into hematopoietic stem cells.** The scheme illustrates the different strategies used by various laboratories to attempt generating HSCs, including different initial somatic cell types (fibroblasts, endothelial cells, B cells and lineage committed blood progenitors), different TF combinations and the species to start with reprogramming. HSC- Hematopoietic stem cell; CMP- Common myeloid progenitor.

A new route for somatic cell reprogramming to hematopoietic lineages was designed when Pereira and colleagues screened a pool of specific transcription factors for hematopoiesis in mouse embryonic fibroblasts (MEFs) harboring a human CD34 reporter with H2BGFP (Pereira et al., 2013). This screening unveiled four TFs - Gata2,

Gfi1b, cFos and Etv6 – that efficiently activates the human CD34 reporter and induces an initial population of endothelial-like precursors (expressing Sca1+ Prom1+) that further give rise to hematopoietic cells with features of nascent HSCs (Prom1+Sca1+CD34+Cd45-). Interestingly, a minimal combination of Gata2, Gfi1b and cFos was sufficient to induce hematopoietic colonies from fibroblasts though, Etv6 was shown to increase the reprogramming efficiency up to 6%. Re-aggregation with placental cells allowed the emergence of hematopoietic CD45+ cells with colony-forming potential (Pereira et al., 2013). These results support the idea that definitive hematopoiesis emerge via an endothelial-to-hematopoietic transition and, it is the first study demonstrating that a simple cocktail of four TFs holds the potential to unveiling a complex developmental process *in vitro* (Figure 1.5 and Figure 1.6).

Following successful induction of a hemogenic program in mouse fibroblasts using Gata2, Gfi1b, cFos and Etv6, in this thesis I will extend the same strategy to human fibroblasts. This system can now be used to understand the biology of blood origins and it will facilitate the clinical translation of these cells.

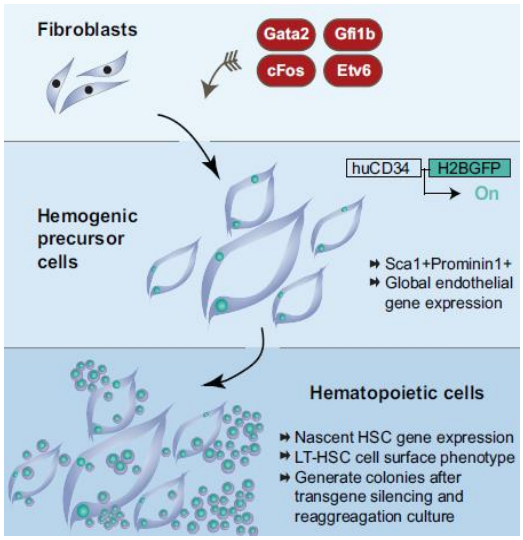



Figure 1.6 - **Stepwise induction of hemogenesis using direct reprogramming.** Hemogenic program in mouse fibroblasts is induced by the combination of Gata2, Gfi1b, cFos and Etv6. Subsequently, emergent hematopoietic cells progress via an endothelial-to hematopoietic transition (Adapted from Pereira et al., 2013).

Upon this finding, similar approaches have emerged for the generation of functional HSCs (Figure 1.5). In 2014, three independent studies provided alternative strategies to induce somatic cells into hematopoietic committed cells. The combined

overexpression of the TFs Gata2, Erg, Lmo2, Runx1 and Scl was shown to rapid convert fibroblasts into hematopoietic progenitor cells. These progenitors revealed an endothelial intermediate state, possessed multilineage clonogenicity and short-term reconstitution when co-cultured with OP9 and OP9-DL1 cell lines. In addition, loss of p53 function in induced cells revealed increased reprogramming efficiency but no improvement in the short-term engraftment (Batta et al., 2014).

A further step towards the generation of transplantable HSCs was achieved by combining direct lineage reprogramming with supportive niche cells. Transient overexpression of Runx1t1, Lmo2, Prdm5, Hlf and Zfp37 in myeloid progenitors led to the formation of HSCs-like cells by transplanting the reprogrammed cells directly into immunocompromised mice allowing the reprogrammed cells to maturate *in vivo*. All hematopoietic lineages were detected in primary and secondary recipients and the addition of Meis1 and Mycn improved the reprogramming efficiency (Riddell et al., 2014). This study illustrates the importance of the hematopoietic niche in HSCs development and despite promising future experiments would be imperative to assess the risk of tumor formation as most TFs used in this study are oncogenes. Recently, two studies from the same group, used a specific vascular niche capable of expanding HSCs to support the reprogramming of either human endothelial cells overexpressing FOSB, GFI1, RUNX1 and SPI1 (Sandler et al., 2014) or mouse endothelial cells overexpressing the respective TFs orthologous (Lis et al., 2017). In both studies, reprogrammed cells engrafted primary and secondary immunocompromised mice; however, lymphoid reconstitution was only achieved in the mouse study and no indication of leukemia post engraftment was found (Lis et al., 2017).

Overall, these studies indicate that direct lineage reprogramming can recapitulate early stages of hematopoietic development allowing the study of a complex developmental process “in a dish”.

### 1.3.3. GATA2, GFI1B and FOS networking and role in transcription regulation

Hematopoiesis is a dynamic and complex process governed by regulatory transcription factors that control specific gene expression. Several transcription factors have been identified in hematopoietic lineages and have been used in multiple direct hemogenic reprogramming studies. As seen in Figure 1.5, key hematopoietic transcription factors overlap in several studies including GATA2, GFI1B and FOS; however, their precise role in hemogenic reprogramming remains yet to be determined. In this section, I will discuss the role of GATA2, GFI1B and FOS based on their networking and their known roles in hematopoiesis.

GATA2 or GATA binding protein 2 is a key transcription factor for EHT as described in section 1.2.5, and during hematopoietic differentiation Gata2 also plays important functions. GATA2 is expressed in megakaryocytes and mast cells but dispensable for erythroid and myeloid terminal differentiation (Johnson et al., 2012; Tsai and Orkin, 1997). During hematopoietic differentiation, GATA2 expression declines once erythroid progenitors commit, in contrast, GATA1 expression increases along erythropoiesis, this dynamic transition is called GATA switching (Doré et al., 2012; Gao et al., 2016). Multiple binding sites for GATA1 and GATA2 are found within the regulatory regions of GATA2, including the -110 kb and +9.5 kb intronic enhancers (Gao et al., 2013; Grass et al., 2006; Lim et al., 2012; Martowicz et al., 2005; Snow et al., 2011). In addition, GATA2 interacts with FOG (Chang et al., 2002), AP1 (Kawana et al., 1995), PU.1 (Zhang et al., 1999), PPARγ (Tong et al., 2000), and HDAC3 and HDAC5 (Ozawa et al., 2001) among others. In addition, many factors regulate GATA2 transcription including CEBPA, HOXA9, ETS1, BMP4, NOTCH1, EVI1 and SPI1 and by cytokines TNFα and IL1 (Vicente et al., 2012).

With the advances in genome-wide analysis, some studies report the genome binding patterns of ten hematopoietic stem and progenitor regulators (SCL/TAL1, LMO2, LYL1, GATA2, RUNX1, MEIS1, PU.1, ERG, FLI-1 and GFI1B) in multipotent progenitor cells (MPP). They identified a complex network of interactions between a heptad of transcription factors - GATA2, SPI1, TAL1, LMO2, RUNX1, ERG and LYL1-

that specify early lineage commitment (Wilson et al., 2010; Beck et al., 2013). Most of heptad-bound regions have a GATA motif and approximately 40% of them contain a RUNX motif, suggesting an interaction where Gata2 binding might be required for the recruitment of Runx1 to promoters and enhancers (Wilson et al., 2010). Single-cell analysis revealed the dynamical function of the heptad by identifying two sets of TFs responsible for specific lineages, Fli1/Erg promote the endothelial lineage while Gata2/Runx1 promote the hematopoietic lineage (Bergiers et al., 2018). Detailed analysis revealed combinatorial interactions between GATA2, SCL and RUNX1 in controlling a set of genes responsible for the balance between proliferation and quiescence of HSCs (Wilson et al., 2010), consistent with the description from the previous section. Recently, Shu and colleagues demonstrated that all 6 GATA family members can replace Oct4 to induce pluripotency by inhibiting ectodermal specific genes (Shu et al., 2015).

GFI1B, growth factor independence 1, is differentially expressed within the hematopoietic system, including megakaryocytes, B cell progenitors, HSCs and erythrocytes (Vassen et al., 2007). Its expression is also dose and time dependent, during erythropoiesis Gfi1B expression is induced by a combination of GATA1 and NF-Y (Huang, 2005; Huang et al., 2004; Rodriguez et al., 2005). Gfi1B gene expression can be controlled through an auto-regulatory mechanism by binding to its own promoter (Anguita et al., 2010; Vassen et al., 2005). Based on its biochemical function as a transcriptional repressor, Gfi1b can occupy specific target gene promoters or it can modify chromatin structure by associating with LSD1, CoREST, histone methyltransferases G9A and SUV39H1 and histone deacetylases HDAC1-2 (Chowdhury et al., 2013; Laurent et al., 2012; Saleque et al., 2007; Thambyrajah et al., 2016; Vassen et al., 2006). Remarkably, Gfi1B can also bind y-satellite sequences at foci of pericentric heterochromatin region (Vassen et al., 2006) and several target genes have been identified including Bcl-LX, Socs1, Socs3, GATA3, Meis1 and Rag1/2 (Chowdhury et al., 2013; Jegalian and Wu, 2002; Kuo and Chang, 2007; Schulz et al., 2012; Xu and Kee, 2007).

Single cell transcription factor analysis has recently revealed and validated a dynamic regulatory relationship between Gata2, Gfi1 and Gfi1B in blood stem and progenitor cells. This putative transcription factor triad was predicted by bioinformatic analysis, in which Gata2 can activate expression of both Gfi1 and Gfi1B, whereas Gfi1 can repress

expression of Gata2 and as previously reported Gfi1 and Gfi1B can be mutually inhibited (Doan, 2004; Moignard et al., 2013). However, in MPP, Gata2 and Gfi1B binding patterns showed few overlap (Wilson et al., 2010).

FOS

The members of Fos family (Fos, FosB, Fra1, Fra2) can dimerize with proteins of the Jun family (Jun, JunB, JunD) to form the transcription factor complex Activator Protein 1 (AP1) which is involved in differentiation, apoptosis and cellular proliferation (Angel and Karin, 1991). It is thought that Fos proteins cannot form homodimers; however, recent evidence combining fluorescence microscopy and computer modelling confirmed that Fos proteins stably homodimerize and could bind to the chromatin (Szalóki et al., 2015). The AP1 proteins share a highly conserved basic leucine zipper (bZIP) domain and controls several genes containing AP1 sites (consensus sequence 5′-TGAG/CTCA3′) (Sassone-Corsi et al., 1988; Shaulian and Karin, 2001; Turner and Tijan, 1989).

Fos knockout mice are viable; however, some defects were observed related with growth, bone development, tooth eruption and also altered hematopoiesis (Johnson et al., 1992; Wang et al., 1992). In addition, Johnson and colleagues reported that null embryos showed reduced fetal weights and reduced placentas suggesting a role in extraembryonic tissues (Johnson et al., 1992). In fact, absence of Fos expression in mice placentas showed low levels of HSC activity (Ottersbach and Dzierzak, 2005). Fos deficient mice show a normal T-cell function (Jain et al., 1994) and Fos promotes angiogenesis *in vivo* and *in vitro* (Marconcini et al., 1999). In Xenopus, AP1 was shown to induce hematopoiesis via BMP4 during ontogeny (Lee et al., 2012). In zebrafish, Fos-Vegfd signaling regulates negatively the HSCs formation (Wei et al., 2014). Activation of AP1 is also shown to stimulate engraftment of HSCs by epoxyeicosatrienonic acids (Li et al., 2015).

Mckinney-Freeman and colleagues defined emergent and adult HSCs through transcriptome analysis which seem to adopt three different transcriptional states during ontogeny, designated as ‘‘Yolk Sac-like’’ while the most immature cells, ‘‘Specifying HSCs’’ and ‘‘Definitive HSCs’’. Genes such as *Fos* and *Fosb* were abundantly expressed in nascent but not adult HSCs (McKinney-Freeman et al.,

2012). During mouse ESC differentiation, it was shown that AP1 motifs were enriched in open chromatin regions and co-localized with TF-binding sites that were specific to HE cells (Goode et al., 2016). In endothelial cells, GATA2 co-occupied sites bound by JUN and FOS and gene ontology analysis identified the co-targeted genes as inflammatory genes (Kanki et al., 2011; Linnemann et al., 2011).

In summary, Gata2, Gfi1b and Fos are expressed in the early embryo and show important roles in hematopoietic fate. How the TFs function during direct reprogramming is yet unknown. Whether the transcription factors act sequentially or all at the same time, whether they regulate each other, and what are their respective targets will be the focus of this thesis.

### 1.3.4. Epigenetic modifications during reprogramming

Induced reprogramming is essentially an epigenetic phenomenon. Ectopic expression of the same transcription factors in somatic cells do not always reprogram to the same fate. Epigenetic modifications inform how the gene expression state of a cell is altered without changing the DNA sequence. Consequently, a fully differentiated cell and a stem cell within an organism hold the same genome but a diverse epigenome. The stability of a phenotype results from a combination of multiple actors including cis-acting epigenetic modifications like post-translational histones modifications, DNA methylation, nucleosome positioning, and trans-acting factors like noncoding RNAs, chromatin remodeling complexes, transcriptional coactivators and sequence-specific DNA binding transcription factors (reviewed in Vierbuchen and Wernig, 2012).

One of the earliest studies highlighting the importance of epigenetics in nuclear reprogramming was performed through cell fusion. HeLa cells were treated with 5-azacytidine (5-AzaCR), an inhibitor of DNA methyltransferases, prior to fusion with muscle cells and the resulting heterokaryon showed activation of muscle genes (Chiu and Blau, 1985). Shortly after, the skeletal muscle gene, MyoD, was identified using the same treatment (Lassar et al., 1986) and its cDNA was further transfected into fibroblasts treated with 5-AzaCR, converting them into stable myoblasts (Davis et al., 1987). These studies indicate that the pre-existing state of the initial cell is determinant for an efficient cellular conversion.

Since the discovery of transcription factor mediated conversion to pluripotency, iPSCs, the understanding of epigenetic states has been enhanced and, in some cases, well documented. For a fully pluripotent reprogramming to occur, the somatic epigenetic signature must be removed and a new epigenome must be established. Successfully reprogrammed cells exhibit some modifications such as reactivation of endogenous genes and the silent X chromosome, reactivation of telomerase activity, DNA methylation of lineage-specific genes, and reorganization of post-translational histone marks (Mikkelsen et al., 2008; Polo et al., 2012; Stadtfeld et al., 2008; Takahashi and Yamanaka, 2006; Wernig et al., 2007). Temporal analysis of large gene expression datasets in iPSC allowed the identification of three distinct phases – initiation, maturation and stabilization - during reprogramming (Samavarchi-Tehrani et al., 2010). The most well characterized phase is the initiation phase in which morphological changes, involving mesenchymal-to-epithelial transition (MET) by silencing the transcription factors Snail1/2 and Zeb1/2 and the subsequent upregulation of epithelial markers E-cadherin (CDH1) and Epcam, and apoptosis occurs within the first 48h (Mikkelsen et al., 2008; Samavarchi-Tehrani et al., 2010; Stadtfeld et al., 2008). Changes in histones marks have also been reported (Table 1.2), and at this stage, modification of H3K4me2 occurs by one cell division as well as transcriptional changes arise at genes marked by H3K4me3 (Koche et al., 2011). In addition to genes of EMT, such as Snail, also fibroblast genes are associated with the activation marks H3K4me2 and H3K4me3 and are defined by an open chromatin state (Soufi et al., 2012). During the initiation phase, the acquisition of pluripotent-associated genes associated with repressive marks like H3K27me3 is essential for induced pluripotency as well as gaining resistance to cell senescence or apoptosis (David and Polo, 2014; Mikkelsen et al., 2008; Soufi et al., 2012). The transcription factors Oct4, Klf4 and Sox2 were shown to bind to closed chromatin within the first 48h by occupying distal regulatory elements that have gained *de novo* the activation mark H2K4me2, while cMyc strictly binds to open chromatin regions (Koche et al., 2011; Soufi et al., 2012). Despite not being absolutely required for reprogramming, cMyc is thought to serve as facilitator of the transcriptional action of OSK factors and to have a role in the induction of MET whereas the other TFs act as pioneers to shift nucleosomes in closed chromatin and recruit other TFs (Soufi et al., 2012, 2015). Recent work has been shown that regardless of their pioneer potential, OSK bind to the chromatin in a cooperative manner (Chronis et al., 2017). In addition, the

acquisition of pluripotency was demonstrated to be a stepwise process in which OSK silence at early stages of reprogramming the pre-established fibroblast program and bind to pluripotency-related enhancers to induce reprogramming (Chronis et al., 2017). The acquisition of pluripotent markers was also shown to be gradual with some genes being expressed in the earliest phase and others in the latest phase (Buganim et al., 2012; Polo et al., 2012). The major bottleneck occurs between the initiation and the maturation phase of reprogramming. In this phase the first pluripotent genes are activated including Nanog, Esrrb and endogenous Oct4 (David and Polo, 2014). In the stabilization phase, acquisition of a pluripotent program is maintained without the enforced expression of reprogramming factors (Wernig et al., 2007).

Aside from Yamanaka factors, other molecules have been implicated with a higher iPSC reprogramming efficiency such as, inhibitors of histone deacetylases, histone demethylases and DNA methyltransferases. Studies have been shown that these molecules can be either combined with OSKM or even replace some of the TFs (Feng et al., 2009; Krishnakumar and Blelloch, 2013; Nie et al., 2012). With an ever-growing list of post-translational modifications identified in recent years (Bannister and Kouzarides, 2011), in Table 1.2 selected histone marks that have been playing a significant role during TF mediated reprogramming are listed.

Table 1.2 - List of selected histone marks during TF mediated reprogramming.

| Histone mark | Function associated | Reprogramming fate | References |
|---|---|---|---|
| H3K4me1 | Marks active or poised enhancers | Neuronal reprogramming; iPSCs | (Chronis et al., 2017; Wapinski et al., 2013) |
| H3K4me2 | Marks active and poised promoters and enhancers | iPSCs | (Koche et al., 2011; Soufi et al., 2012) |
| H3K4me3 | Marks TSS of active and poised genes | iPSCs; Cardiac reprogramming; | (Ieda et al., 2010; Koche et al., 2011; Polo et al., 2012; Soufi et al., 2012) |
| H3K9me3 | Marks heterochromatin regions | iPSCs; Neuronal reprogramming | (Soufi et al., 2012; Wapinski et al., 2013) |
| H3K27ac | Marks open chromatin and active enhancers | iPSCs; Neuronal reprogramming; Trophoblast reprogramming | (Chronis et al., 2017; Rhee et al., 2017; Wapinski et al., 2013) |
| H3K27me3 | Marks repressed genes | iPSCs; Cardiac reprogramming; Neuronal reprogramming | (Ieda et al., 2010; Koche et al., 2011; Polo et al., 2012; Wapinski et al., 2013) |
| H3K36me3 | Marks active genes | iPSCs | (Koche et al., 2011) |

As seen in Table 1.2 histone modifications have been reported mainly in induced pluripotency studies, recently, the mechanisms for direct reprogramming towards neurons were also studied, where they reveal a hierarchical induction with Ascl1 as a pioneer factor by binding to “on target” genomic sites when compared to neural precursor cells, and by recruiting Brn2 and Myt1l (Wapinski et al., 2013). A trivalent chromatin signature was identified for H3K4me1, H3K27ac and H3K9me3 (Table 1.2). This signature was associated with Ascl1 binding sites and was shown to predict the permeability of chromatin and consequently the reprogramming fate (Wapinski et al., 2013). In cardiac direct reprogramming, the enrichment of the repressive mark H3K27me3 and the active mark H3K4me3 revealed a “bivalent” state in which the cardiac genes promoters showed an enrichment in H3K4me3 and depletion of H3K27me3 (Ieda et al., 2010).

Overall, the epigenetic machinery seems to be a key player in the establishment and maintenance of the reprogrammed fate. Taking into account that epigenetic processes also plays a fundamental role in cell state transition during embryonic development and despite no epigenetic study showed, up to date, its role in hemogenic reprograming, thus, the epigenetic landscape during hemogenic reprograming will be address in this thesis.

## 1.4. Aims of this study

Hematopoiesis, the process of blood formation, is dependent on self-renewing and multipotent HSCs. The generation of autologous long term self-renewing HSCs *in vitro* as replacement therapy for blood disorders has been considered the holy grail of regenerative medicine. To achieve this, it is vital to understand the complex signaling cascades and the transcriptional networks underlying HSC emergence during development.

Direct conversion of a cell type into another by overexpression of specific TFs offers a powerful tool to generate autologous cells from unrelated cell-types, to interrogate lineage specification, and to investigate underlying mechanisms of TF mode of action. The goal of this project is to establish direct hematopoietic reprogramming as a comprehensive model for molecular dissection of HSC specification and definitive EHT in both mouse and human systems. For this purpose, the third chapter will focus on using direct reprogramming to inform *in vivo* developmental hematopoiesis and HSC emergence. In this chapter mouse placentas will be used to isolate the immediate precursor cell of HSCs during embryogenesis based on a phenotype revealed by cellular reprogramming. The fourth chapter will focus on translating hemogenic reprogramming to human cells and the comprehensive characterization of induced cells. Chapter five will focus on the elucidation of the underlying mechanisms of human hemogenic reprogramming. By using induction from human fibroblasts, I uncover how hemogenic TF interact and cooperatively set in motion hematopoietic reprogramming. Finally, chapter six concludes with a discussion of our studies and implications.

This work might provide valuable insights for the understanding of transcriptional regulation of definitive hematopoiesis while helping accelerate the pace of HSC induction potential for therapeutic application.

# Chapter 2 – Materials and Methods

## 2.1 Materials

Table 2.1 - Antibodies used in this study.

| Antibody/Antigen | Clone | Conjugate | Source | Application |
|---|---|---|---|---|
| mCD133 (Prom1) | 13A4 | APC, Biotin, PE | eBioscience | FACS |
| mLy-6A/E (Sca1) | D7 | Pacific blue, APC-Cy7 | Biolegend | FACS |
| mCD34 | RAM34 | Alexa Fluor 700 | eBioscience | FACS |
| mCD45 | 30-F11 | Pac Blue | Biolegend | FACS |
| mCD49f | GoH3 | PE | eBioscience | FACS |
| mCD31 (PECAM-1) | 390 | PE | eBioscience | FACS |
| mCD105 | MJ7/18 | PE | eBioscience | FACS |
| mCD144 (VE-CAD) | eBioBV13 | APC | eBioscience | FACS |
| mCD202b (Tie2) | TEK4 | PE | eBioscience | FACS |
| mCD41 | eBioMWReg30 | APC | eBioscience | FACS |
| mCD117 | 2B8 | PE, APC | eBioscience, Biolegend | FACS |
| hCD34 | 581; 8G12 | PE-Cy7; PE | BD Biosciences | IF, FACS |
| hCD49f | GoH3 | PE; PE-Cy5 | BD Biosciences | IF, FACS |
| hCD45 | HI30 | V500 | BD Biosciences | FACS |
| hCD45RA | HI100 | FITC | BD Biosciences | FACS |
| hACE | BB9 | PE | BD Biosciences | FACS |
| hCD90 | 5E10 | APC | E-Bioscience | FACS |
| hCD38 | HIT2 | AF700 | BD Biosciences | FACS |

| | | | | |
|---|---|---|---|---|
| hCD133 | AC133 | Biotin | Miltenyi Biotec | FACS |
| hCD3 | UCHI1 | APC-eFluor 780 | eBioscience | FACS |
| hCD19 | HIB19 | PE | Biolegend | FACS |
| hCD11c | 3.9 | Alexa 488 | Biolegend | FACS |
| hCD14 | HCD14 | PE-Cy7 | Biolegend | FACS |
| hCD34 | Diamond CD34 Isolation kit | FITC | Miltenyi Biotech | MACS |
| Lineage - | Diamond CD34 Isolation kit | Biotin | Miltenyi Biotech | MACS |
| FLAG | M2 | - | F1804, Sigma | WB, ChIP, IF |
| HA | 4C12 | - | Ab9110, Abcam | WB,ChIP, IF |
| FOS | 4 | - | Sc-52, Santa Cruz Biotechnology | WB, ChIP, IF |
| Rabbit IgG | Sc-2027 | - | Santa Cruz Biotechnology | ChIP, IP |
| β-Actin | A2066 | - | Sigma | WB |

Table 2.2- Plasmids used in this study.

| Plasmid | Name | Source/Reference |
|---|---|---|
| Empty backbone | pFUW-TetO-mOrange | (Pereira et al., 2013) |
| Transactivator | pFUW-TetO-M2rtTA | (Pereira et al., 2013) |
| Envelope plasmid | pMD2 | (Pereira et al., 2013) |
| Packaging plasmid | pPAX2 | (Pereira et al., 2013) |
| HA epitope tag plasmid | pLV-TRE-HA | Lemischka's group |
| 3xFLAG epitope tag plasmid | pJW321-3xFLAG | Wang's group |

Table 2.3 - Cell lines used in this study.

| Cell line | Name | Source |
|---|---|---|
| HDF | Human Dermal Fibroblast | ScienCell |
| BJ | Neonatal Foreskin Fibroblast | ATCC |
| HEK293T | Human embryonic kidney cells | ATCC |
| OP9 | Murine bone marrow stroma | ATCC |
| OP9-DL1 | Murine bone marrow stroma transduced with Delta-1 | ATCC |
| ATF024 | Murine fetal liver stroma | ATCC |

Table 2.4 - Public datasets analyzed in this study.

| | GEO accession | Cell type | Reference | Experiment |
|---|---|---|---|---|
| H3K4me3 | GSM733650 | HDFs | (Dunham et al., 2012) | ChIP-seq |
| H3K4me1 | GSM1003526 | HDFs | (Dunham et al., 2012) | ChIP-seq |
| H3K27ac | GSM733662 | HDFs | (Dunham et al., 2012) | ChIP-seq |
| H3K27me3 | GSM733745 | HDFs | (Dunham et al., 2012) | ChIP-seq |
| H3K9me3 | GSM1003553 | HDFs | (Dunham et al., 2012) | ChIP-seq |
| H3K36me3 | GSM733733 | HDFs | (Dunham et al., 2012) | ChIP-seq |
| DNase I | GSM736567 | HDFs | (Thurman et al., 2012) | ChIP-seq |
| H3K4me3 | GSM733680 | K562 | (Dunham et al., 2012) | ChIP-seq |
| H3K4me1 | GSM733692 | K562 | (Dunham et al., 2012) | ChIP-seq |
| H3K27ac | GSM733656 | K562 | (Dunham et al., 2012) | ChIP-seq |
| H3K27me3 | GSM733658 | K562 | (Dunham et al., 2012) | ChIP-seq |
| H3K9me3 | GSM733776 | K562 | (Dunham et al., 2012) | ChIP-seq |
| H3K36me3 | GSM733714 | K562 | (Dunham et al., 2012) | ChIP-seq |
| GATA2 | GSM935373 | K562 | (Pope et al., 2014) | ChIP-seq |
| FOS | GSM935355 | K562 | (Pope et al., 2014) | ChIP-seq |
| GFI1B | GSM1278242 | Proerythroblast | (Pinello et al., 2014) | ChIP-seq |

| | | | | |
|---|---|---|---|---|
| Hemogenic reprogramming | GSE47497 | MEFs | (Pereira et al., 2013) | mRNA-seq |
| Human HSCs | GSE29105 | UCB HSCs | (Notta et al., 2011) | Microarray |

## 2.2. Methods

### 2.2.1 SubCloning

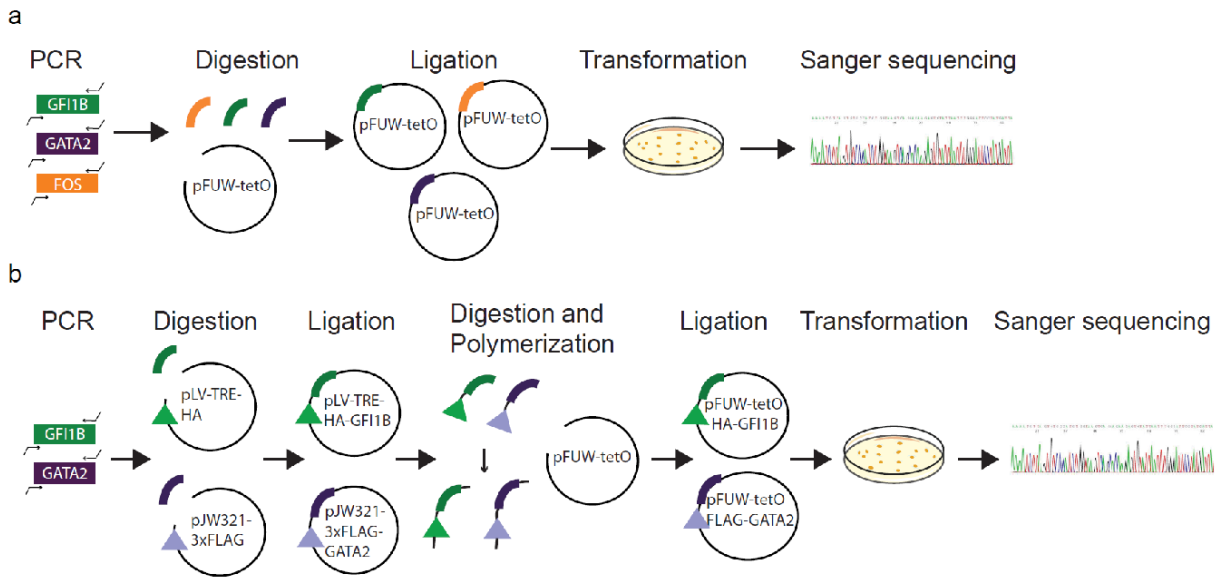



Figure 2.1 - **Strategy for subcloning human transcription factors coding sequences in inducible lentiviral vectors**. **(a)** GFI1B (green), GATA2 (purple) and FOS (orange) PCR products were digested and ligated with pFUW-tetO plasmid vector. Positive bacterial colonies of competent bacteria with constructs were screened by PCR and confirmed by Sanger sequencing. (**b**) GFI1B and GATA2 were subcloned into vectors containing epitope tags HA and 3xFLAG, respectively. The tags and constructs were then digested, treated with Klenow polymerase I and then subcloned into pFUW-tetO plasmid vector. Positive colonies were screened by PCR, and their sequences were confirmed by Sanger sequencing.

#### 2.2.1.1 Polymerase Chain Reaction (PCR)

To amplify the DNA fragment from a plasmid template DNA to further clone them, a PCR reaction was assembled according to Table 2.5 using Phusion Flash (Thermo Scientific) high-fidelity PCR Master Mix (30 cycles of 98°C for 10 sec; 62°C for 15 sec and 72°C for 1 min) with primers listed on Table 2.6. Subsequently PCR products were purified by gel electrophoresis if needed followed by a DNA clean-up.

Table 2.5 - PCR reaction mix to amplify fragments from plasmid DNA.

| Volume (µl) | Reagent |
|---|---|
| 15 | Phusion Master Mix (2x) |
| 8 | $ddH_2O$ |
| 3 | Forward Primer (5µM) |
| 3 | Reverse Primer (5µM) |
| 1 | Template DNA |

Table 2.6 - Primer list used for cloning of human coding sequences.

| TF | Primer Forward | Primer Reverse |
|---|---|---|
| GFI1B | AATATCGAATTCATGCCACGCTCCTTCCTG | GGTATCGAATTCTCACTTGAGATT GTGCTGGCT |
| FOS | AATATCGAATTCATGATGTTCTCGGGCTTCAACGCAG | GGTATCGAATTCTCACAGGGCCAGCAGCGTGGG |
| GATA2 | AATATCGAATTCATGGAGGTGGCGCCCGAGCAGCCG | GGTATCGAATTCTCACTAGCCCATGGCGGTCAC |
| HA-GFI1B | ATCGAATTCACCATGGGCCACCATGTACCCATAC | GGTATCGAATTCTCACTTGAGATTGTGCTGGCT |
| 3xFLAG-GATA2 | AATATCGCTAGCATGGAGGTGGCGCCCGAGCAGCCG | GGTATCTTAATTAATCACTAGCCAGGCGGTCAC |
| TRE (sequencing primer) | TCC ACG CTG TTT TGA CCT CC | - |
| CR17AS | GGGATAATTTCAGCTGACTAAACAG | TTCCGTTTAGTTAGGTGCAGTTATC |

#### 2.2.1.2 Gel Electrophoresis

Gel electrophoresis was performed using 1% agarose gel in TAE buffer with 0.5 µg/ml EtBr. Orange loading dye (6x) was added to PCR products and both DNA ladder and samples were loaded into the pouches of the agarose gel and a voltage of 120V was applied. The fragments were cut out of the gel and DNA clean-up was subsequently performed.

#### 2.2.1.3 DNA gel extraction

PCR products or DNA fragments cut out of an agarose gel was performed using QIAquick Gel Extraction kit from Qiagen according to its protocol. In the elution step, 30 µl of TE buffer or dd$H_2O$ were used.

#### 2.2.1.4 Restriction digestion

30 µl of PCR product and 1µg of plasmid vector were restricted with their respective enzymes. The digests were incubated for 1-2h at 37°C and the plasmid vector was treated with alkaline phosphatase calf intestinal (CIP, Roche) for 15 min at 37°C to avoid circularization of the vector (Figure 2.1).

#### 2.2.1.5 Generating blunt end DNA fragments

Blunting reaction for tagged constructs were performed using 1 unit of DNA Polymerase I Large Klenow Fragment (M0210 – New England Biolabs), 1x NEB Buffer supplemented with 33 µM dNTPs in a total reaction volume of 30 µl. The reaction mixture was incubated for 15 min at room temperature and subsequently purified (Figure 2.1).

#### 2.2.1.6 Ligation

DNA ligations were performed using 1 µl of T4 DNA Ligase (Thermo Scientific), 2µl 10x T4 DMA Ligase Buffer and linear vector DNA in a 3:1 molar ratio of insert over vector DNA in a total reaction volume of 20µl. The reaction mixture was incubated for 60 min at room temperature. A ligation without inserts was added as a control.

#### 2.2.1.7 Heat shock transformation

Competent *E.coli* were thawed on ice before 10µl ligation reaction mixture were added to 50µl competent cells, gently mixed an incubated for 30 minutes on ice. Heat shock was performed for 1 minute at 42°C in a water bath and immediately incubated on ice for 5 minutes. Bacteria were plated on LB agar plates supplemented with 100 µg/ml ampicillin and incubated overnight at 37°C.

#### 2.2.1.8 Plasmid preparation (Mini and Maxi)

For small scale plasmid preparations, 3 ml LB medium supplemented with antibiotic (100ug/ml ampicillin) was inoculated with a single colony and incubated overnight at 37°C and 200rpm. 1 ml of the suspension was then transferred to a 1,7ml tube, centrifuged for 10 min at 6,000rpm, the supernatant was discarded and QIAprep Minirep kit was used for Plasmid DNA Purification. 250ul of P1 buffer was added and the pellet was dissolved completely by vortexing. Then, 250ul P2 buffer was added, the solution was mixed by inverting 6 times and the mixture was incubated for 5 min at room temperature. In the following step, 350 µl N3 was added, the solution was mixed by inverting 6 times before the mixture was centrifuged for 10 min at 13,000 rpm. The supernatant was carefully transferred to the spin column avoiding any transfer of the precipitant and centrifuged for 1min. The spin column was washed with 750 µl of PE buffer and centrifuged for 1 min. The flow-through was discarded and the spin column was then placed in a new tube. The elution buffer (EB) was added, let stand for 1 min and centrifuged for another minute. The purified plasmid DNA was stored at 4°C.

For large scale plasmid preparation 250ml LB medium supplemented with the appropriate antibiotic (100 µg/ml ampicillin) was inoculated with 1 ml bacterial culture and incubated overnight at 37°C and 200 rpm.

The plasmid purification was performed using the GenElute High Performance (HP) Plasmid Maxiprep kit (Sigma) with the following alterations to the manufacturer's protocol: first, the bacterial culture was harvested by centrifugation at 4000 rpm for 10 min. Second, the spin format was preferred and a centrifugation step at 3000g for 1 min after addition of Neutralization Solution was done before applying the solution to the column. Third, the elution plasmid DNA step was performed for maximum

concentration of plasmid. Last, DNA precipitation was performed at 4000 rpm for 60 min instead of 15000g for 30 min at 4°C.

#### 2.2.1.9 Sanger Sequencing

Sequencing of DNA was done by GENEWIZ with 10µl of plasmid DNA (300 ng/µl) including 10pmol of the vector primer – TRE primer (Table 2.6). Sequencing results were aligned with the consensus CDS from NCBI with NTI software.

### 2.2.2 Cell isolation

C57BL/6 pregnant mice (The Jackson Laboratory) were used to dissect placentas between E10.5 and E12.5. The day of copulation plug was described as E0.5. Placentas were dissected and separated from the umbilical cord and maternal decidua. Tissues were kept on ice, washed in PBS, dissociated mechanically through an 18G needle, and treated with 0.2%fetal bovine serum (Benchmark) for 1.5hr at 37°C, followed by passages through 20G and 25G needles. Single cell suspensions were filtered through 70µm cell strainers (BD Falcon). Placentas from the same litter were combined for cell isolation. Animal experiments and procedures were approved by the Institutional Animal Care and Use Committee and conducted in accordance with the Animal Welfare Act.

### 2.2.3 Cell and Cell Culture

Placental-derived PS34CD45- cells were cultured in gelatin-coated dishes. Cultures were maintained in MyeloCult media (M5300; Stem Cell Technologies) supplemented with hydrocortisone (10-6M; Stem Cell Technologies) and 100 $ngml^{-1}$ SCF, 100 $ngml^{-1}$, Flt3L, 20 $ngml^{-1}$ IL-3, 10 $ngml^{-1}$ TPO, and 20 $ngml^{-1}$ IL-6 (R&D). The AFT024 cell line was cultured in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) containing 10% fetal bovine serum (FBS; Benchmark), 1mM L-Glutamine and penicillin/streptomycin (10 $\mu gml^{-1}$; Invitrogen) at 32°C and mitotically inactivated by irradiation as previously described (Moore et al 1997). The stromal cells from mouse embryos bone marrow OP9 and OP9-DL1 were co-cultured as feeder layers with PS34CD45- cells in the presence or absence of 100 $ngml^{-1}$ SCF, 100 $ngml^{-1}$ Flt3L, 20 $ngml^{-1}$ IL-3, 10 $ngml^{-1}$ TPO, and 20 $ngml^{-1}$ IL-6. The OP9 and OP9-DL1 cell lines were maintained in Minimum Essential Medium (MEM) Alpha (Gibco) containing 20%

fetal bovine serum (FBS; Hyclone), 1mM L-Glutamine and penicillin/streptomycin (10µgml$^{-1}$; Invitrogen) in 5% $CO_2$ at 37°C. OP-9 and OP-9-DL1 were never allowed to reach 100% confluency. 293T cells (ATCC) were maintained in DMEM (Dulbecco's Modified Eagle Medium, Invitrogen), 1mM L-Glutamine (Invitrogen, 10% of Fetal Bovine Serum (FBS) (Benchmark) and 1mM penicillin/streptomycin (Invitrogen). Human adult dermal fibroblasts (HDF, ScienCell), neonatal foreskin fibroblasts (BJ, ATCC) were grown in gelatin coated dishes with fibroblast media DMEM (Invitrogen) containing 10% FBS (Benchmark), 1mM L-Glutamine and penicillin/streptomycin (10 µgml$^{-1}$, Invitrogen) in 5% $CO_2$ at 37ºC until confluent.

### 2.2.4 Lentivirus Production

293T cells were used as packaging cells to produce 2$^{nd}$ generation lentiviral vectors pFUW-TetO lentiviral vector where expression is under the control of the tetracycline operator and a minimal CMV promoter. Lentiviral vectors containing the reverse tetracycline transactivator M2rtTA under the control of a constitutively active human ubiquitin C promoter (FUW-M2rtTA) (Pereira et al., 2013). The cells were transiently transfected with a standard calcium phosphate precipitation protocol where 293T cells were seeded onto 10 cm tissue culture dishes to reach up to 60-70% confluency. 84µg of transfer plasmid pFUW-TetO, 84µg of pPAX2 and 42µg of pMD2 plasmids were mixed in a 15 ml tube and adjusted with Molecular Grade Water (Corning) to a final volume of 2ml. 250µl of $CaCl_2$ (Sigma) and 2ml of BES-buffered saline solution (Sigma) were added to the previous solution and incubated for 15 min. DMEM (Invitrogen) containing 10% FBS (Benchmark), 1mM L-Glutamine and 25µM of chloroquine replaced D10 medium and 1 ml of DNA mixture was added directly to 293T cells dropwise and uniformly through the medium. Cells were incubated overnight at 37°C and medium was changed and cells incubated at 32°C. Viral supernatants were harvested after 36, 48 and 72 hours, filtered (0.45 µm), concentrated 40-fold with Amicon ultra centrifugal filters (Millipore) and stored at -80°C. Transductions with mOrange in pFUW-tetO resulted in >95% efficiency.

### 2.2.5 Methylcellulose clonogenic assays

Clonogenic progenitors were assayed in 1% methylcellulose media (Methocult M3434, Stem Cell Technologies) supplemented with TPO (10 ngml$^{-1}$). Hematopoietic colonies

were scored and counted after 7-10 days of culture in 5% $CO_2$ at 37ºC. CFU-Mix, CFU-GM, CFU-G, CFU-M and CFU-E were identified.

### 2.2.6 Immunofluorescence

Live immunofluorescence was performed with Phycoerythrin (PE)-conjugated sterile rat monoclonal antibodies against CD49f and CD34 (Table 2.1) at a 1:20 dilution. Emergent colonies were washed once with PBS 5% FBS and incubated with conjugated antibodies for 30 min at room temperature in the presence of mouse serum. Cultures were then washed twice with PBS 5% FBS to remove unbound antibody. For pinocytosis experiments 5(6)-Carboxyfluorescein (Sigma) was incubated for 16 hours in transduced HDFs and washed twice with PBS 5% FBS. For detection of tagged transcription factors, fibroblasts were fixed with 2% PFA for 20 min, permeabilized with 0.4% Triton X-100, blocked for 30 min and incubated with anti-FLAG, anti-HA, anti-FOS antibodies (Table 2.1) at 1:200 dilution for 2 hours. Cells were washed, incubated with secondary antibodies conjugated with Alexa Flour 488 (Invitrogen, A12379) and nuclear counterstained with 4,6-diamidino-2-phenylindole (1 $\mu g ml^{-1}$, Sigma). Cells were visualized on a Leica DMI4000 microscope and processed with Leica software and Adobe Photoshop.

### 2.2.7 Genomic PCR

Genomic DNA was isolated using Easy DNA extraction kit (Invitrogen). Presence of human sequences was checked by PCR using Phusion Flash (Thermo Scientific) high-fidelity PCR Master Mix (30 cycles of 98°C for 1 sec; 60°C for 5 sec and 72°C for 15 sec) with primers for the chromosome 17 alpha satellite sequences (Table 2.6).

### 2.2.8 Flow Cytometry Analysis and Fluorescence-Activated Cell Sorting (FACS)

Mouse PS34 CD45- cell populations and human CD34CD49f cell populations were dissociated with TrypLE Express or Accutase Cell detachment solution (Innovative Cell Technologies, Inc) and stained with fluorochrome-coupled antibodies (Table 2.1). Cell populations were isolated on an InFlux cell sorter (BD Biosciences) and immediately lysed in Trizol (Ambion) for RNA extraction, cultured on 0.1% gelatin coated 6-well plates in Myelocult media or transplanted. Flow cytometric analysis was performed on a 5-laser LSRII with Diva software (BD Biosciences) and further

analyzed using FlowJo software. DAPI (1 μgml$^{-1}$) was added before analysis to exclude dead cells.

### 2.2.9 Cytospin and Benzidine Staining

3,000-100,000 cells were washed twice in cold 5% FBS in PBS and diluted in 200 μl of cold 1% FBS in PBS. The samples were loaded into the appropriate wells of the Cytospin centrifuge equipment. Samples were spun at 500 r.p.m. for 3 min to allow cellular adherence to the slides. Slides were fixed and stained using DIFF Stain Kit (IMEB, inc), followed by 10 min in PBS and a quick wash in distilled water. Slides were viewed on a Nikon Eclipse TE2000-U microscope and images acquired using Nikon ACT-1 software. For Benzidine staining of hemoglobin, 1 tablet of 3,3',5,5'-Tetramethylbenzidine (Sigma) was dissolved in 10 ml of PBS, filtered and supplemented with 10 μl of 30% $H_2O_2$ before use. To directly stain mixed colonies in the methylcellulose media, 1 ml of staining solution was added and incubated at 37°C for 30 minutes.

### 2.2.10 Long-term Repopulation Assays

Rag2/IL2rγ (B10; B6-*Rag2tm1Fwa Il2rgtm1Wjl*, Taconic) were used as recipients for directly isolated placental cells or cells co-cultured with stromal cells. After 3-4 days of co-culture,

cells were dissociated with Accutase Cell detachment solution (Innovative Cell Technologies). Animals were transplanted with cells intravenously by retro-orbital injection. Up to 6 hours before transplantation with placental derived cells Rag2/IL2rγ or SJL mice received a sublethal total body irradiation dose of 600 cGy. The number of cells transplanted per animal was in the range of 2-4 embryo equivalents cells per animal. Starting 4 weeks after transplantation, mice were bled on the orbital venous plexus and donor contribution was assessed by FACS with anti-CD45.1/CD45.2 (for SJL) and anti- CD19/CD4/CD8 (for Rag2/IL2rγ) antibodies. Red blood cells were lysed with Pharm Lyse Buffer (BD) for 5 min on ice. Dead cells were excluded by staining with 4,6-diamidino-2- phenylindole (DAPI, 1 μg/mL, Sigma).

NSG (NOD.Cg-*Prkdc*$^{Scid}$*Il2rg*$^{tm1Wjl}$/Sz, Jackson laboratories) mice were used as recipients for human induced cells. For isolating a population that include both CD49f$^+$ and CD34$^+$ cells, cultures were dissociated with Accutase 25 days after transduction

and stained with PE-conjugated anti-CD49f and PE-conjugated CD34. The PE-positive cell population was isolated by FACS sorting. Animals were transplanted with cells intravenously by retro-orbital injection. Up to 6 hours before transplantation with human cells 4-week old female NSG mice received a sublethal total body irradiation dose of 200 cGy. The numbers of $PE^+$ cells transplanted were in the range of 100,000 cells per animal. Starting 3-4 weeks after transplantation, NSG mice were bled on the orbital venous plexus and human contribution was assessed by FACS with anti-CD45 antibodies (Table 2.1) and by PCR with human specific primers. Animal experiments and procedures were approved by the Institutional Animal Care and Use Committee and conducted in accordance with the Animal Welfare Act.

### 2.2.11 Isolation of Tissues from Recipient Mouse and Isolation of HSCs from Bone Marrow

At the end of each experiment mice were sacrificed and peripheral blood, bone marrow (BM), and spleen of engrafted mice were further analyzed by FACS. The BM from primary recipients was used in secondary transplantation experiments. Total BM cells were harvested from long bones (tibias and femurs) by crushing with pestle and mortar in PBS supplemented with 5% FBS (Benchmark). Bone debris was filtered away with 70 μm cell strainers (BD). Red blood cells were lysed with Pharm Lyse Buffer (BD) for 5 min on ice and further filtered through 45 μm cell strainers to obtain a single-cell suspension. Dead cells were excluded by staining with 4,6-diamidino-2-phenylindole (DAPI, 1 μg/mL, Sigma). For HSC isolation, BM was harvested from the long bones and hips of primary transplant recipients as described. BM was depleted of lineage marker- (B220, CD3ε, Ter119, CD11b, Gr-1) and CD48-expressing cells via magnetic separation using a cocktail of biotinylated antibodies and streptavidin-coated magnetic beads (Dynabeads, Life Technologies). Lineage and CD48 depleted BM was then further stained for HSC markers prior to FACS sorting. Lineage-negative, CD48-negative, Sca-1-positive, cKit-positive and CD150-positive (LSK48-150+) HSCs were sorted from engrafted primary mice. 1,300-2,000 LSK48-150+ cells were transplanted per mouse for secondary transplantation assays.

### 2.2.12 Lineage Tracing

We crossed homozygous Prom1-CreERT2 mice (Zhu et al., 2009) with R26StopYFP mice (Cooley et al., 2014) and activated Cre with a 4-Hydroxytamoxifen (Sigma) injection (IP, 300 ng in 150µl per mouse) at E10.5. We analyzed the presence of YFP-positive cells in the peripheral blood of the progeny of injected and non-injected mice at 2 and 15 months of age. Analysis of the bone marrow and spleen was performed at 15 months of age. As an additional control we injected 4-Hydroxytamoxifen (IP) in adult heterozygous mice (Prom1-CreERT2 x R26StopYFP) and verify the lack of YFP positive cells in the peripheral blood 2 months after injection.

### 2.2.13 Immunoprecipitation (IP) and Co-IP

Nuclear extracts were prepared from HDFs with ectopic expression of 3xFLAG-tagged GATA2, HA-tagged GFI1B and FOS and incubated with 5 µg of each antibody: anti-FLAG (Sigma, F1804), anti-HA (Abcam, ab9110), anti-FOS (SantaCruz Biotechnology, sc-52) and protein G Agarose beads (Sigma) (Table 2.1). The immune complexes were then washed four times with the lysis buffer by centrifugation. Unless otherwise specified, all IP/co-IP were performed using approximately 5% of input samples. For the control IP, we used 5 µg of rabbit IgG (Santa Cruz Biotechnology, sc2027) (Table 2.1). Samples were heated in SDS sample buffer and processed by western blotting.

### 2.2.14 Western Blot Analysis

Cells were lysed in RIPA-B buffer (20 mM Na2HPO4 [pH 7.4], 150 mM NaCl, 1% Triton X-100) in the presence of protease inhibitors (3 µg/ml aprotinin, 750 µg/ml benzamidine, 1 mM phenylmethylsulfonyl fluoride, 5 mM NaF and 2 mM sodium orthovanadate) and incubated on ice for 30 min with occasional vortexing. Samples were centrifuged to remove cell debris and heated in SDS sample buffer. For immunoblotting, membranes were blocked with TBST buffer (10 mM Tris-HCl (pH 7.9), 150 mM NaCl, and 0.05% Tween 20) containing 3% milk, incubated with primary antibodies, washed three times with TBST, incubated with HRP-conjugated secondary antibodies, washed three times with TBST and subsequently detected by ECL (Thermo Scientific) or Femto (Thermo Scientific).

### 2.2.15 Chromatin Immunoprecipitation analysis

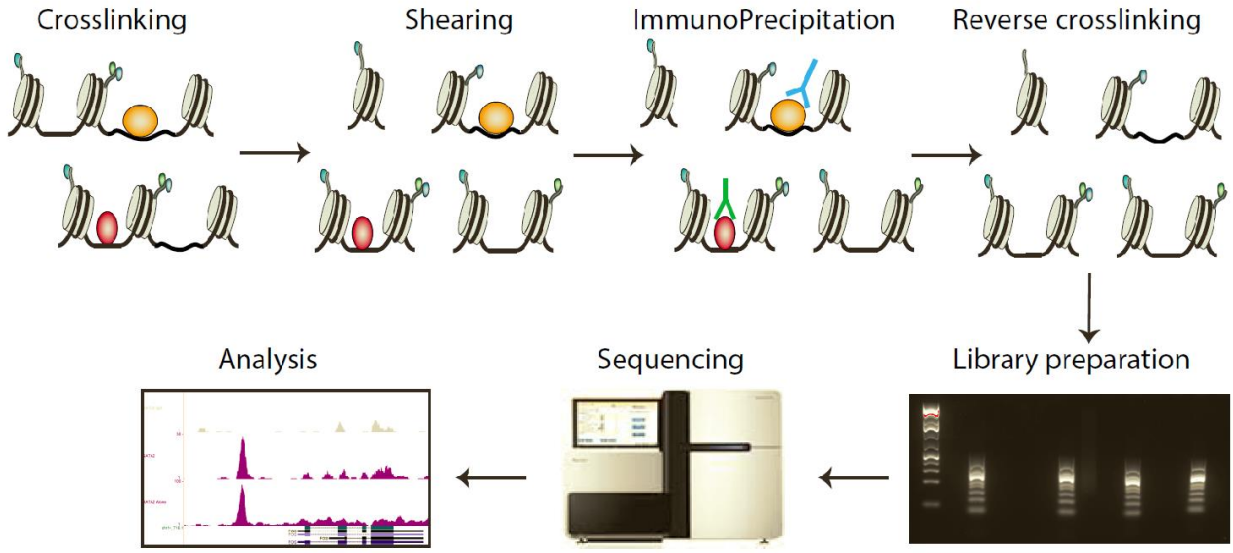



Figure 2.2 - **Overview of ChIP-seq for DNA-binding transcription factors.** The protocol starts by cross-liking DNA and proteins with formaldehyde, followed by cell lysis, sonication of chromatin to fragment DNA, immunoprecipitation with a specific antibody and reverse cross-linking is followed by purification of ChIP DNA. Library preparation comprises end repair, purification, A-tailing, adapter ligation and size selection in preparation for sequencing. Data generated is further analyzed using bioinformatic tools.

#### 2.2.15.1 Chromatin Immunoprecipitation (ChIP)-seq

ChIP assays were performed in HDFs transduced with a pool of 3xFLAG-tagged-GATA2, HA-tagged-GFI1B and FOS (Table 2.1) and the transgenes were induced with Doxycycline. A summary of the overall procedure is illustrated on Figure 2.2. After 48hr, 20-50x10^6 cells were used for each experiment and crosslinking conditions were optimized for each factor. GATA2 and GFI1B ChIP was performed as previously described (Lee et al., 2006), briefly, cells were fixed with 11% formaldehyde (Sigma) at room temperature on a rotating platform for 10 min. Formaldehyde was quenched by adding of 125 nM of glycine on a rotating platform for 5 min at room temperature and cross-linked cells were washed twice in ice-cold PBS. Chromatin shearing was done using the E210 Covaris to a 150-350bp range with the follow conditions: 7 minutes, duty cycle: 5%, Intensity 4, Cycles per burst: 200, Temperature: 4°C (Figure 2.3). Insoluble debris was centrifuged, then sheared chromatin fragments were incubated overnight at 4°C with antibodies coupled to 50 µl Protein G dynabeads (Invitrogen). For FLAG and HA 10µg of antibody per 20-50x10^6 cells. Beads were washed five times with RIPA buffer and once with TE containing 50 mM NaCl, and

complexes eluted from beads in elution buffer by heating at 65°C and shaking in a Thermomixer. Reverse cross-linking was performed overnight at 65°C. Whole cell extract DNA was treated for cross-link reversal. Immunoprecipitated and whole cell extract DNA were treated with RNaseA, proteinase K and purified using Phenol:Chloroform:Isoamyl Alcohol extraction followed by ethanol precipitation.
For FOS ChIP 3 µg of antibody was used per 5-10x10^6 cells were double crosslinked. First, cells were crosslinked in PBS supplemented with Di(N-succinimidyl) glutarate (DSG, ThermoFisher Scientific 20593) at a final concentration of 2 mM for 45 min at room temperature on a rotating platform. After 3 washes in PBS, formaldehyde crosslinking of proteins and DNA was done for 10 min at room temperature at a concentration of 11% formaldehyde (Sigma) in PBS. Formaldehyde was quenched by adding of 125 nM of glycine on a rotating platform for 5 min at room temperature and crosslinked cells were washed twice in ice-cold PBS. Further, nuclei isolation, sonication and ChIP were performed as previously described (Lee et al., 2006).

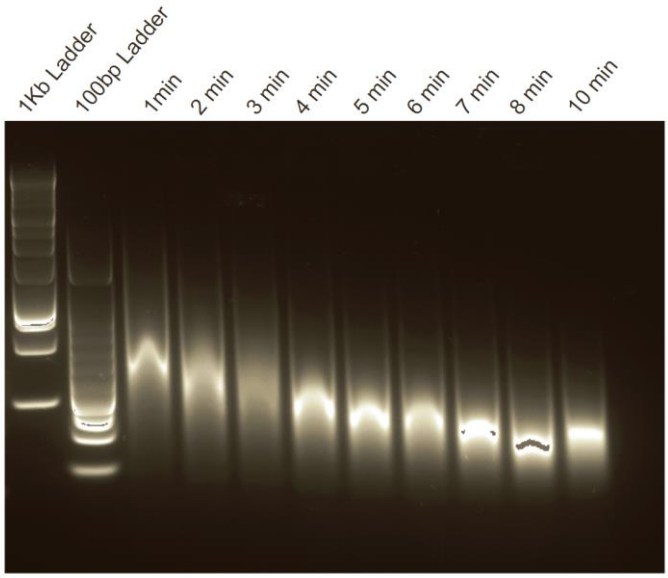



Figure 2.3 – **Optimization of sonication time for ChIP.** Chromatin was sheared to a size of 150 to 300 bp for an optimized period of 8 min using Covaris E220 Focused Ultrasonicator (Adaptive Focused Acoustics™). 500ng of DNA were analyzed on a 2% agarose gel with a 1Kb ladder and a 100bp ladder.

### 2.2.15.2 ChIP-seq library preparation and high throughput sequencing

Libraries were prepared using either KAPA Hyper Prep Kit or NEBNext ChIP-seq Library Prep Master Mix Set for Illumina according to the manufacturer's guidelines.

Libraries were size-selected on a 2% agarose gel for a 200-400bp fragments and were sequenced on Illumina HiSeq 2000.

#### 2.2.15.3 ChIP-seq analysis

ChIP-seq analysis was done on the raw FASTQ files. The FASTQ files were mapped to the human *hg19* genome using Bowtie 2 (Langmead and Salzberg, 2012) program allowing for 2 base pair mismatches. The mapped output files were processed through MACS1.4 analysis software to determine peaks. Homer software package (http://homer.salk.edu/homer/index.html) (Feng et al., 2012) was used for peak annotation and further analysis was performed using the Galaxy server https://usegalaxy.org) (Afgan et al., 2016) and R statistical software.

To produce the heatmaps, each feature (such as peaks of a transcription factor, histone marks) was aligned at GATA2 or GFI1B summits and tiled the flanking up- and downstream regions within ± 4kb in 100bp bins. To control for input in our data, we computed at each bin a input-normalized value as $\log2(RPKM_{Treat}) - \log2(RPKM_{Input})$, where $RPKM_{Treat}$ is RPKM of the corresponding TF or histone and $RPKM_{Input}$ is RPKM of the corresponding whole genome 'Input'. The density of Dnase-seq signal was plotted within ± 1kb around the center of GATA2 or GFI1B summits and compared it to the resistant sites, which were resized to be in the same range as GATA2 or GFI1B summits.

#### 2.2.15.4 ChIP-seq Data Visualization

University of California Santa Cruz (UCSC) Genome Browser (http:// genome.ucsc.edu) (Kent et al., 2002) was used for visualizing the original ChIP-seq data and the MACS output on a local computer.

#### 2.2.15.5 Chromatin State Fold-Enrichment

Using the ChromHMM Overlap Enrichment (Ernst and Kellis, 2012), the enrichment scores for genomic features was calculated, such as GATA2 and GFI1B Chip-seq peaks and histone marks based on public segmentation. ChromHMM segmentation which contains 18 different chromatin states was used and, it was downloaded from Roadmap website; The enrichment scores were calculated as the ratio between the observed and the expected overlap for each feature and chromatin state based on their sizes and the size of the human genome.

#### 2.2.15.6 Motif Analyses

For *de novo* motif discovery, the findMotifsGenome.pl procedure from Homer (Heinz et al., 2010) was used on GATA2 and GFI1B separately. The regions that are co-bound by GATA2 and GFI1B were found using bedtools (Quinlan and Hall, 2010). The following regions were afterwards used for *de novo* motif discovery using Homer and CCAT (Jiang and Singh, 2014), which allowed us to evaluate combinatorial interplay among transcription factors. In order to evaluate similarity of the two sets based on the intersections between them Jaccard statistic was used (Vorontsov et al., 2013).

### 2.2.16 Population Whole Genome transcriptional analysis

#### 2.2.16.1 GPSforGenes

Gene expression data was downloaded from BioGPS database (GeneAtlas U133A, http://biogps.org), transformed to log-space and normalized to bring the expression values to 0-1 range for each gene across different samples (Wu et al., 2016). The resulting data was then searched for samples with the highest averaged expression for GATA2 + cFOS + GFI1B and GATA2 + cFOS + GFI1B + ETV6.

#### 2.2.16.2 mRNA-seq Library Preparation and Sequencing

PS34 cKit$^+$, PS34cKit$^-$, PS34 cKit$^+$Cd45$^+$, HSPCs (CD34$^+$CD45$^+$cKit$^+$), MBC (CD34$^-$CD45$^+$cKit$^-$) mouse population cells, and day 15 CD49f$^+$, day 25 CD49f$^+$ and day 25 CD34$^+$CD49f$^+$ human population cells were FACS isolated. HDF and BJ were used as unsorted controls. Cells were lysed in Trizol (Ambion) and RNA integrity was evaluated using a Eukaryotic RNA 6000 Nano chip on an Agilent 2100 Bioanalyzer (Agilent Technologies). Up to 1 µg of total RNA from each sample was used for library preparation with the TruSeq RNA Sample Preparation Kit (Illumina) according to manufacturer's instructions. A common adapter was used for all samples and barcode sequences present in the reverse primer were introduced by 12-20 cycles of amplification as described (Pereira et al., 2013). Each library was assessed for quality and size distribution using an Agilent High Sensitivity Assay bioanalyzer chip and quantified by real-time PCR. Equimolar amounts of each barcoded library were mixed and single-end sequenced on an Illumina HiSeq Sequencing System.

### 2.2.16.3 mRNA-seq analysis

For each sample 4.5-26.5 M 100-nt reads were obtained, pre-processed with the FASTX-toolkit suite (http://hannonlab.cshl.edu/fastx_toolkit/) and aligned to the human genome (*Homo sapiens* hg19 assembly) using TopHat mapper. Post alignment with TopHat release 1.4.1 (Langmead et al., 2009; Trapnell et al., 2009, 2010) against *Mus musculus* mm9 assembly and *Homo Sapiens* hg19 assembly using the known transcripts option.

For mouse samples, transcript assembly and expression estimation were conducted with the Cufflinks package v2.1.1 and v2.2.1 (Roberts et al., 2011; Trapnell et al., 2010, 2012) for known transcripts using geometric library normalization to obtain gene FPKM report and differential expression tables. Inter-sample FPKM expression correlation was conducted by Spearman correlation and hierarchical clustering and plotted via the R v3.1.0 statistical package. Principal component analysis (PCA) was performed on the sample FPKM values scaled to unit variance and zero centered, PC1 and PC2 were displayed in Rv3.1.0. Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was performed using the gene FPKM output from Cufflinks for hemogenic precursors (E10.5_PS34cKit- and E12.5_PS34cKit-) and HSPC (E10.5_HSPCs and E12.5_HSPCs) sample classes and analysed against the Molecular Signatures Database version 2.0 (Subramanian et al., 2005) curated gene sets and pathway gene sets assembled from the Netpath-annotated immune signaling pathways (Kandasamy et al., 2010) for gene set sizes 0-5000 and ranked by Ratio_of_Classes. Bedgraph files from TopHat aligned reads were generated with Bedtools v2.20.1 and visualized on the UCSC Genome Browser (Kent et al., 2002) for examination of the mm9 assembly chr7 (7qE3) region for β-globin gene expression in $Ter119^{+}$ sorted samples.

For human samples, all resultant .bam files were processed using Samtools version 0.2.5 (Li et al., 2009) and Bedtools version 2.16.2 (Quinlan and Hall, 2010) and visualized on the Integrated Genome Browser version 2.1 (Robinson et al., 2011) or the UCSC Genome Browser. Transcript assembly and expression estimation was conducted with Cufflinks release 1.3.0 (Roberts et al., 2011; Trapnell et al., 2009, 2012) using hg19 reference annotation and upper quartile normalization. Cufflinks assemblies were merged and processed through Cuffdiff for gene FPKM reporting and differential expression analysis. Each library was treated as a separate non-replicate sample. Gene transcript count data from the mRNA-seq analysis was obtained by

reprocessing the data through TopHat release 2.0.0 and Cufflinks and Cuffdiff release 2.0.0. Gene set enrichment analysis (GSEA) between HDF or BJ and $CD34^{+}CD49f^{+}$ was performed using the genes.fpkm.tracking file output from Cufflinks release 1.3.0 run against the Molecular Signatures Database version 2.0 (Subramanian et al., 2005) curated gene sets (Gene set sizes 0-5000) ranked by Ratio_of_Classes. Non-negative Matrix Factorization (NMF) (Brunet et al., 2004) of the FPKM values obtained from RNA sequencing was performed on the GenePattern Platform (Reich et al., 2006) using the NMF consensus analysis module at *k. initial* = 2 and *k. final* = 5, and metagene results for *k* = 4. Visualization of FPKM expression density and inter-sample FPKM correlation was conducted in R version 2.15.0 with the CummeRbund package (Trapnell et al., 2012). Gene list enrichment analysis with gene set libraries created from level 4 of the MGI mouse phenotype ontology was performed with Enrichr (http://amp.pharm.mssm.edu/Enrichr/) (Kuleshov et al., 2016).

### 2.2.17 Single-cell mRNA-seq using C1 Auto-Prep System by Fluidigm

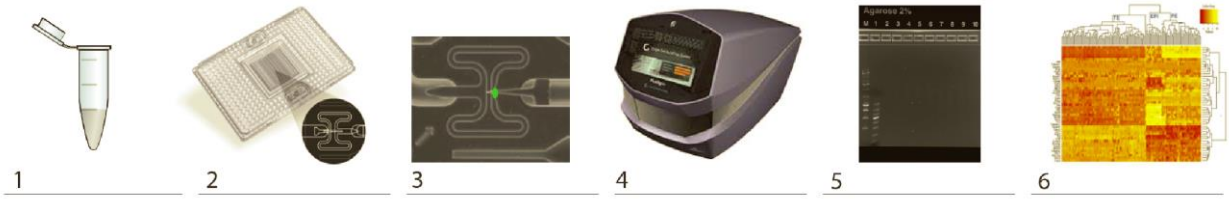



Figure 2.4 - **The Single Cell mRNA-seq workflow using the microfluidic system by Fluidigm**. C1 Single Cell Auto Prep System automates single cell isolation (2), washing and live-dead staining (3), lysis, reverse transcription and PCR amplification (4) to provide high quality cDNA for further analysis. The harvested cDNA is subjected to library preparation (5) and high throughput sequencing and analysis (6). 148 mouse and 286 human single cells were sequenced.

#### 2.2.17.1 Single cell suspension preparation

PS34CD45- cells at E10.5 and E12.5, and HSPCs ($CD45^{+}CD34^{+}cKit^{+}$) were FACS sorted and collected. Human fibroblasts and reprogrammed cells were harvested by TrypLE Express (Thermo Fisher Scientific) and day 15 ($CD34^{-}CD49f^{+}$) and day 25 ($CD34^{+}CD49f^{+}$) were sorted and collected. Umbilical cord blood cells were obtained from the New York Blood Center and Lin-$CD34^{+}$ cells were isolated using Diamond CD34 Isolation Kit (Miltenyi Biotech) according to manufacturer's instructions. After isolation, Lin-$CD34^{+}$ cells were stored in liquid nitrogen until use (Figure 2.4).

### 2.2.17.2 Single cell cDNA synthesis and amplification

For each condition 500-1000 cells/µl were mixed with C1 Cell Suspention Reagent (3:2 ratio) prior loading on to a 10-17µm C1 Single Cell mRNAseq IFC chip (Fluidigm) (Figure 2.4-2), and cell capture was performed on a C1 Auto Prep System according to the manufacturer's instructions. The capture efficiency was determined using a microscope to exclude samples with no cell or more than one cell captured or to exclude samples where cellular debris was visible (Figure 2.4-3). To determine the viability of the captured cells, LIVE/DEAD kit (Life Technologies) was prepared by adding 2.5µl ethidium homodimer-1 and 0.625 µl calcein AM to 1.25 ml C1 Cell Wash Buffer and 20 µl were loaded onto the C1 chip (Figure 2.4-3). Upon capture, single cells were lysed using lysis buffer and, reverse transcription and cDNA preamplification were performed in the chip using the SMARTer Ultra Low RNA Kit for Illumina Sequencing (Clontech). During reverse transcription, adaptors are incorporated within the primers allowing amplification of full-length transcript by PCR. Full length amplified cDNA was harvested and stored at -20°C.

### 2.2.17.3 Single cell cDNA Quantification PicoGreen

The cDNA reaction products were quantified using the Quant-iT PicoGreen double-stranded DNA Assay Kit (Life Technologies) and read on fluorescence MicroPlate Reader with an excitation ~480 nm and emission ~520 nm. The samples were then diluted to a final concentration of 0.15–0.30 ngµl$^{-1}$ using C1 Harvest Reagent.

### 2.2.17.4 Single cell mRNA-seq Library Preparation

Sequencing libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina) modified for single cell mRNA sequencing. Samples were simultaneously tagged with adapters and fragmented enzymatically using a transposase. Tagmented cDNA was amplified using the dual index primers (Nextera XT DNA Index kit 96 indices) and 48 or 96 single-cell libraries were pooled together. Clean up reactions were done with Ampure XP beads (Beckman Coulter Genomics). Libraries were visualized on a 2% Agarose Gel (E-Gel, Thermo Fisher Scientific) (Figure 2.4-5). The concentration of the pooled libraries was determined using Agilent Bioanalyzer using the High Sensitivity DNA analysis kit.

### 2.2.17.5 Single cell mRNA-seq Analysis

For single cell mRNA-Seq analysis the raw fastq files were aligned against the Ensemble GRCh38 genome using the Gencode v25 gene annotation. Gene level read counts were then calculated using featureCounts (Liao et al., 2014) from the Subread package. Raw counts were log transformed and quantile normalized. For hierarchical clustering Pearson correlation was used as distance metric. Hierarchical clustering was plotted using the dendextend library in R (Galili, 2015) (Figure 2.4-6).

### 2.2.17.6 Gene List Enrichment Analysis with Single Cell Data

Differential gene expression was calculated for all pairwise sample groups (HDF, Day 2, $CD34^+$ $CD49f^+$, $CD49f^+$, UCB $CD34^+$) using variance-stabilizing transformation combined with limma (Voom) (Ritchie et al., 2015). To calculate gene set enrichment the top 500 and bottom 500 genes for each differential expression gene signature were uploaded to Enrichr (Kuleshov et al., 2016). The significance of gene set overlap was measured using the Fisher Exact test.

## 2.2.18 Integration of Independently obtained Gene Expression and Genome Location Datasets

Data from microarray and RNA-seq experiments were adjusted using quantile normalization; genes insignificantly changing their expression across samples were removed using ANOVA with the consequent adjustment for multiple testing (Benjamini-Hochberg correction, $p < 0.05$). Data for the remaining genes was converted to Z-scores. Hierarchical clustering and PCA of the integrated dataset was performed using Cluster 3.0 and visualized with JAVA treeview. For differential gene expression analysis, DESeq2 (Love et al., 2014) was used in HDF, day 15 $CD49f^+$, day 25 $CD49f^+$ and day 25 $CD34^+CD49f^+$ and genes were selected using the following criteria: 1) DESeq2 differential calls with an adjusted p-value < 0.05 (FDR used as adjusted p-value) between either of four groups of sample; 2) Absolute changes in expression between minimal and maximal expression > 1.5 fold. Intersection between the genes identified by ChipSeq for TFs: GATA2 and GFI1B (genes were identified as targets based on binding peaks within ± 5kb around the transcriptional start site of genes) and the selected set of genes emerged from mRNA-seq data identified 1,425

genes, which were clustered and visualized. Those 1,425 were divided into 3 groups: 1) Genes bound only by GATA2; 2) Genes bound only by GFI1B; 3) Genes co-bound by GATA2 and GFI1B.

### 2.2.19 Accession Numbers

Population and single cell mRNA-seq from mouse placental precursors were deposited in the GEO database with the accession number GSE54574. Population, single cell mRNA-seq and ChIP-seq data from human hemogenic reprogramming were deposited in the GEO database with the accession number GSE51025. The public datasets used in this study can be found in Table 2.4.

# Chapter 3 – *In vitro* hemogenic reprogramming guides *in vivo* identification of hemogenic endothelial precursors in placenta

This chapter of the dissertation was published in the following research article:

Hematopoietic Reprogramming *in vitro* informs *in vivo* identification of hemogenic precursors to definitive hematopoietic stem cells

Carlos-Filipe Pereira, Betty Chang, Andreia Gomes, Jeffrey Bernitz, Dmitri Papatsenko, Xiaohong Niu, Gemma Swiers, Emanuele Azzoni, Marella F.R.T. de Bruijn, Christoph Schaniel, Ihor R. Lemischka, and Kateri A. Moore - Developmental Cell, 2016

The use of cell tracing methodologies allowed the characterization of EHT, in which HSC are generated from the specialized hemogenic endothelium cells (Boisset et al., 2010; Eilken et al., 2009; Kissa and Herbomel, 2010; Lancrin et al., 2009). However, the phenotypic identification of these cells during embryonic development is currently not fully established due mostly to the lack of exclusive markers. In addition, the functional definition of earlier HSC precursor cells is hampered by the difficulty in isolating them to purity (Boisset et al., 2015).

The development of iPSCs and directly reprogrammed cells caused a paradigm shift in the way we think about cell fate stability in multipotent and unipotent somatic cells. This led to the identification of minimal TF networks that kick-start lineage reprogramming (Pereira et al., 2012), and provided mechanistic insights into TF modes of action (Soufi et al., 2012; Wapinski et al., 2013). However, it is not known if direct lineage reprogramming recapitulates developmental lineage specification (Pereira et al., 2013), or if it leads to aberrant cell types without a clear *in vivo* equivalent (Doulatov et al., 2013).

Our direct reprogramming seems to recapitulate developmental hematopoiesis and traverses through an HP cell with a specific phenotype. Therefore, by using an *in vitro* model that resembles embryonic hematopoietic environment we asked if this information could provide insights into HSC ontogeny *in vivo*. To access the presence of a specific phenotype *in vivo*, previously identified as hemogenic precursor cells *in vitro*, we used mouse placentas at different gestation periods.

The chapter is sectioned into three main topics: a) identification of the hemogenic phenotype in mouse placenta, b) transcriptional characterization of hemogenic cells from placenta and c) engraftment capacity of phenotypic hemogenic cells. This work was performed in collaboration with Filipe Pereira and Betty Chang.

## 3.1 **Identification of the hemogenic phenotype in mouse placenta.**

Since Sca1⁺Prom1⁺CD34⁺ was identified as hemogenic precursor phenotype during in vitro reprogramming (Pereira et al., 2013), we first aimed to investigate the presence of similar phenotype in mouse placentas as shown in Figure 3.1. For that purpose, different gestation stages were selected corresponding to the onset of HSC activity and its pool expansion in the placenta (Gekas et al., 2005; Ottersbach and Dzierzak, 2005).

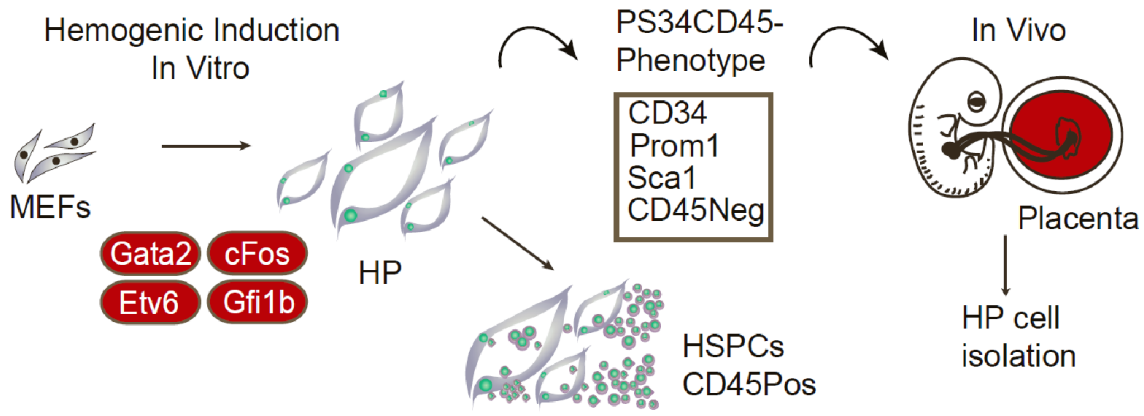



Figure 3.1 - **Strategy used for HP isolation *in vivo* using the phenotype described from *in vitro* hemogenic induction studies**. Gata2, cFos, Gfi1b and Etv6 efficiently induced an hemogenic precursor expressing Prominin 1, Sca1, CD34 (termed PS34) and negative for CD45 on MEFs. This phenotype is further investigated *in vivo*. Bona fide HSPCs are used as controls.

Mouse placentas were collected at E10.5, E11.5 and E12.5 and were detached from the umbilical cord and any maternal tissue (Figure 3.2a). Upon dissociation, single cells were analyzed by flow cytometry for the expression of Prom1, Sca1, CD34 and CD45. Among Sca1⁺ cells, up to 8.8% were also Prom1⁺ (8.8% for E10.5 and 1.9% for E12.5). Prom1⁺Sca1⁺ cells co-expressed CD34 and most of those did not express the pan-hematopoietic marker CD45 (80.7% for E10.5 and 67.8% for E12.5) (Figure 3.2b). This represents the same cell surface phenotype previously identified *in vitro* by introduction of TFs into fibroblasts (Pereira et al., 2013). Furthermore, the population with the Prom1⁺Sca1⁺CD34⁺ (PS34) phenotype is more prevalent at 10.5 and declines at E11.5 and E12.5 (5.5% ± 2.4%, 1.8% ± 0.5%, and 1.3% ± 0.6%, respectively; Figure 3.2c). We found that the subpopulation CD45- PS34 decreases with gestation time (80.6% ± 7.0%, 71.1% ± 6.2%, and 69.7% ± 14.2%, respectively; Figure 3.2c). These

results suggest that a population expressing this phenotype can be identified *in vivo* guided by the previous *in vitro* study through enforced expression of specific TFs into somatic cells.

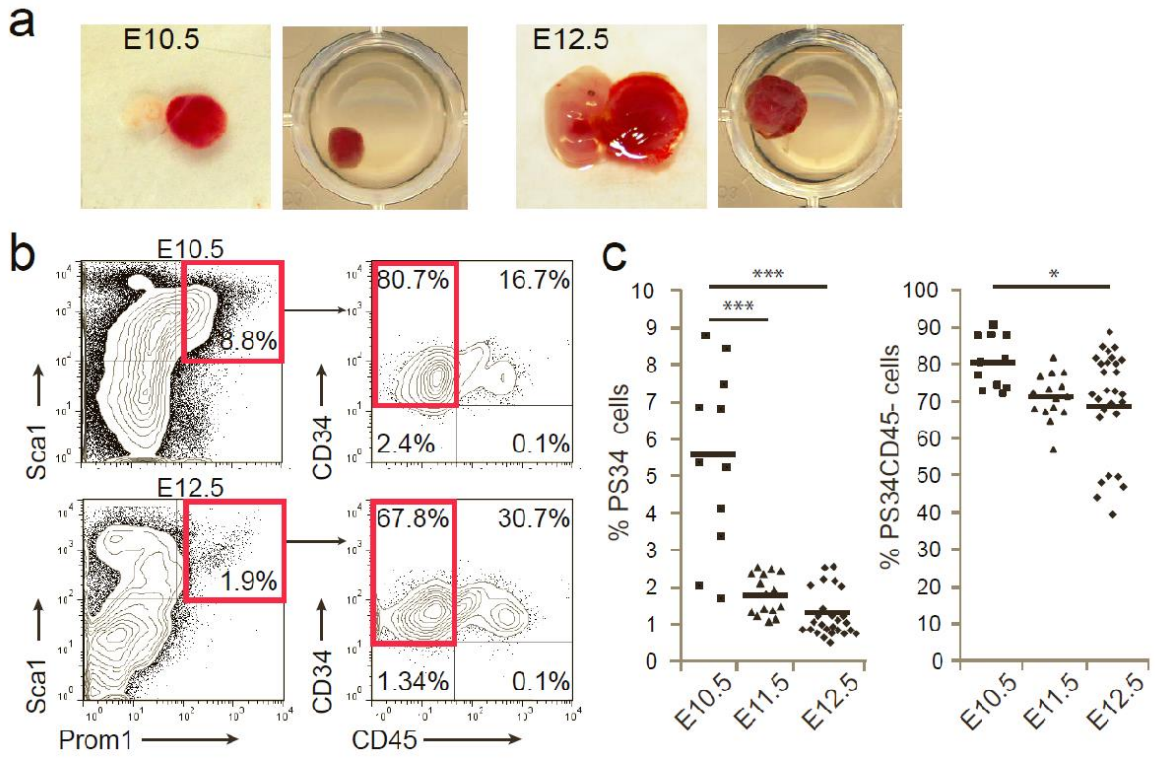



Figure 3.2 - **Isolation and phenotypic characterization of HP cells from mouse placentas**. (**a**) E10.5 and E12.5 were isolated and dissociated to a single-cell suspension. (**b**) Sorting strategy for isolation of $Sca^{+}Prom1^{+}CD34^{+}CD45^{-}$ population for E10.5 and E12.5 highlighted in red boxes. (**c**) Expression of Sca1, Prom1 and CD34 cells from litters at E10.5, E11.5 and E12.5.

We sought to determine whether PS34CD45- cells retain hematopoietic potential. PS34CD45- cells were sorted and cultured on gelatin-coated dishes supplemented with the high expansion cytokine combination SCF, IL-3 and Flt3L (Figure 3.3a). Upon 7 days on culture, we observed the emergence of $CD45^{+}$ cells in association of with large adherent flat cells (Figure 3.3a and b). To mimic the interaction between cells from the hematopoietic niche and HSCs, PS34CD45- cells were co-cultured with a stromal cell layer capable of HSCs maintenance. When plated on AFT024 stroma (Moore et al., 1997), $PS34CD45^{-}$ cells generated $CD45^{+}$ cobblestone-like colonies after 4 weeks as shown in Figure 3.3c. Taken together, these results demonstrate that direct reprogramming can be used as tool that facilitate the interrogation of

developmental processes and suggest that the PS34CD45$^{-}$ phenotype identified *in vitro* points to the isolation of hematopoietic progenitors *in vivo*.

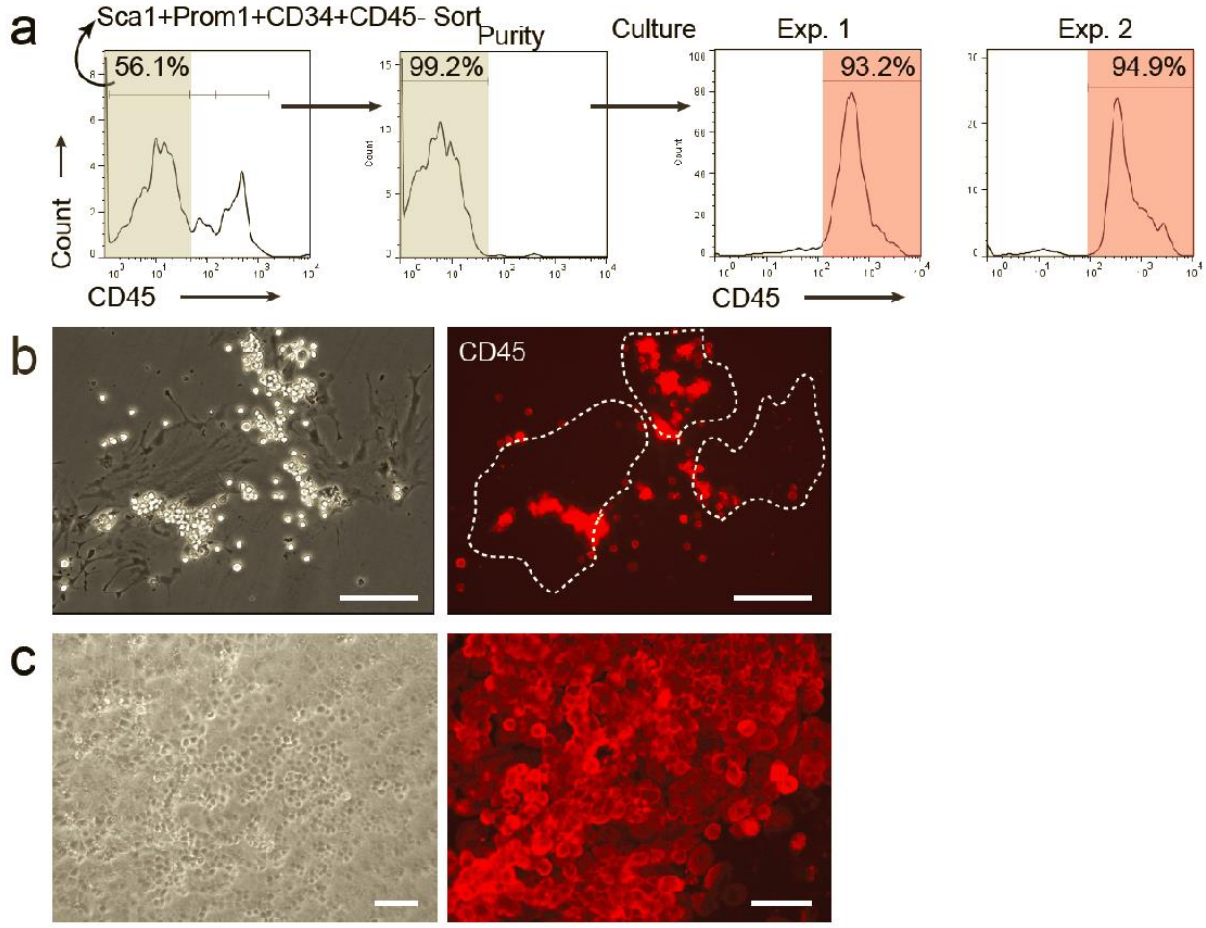



Figure 3.3 - **Hematopoietic competence of the PS34CD45- cells.** (**a**) PS34CD45$^{-}$ cells were sorted at E12.5 and plated on gelatin-coated plates in MyeloCult supplemented with SCF, IL-3 and Flt3L for 7 days, stained for CD45 and analyzed by flow cytometry or (**b**) immunofluorescence. Dashed lines highlight large adherent cells associated with round non-adherent CD45$^{+}$ cells. Scale bar, 100µm. (**c**) PS34CD45$^{-}$ cells were cultured on AFT024 stroma for 4 weeks and stained for CD45.

## 3.2 **Transcriptional characterization of hemogenic cell from placenta.**

To investigate the transcriptional program that underlies placental hemogenic cells, we performed mRNA-seq on PS34 populations at E10.5 and E12.5. Two or three biological replicates were sorted from PS34 subsets including the cKit$^{-}$CD45$^{-}$, cKit$^{+}$CD45$^{-}$, and cKit$^{+}$CD45$^{+}$ populations. The population of HPs is heterogenous in the expression of cKit consequently the PS34CD45$^{-}$ cells was separated by cKit

expression. In addition, placental HSPCs ($CD45^+cKit^+CD34^+$) and mature blood cells (MBC $CD45^+cKit^-CD34^-$) were profiled as controls (Figure 3.4a). Spearman rank correlation was used to evaluate the association between populations and between replicates. The dataset showed that biological replicates correlate with each other and, the hierarchical clustering separates MBCs from the HSPCs and Precursors populations, as expected, with MBCs displaying the most distant gene expression profile (Figure 3.4b). Moreover, $CD45^-$ ($cKit^-$ and $cKit^+$) and $CD45^+$ cells ($PS34CD45^+$ and HSPCS) generate two very distinct clusters. $PS34CD45^-$ cells cluster by phenotype ($cKit^-$ and $cKit^+$) but not by gestation time, suggesting that the PS34 resembles HPS independently of gestation time. To determine the background biological variability, principal component analysis (PCA) was performed for E10.5 and E12.5. The analysis showed that all replicates from E10.5 and E12.5 cluster closely, except for PS34 $cKit^-$ at E10.5. Furthermore, PS34 $cKit^-$ and PS34 $cKit^+$ are distant by principal component 2 in both gestation periods, and $PS34CD45^+$ and HSPCs are further apart from $CD45^-$ HPs, showing an order in the clusters from $cKit^-$ to $cKit^+$ to $CD45^+$ (Figure 3.4c). The putative trajectory is consistent with the phenotypic changes during EHT, which suggest that the developmental progression is kept within the PS34 compartment.

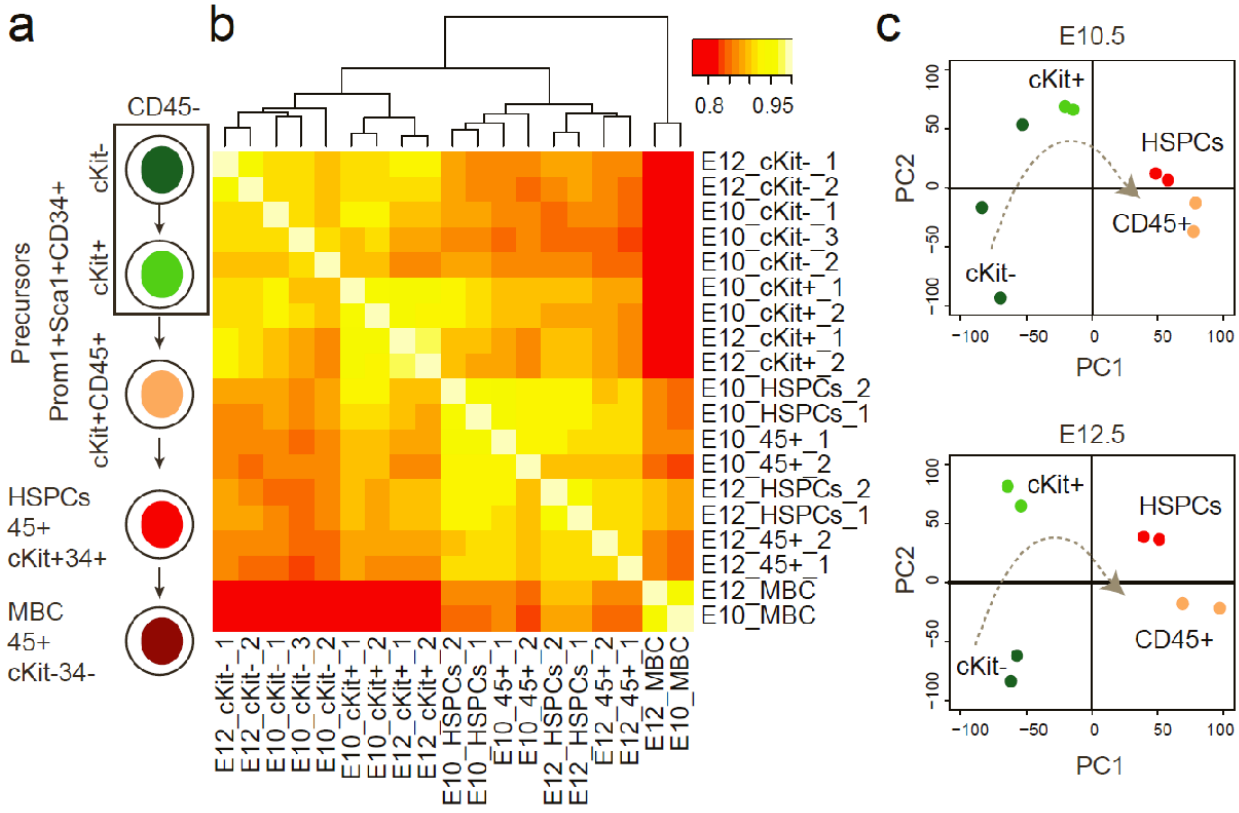



Figure 3.4 - **Global gene expression analysis of placental hemogenic cells.** (**a**) Hemogenic PS34 cells from E10.5 and E12.5 placentas were fractioned into CD45$^-$cKit$^-$, CD45$^-$cKit$^+$, and CD45$^+$cKit$^+$ populations. HSPCs and mature blood cells (MBC) were also sorted for comparison. (**b**) Global gene expression levels were profiled by mRNA-seq (biological replicates: 1, 2, and 3). Spearman rank correlation heatmap and hierarchical clustering dendogram of the expression profiles are displayed. (**c**) PCA shows the relative distances between samples and a hypothetical temporal trajectory at E10.5 (upper panel) and 12.5 (lower panel).

Transcription factors intrinsically regulate gene expression and can modulate the cell fate. Gata2, cFos, Gfi1b and Etv6 integrate the TF core used for the conversion fibroblasts into hemogenic cells (Pereira et al., 2013). Expression profiles of the indicated genes were determined by alignment of reads at individual gene loci and, genes were quantified by fragments kilobase of exon per million fragments mapped (FPKM) values. During EHT, *Gata2* and *cFos* were highly expressed in placental PS34 (Figure 3.5). *Etv6* was expressed in all populations and *Gfi1b* only showed expression at E12.5 in HSPCs which is consistent with the role of *Gfi1b* at later stages of EHT in repressing endothelial genes (Lancrin et al., 2012).

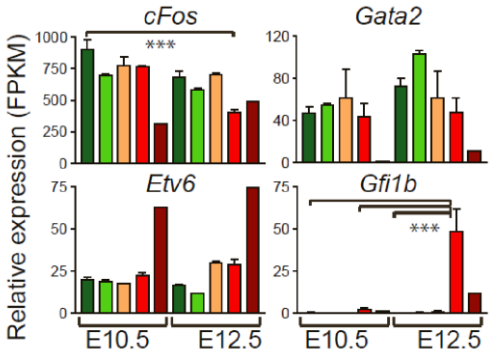



Figure 3.5 - **Relative expression of TFs at E10.5 and E12.5.** Expression levels of programming TFs cFos, Gata2, Etv6 and Gfi1b shown as FPKM mean values ± SD. Bars colors are coordinated with populations shown in Figure 3.4 (**a**). *p<0.05, **p<0.01, *** p<0.001.

To understand the molecular program involved in EHT, we looked at the expression profile of key genes such as *Sox17* that has been identified in HE cells from mice and human cells (Clarke et al., 2013; Nakajima-Takagi et al., 2013; Nobuhisa et al., 2014). Remarkably, *Sox17* is expressed at the highest levels in PS34CD45- cells with a similar pattern of Prom1 expression. *Sox17* continues to be expressed at lower levels in HSPCs while Prom1 expression decreases more markedly, suggesting that this gene product more specifically labels HPs and possibly excludes downstream HSPCs. As expected from their roles in the specification of definitive hematopoiesis, *Scl*/*Tal1* and *Runx1* are also expressed in PS34 cell populations (Figure 3.6a). The expression of other reported markers of hemogenic cells such as *Podxl* (Hara et al., 1999) and *Kitl* (Pereira et al., 2013) is also enriched in PS34CD45$^{-}$ hemogenic cells (Figure 3.6a). During EHT, other transcriptional regulators associated with HSCs are expressed at distinct stages (Figure 3.6b), some of the TFs are expressed early (e.g. *cFos*, *Hoxb4*, *Erg*, *Lmo2*, *Id1*, and *Msi2*) and decrease upon maturation whilst TFs such as Gata2, Tal1, Sox4 are continuously expressed from HP cells to HSPCs. In addition, some TFs are only expressed at later stages (e.g. *Gfi1b*, *Myb*, *Cebpα*, *Pu.1*, *Nfe2*, and *Ikzf1*) (Figure 3.6b). Taken together, these results show transcriptional changes of specific EHT markers.

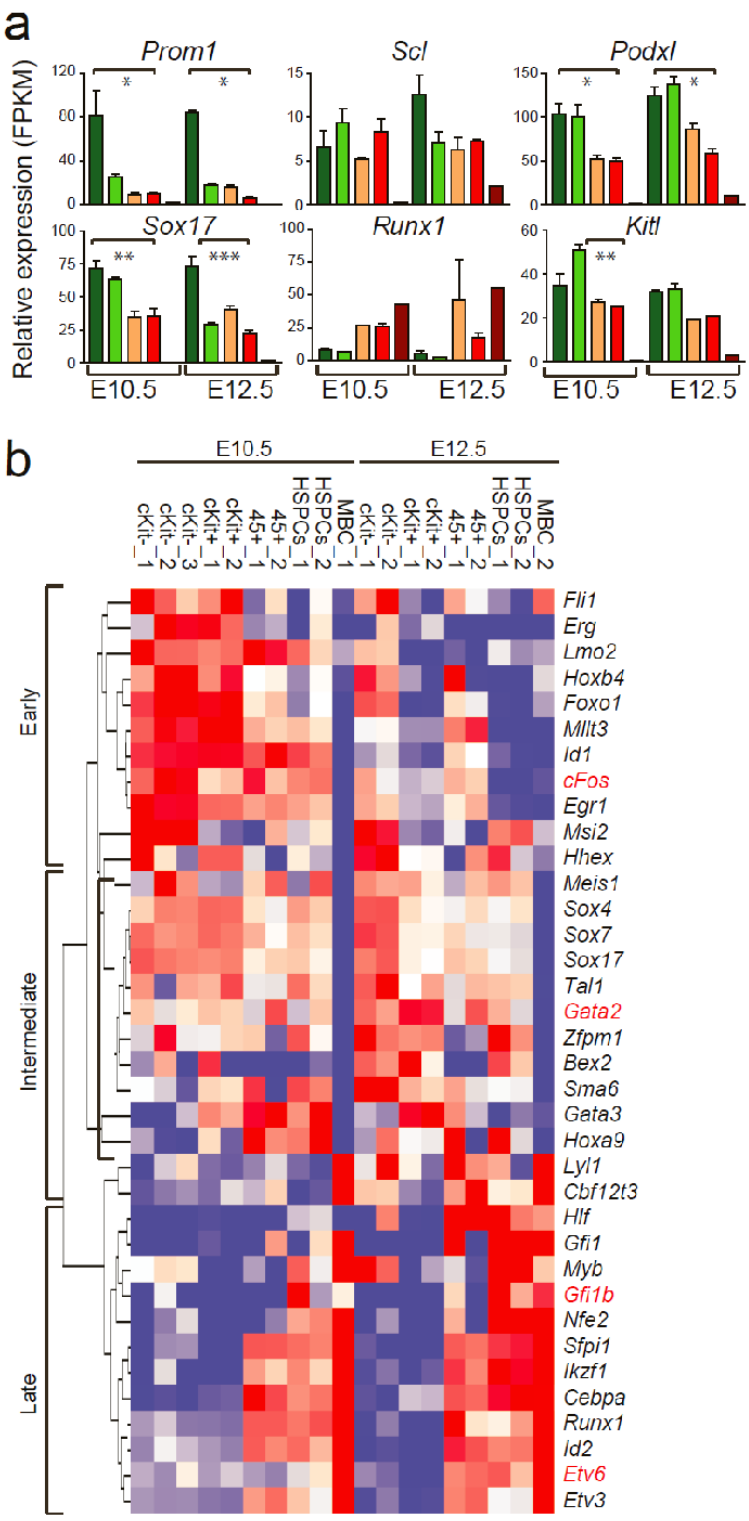


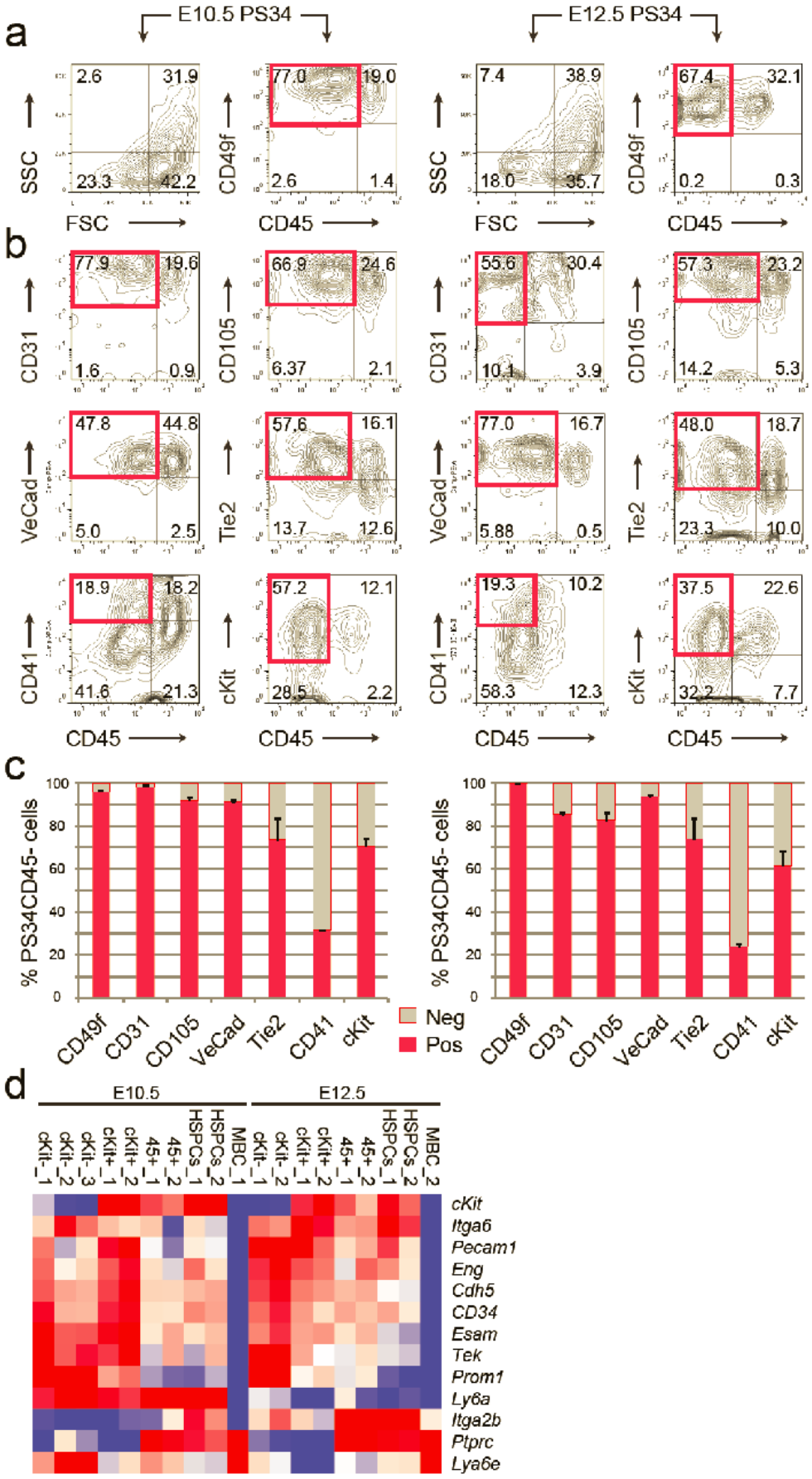

Figure 3.6 - **Relative expression of EHT transcription factors at E10.5 and E12.5.** (**a**) EHT-associated genes are shown as FPKM mean values ± SD. Bars colors are coordinated with populations shown in figure 3.4 (**a**). *p<0.05, **p<0.01, *** p<0.001. (**b**) Heatmap showing the expression of transcription factors associated with HSC biology. Unsupervised hierarchical clustering was applied. Transcriptions factors are grouped according to the stage of endothelial-to-hematopoietic transition. Early, intermediate and late groups are highlighted. Gene expression data were analyzed by Cluster 3.0 and displayed by Treeview. Red indicates increased expression, whereas blue indicates decreased expression over the mean. The hemogenic programming transcription factors Gata2, cFos, Gfi1b and Etv6 are highlighted in red.

## 3.3 **PS34 cell express Endothelial and Hematopoietic markers**

As HSCs and endothelial cells share an ontological common precursor via EHT during embryogenesis, we addressed whether PS34 cells express endothelial and hematopoietic markers associated with HE cells. We analyzed the expression of specific markers such as cKit, Integrin-a6 (CD49f), Pecam1 (CD31), Endoglin (CD105), Cdh5 (VE-cadherin), Tek (Tie2), Integrin-a2b (CD41), Lya6a/e (Sca1) and Ptprc (CD45) within the PS34 population from E10.5 and E12.5 placentas (Figure 3.7). At E10.5 and E12.5 most of the PS34CD45-cells express the endothelial markers Ve-cadherin, Cd31, CD105, and Tie2 (Figure 3.7b and c). PS34CD45$^-$ also express CD49f (Figure 3.7a and c), a marker previously identified on vitro reprogramming hemogenic cells (Pereira et al., 2013). PS34CD45$^+$ cells express VE-cadherin, CD31, Cd105, Tie2 and CD49f consistent with the retention of endothelial markers by emergent HSPCs (Taoudi et al., 2008). Interestingly, PS34CD45- cells are heterogeneous for the early hematopoietic markers CD41 (31.5% ± 0.3% at E10.5 and 24.1% ± 1.1% at E12.5) and cKit (70.9% ± 3.6% at E10.5 and 61.4% ± 6.9% at E12.5), showing that the PS34 marker combination captures a heterogeneous population of cells undergoing EHT (Figure 3.7b and c). This is consistent with the reported placental expression of CD41 (Rhodes et al., 2008) and the requirement for cKit signaling during HE cell specification (Marcelo et al., 2013). In addition, we also confirmed the expression of endothelial and hematopoietic markers on PS34 cells at the transcriptional level (Figure 3.7d).

*(figure on the next page)*

Figure 3.7 - **PS34CD45- cells express endothelial and hematopoietic markers.** (**a** and **b**) E10.5 (left panels) and E12.5(right panels) placentas were dissociated to a single-cell suspension and analyzed. Prom1$^+$Sca1$^+$CD34$^+$ cells are displayed for (**a**) forward-scattered light (FCS)/side-scattered light (SSC), CD49f and (**b**) CD31, CD105, VE-cadherin, Tie2, CD41, and cKit expression. Red boxes highlight the phenotype Prom1$^+$Sca1$^+$CD45$^-$ cells. (**c**) Quantification of the percentage of PS34CD45$^-$ cells positive for CD49f, CD31, CD105, VE-cadherin, Tie2, CD41, and cKit (mean ± SD, n = 2). (**d**) Global gene-expression levels were profiled by mRNA-seq (biological replicates: 1, 2 and 3). Heatmap shows the expression of endothelial and hematopoietic cell surface markers. Red indicates increased expression, whereas blue indicates decreased expression over the mean.


a
E10.5 PS34
E12.5 PS34
SSC
FSC
CD49f
CD45
b
CD31
CD105
VeCad
Tie2
CD41
cKit
c
% PS34CD45- cells
Neg
Pos
CD49f
CD31
CD105
VeCad
Tie2
CD41
cKit
d
E10.5
E12.5
cKit_1
cKit_2
cKit_3
cKit+_1
cKit+_2
45+_1
45+_2
HSPCs_1
HSPCs_2
MBC_1
MBC_2
cKit
Itga6
Pecam1
Eng
Cdh5
CD34
Esam
Tek
Prom1
Ly6a
Itga2b
Ptprc
Ly6e

In addition, HE cells activate an arterial identity program characteristic of sites such as the roof of the dorsal aorta in the embryo (Wilkinson et al., 2009). We asked whether the expression of the arterial genes *Foxc1*, *Jag1*, *Kdr*, *Hey2*, *Foxc2*, *Notch1*, *Notch4*, *Dll4*, and *Efnb2* was activated in PS34 cells. The expression of this gene set was mostly enriched in PS34CD45$^{-}$ cells (Figure 3.8a). We used gene set enrichment analysis (GSEA) to compare the transition from PS34CD45$^{-}$cKit$^{-}$ precursors to HSPCs with published gene sets (Figure 3.8b and c). GSEA showed significant enrichment for Long Term (LT) and Short Term (ST) HSC gene sets in PS34CD45$^{-}$cKit$^{-}$ cells (Figure 3.8b, 3 out of 3 ST-HSC and 3 out of 3 LT-HSC gene sets; false discovery rate <0.25). This result suggests that many HSC signature genes are already expressed in HPs. Other enriched and downregulated gene sets included Hedgehog signaling, Tight Junction, Blood Clotting cascade, and epithelial mesenchymal transition (EMT) (EMT_DN). Conversely, in HSPCs, gene sets for hematopoietic progenitors and genes that are upregulated during EMT (EMT_UP) are enriched. We next determined whether immune signaling pathways were enriched in the PS34CD45$^{-}$cKit$^{-}$ precursors and HSPCs. Curated immune signaling pathways (21 out of 32) were enriched in HSPCs including cKit, interleukin 3 (IL-3), and B and T cell receptor signaling, in agreement with hematopoietic specification (Figure 3.8c). In PS34CD45$^{-}$cKit$^{-}$ precursors, only 9 out of 32 immune pathways were enriched as expected in endothelial-like precursor cells. The enriched signaling pathways include α6β4 Integrin, Egfr1, Id, Hedgehog, Wnt, and Notch, consistent with their roles in hemogenesis and the pathways previously identified in HP cells programmed *in vitro* (Pereira et al., 2013).

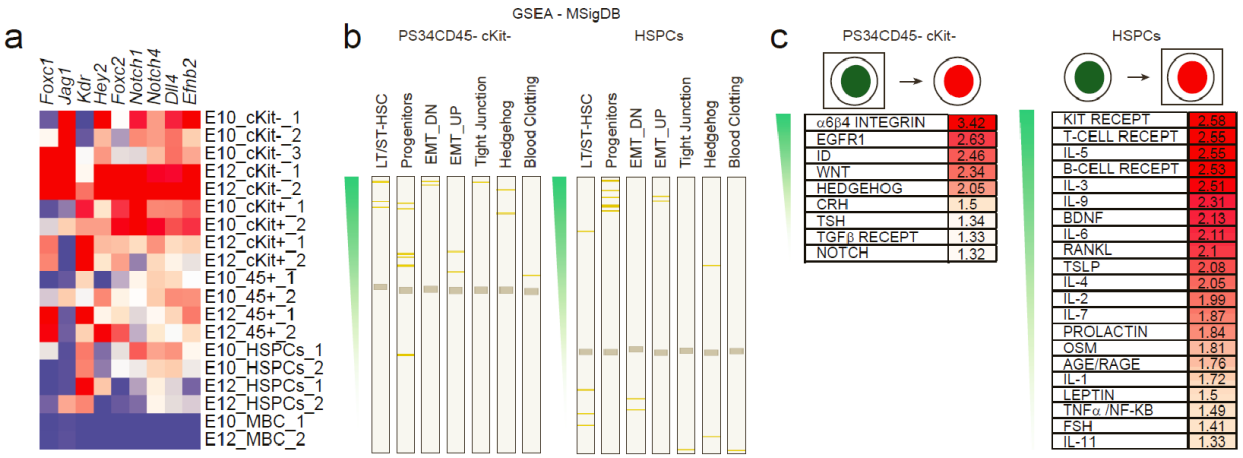



Figure 3.8 - **Gene expression and enrichment analysis of PS34CD45$^-$ cells**. (**a**) Heatmap showing expression of arterial markers. Red indicates increased expression, whereas blue indicates decreased expression over the mean. (**b**) Gene set enrichment analysis (GSEA) for hemogenic PS34CD45$^-$cKit$^-$ cells and HSPCs. Gene-expression lists were analyzed for enrichment of genes sets present in the Molecular Signature Database (MSigDB; 1,888 gene sets, gene size 0-5,000). Gene sets are ordered according to the normalized enrichment score (NES) and the grey horizontal bar shows the enrichment cut-off (FDR=0.25). Orange lines highlight the labelled gene sets. (**c**) GSEA for PS34CD45$^-$cKit$^-$ cells (right panel) and HSPCs (left panel) of NetPath-annotated signaling pathways. Pathways are ordered according to the normalized enrichment score (NES) and only enriched pathways are shown (false discovery rate <0.3, NES >1.3).

### 3.4 Integration of induced and placental hemogenic progenitor cells establish a signature

Due to the remarkable similarities between placental HP cells and *in vitro* programmed cells, we compared gene expression datasets to generate a robust signature for HPs. We reasoned that the overlap between programmed cells and placental derived precursors may better illustrate the genes and pathways required for HP cell function (Figure 3.9). Interestingly, PS34CD45$^-$ cells, in particular PS34CD45$^-$cKit$^-$ cells at E10.5 and E12.5 showed a robust clustering with programmed cells isolated at day 20 (Figure 3.9a). In contrast, programmed CD45$^+$cKit$^+$ and CD45$^+$cKit$^-$ cells show similarities with HSPCs and PS34CD45$^+$ cells from the placenta especially at E12.5 (Figure 3.9a). We used K-means clustering of the integrated gene list to extract clusters of genes representative of HPs and HSPCs. Ten clusters were identified using this analysis: six overrepresented in HPs and four enriched in HSPCs (Figure 3.9b). This analysis revealed the dynamic nature of hemogenic cells as they progress from cKit$^-$ to cKit$^+$ (Figure 3.9b). We performed pathway analysis using the PANTHER classification system. Interestingly, HPs are enriched for integrin signaling,

cytoskeletal regulation by Rho GTPases, Wnt signaling, angiogenesis, Vegf signaling, and others (Figure 3.9c). In contrast, HSPCs are enriched for pathways involved in immune cell function such as inflammation, Toll receptor signaling, and B and T cell activation (Figure 3.9c). We also used co-expression clustering to confirm gene allocation to HE or HSPCs clusters (Figure 3.9d). Co-expression clustering generated two distinct gene networks, which overlapped with the HP and HSPCs clusters generated using K-means clustering. As previously observed, the HP cluster was more heterogeneous than the HSPC cluster (Figure 3.9d). Genes identified in the HP cluster include *F11r*, recently shown to be expressed in HSC precursors and required for HSC generation in zebrafish through Notch signaling (Kobayashi et al., 2014). Other identified genes include the polarity regulators *Pard3*, *Par6b*, *Amotl1*, *Prkcz*, and *Amotl2*; Notch signaling mediator *Hes1*; semaphorins *Sema3c* and *Sema7a*; and aldehyde dehydrogenase *Aldh1a1*.

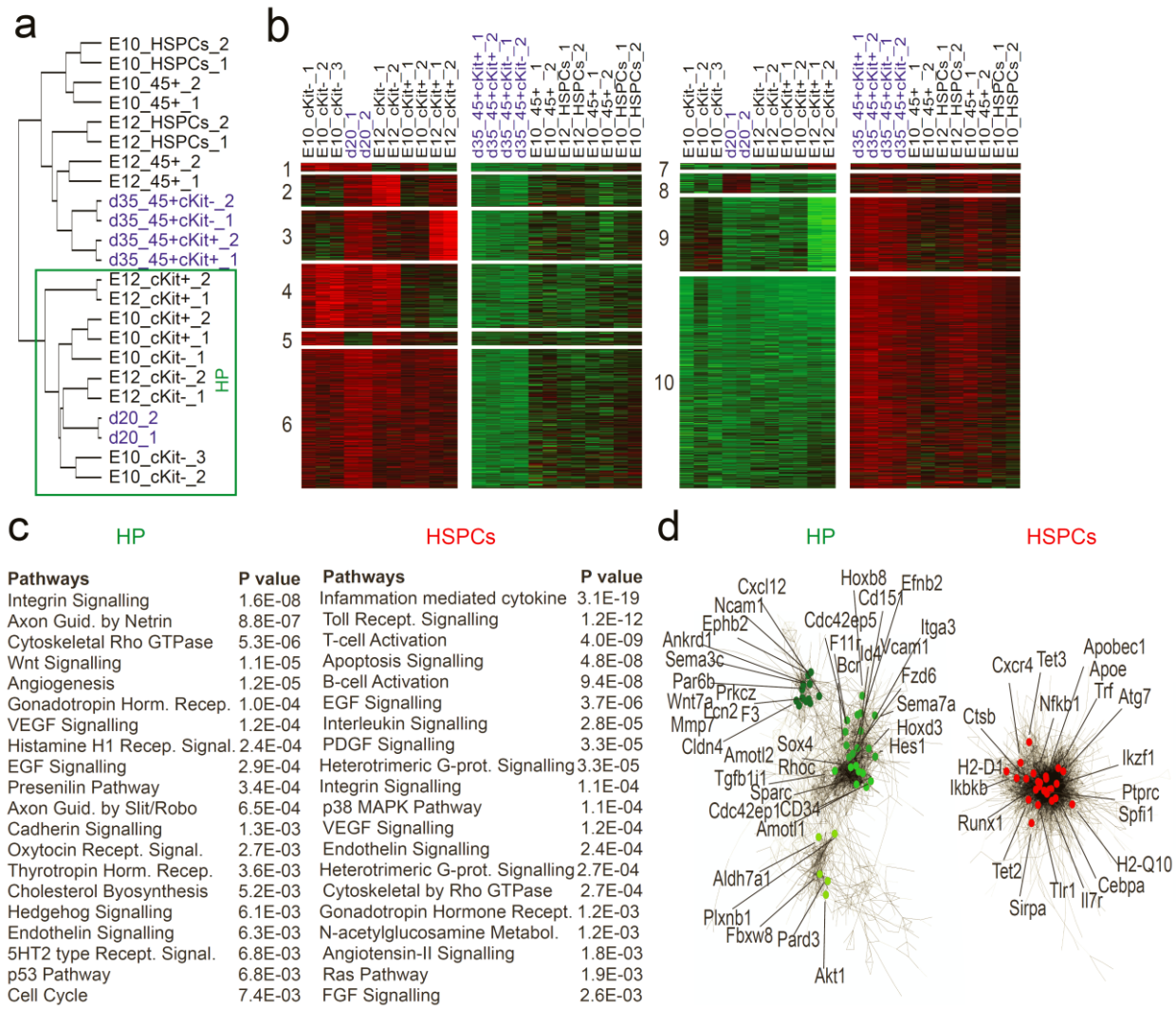



Figure 3.9 - **Hemogenic gene expression signature defined by integration of placental hemogenic cells and induced hemogenic cells.** (**a**) Hierarchical clustering showing the integration of gene-expression data from programmed cells (blue; data from Pereira et al., 2013) with placental HPs and HSPCs. HP cell clustering is highlighted. (**b**) K-means clustering of integrated data reveals 10 clusters of genes enriched in hemogenic cells (cluster 1-6) and HSPCs (cluster 7-10). Red indicates increased expression and green decreased expression relative to the mean. (**c**) Pathway enrichment analysis was performed on genes enriched in the HP and HSPCs clusters using Panther software (www.pantherdb.org). Enriched terms and corresponding p values are shown. (**d**) Co-expression networks show the HP cell network (left panel) and HSPCs enriched network (right panel). Specific genes of the network are highlighted.

Transcription regulator prediction using ENCODE Chip-sequencing data focused on genes activated in HPs (Figure 3.10a, left panel) showed highest enrichment for Gata2 targets ($p = 1.02\ 3\ 10^{-7}$) and the components of the AP-1 TF complex cFos, cJun, and JunD. This result suggests cooperation between Gata2 and AP-1 in the specification of HP identity. In contrast, in HSPCs, a completely different set of transcriptional

regulators was identified, including Ikzf1 ($p = 1.41 \times 3 \times 10^{-38}$), Pu.1 ($p = 3.35 \times 3 \times 10^{-23}$), and Nfkb1 ($p = 2.023 \times 3 \times 10^{-17}$) (Figure 3.10a, right panel). Analysis of genes upregulated in the HSPC cluster using the Mouse Genome Informatic (MGI) mouse mutant phenotype database showed that genetic perturbations caused largely hematopoietic phenotypes (Figure 3.10b, right panel). In contrast, HP upregulated genes affect blood vessel and embryo development (Figure 3.10b, left panel).

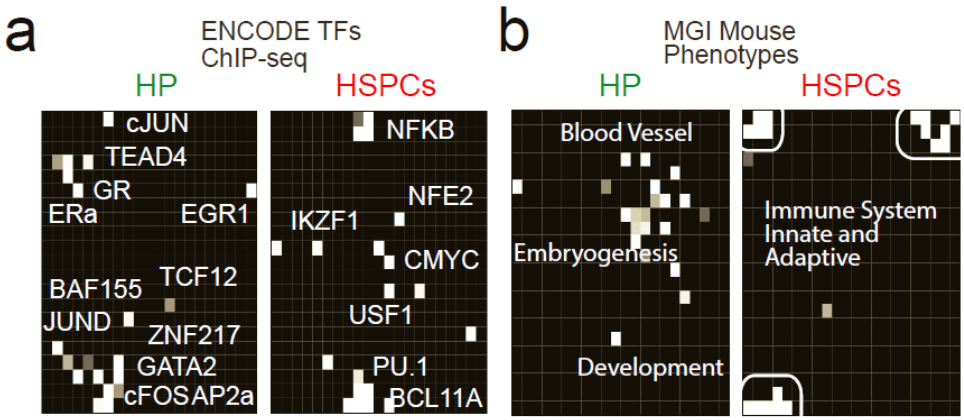



Figure 3.10 - **Transcription regulator prediction in HP and HSPCs.** HP and HSPCs gene clusters were analyzed for gene-list enrichment with gene-set libraries created from ENCODE transcription factor Chip-Seq (**a**) and level 4 of the MGI mouse phenotype ontology (**b**). Enriched terms are highlighted, and brightness represents increasingly significant P values.

We also performed comprehensive gene ontology (GO) analysis in HPs and HSPCs using DAVID clustering (Table 3.1). The top molecular function, biological process and cellular component categories in HP were cytoskeletal regulation, adherens and tight junctions, cellular adhesion, embryonic morphogenesis, cell motion and angiogenesis, and small GTPase regulator activity (**Error! Reference source not found.**). These data are consistent with the recent implication of cellular adhesion and migration during hemogenesis (Lie et al., 2014). GO categories in HSPCs are completely distinct, lysosome, inflammatory response and innate immunity are the most enriched terms.

Table 3.1 - Gene ontology functional annotation clustering in HPs and HSPCs.

| Cluster | ES | HPs (326 clusters) | ES | HSPCs (334 clusters) |
|---|---|---|---|---|
| 1 | 10.2 | Cytoskeleton | 26.9 | Lysosome |
| 2 | 8.19 | Embryonic Development | 13 | Inflammatory response |
| 3 | 7.42 | Actin-Binding | 12.6 | Innate Immunity |
| 4 | 6.52 | PDZ Domains | 10.9 | PH domain |
| 5 | 6.37 | Adherens Junction | 10.5 | Lymphocyte activation |
| 6 | 5.92 | Tight Junction | 8.72 | Membrane lipid metabolic process |
| 7 | 5.64 | Cell Adhesion | 8.48 | Endocytosis |
| 8 | 5.42 | Cytoskeleton Organization | 7.9 | Vesicle |
| 9 | 4.76 | Cell Motion | 6.73 | Activation of immune system |
| 10 | 4.68 | Embryonic Morphogenesis | 6.62 | GTPase regulator activity |
| 11 | 4.28 | Microtubule | 6.06 | Regulation of chemokine production |
| 12 | 4.15 | Endoplasmic Reticulum | 5.72 | Cell chemotaxis |
| 13 | 4.14 | Tube Morphogenesis | 5.23 | T cell activation |
| 14 | 3.94 | SH3 Domain | 5.15 | Adaptive immune response |
| 15 | 3.78 | Angiogenesis/Cell Motion | 5.08 | Regulation of apoptosis |
| 16 | 3.73 | PH Domain | 4.73 | Apoptosis |
| 17 | 3.7 | Actomyosin | 4.35 | NF-kappaB transcriptional activity |
| 18 | 3.61 | Negative Regulation of Cell Differentiation | 4.28 | Cell membrane |
| 19 | 3.48 | LIM Domain | 4.25 | Interleukin-6 production |
| 20 | 3.48 | Chaperone | 4.19 | GTPase activation |
| 21 | 3.31 | GTPase regulator activity | 3.94 | ATP metabolic process |
| 22 | 3.27 | ATP binding | 3.93 | SH2 domain |
| 23 | 3.26 | Skeletal System Morphogenesis | 3.88 | ATP binding |
| 24 | 3.05 | CRIB Domain | 3.86 | Late endosome |
| 25 | 2.78 | Signalosome | 3.74 | Lymphocyte homeostasis |
| 26 | 2.75 | Small GTPase regulator activity | 3.67 | Regulation of adaptive immune response |

| | | | | |
|---|---|---|---|---|
| 27 | 2.73 | Apoptosis | 3.57 | Lipid binding |
| 28 | 2.66 | Protein Stabilization | 3.55 | SH3 domain |
| 29 | 2.64 | Regulation of cell morphogenesis | 3.55 | Blood Cell Homeostasis |
| 30 | 2.53 | Protein folding | 3.54 | SH3 domain binding |

### 3.5 Single cell reconstruction reveals the trajectory of placental hemogenic cells

Understanding the cellular heterogeneity of cells during EHT is important to decipher its molecular program, for this reason we next performed single cell mRNA-Seq analysis of the placental PS34CD45$^-$ population (at E10.5 and E12.5) and HSPCs (at E12.5) to complete the molecular profiling. We have profiled 48 cells for each population, a total of 144 single cells. Using genome-wide unsupervised hierarchical clustering, we show that PS34CD45$^-$ cells cluster together independently of their gestation time (E10.5 and E12.5), and HSPCs form a separate cluster (Figure 3.11a). Regarding programming factor expression, we show that *Gata2* is similarly expressed in the three populations profiled and *Etv6* increases in HSPCs. However, this analysis revealed that the expression of *cFos* decreases from E10.5 PS34CD45$^-$ to E12.5 PS34CD45$^-$ to HSPCs (Figure 3.11b). This result is much more evident at the single cell than population analysis (Figure 3.5). We also observed that *Gfi1b*, although not detected in the PS34CD45$^-$ populations, could be detected in some PS34CD45$^-$ cells at E12.5 and further increased in the HSPC. These data support the role of *Gfi1b* in the later stages of EHT. We then ordered those cells according to gene expression to reconstruct the trajectory of EHT (Figure 3.11c). Endothelial genes are downregulated (*Cdh5*, *Pecam1*, *F11r*) and hematopoietic genes are upregulated (*Ikzf1*, *Pu*.1, *Myb*) according to the pseudo-ordering of cells. When reconstructing the trajectory using PCA and pseudo-ordering, we can observe the overlap between *Prom1* expression and genes expressed at later stages of EHT such as Runx1 and Gfi1b (Figure 3.11d). Using single-cell data, we confirmed the embryonic and endothelial arterial program of PS34 cells by addressing the frequency of cells expressing endothelial markers (*CD31*/*Pecam1*, *CD34*, 98%– 100%), cytokeratins expressed by trophoblasts (*Krt7*, *Krt19*, and *Krt17*, 0%–15%), the trophoblast TF *Hand1* (2%–4%), the venous marker *Nr2f1* (2%), and the arterial marker *Efnb2* (81%–96%). Together these analyses provide a molecular profile for the HP cell program and EHT.

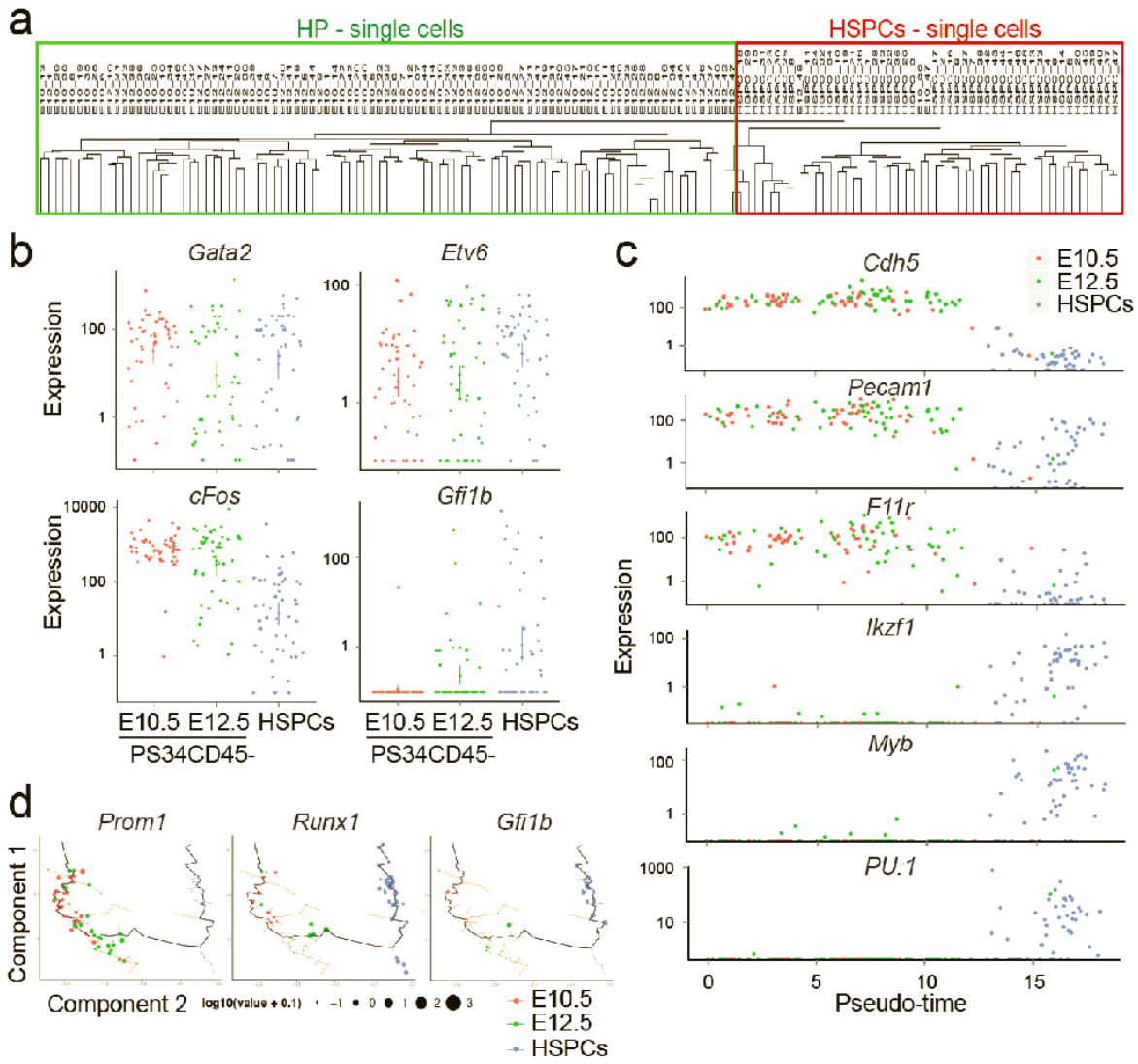



Figure 3.11 - **Single-cell global gene expression of placental hemogenic cells and induced hemogenic cells.** (**a**) Non-supervised hierarchical clustering showing the integration of genome-wide gene-expression data from placenta-derived single cells. Clustering between PS34CD45 at E10.5 and E12.5 (green) and HSPCs (red) are highlighted. (**b**) Expression levels of programming factors in placental single cells. Each dot represents FPKM values for individual cells. (**c**) Expression levels of endothelial and hematopoietic genes in placental single cells ordered with Monocle software (Pseudo-time). Each dot represents FPKM values for individual cells. (**d**) Cell expression profiles (dots) in a two-dimensional independent component space. Lines connecting points represent edges of the minimum-spanning tree constructed by Monocle. Solid black line shows pseudo-time ordering.

## 3.6 **PS34CD45- engraft immunodeficient mice upon Notch activation**

Clonogenic capacity of hematopoietic cells can be evaluated *in vitro* by colony forming unit (CFU) assay. To further assess the ability of PS34 cells to proliferate and differentiate into colonies with both myeloid-erythroid and lymphoid potential *in vitro*,

CFU assay was performed. PS34CD45$^-$ and PS34CD45$^+$ cells at E10.5 and E12.5 were plated in methylcellulose. Both PS34CD45$^-$ and PS34CD45$^+$ cells at E10.5 and E12.5 do not generate CFUs in contrast to the E12.5 cKit$^+$CD34$^+$CD45$^+$ HSPC population (Figure 3.12).

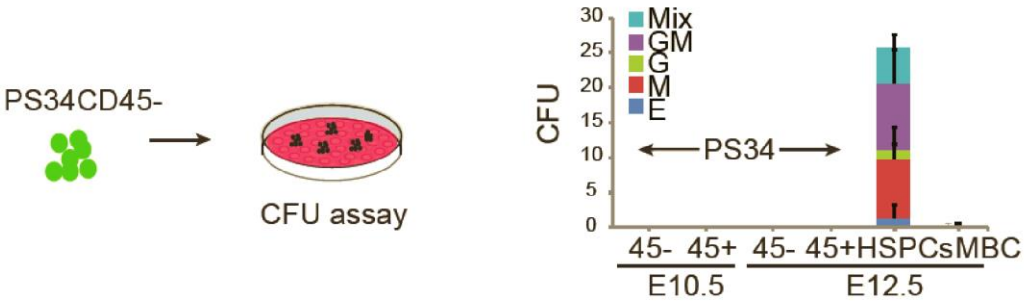



Figure 3.12 – **PS34CD45- and PS34CD45+ cells were isolated form placentas at E10.5 and E12.5 and assayed for hematopoietic colony formation.** HSPCs (CD45$^+$cKit$^+$CD34$^+$) and mature blood cells (CD45$^+$cKit$^-$CD34$^-$) were included as controls. Color code represents the type of hematopoietic colony (mean SD, n=2). GM= Granulocyte, macrophage; G= granulocyte; M= macrophage; E=Erythroid.

In order to acquire clonogenic activity, we hypothesized that placental PS34CD45$^-$ cells would require maturation. Therefore, we performed co-cultures of PS34CD45 cells with OP-9 stromal cells and cytokines. After 4 days of co-culture, cells with an HSPC phenotype (CD45$^+$CD34$^+$cKit$^+$) were generated and subsequently tested in CFU assays Figure 3.13). PS34CD45$^-$ derived cells generated large mixed colonies that were similar to those from HSPCs (Figure 3.13b and c). The presence of hematopoietic cytokines (stem cell factor [SCF], IL-3, IL-6, FltL3, and thrombopoietin) was required to generate large numbers of hematopoietic colonies (>50 CFU per plate; Figure 3.13c). As controls, we plated freshly isolated PS34CD45, which are unable to generate colonies, and HSPCs, which do generate colonies (Figure 3.13b). Mixed colonies generated from both PS34CD45 and HSPCs were large, contained hemoglobinized erythroid cells, and displayed mixed cellular morphologies (Figure 3.13c).

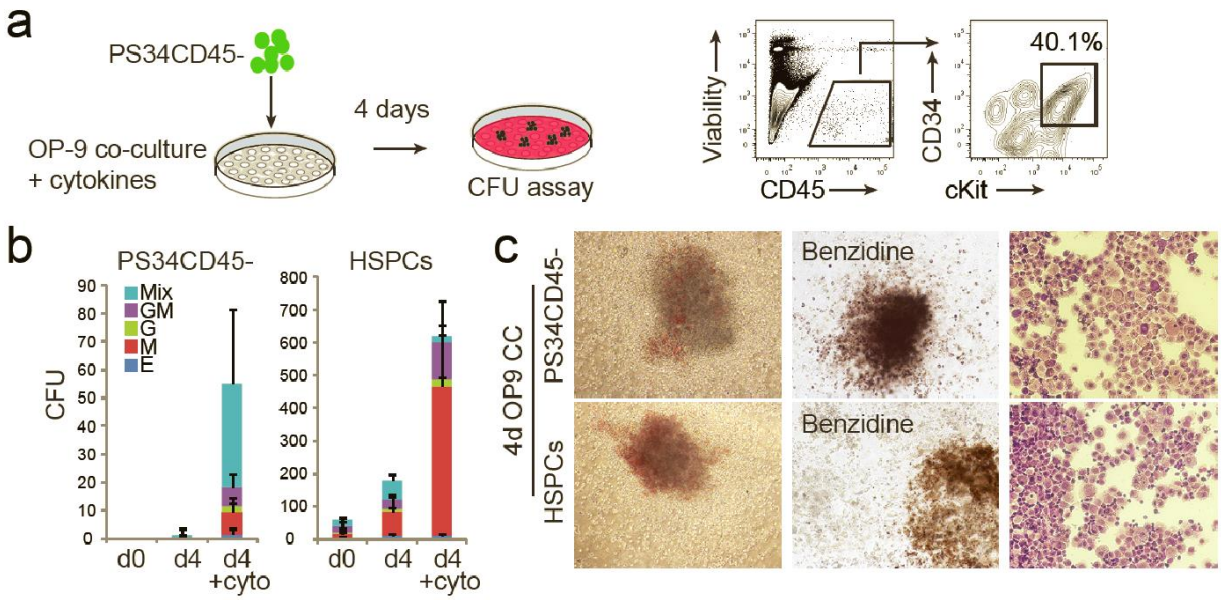



Figure 3.13 - **Acquisition of clonogenic activity by co-culture PS34CD45- with OP9 cells.** (**a**) E12.5 PS34CD45$^-$ cells were co-cultured on OP-9 in the presence of cytokines (SCF, Fltl3L, IL-3, IL-6, and TPO) for 4 days. FACS analysis shows the generation of CD45$^+$CD34$^+$cKit$^+$ cells that were tested for clonogenic activity. (**b**) Number and type of colonies generated at day 0 and after co-culture on OP9 of PS34CD45$^-$ or HSPCs (CD45$^+$cKit$^+$CD34$^+$) with or without cytokines (mean ± SD, n = 3). (**c**) Pictures of representative mixed colonies (left), benzidine hemoglobin staining (middle) and modified Giemsa staining (right).

We next asked whether an *in vitro* co-culture step with OP-9 would confer engraftment potential to the emergent hematopoietic progenitors. We isolated PS34CD45 cells at E12.5 and performed co-cultures with OP-9 for 3 or 4 days. We did not detect engraftment after OP-9 co-culture with cytokines (Figure 3.14). This result suggests that hematopoietic progenitors may be pushed to differentiate in the culture conditions employed, and activation of a critical signaling pathway(s) may be required to maintain self-renewal of *in vitro* generated HSPCs. Notch signaling is essential for the specification of HSC and definitive hematopoiesis (Kumano et al., 2003; Marcelo et al., 2013).

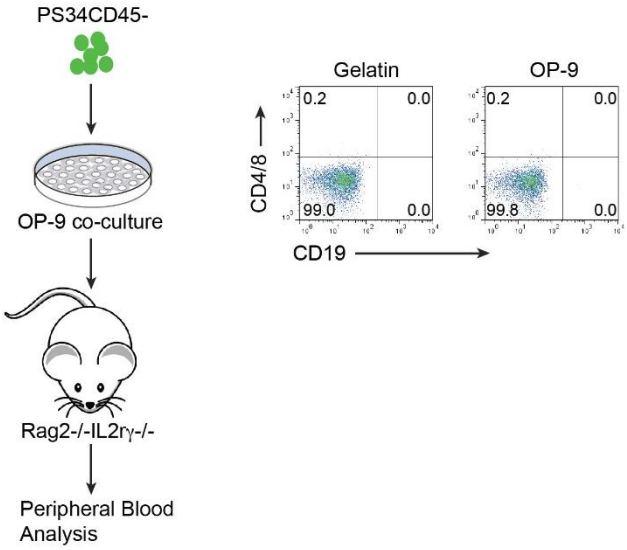



Figure 3.14 - **PS34 Cells Do Not Directly Engraft Adult Mice.** E12.5 placentas were dissociated and PS34CD45- cells isolated and cultured in gelatin-coated dishes or co-cultured with OP-9 cells in the presence of cytokines for 4 days. Co-cultures were dissociated and transplanted intravenously into RAG2/IL2rγ immunodeficient mice. After 12 weeks the presence of B (CD19+) and T (CD4+/CD8+) cells in peripheral blood was analyzed by flow cytometry.

During development, close contact between Notch signal-emitting cells and precursors is essential for successful HSC specification and maturation (Clements et al., 2011; Kobayashi et al., 2014). We hypothesized that providing a Notch signal *in vitro* would confer engraftment potential to PS34CD45 cells. To test this, we isolated PS34CD45 cells from E12.5 placentas and performed 4-day co-cultures on gelatin-coated dishes, OP9-DL1 cells, which express the Notch ligand Delta1 (Figure 3.15a). Twelve weeks after transplantation into Rag2-/-/IL2Rgc mice, we detected a large population of B and T cells in peripheral blood only when PS34CD45 cells were co-cultured with OP9-DL1 (Figure 3.15b, upper panel, and 3.15c). The populations of B and T cells are maintained in the peripheral blood for up to 8 months after transplantation (Figure 3.15b). We next asked if cells derived from PS34CD45 cells reestablish an HSPC compartment in the bone marrow and maintain self-renewal. To address this, we sacrificed transplanted mice at 14 weeks and isolated purified HSCs ($LSKCD48CD150^{+}$) from bone marrow and transplanted secondary Rag2-/-/IL2Rγc mice. We first confirmed that both the spleen and bone marrow of transplanted mice contained populations of B and T cells at 14 weeks (Figure 3.15c, lower panel). At 8 months after secondary transplantation, we could still detect B and T lymphocytes in

peripheral blood, bone marrow, and spleen of these animals (Figure 3.15b and d). These results demonstrate that the PS34CD45 phenotype marks early precursors of bona fide HSCs, and that Notch signaling is crucial for a productive EHT *in vitro*. In summary, based on the phenotype and gene expression data from our previous *in vitro* programming studies (Pereira et al., 2013), we have identified and characterized in detail HSC precursor cells.

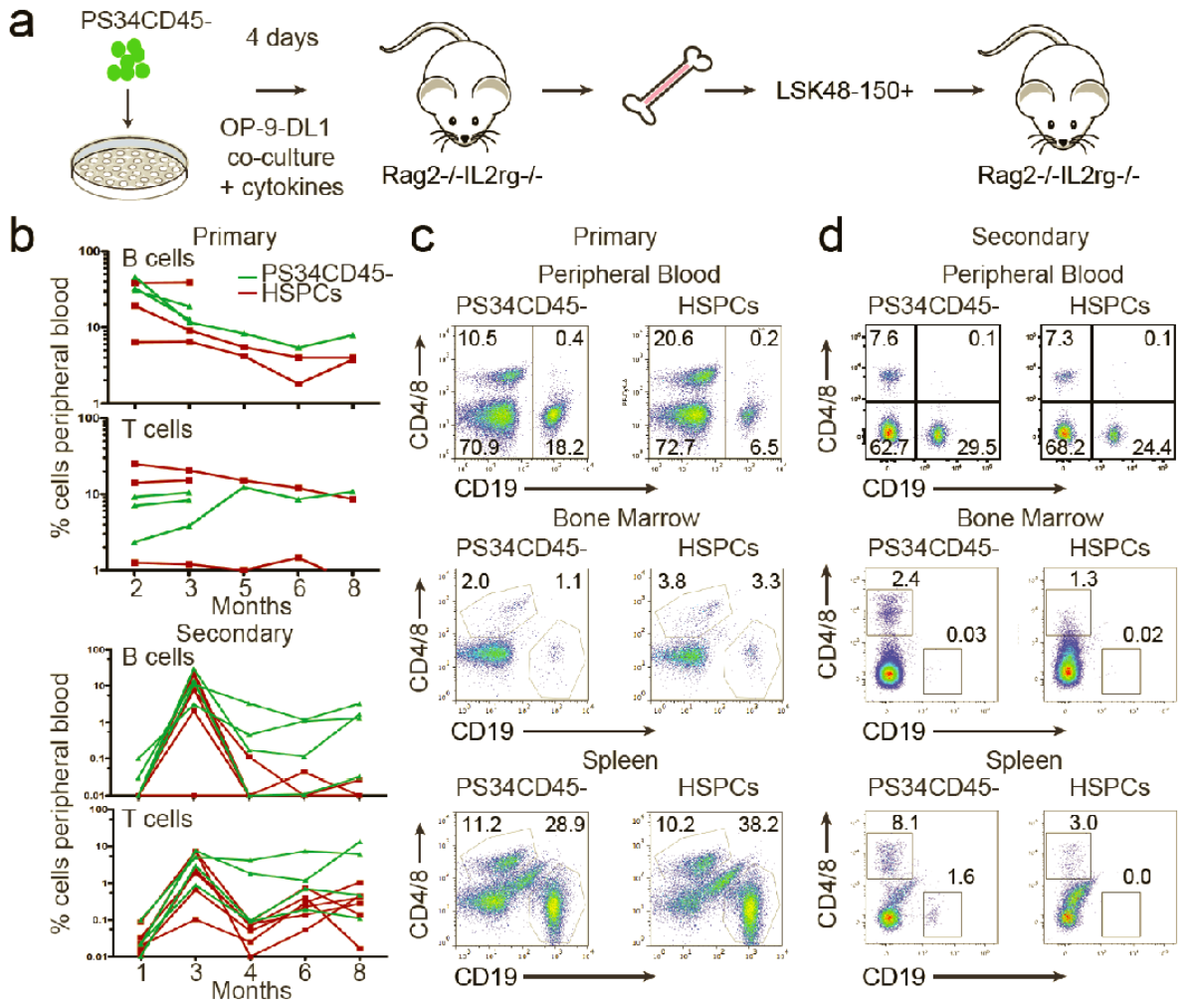



Figure 3.15 - **Placental hemogenic cells engraft immunodeficient mice after OP9-Dl1 co-culture.** (**a**) E12.5 placentas were dissociated and PS34CD45 cells were isolated and co-cultured with OP9-DL1 cells in the presence of cytokines for 4 days. Co-cultures were dissociated and transplanted intravenously into Rag2-/-/IL2Rgc-/- mice. (**b**) Mice were bled at 2–8 months (experiment 1) or 2–3 months (experiment 2) after transplantation (upper panel) with PS34CD45- co-cultured with OP9-DL1 (green lines). Additional mice were transplanted with freshly isolated HSPCs (red lines). After 14 weeks or 8 months of primary transplant, engrafted mice were sacrificed, and HSCs (LSK48-150$^+$) were purified from bone marrow and transplanted intravenously into secondary Rag2-/-/IL2Rgc-/- mice (lower panel). (**c**) After 12–14 weeks, the presence of B cells (CD19$^+$) and T cells (CD4$^+$ /CD8$^+$) in peripheral blood, bone marrow, and spleen was analyzed by flow cytometry. (**d**) Analysis for the presence of B and T cells in peripheral blood after 3 months and in spleen and bone marrow after 8 months of secondary transplantation.

## 3.7 **Discussion**

In developmental hematopoiesis, there is no surface marker definition for the precursor cells to HE. CD34 is not expressed (or at low levels) in LT-HSC from bone marrow and fetal liver (Gekas et al., 2005; Osawa et al., 1996; Sánchez et al., 1996) but is highly expressed in HSCs from the placenta, AGM (Gekas et al., 2005; Kumano et al., 2003; Sánchez et al., 1996), and in repopulating cells from E9.5 yolk sacs (Yoder et al., 1997). CD34 is also expressed in the HE derived from human PSCs (Nakajima-Takagi et al., 2013; Sturgeon et al., 2014). The *Ly6a* gene that encodes the Sca1 antigen is expressed before HSC emergence in the dorsal aorta and is a marker of definitive but not primitive HE (Chen et al., 2011; Ling et al., 2004). Placental and embryonic HPs are also marked by the expression of the *Prom1* gene. In humans, *Prom1* is a marker for endothelial progenitor cells and HSCs. Indeed, it has been shown that a $CD34^+$ $PROM1^+$ population from human umbilical cord blood can give rise to endothelial and hematopoietic cells (Wu et al., 2007). In the mouse, however, Prom1 is not expressed in LT-HSCs isolated from bone marrow (Arndt et al., 2013). We showed that Prom1 expression rapidly declines during EHT allowing the exclusion of specified, mature HSPCs. These three markers showed peak expression in PS34CD45cKit cells, in which we also detected other genes previously implicated in HE or hemogenic programming (*Sox17*, *Scl*, *Podxl*, *Kitl*, *cFos*, *Gata2*) (Clarke et al., 2013; Pereira et al., 2013). Runx1 and Gfi1b, critical genes during EHT (Lancrin et al., 2012; Swiers et al., 2013), were only detected at low levels, implying that PS34CD45 cells represent early-stage HPs. Furthermore, along with the expression of endothelial markers, we detected the expression of CD49f, an antigen identified in induced HPs (Pereira et al., 2013) and human HSCs (Notta et al., 2011). CD49f is expressed in both PS34 and emergent HSPCs. Taken together, these data highlight the relevance of the phenotype defined by Prom1, Sca1, and CD34 for the prospective isolation of early definitive HPs *in vivo*. In the E10.5–E11.5 AGM region, two types of HSC precursors that sequentially develop into HSCs have been described: type I pre-HSCs (VE-$caderin^+$ CD45CD41low) and type II pre-HSCs (VE-$cadherin^+$ $CD45^+$) (Rybtsov et al., 2011; Taoudi et al., 2008). In the E9.5 mouse embryo, an earlier HSC precursor (pro-HSC, VE-$cadherin^+$ CD45CD41lowCD43) that is SCF dependent has also been described (Rybtsov et al., 2014). As suggested by the genome-wide gene expression data, PS34 cells represent an earlier precursor, where hematopoietic markers such

as CD45, CD41, and Runx1 are not highly expressed. It will be interesting to investigate the maternal and fetal factors (Rossant and Cross, 2001) that affect the specification and maturation of HPs into HSCs. Defining such factors may facilitate the efficient maturation of hemogenic cells *in vitro*. We show that a co-culture step with OP9 supplemented with cytokines allows the specification of HPs into definitive HSPCs with myeloid and lymphoid clonogenic activity. The activation of the Notch pathway (by co-culture with OP9-DL1) endows serial transplantation capacity in Rag2/IL2Rγc mutant mice, highlighting the role of Notch and a permissive environment for precursor cell maturation (Hadland et al., 2015). We show that the transcripts for the adhesion molecule *F11r* (also known as Jam1a) (Kobayashi et al., 2014) and the TFs *Foxc2* (Jang et al., 2015) and *Hes1* (Bertrand et al., 2010) are upregulated both in placental PS34 cells and programmed hemogenic cells. Interestingly, in zebrafish, Notch-dependent Gata2b expression within the hemogenic cell compartment is required to initiate Runx1 expression (Butko et al., 2015), supporting the notion that Gata2 functions in HPs before Runx1 (Ditadi et al., 2015). The ability to isolate HPs in large quantities from developing placentas and subsequently mature them *in vitro* into functional HSPCs provides an opportunity to dissect EHT and stem cell generation. Using this system, it will be interesting to address the impact of inflammatory signals (Li et al., 2014), Wnt ligands (Ruiz-Herguido et al., 2012), and other substrates as well as soluble factors on HP maturation and the establishment of stem cell self-renewal. Our PS34 dataset and integrative analysis between programmed precursors and *in vivo* precursor cells provides a stringent signature for HP cells. By integrating the two datasets, we eliminate concerns of contamination with other cell types (ex vivo isolation) and also aberrant gene activation from TF overexpression. Our data show that programmed precursors from fibroblasts and placental PS34CD45$^{-}$cKit$^{-}$ cells are strikingly similar and correlate with 23GFP$^{+}$ cells from AGM. This generated signature can be used to test the relevance of specific pathways (such as integrin signaling, epithelial to mesenchymal transition, angiogenesis, cytoskeletal regulation by Rho GTPases, Netrin signaling, and others) for HP specification. It will be interesting to assess the role of individual or a combination of genes enriched in HP clusters during EHT and stem cell generation (e.g., *Aldh7a1*, *Cxcl12*, *Kitl*, *Sema3c*, *Par6b*, *Wnt7a*, *Vcam1*). HP lineage divergence from endothelium may occur before extensive formation of intra-aortic clusters (Rybtsov et al., 2011; Swiers et al., 2013). The enrichment of

angiogenesis, migration and adhesion-related genes in PS34CD45 cells also suggests that HP cells do not represent a cohort of mature endothelium but more likely a different lineage with endothelial features that is committed to the hematopoietic route (Ditadi et al., 2015). This result is also supported by the recent implication of cell migration as a feature of HPs in zebrafish (Kobayashi et al., 2014) and Runx1 targets in mice (Lie et al., 2014). In summary, we used information generated during direct programming into hemogenic cells (Pereira et al., 2013) to provide insights into the specification of HSCs in the placenta. Direct cell reprogramming is therefore not only a methodology for generating specific cell types but also can provide the minimal TF network to kick start the specification of cell identity and cellular markers and phenotypes to clarify developmental specification. This approach can be complementary to traditional developmental approaches especially in scenarios where genetic studies might fail; for example, when genes functionally compensate each other or due to the complexity of embryonic tissue and spatial-temporal constraints. Previous studies of AP-1 in HSC specification illustrate the advantage of this approach. Deletion of cFos did not severely impair placental HSCs (Ottersbach and Dzierzak, 2005). However, *cFos* overexpression was absolutely required for hemogenesis from fibroblasts (Pereira et al., 2013) or the homologous gene FOSB, from human ECs (Sandler et al., 2014). In light of the recent implication of inflammatory pathways as major players during HSC specification (Li et al., 2014), it would be interesting to investigate the link between AP-1 activation, NF-kB, and inflammation (Linnemann et al., 2011; Natoli, 2010).

Collectively, our results show that Prom1, Sca1, and CD34 mark HP cells. These precursors can be matured to serially engrafting HSCs *in vivo*. Our results support the view that HP lineage divergence is an early developmental event and underscore the requirement of Notch activation for their maturation into functional HSCs. These studies provide an *in vitro* platform for the molecular dissection of definitive EHT and HSC emergence. The derived gene expression platform will work both ways, suggesting new ways to improve *in vitro* hematopoietic reprogramming. Taken together, our studies suggest that both processes will provide insights into the expansion of HSCs from both reprogrammed and donor/patient-derived HSCs for clinical transplant and regenerative medicine.

# Chapter 4 - Reprogramming human fibroblast into hemogenic cells

Our understanding of hematopoietic development has been formed by many decades of research using the model organisms available, including fish, chick, frog and mouse, which provides valuable insights into the origins of the adult hematopoietic system and a wider understanding of embryonic hematopoiesis (Medvinsky et al., 2011). Most of the studies aiming to elucidate the hematopoietic system are done in mice due to their small size, short reproductive cycle, ease maintenance and sharing of genomic and physiological characteristics with humans (Walsh et al., 2017). Ultimately, animal models are used as substitutes of the human organism due to the ethical and logistical limitations. However, mice and humans diverge in size, lifespan and among other characteristics they differ on their immune system (Doulatov et al., 2012). Concerning HSCs markers, unlike mouse HSCs, human cells are $CD34^+$ $FLT3^+$ (Ogawa et al., 2001; Sitnicka, 2003) and do not express CD150 (Larochelle et al., 2011). In addition, the identification of human hematopoietic progenitors and stem cell markers like integrin alpha 6 (CD49f) and angiotensin-converting enzyme (ACE) allowed a better understanding of human HSC specification (Jokubaitis et al., 2008; Notta et al., 2011; Sinka et al., 2012) with the expression of ACE being associated with hematopoietic precursors in human AGM (Sinka et al., 2012). Despite these advances, the study of HSC emergence is still a challenge in humans. Alternative strategies to study the regulation of hematopoietic development in humans are warranted. In Chapter 3 we have demonstrated that lineage reprogramming can be employed to inform HSC specification in the mouse. We hypothesize that the basic cellular and genetic mechanisms underlying HSC specification can be investigated using direct lineage conversion.

In this chapter, we aimed to establish the generation of human hemogenic cells and to define these cells identity as they change during reprogramming.

## 4.1 **Transferring hemogenic reprogramming to the human system**

To assess the feasibility of using the same TFs we used in the mouse hemogenic reprogramming we determined the pattern of expression in the human system. We found that both the minimal (GATA2, GFI1B, and FOS; GGF) and optimal (GATA2, GFI1B, FOS, and ETV6; GGFE) combinations of TFs were highly expressed in human $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) when compared to a broad collection of other cell-types and tissues (Figure 4.1). We next expressed the human

TFs in adult human dermal (HDF) and neonatal foreskin (BJ) fibroblasts using a Doxycycline (Dox) inducible vector system (Pereira et al., 2013). Accordingly, the TFs were highly expressed in human fibroblasts in the presence of Dox and upon its removal expression was undetectable within 6 days. Therefore, this lentiviral system is similarly robust in human as in mouse cells for inducible and controllable transgene expression.

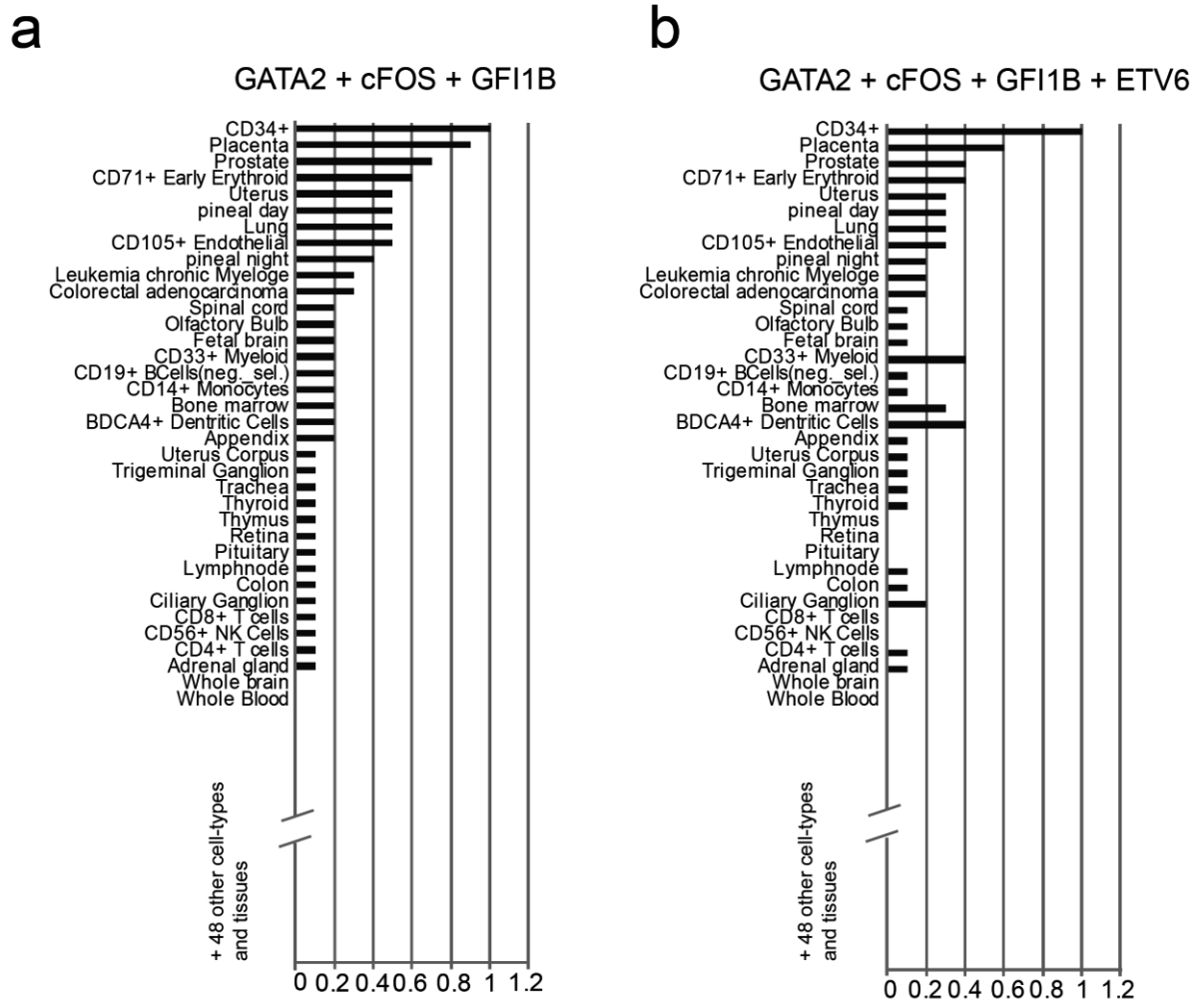



Figure 4.1 - **Analysis of TF gene expression in human cells/tissues.** The combination of either (**a**) GATA2, FOS and GFI1B or (**b**) GATA2, FOS, GFI1B and ETV6 are mostly enriched the most in human CD34$^{+}$ HSPCs among 84 tissues and cell-types. Gene expression data from different human tissues and cell-types, were obtained from the GeneAtlas U133A database. The “GPSforGenes” program was written and used to classify the tissues where combinations of TFs were most enriched (best fit = 1).

## 4.2 **Induction of human CD34⁺ and CD49f⁺ cells by GATA2, GFI1B and FOS**

We next transduced both HDF and BJ cells with viruses expressing GGFE and GGF. Two days after transduction Dox was added and the cultures were analyzed for over 25 days (Figure 4.2). An analysis of both colony number and expression of CD34 revealed that 8.4-13.6% of the colonies contained CD34⁺ cells (Figure 4.3a). Interestingly, in human cells the GGF minimal cocktail appears more effective than GGFE (Figure 4.3a). Similar morphological changes were observed in both HDF and BJ transduced cultures. The cells appeared to round up in cobblestone like areas during the midpoint of the culture with semi-adherent and non-adherent round cells appearing later (Figure 4.3b). Round hematopoietic-like cells that express CD34 and CD49f were observed and confirmed by immunofluorescence (Figure 4.3c) and quantified by flow cytometry (Figure 4.3d). These changes defined the induction period with a large population of cells expressed CD49f⁺ (20-25%) while CD34 expression was more restricted (0.2-2.0%). The CD49f⁺ population emerges early (day 12) and is maintained, while CD34⁺ cells peak at day 25 and subsequently decline but do persist. No colonies, morphologic changes, or CD49f⁺ and CD34⁺ cells were observed after transduction with control virus (Figure 4.4).

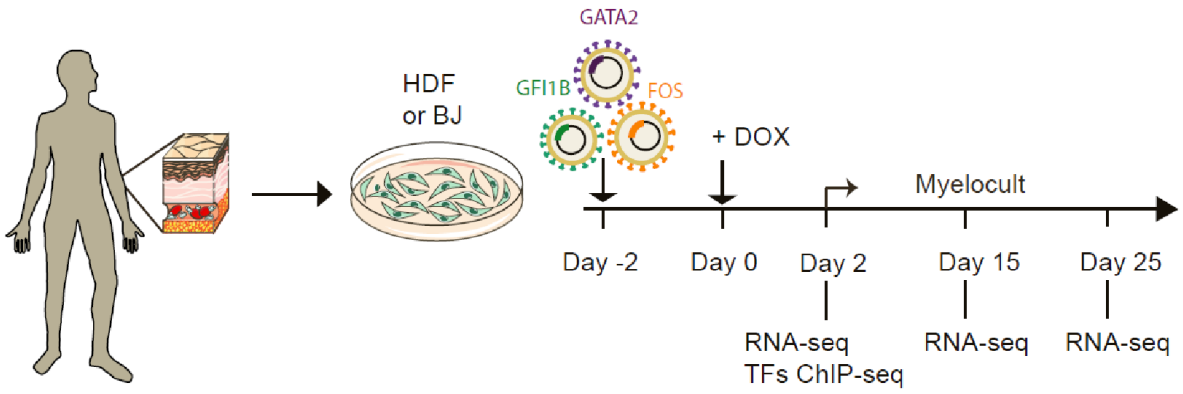



Figure 4.2 - **Strategy for inducing hemogenesis in human fibroblasts.** Adult human dermal fibroblasts (HDF) and Neonatal foreskin (BJ) were transduced using an inducible lentiviral system with human coding sequences for GATA2, FOS and GFI1B. The transgene expression starts upon addition of doxycycline in the media. After 2 days cells will be cultured in Myelocult media for up to 30 days. Analysis will be performed at different time points.

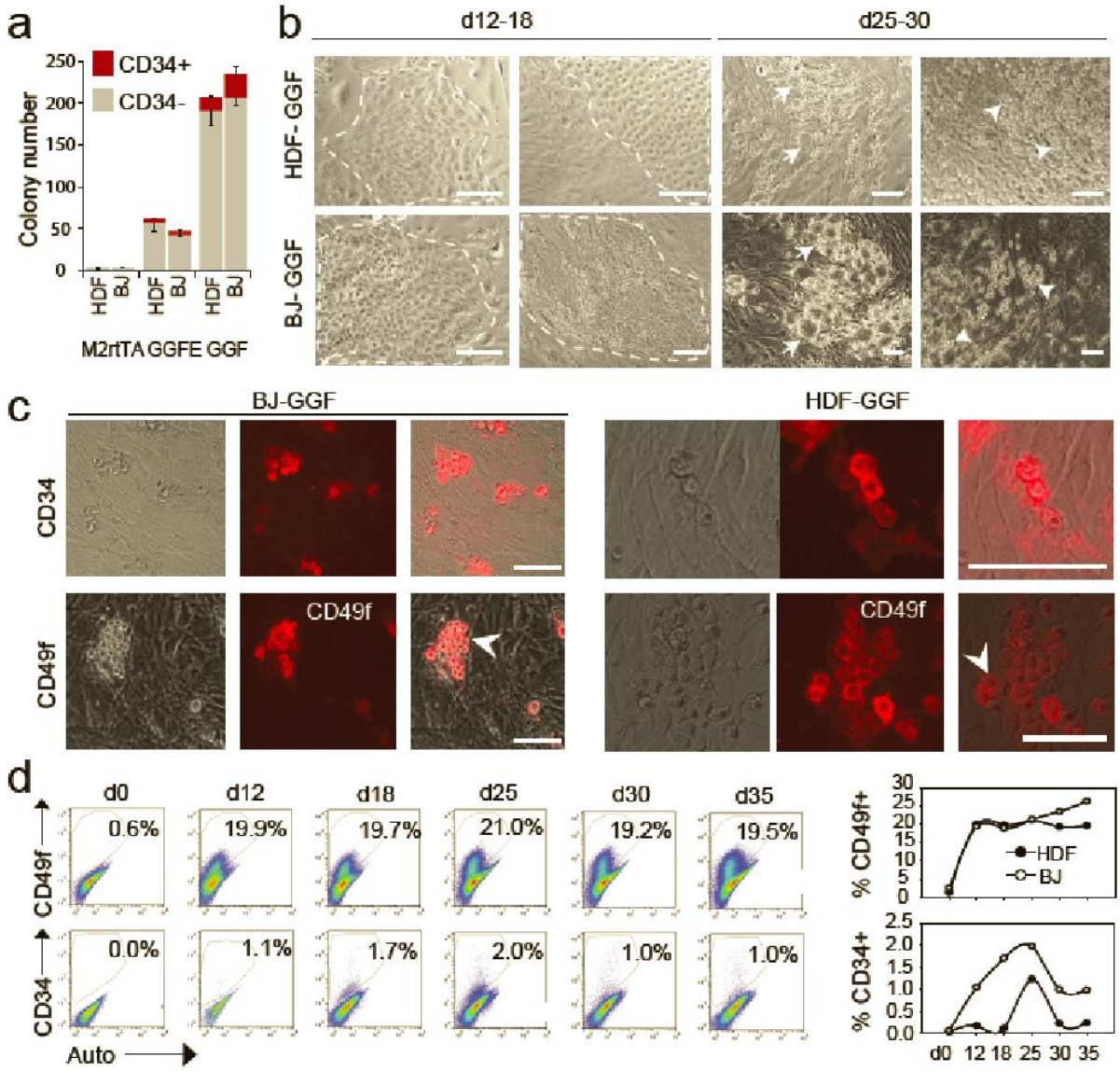



Figure 4.3 – **GATA2, GFI1B and FOS without ETV6 induce CD34⁺ and CD49f⁺ colonies in human fibroblasts.** (**a**) Cells were transduced with GATA2, GFI1B, FOS, ETV6 (GGFE); GATA2, GFI1B, FOS (GGF); or control M2rtTA viruses and cultured with Dox for 30 days. Colonies were stained for CD34 and counted. Colony numbers are per 10,000 transduced fibroblasts (mean ± SD). (**b**) Colony morphology 12-18 days (left panels, dashed line) and 25-30 days (right panels) after induction. (**c**) Colonies were assayed by immunofluorescence for CD34 (upper panels) or CD49f (lower panels) 30 days after transduction. Arrows and arrowheads highlight distinct induced cellular morphologies. Scale bars, 100 µm. (**d**) Kinetics of induction in days of CD34 and CD49f-positive populations.

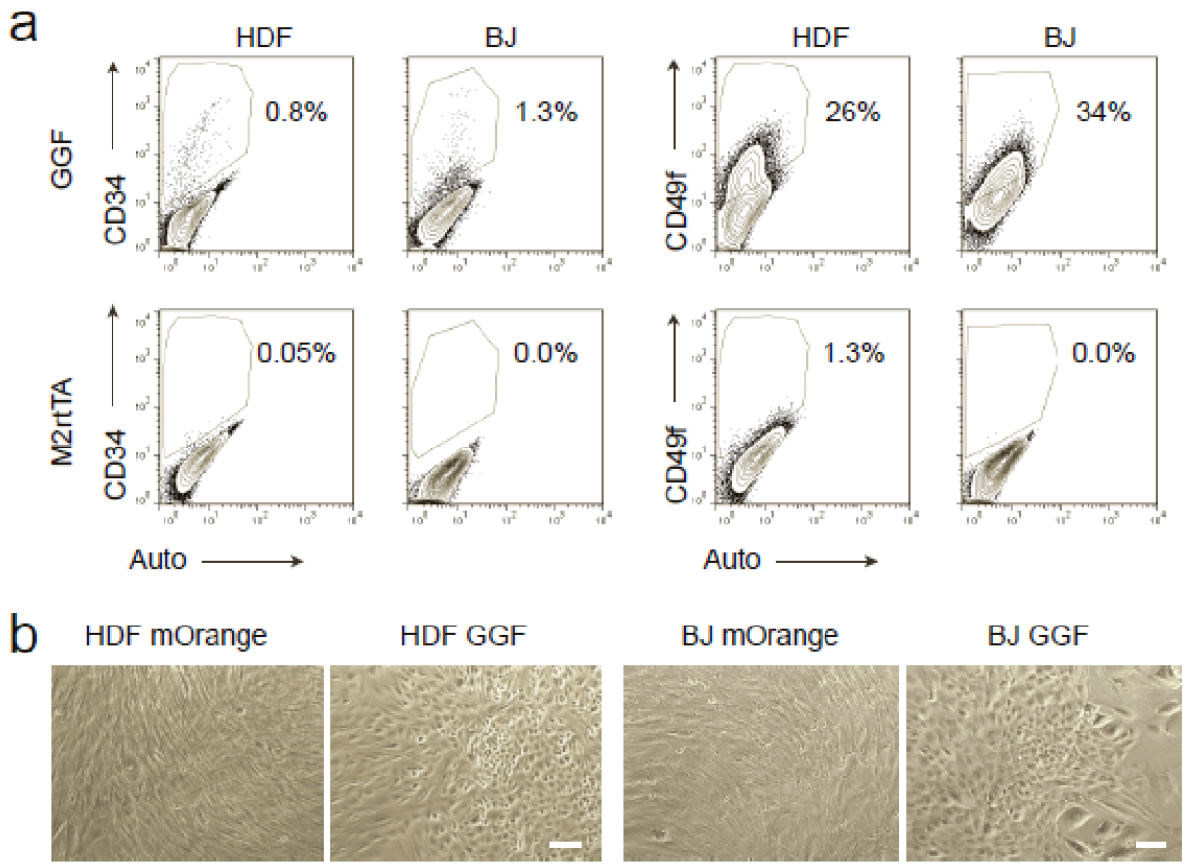



Figure 4.4 - **Characterization of human induced cells.** (**a**) Flow cytometry analysis of CD34 and CD49f expression in HDF and BJ fibroblasts 26 days after transduction with GGF (upper panels) or M2rtTA as a control (lower panels). (**b**) Morphological analysis of HDF and BJ fibroblasts 26 days after transduction with GGF or mOrange as a control.

### 4.3 **Induced colonies contain cells with human HSPC surface phenotype**

We further examined the induced colonies for the co-expression of markers that define a human HSC phenotype. We first showed that these cells co-express CD34 and CD49f in both HDF and BJ cells (Figure 4.5a). Transduced BJ cells generated greater numbers of $CD49f^+$ and $CD34^+CD49f^+$ double-positive cells (Figure 4.5b) possibly due to their neonatal origin. A subpopulation of $CD34^+CD49f^+$ cells expressed CD133 (Prominin1), and low levels of CD45, which is also expressed at low levels in human HSCs (Figure 4.5c) (Jay et al., 2004). Using multicolor flow cytometry, we showed that $CD34^+CD49f^+$ cells express CD90 (Thy1) and among these 9-14% also express CD133 and low levels of CD45 (Figure 4.5). These are all positive markers of human HSPCs (Notta et al., 2011). The negative markers CD38 and CD45RA were not

detected in HDF or BJ-derived reprogrammed cells (Figure 4.5e). These data fit well with extensive phenotypic definitions of human HSCs in UBC. One study has reported engraftment of NOD-scid IL2Rγ-null (NSG) mice with single $CD49f^{+}CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}$ UBCs (Notta et al., 2011). Interestingly, GATA2 or FOS removal from the three-factor combination abolished the generation of $CD34^{+}CD49f^{+}$ cells while removal of GFI1B increased the number of double positive cells (Figure 4.6a) but diminishes the percentage of $CD45^{+}$ and $CD133^{+}$ cells (Figure 4.6b) consistent with the role of GFI1B at the later stages of EHT in the mouse (Pereira et al., 2016; Thambyrajah et al., 2016). These results demonstrate that GATA2, GFI1B and FOS induce a human HSPC phenotype in two different types of human fibroblasts.

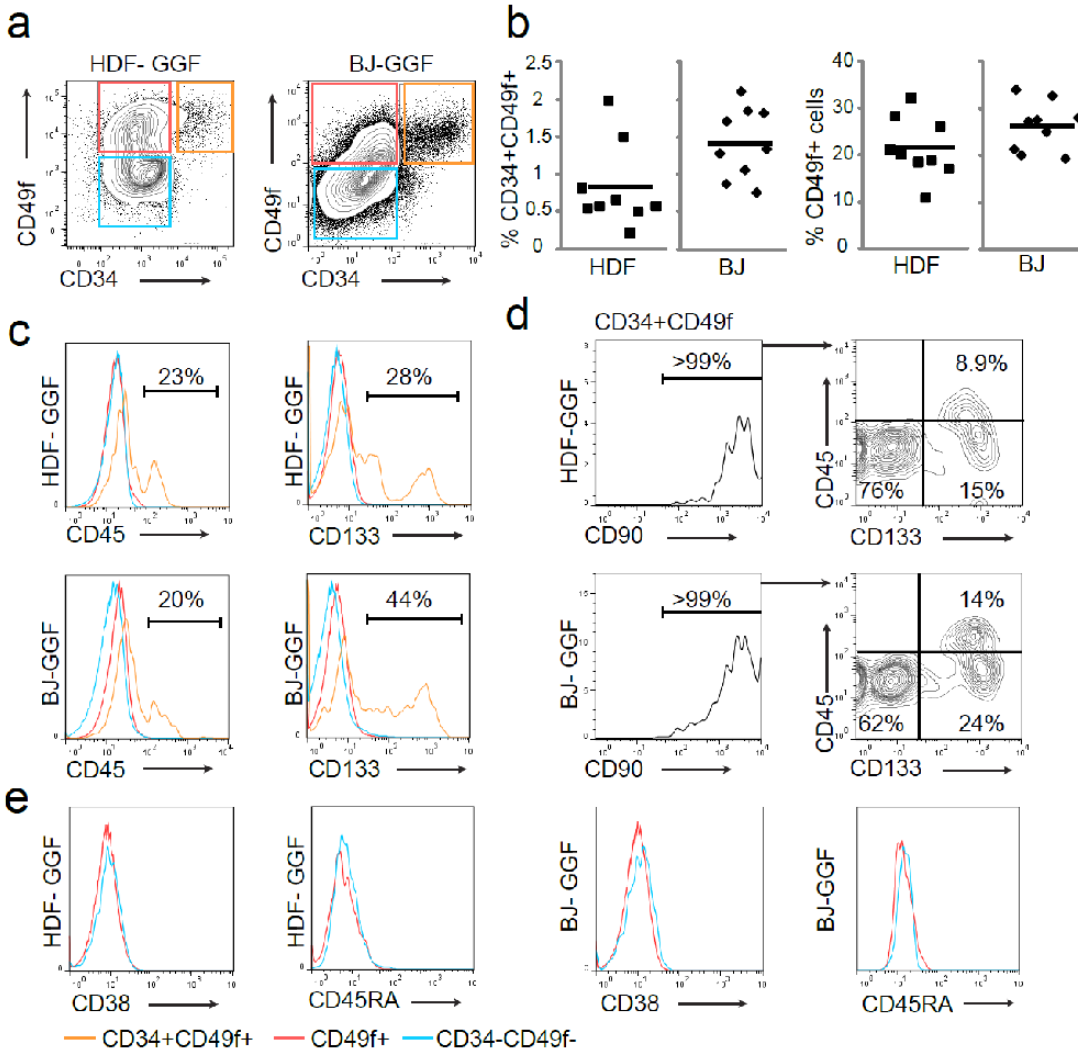



Figure 4.5 - **A human HSC-like cell surface phenotype is induced during reprogramming.** (**a**) Analysis of CD34 and CD49f expression 26 days after transduction with GGF. (**b**) Quantification of $CD34^{+}CD49f^{+}$ and $CD49f^{+}$ cell populations. Each symbol represents an experiment. (**c**) Expression of CD45 and CD133 within $CD34^{+}CD49f^{+}$ (orange

lines), CD49f$^+$ (red lines) and double negative population (blue lines) from two independent experiments. (**d**) Expression of CD90, CD45 and CD133 within the CD34$^+$CD49f$^+$ population. (**e**) Lack of expression of CD38 and CD45RA in both CD34$^+$ (red lines) and CD34$^-$ (blue lines) populations from two independent experiments.

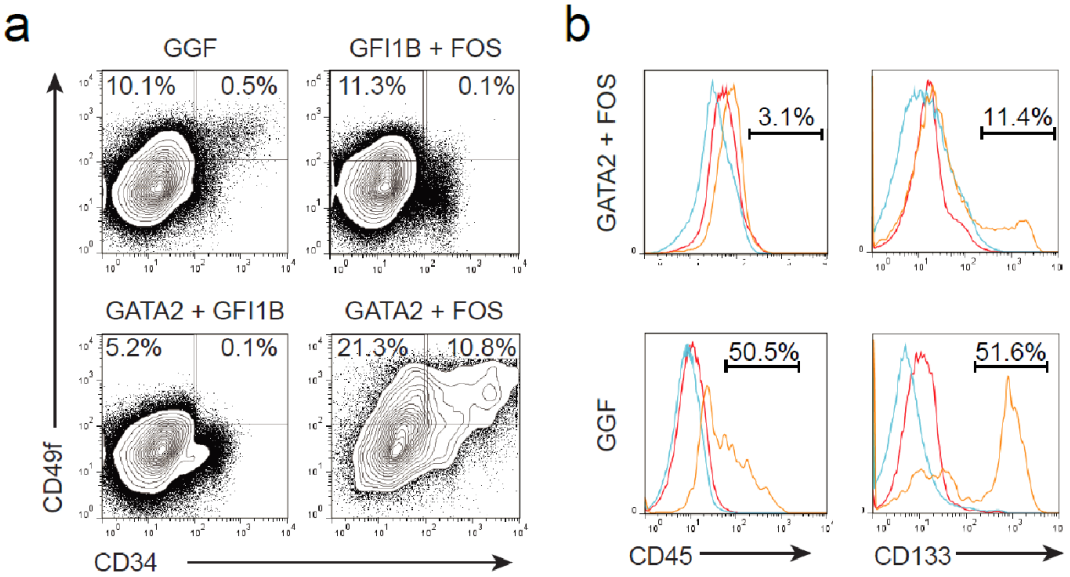



Figure 4.6 - **A combination of GATA2, GFI1B and FOS is required to induce an HSC cell surface phenotype.** (**a**) Flow cytometry analysis of CD34 and CD49f expression 27 days after Dox induction in fibroblasts transduced with GATA2, GFI1B and FOS (GGF) or with combinations of two TFs (orange box CD34$^+$CD49f$^+$, red box CD34$^-$CD49f$^+$, blue box CD34$^-$CD49f$^-$). (**b**) Expression of CD45 and CD133 is found within the CD34$^+$CD49f$^+$ population (orange lines) from GGF but not GATA2$^+$FOS transduced fibroblasts. CD49f$^+$ (red lines) or double negative populations (blue lines) do not contain significant populations of CD45$^+$ or CD133$^+$ cells from either transduction.

## 4.4 **Comprehensive gene expression analyses during reprogramming**

To further interrogate the gene expression changes occurring during the reprogramming process we performed mRNA-sequencing (RNA-seq) analysis at different times after transduction of HDF and BJ cells. We sorted 3 biological replicates of non-transduced fibroblasts, day 15 CD49f$^+$CD34$^-$, day 25 CD49f$^+$CD34$^-$ and day 25 CD49f$^+$CD34$^+$ cells. All biological replicates correlate well in contrast to comparisons between different samples (Figure 4.7). We used non-negative matrix factorization (NMF) coupled with consensus clustering to measure sample diversity (Brunet et al., 2004).

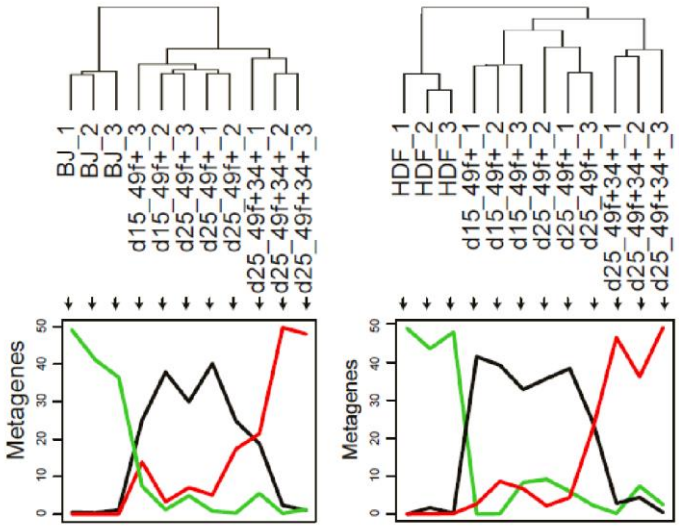



Figure 4.7 - **Metagene expression patterns during reprogramming.** Populations of non-transduced fibroblasts (BJ and HDF) and GGF transduced day 15 CD49f+, day 25 CD49f+ and day 25 $CD34^{+}CD49f^{+}$ cells were profiled by RNA-seq (3 biological replicates). Ordered tree linkage displays clustering of the profiled samples (upper panel) and the metagenes that represent most of the variability associated with each cellular transition (lower panel).

Metagene analyses showed: 1) sets of genes expressed in fibroblasts and silenced in all other samples, 2) genes transiently expressed in $CD49f^{+}$ cells and 3) genes that start to be expressed in $CD34^{+}CD49f^{+}$ cells (Figure 4.7). *MMP1*, a fibroblast-associated gene is silenced in the sorted $CD49f^{+}$ and $CD34^{+}$ populations. *ANGPTL4*, a gene implicated in the positive regulation of angiogenesis, is transiently activated in $CD49f^{+}$ cells. ACE, the gene encoding the angiotensin-converting enzyme is upregulated in $CD49f^{+}$ cells and continues to be expressed in $CD34^{+}$ cells in HDFs and BJ-derived induced cells (Figure 4.8a). Interestingly, *ACE* expression has been proposed to identify HSC precursors in the human AGM (Sinka et al., 2012). We confirmed that ACE and CD49f were expressed on the cell surface in transduced cells and that $CD34^{+}$ cells are contained in this subset (Figure 4.8b).

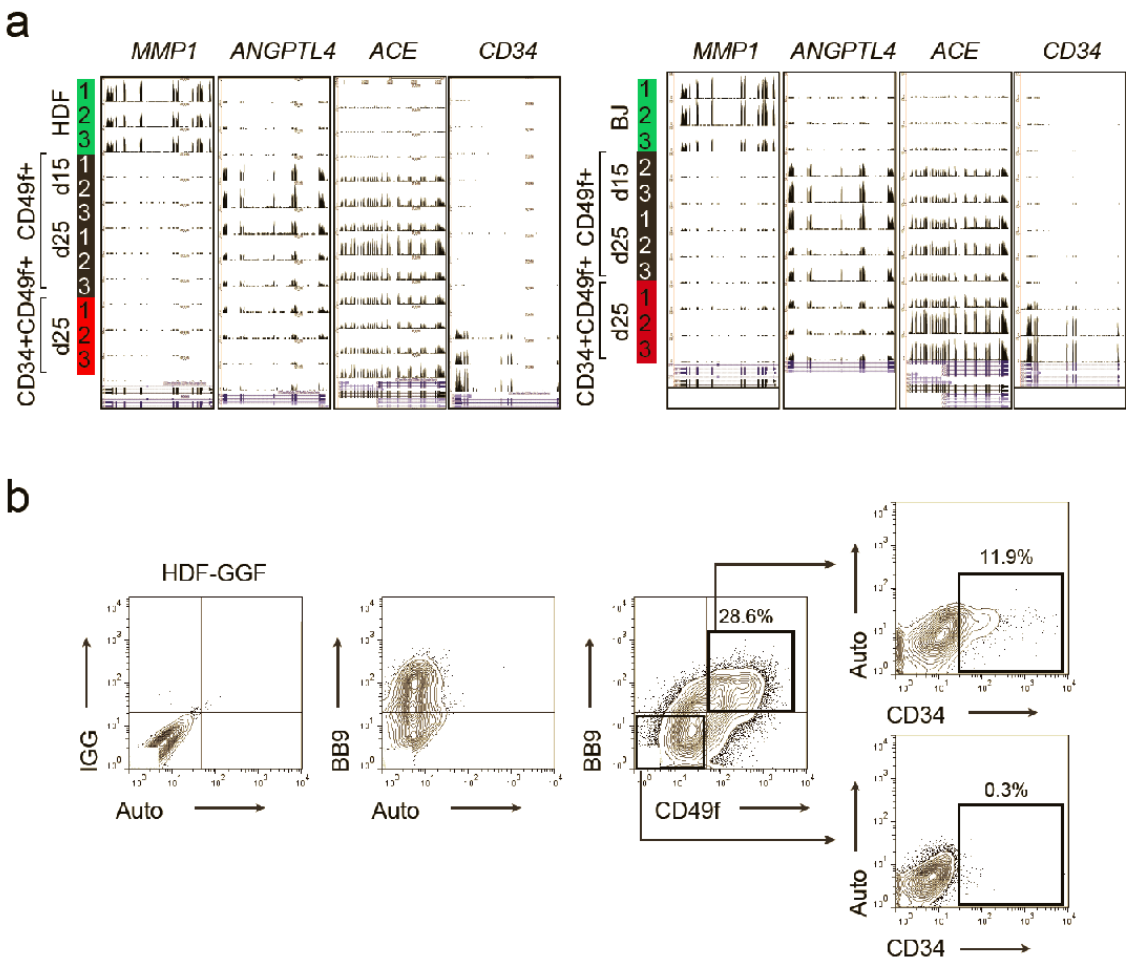



Figure 4.8 - **Key hemogenic genes are induced during reprogramming.** (**a**)HDF and BJ fibroblasts were transduced with GATA2, GFI1B and FOS (GGF) and global gene expression levels in non-transduced fibroblasts (BJ), day 15 CD49f$^+$, day 25 CD49f$^+$ and day 25 CD34$^+$CD49f$^+$ were profiled by RNA-seq (biological replicates: 1, 2 and 3). Reads were aligned to the human genome and those that mapped to the *MMP1*, *ANGPTL4*, *ACE* and *CD34* genes are displayed as maximum read heights. (**b**) Flow cytometry with the BB9 antibody to confirm ACE expression in HDF fibroblasts 25 days after transduction of HDF with GGF. Percentages of CD49f$^+$BB9$^+$ and CD49f$^+$BB9$^+$CD34$^+$ cells and Isotype control for BB9 are shown.

The RNA-seq datasets generated from transduced BJ and HDF cells were integrated. At days 15 and 25 the fibroblast-associated gene signature including GREM1, PSG5, FGF5, LUM, HAPLN1, and MAB21L is silenced (Figure 4.9a and c). These genes are also silenced during induced Pluripotent Stem Cell (iPSC) reprogramming (Yu et al., 2007). Among upregulated genes in CD49f$^+$CD34$^+$ cells we identified those encoding proposed markers of AGM HSC precursors such as ACE, F11R, and EPCAM, the transcription factors RUNX1, SCL and FOXO1 and homing molecules such CD9.

Several pro-angiogenic factor genes such as JAG1, SEMA4D, VWF, FOXC2 and ETS2 are expressed in the CD49f$^+$ cells at both day 15 and 25 (Figure 4.9b and c).

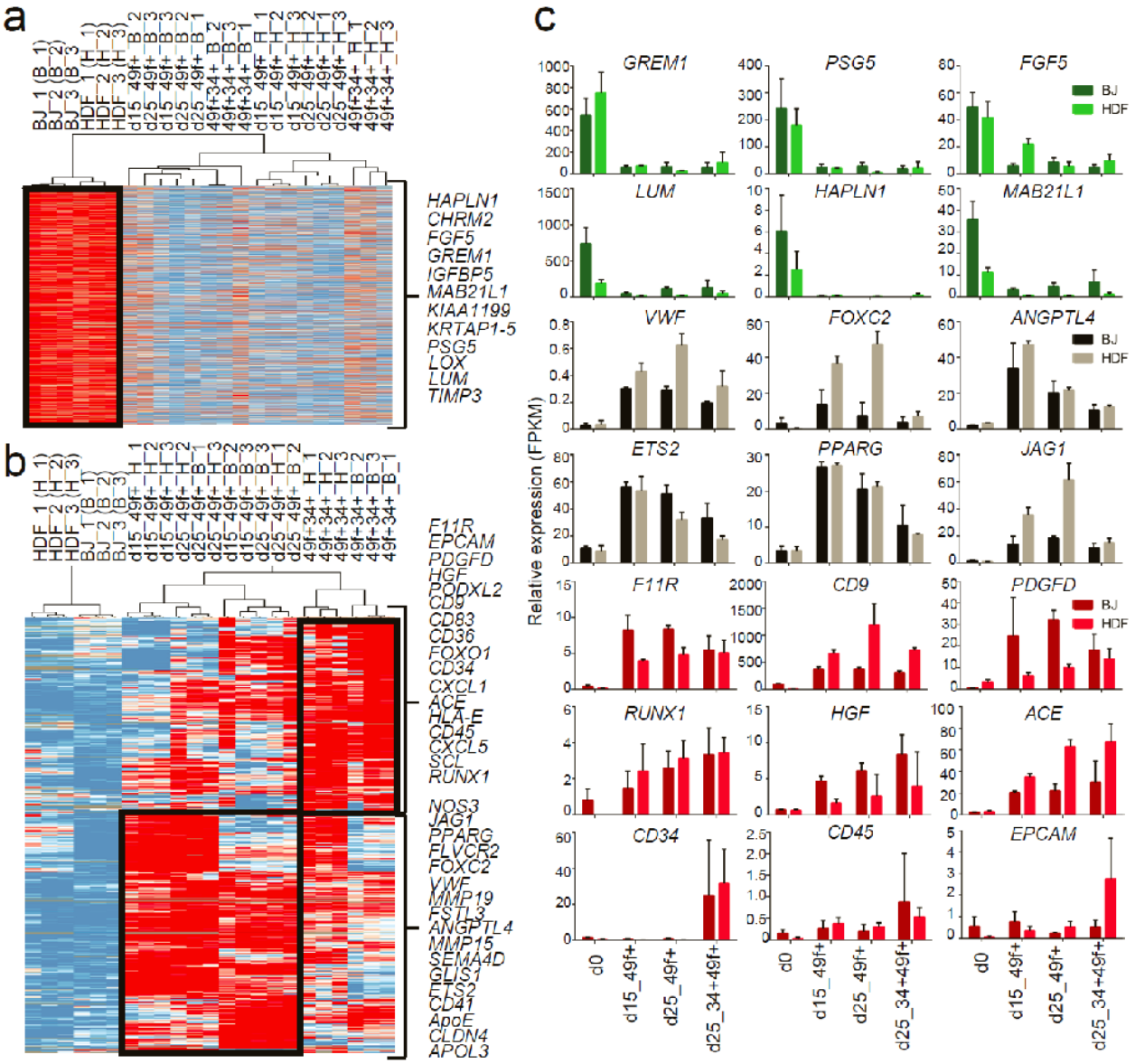



Figure 4.9 - **Dynamic activation of endothelial and hematopoietic gene expression in human hemogenic colonies.** Non-transduced fibroblasts and GGF transduced day 15 CD49f$^+$, day 25 CD49f$^+$ and day 25 CD34$^+$CD49f$^+$ cells were profiled by RNA-seq (3 biological replicates). (**a**) Heatmap of genes expressed in fibroblasts and silenced in CD49f$^+$ and CD34$^+$CD49f$^+$ cells. (**b**) Heatmap of genes activated in CD49f+ and CD34$^+$CD49f$^+$ cells. Black boxes highlight the stage-specific expression of gene sets. Red indicates increased expression and blue decreased expression over the mean. Data were analyzed by Cluster 3.0 and displayed by Treeview. (**c**) The expression levels of fibroblast (green), endothelial (black) and hematopoietic (red)-specific genes are shown as FPKM mean values ± SD.

We asked if an overall angiogenic program is initially activated during reprogramming. We compiled a gene list of positive angiogenic regulators and asked if these were activated in transduced HDF and BJ. The vast majority of these genes are activated

in CD49f$^+$ cells and some continue to be expressed or are activated in CD34$^+$CD49f$^+$ cells (Figure 4.10a and b). Gene set enrichment analysis (GSEA) with angiogenic gene sets from the gene ontology (GO) biological process database revealed enrichment for angiogenesis and regulation of angiogenesis (Figure 4.10c). Endothelial lumens have been shown *in vivo* to originate from intracellular vacuoles generated by pinocytosis (Kamei et al., 2006). As colonies develop from transduced BJ and HDF cells many vacuoles are generated that sequentially form large organized colonies with what appear to be intracellular spaces (Figure 4.10d). We tested whether theses vacuoles represent pinocytic vesicles by the incorporation of the membrane impermeant dye 5,6- carboxyfluorescein. Indeed, vacuoles were fluorescent after 16 hours of culture with the dye (Figure 4.10e). Using immunostaining we detected CD49f$^+$ cells with fluorescein-positive vesicles (Figure 4.10f).

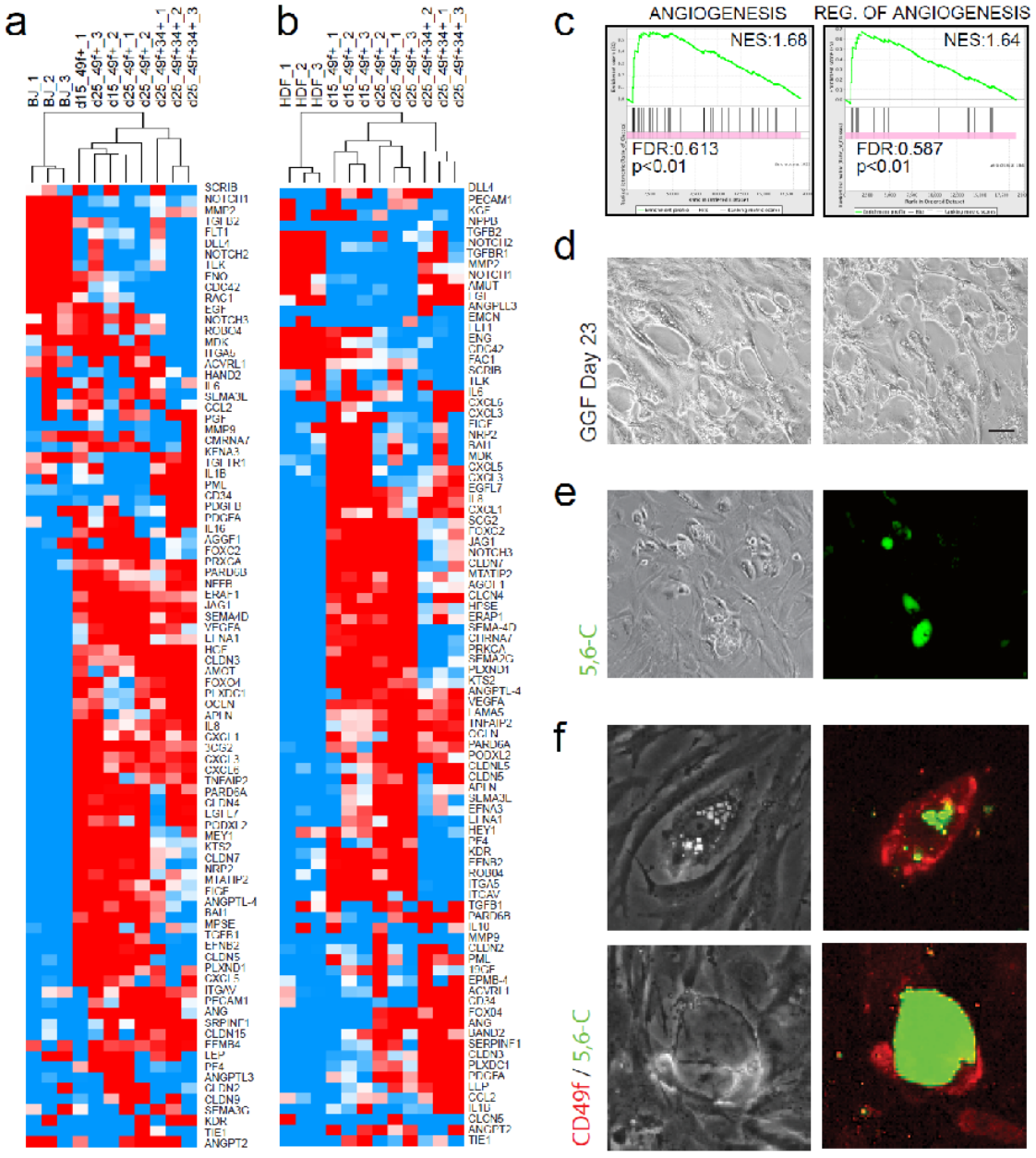



Figure 4.10 - **Induced hemogenic cells display an angiogenic gene expression signature.** The expression of genes implicated in angiogenesis is shown in (**a**) BJ and (**b**) HDF derived cells. Red indicates increased expression and blue, decreased expression over the mean. Data were analyzed by Cluster 3.0 and displayed by Treeview. (**c**) Gene set enrichment analysis was performed for BJ and CD49f$^{+}$ cells. Gene expression lists were analyzed for enrichment of angiogenesis and regulation of angiogenesis gene sets present in the GO Biological process database. (**d**) Colony morphology of BJ and HDF 23 days after induction with GGF. Intracellular vacuoles are numerous. (**e**) The generation of vacuoles by pinocytocis was tested by the incorporation of 5(6)-Carboxyfluorescein (5,6-C, green) for 16 hours in HDF transduced with GGF. (**f**) HDF-derived cells were analyzed by immunofluorescence for CD49f (red) and 5,6-C (green).

Further analyses of genes up- regulated in CD34$^{+}$CD49f$^{+}$ cells using the Mouse Genome Informatics, (MGI) mouse mutant phenotype database showed that their

genetic perturbations result in largely hematopoietic phenotypes (Figure 4.11a). We used GSEA to compare the transition from fibroblasts to $CD34^+CD49f^+$ cells with published gene sets. This showed significant enrichment of HSC gene sets in the $CD34^+CD49f^+$ samples generated from either HDF or BJ fibroblasts (Figure 4.11b, left). Indeed, top enriched gene sets include CD34 TFs as well as the Wnt, Hedgehog and TGFβ signaling pathways, consistent with their hemogenic roles (Figure 4.11b, right) (Medvinsky et al., 2011; Pereira et al., 2013). We next integrated our data with published HSC data from UBC (Notta et al., 2011). Principal component analysis (PCA) was performed with transduced BJ and HDFs datasets and UBC microarray data (Notta et al., 2011). This analysis showed that induced $CD49f^+$ cells are the closest to $CD49f^+$ HSCs while both HDF and BJ fibroblasts are very distinct from all other datasets. The $CD34^+CD49f^+$ cell populations are positioned between HSCs and multipotent progenitors (MPPs) (Figure 4.11c). When compared to our data from mouse hemogenic programming (May et al., 2013), the programmed human $CD34^+CD49f^+$ cells cluster closely with mouse induced HSPCs ($CD45^+cKit^+$) (Figure 4.11d).

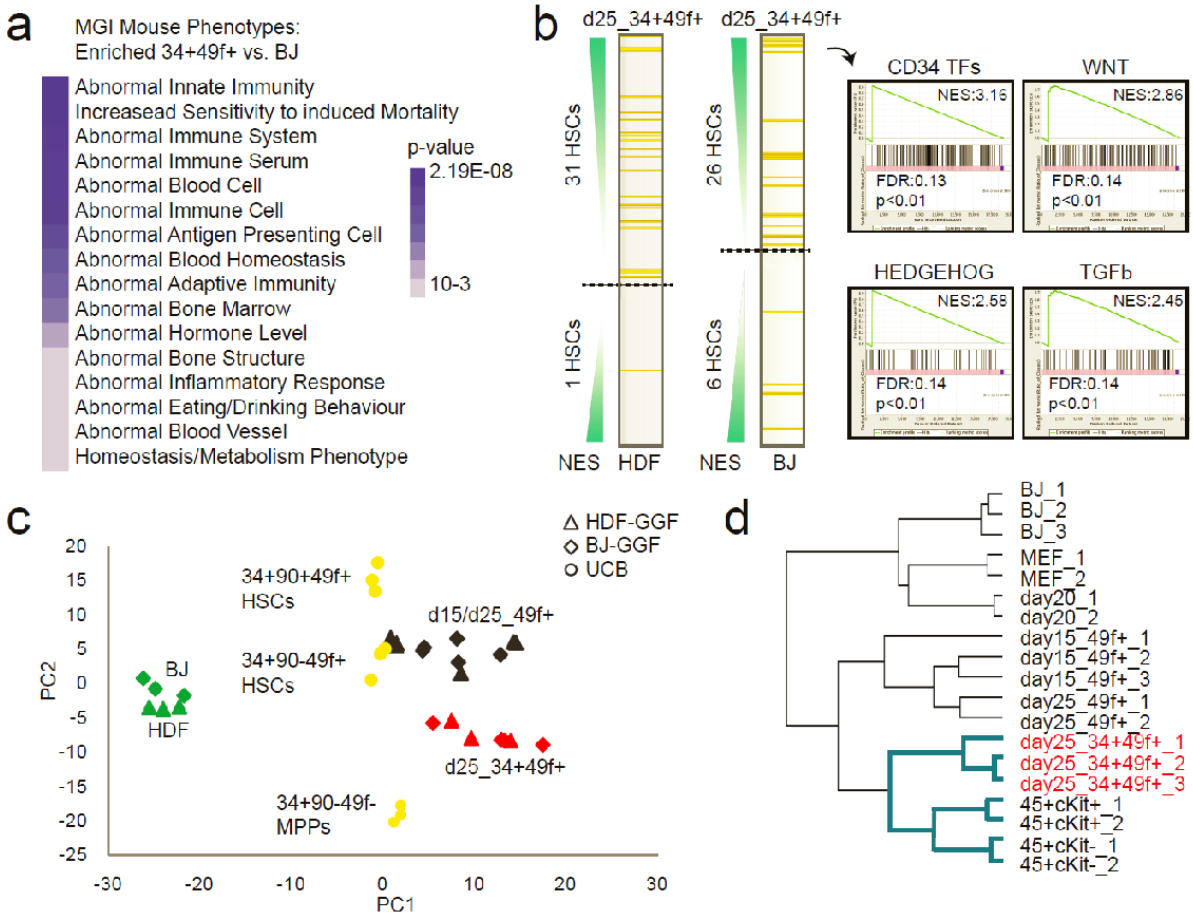



Figure 4.11 - **HSC-like gene expression signatures are displayed by reprogrammed cells.** (**a**) Gene list enrichment analysis with libraries from MGI mutant mouse phenotype ontology for genes upregulated from BJ to CD34$^{+}$CD49f$^{+}$. Heatmap shows enrichment p-values. (**b**) GSEA for HDF and BJ to CD34$^{+}$CD49f$^{+}$ samples. Gene expression lists were analyzed for enrichment of gene sets present in the Molecular Signatures Database (MSigDB) (1888 gene sets, gene size 0-5,000). Orange lines represent HSC datasets ordered according to normalized enrichment score (NES). The dashed line highlights the cut-off FDR = 0.25. Right panels show GSEA for CD34 enriched TFs, Wnt, Hedgehog and TGFβ signaling pathways genesets. (**c**) RNA-seq datasets were integrated with expression data from human UCB HSPCs (from Notta et al., 2011). PCA shows the relative distances between samples. (**d**) Hierarchical clustering integrating data from reprogrammed mouse and human cells; CD34$^{+}$CD49f$^{+}$ cells highlighted in red and blue lines highlight separate cluster for hematopoietic phenotypes (mouse data from Pereira et al., 2013).

To characterize the reprogramming process in more detail, we performed mRNA-seq at the single-cell level at different timepoints of the induction. Human UCB Lin$^{-}$CD34$^{+}$ cells and non-transduced HDFs were profiled as controls. Genome-wide unsupervised hierarchical clustering shows clustering of most single cells according to sample group showing a clear separation of HDF from reprogrammed cells. Surprisingly, only 48h after transgene activation, fibroblasts show a dramatic transcriptional change. In

agreement with the PCA (Figure 4.11c), single CD49f$^{+}$CD34$^{-}$ cells are the closest to human UCB cells (Figure 4.12a). We performed gene list enrichment analysis for pairwise comparisons of HDF, HDF day 2, CD49f$^{+}$CD34$^{-}$ day 15, CD34$^{+}$CD49f$^{+}$ day 25 and UCB. An interactive comparative map between those different groups of single cells was created using the 500 most differentially expressed genes (Appendix I). The hyperlinks for differentially expressed genes and gene set enrichment analysis are included and allow identification of transcription factors, ontologies, pathways among other features. For example, glycolysis and INF-alpha signaling pathways ($p=2.7x10^{-5}$ and $p=1.2x10^{-3}$, respectively) were the top enriched terms for the transcriptional activation occurring in the first 2 days of reprogramming, suggesting that these genes may be direct targets of hemogenic TFs (Figure 4.12b). Taken together, GATA2, FOS and GFI1B direct hemogenic transcriptional changes in multiple types of fibroblasts of both mouse and human origins.

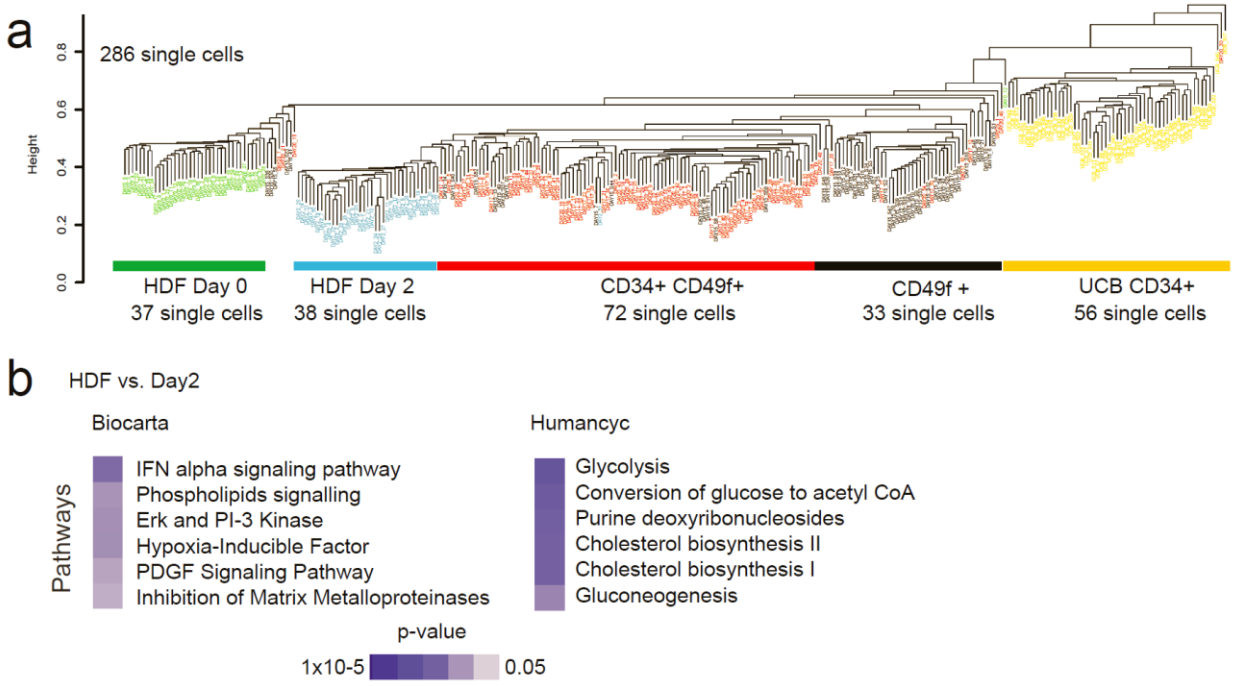


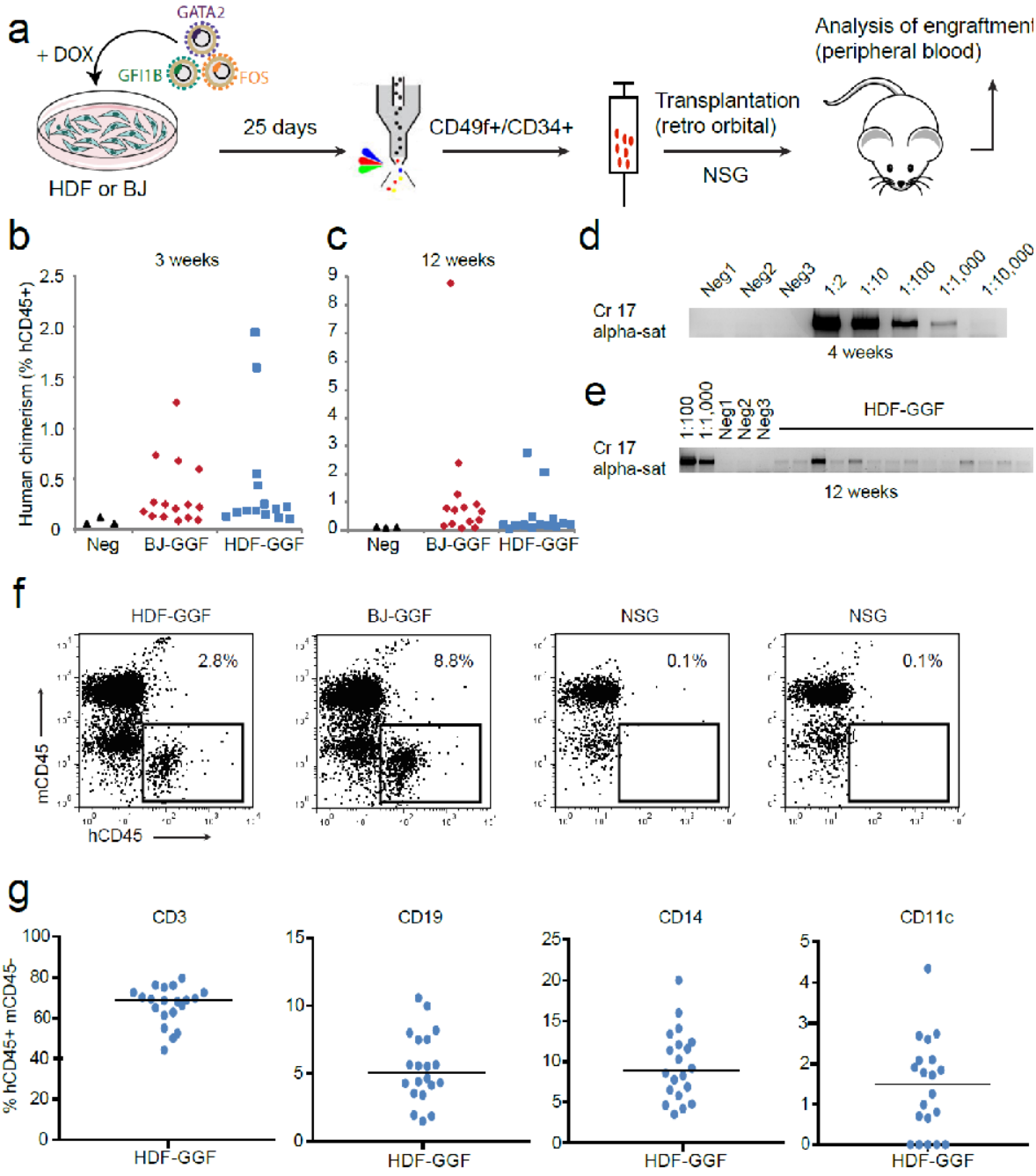

Figure 4.12 - **Single cell transcriptional profile during temporal reprogramming transitions**. (**a**) Non-supervised hierarchical clustering showing genome-wide gene-expression data from HDF-derived single cells UCB CD34$^{+}$ single cells. The number of single cells analyzed and phenotype are detailed. (**b**) Gene list enrichment analysis for the genes upregulated at day 2 when compared to non-transduced HDFs. Biocarta and Humancyc pathway enrichment analyses are shown.

## 4.5 **Induced human cells engraft *in vivo***

The finding that immunodeficient mice could be engrafted with human cells allowed the study of human HSCs in more detail. In order to determine the functional properties of reprogrammed HSPC we asked if they would engraft xenogeneic recipients. Both HDF and BJ cells were transduced with GGF and cultured for 25 days in the presence of Dox. $CD34^+$ and $CD49f^+$ cells were sorted and transplanted (Figure 4.13a). We detected human chimerism in peripheral blood (PB) at both 3 and 12 weeks after transplant (Figure 4.13b and c). We confirmed this result by PCR using primers to human chromosome 17-alpha-satellite sequences (Figure 4.13d and e). A clear population of human $CD45^+$ mouse $CD45^-$ was detected (Figure 4.13f) that contains both lymphoid and myeloid cells (Figure 4.13g). Indeed, the most abundant population detected was $CD3^+$ T-cells. These results demonstrate that inducible expression of GATA2, GFI1B and FOS can convert fibroblasts into human HSPCs.

*(figure on the next page)*

Figure 4.13 - **Reprogrammed cells engraft in vivo after transplantation.** (**a**) Experimental design used to sort and transplant induced cells into NOD-*scid IL2R*γ-*null* (NSG) mice. HDF or BJ fibroblasts were transduced with GGF and cultured with Dox for 25 days. Cells were dissociated and $CD49f^+$ (including $CD34^+CD49f^+$ double positive cells) were sorted and then injected into 4-week old NSG mice. (**b**) Human chimerism in peripheral blood 3 weeks after transplantation with HDF-GGF (n = 15) or BJ-GGF (n = 14). (**c**) % of human CD45 chimerism 12 weeks after transplantation. (**d**) Limit of detection of human engraftment was tested by PCR using human specific primer pairs (Chromosome 17 alpha satellite) from serial dilutions of human HDFs mixed with mouse cells. Mouse cells only were used as negative controls. (**e**) Human chimerism in peripheral blood was tested by PCR. Each lane represents one individual mouse 4 weeks after transplantation with HDF-GGF. Blood from non-transplanted mice were used as controls. (**f**) FACS plots showing individual mice analyzed with human CD45 (hCD45) and mouse CD45 (mCD45) antibodies. (**g**) Lineage marker analysis in gated $hCD45^+mCD45^-$ cells. Scatter plots show expression of human lymphoid and myeloid markers. Data from 20 mice.




## 4.6 **Discussion**

We show that ectopic expression of the TFs GATA2, GFI1B and FOS induce a hemogenic program in human fibroblasts. Induced cells exhibit endothelial and hematopoietic gene expression, HSC surface phenotypes and engraft NSG mice to produce multi-lineage progeny. Mirroring mouse hemogenic induction (Pereira et al., 2013) this transition is dynamic, with initial activation of angiogenic followed by hematopoietic gene signatures.

We show that upon induction with GATA2, GFI1B and FOS we readily detect many endothelial genes including *PPARG, VWF* and *FOXC2* that have defined angiogenic functions. The induction results in the generation of a large population of $CD49f^{+}ACE^{+}$ cells while only a more restricted population activates CD34. This is consistent with the observation that $ACE^{+}CD34^{-}$ cells emerge early during human hematopoietic development and localize beneath the dorsal aorta (Sinka et al., 2012). Our data support the hypothesis that these precursor cells later give rise to $ACE^{+}CD34^{+}$ cells contained in aortic clusters. Indeed, *ACE*, *ITGA6* and *CD34* are GATA2 direct targets during the initial stages of hemogenic reprogramming, providing a direct mechanistic link between human hemogenic precursor phenotype and GATA2.

To faithfully recapitulate HSCs specification major interest lies in defining direct human HSC precursors during ontogeny. Our data provide useful markers: CD49f is present on human LT-HSCs (Notta et al., 2011) and we show co-expression with ACE. The *VNN2* gene that encodes GPI-80 and marks fetal liver HSCs was not detected during reprogramming nor was it a target for GATA2 or GFI1B. This is consistent with the lack of expression of GPI-80 in ESC-derived hemogenic endothelium (Prashad et al., 2015). Whether this gene is activated at later stages of HSC maturation remains to be investigated. It will be interesting in future studies to determine whether $ACE^{+}$ cells present in the AGM (and perhaps in the mid-gestation placenta) co-express CD49f during human embryonic development before they acquire CD34 expression. In addition, our datasets provide other markers that are informative and complement this phenotype such as *F11R*, shown to be present in HSC precursors in zebrafish (Kobayashi et al., 2014). We also found *ANGPTL4* and *JAG1* enriched in $CD49f^{+}$ cells and then down regulated in $CD34^{+}$ cells. These genes may be useful to segregate precursors from emergent human HSCs *in vivo*.

In the mouse, lineage divergence from endothelium may occur before extensive formation of intra-aortic clusters (Rybtsov et al., 2014; Swiers et al., 2013). The enrichment of angiogenesis and cell motility during human hemogenic reprogramming suggests that human hemogenic precursors may not represent a cohort of mature endothelium but more likely a different lineage with endothelial features that is committed to the hematopoietic route. This is also supported by studies of human ESC-derived hematopoiesis (Ditadi et al., 2015). Glycolysis-associated genes were rapidly activated two days after induction and motif enrichment analysis identified HIF1 as a regulator of human hematopoietic reprogramming (Figure 4.12b). This suggests

that a metabolic shift towards glycolysis underlies hemogenic reprogramming and human stem cell formation that may be mediated by HIF and AP-1 as recently suggested in zebrafish (Harris et al., 2013; Zhang et al., 2017).

Several recent publications have induced mouse and human cells into hematopoietic progenitor cells (Batta et al., 2014; Lis et al., 2017; Pereira et al., 2013; Riddell et al., 2014; Sandler et al., 2014). Two recent studies showed the generation of long term reconstituting HSCs derived from mouse endothelial cells or human pluripotent stem cells (Lis et al., 2017; Sugimura et al., 2017). In both cases, reprogrammed cells were taken into either the bone marrow niche or co-cultured on endothelial cells to promote maturation. It will be interesting to address the impact of a supportive niche on our reprogrammed cells to mature these populations. In addition, combining GATA2, FOS and GFI1B specifying factors with factors that may promote HSC maturation should be investigated. A recent paper has demonstrated the generation of serially engraftable murine HSCs by Gata2, Gfi1b and Fos overexpression within teratoma (Tsukada et al., 2017). This suggests that GGF are the instructive factors of HSC identity not only from fibroblasts but also from other cell-types and highlight the importance of the *in vivo* environment as well as the need of suitable mouse models that support the maintenance of human HSCs (Cosgun et al., 2014). It will be of critical importance to develop more-defined culture methods for the controlled maturation of *in vitro*-programmed as well as human embryo-derived nascent HSCs into definitive, fully functional HSCs.

Collectively, our results show that GATA2, GFI1B and FOS are sufficient for the generation of hemogenic cells from human fibroblasts. Our reprogramming system provides a powerful tool to dissect the molecular mechanism of definitive HSC emergence and EHT. To address the molecular mechanisms, in the next chapter, we have performed Chromatin immunoprecipitation sequencing (Chip-Seq) assays at the initial stages of reprogramming and define the chromatin engagement of the three hemogenic TFs.

# Chapter 5 - Human hemogenic reprogramming is mediated by cooperative transcription factor induction

The regulation of biological processes such as gene expression, epigenetic silencing and chromosome stability often depends on the interactions between DNA and proteins. The initial method to identify DNA bound by a specific protein was described in bacteria in 1984 (Gilmour and Lis, 1984). However, it only expanded when large genome-scale methods were developed. Chromatin immunoprecipitation coupled with high throughput sequencing is now a powerful tool and it is widely used to determine how TF and other chromatin related proteins interact with DNA, identify regulatory sequences and epigenetic marks (Wardle and Tan, 2015). Recently, to fully understand the networks of transcriptional regulation other high throughput chromatin profiling assays such as DNaseI hypersensitive sites and DNA methylation based on bisulfite sequencing have been used to determine chromatin accessibility (Ambrosini et al., 2015).

Several TF have been identified to play a crucial role in the emergence of HSCs, however due to lack of understanding of the native mechanisms used by the embryo over the course of hematopoietic fate, the molecular cues remain largely unknown. As shown in the previous chapter, the hemogenic reprogramming system seems to mimic human developmental hemogenesis. In this chapter, taking advantage of this model, we aim to uncover the molecular mechanism of human hemogenic reprogramming by looking at TF-DNA interactions with genome-wide approaches. To determine the genomic occupancy by GATA2, FOS and GFI1B we have used ChIP-seq for the hemogenic core TF (GGF) and we have also determined their “native” target sites for each TF individually. As changes in gene expression also involves profound epigenetic modifications we took advantage of the public data available for histones marks and further integrated the data using bioinformatic tools.

## 5.1 **Generation and validation of TF expressing epitope tags**

ChIP-seq faces the same challenges as many other antibody-based techniques, whereof affinity and specificity in antigen recognition are critical. As the quality of the antibody affects the recognition of the protein of interest and due to the lack of ChIP-grade for GFI1B and GATA2, tagged versions were created with HA and 3xFLAG tag respectively (Figure 5.1a). Expression was confirmed in Western Blots as was confirmed by nuclear and subnuclear localization by immunofluorescence (Figure 5.1 b and c). As expected, GFI1B showed a similar binding pattern consistent with Vassen and colleagues description at the foci of pericentric chromatin (Vassen et al., 2006) (Figure 5.1d).

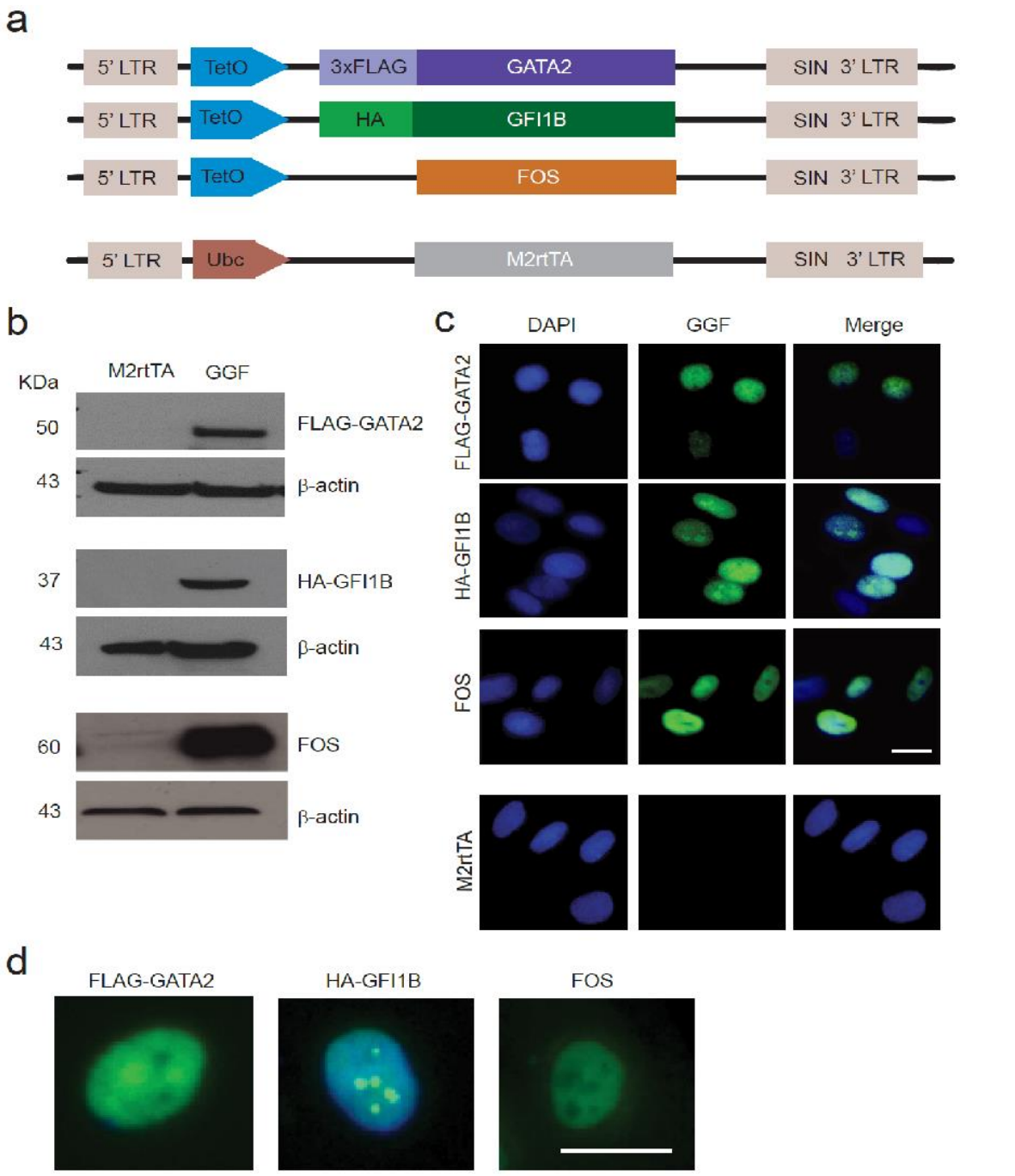



Figure 5.1 - **Strategy for detection and immunoprecipitation of GATA2, FOS and GFI1B in fibroblasts.** (**a**) Schematic representation of the lentiviral vector strategy used to transduce HDFs. (**b**) The specific detection of TFs was validated by western blotting using antibodies against FOS, FLAG or HA tags 48 hours after transducing with GATA2, GFI1B and FOS (GGF). (**c**) Transduced HDFs were assayed by immunofluorescence using antibodies against FOS, FLAG or HA tags where appropriate. Merged pictures show that FOS, FLAG-GATA2 or HA-GFI1B (Green) localize in the nucleus (Dapi, Blue) as expected. (**d**) Shows the sub-nuclear localization of the three TFs and highlight the binding pattern of GFI1B. Scale bars, 20 µm.

## 5.2 GATA2 displays dominant and independent targeting capacity to initiate hemogenic reprogramming

To map where the TFs initially bind to the genome, we performed ChIP-seq 48 hours after transduction with GGF expressed in combination and individually (Figure 5.2a). Out of the three hemogenic TFs, GATA2 showed more extensive binding to the fibroblast genome (6759 peaks), followed by GFI1B (2372 peaks) and FOS (689 peaks) (Figure 5.2b). GATA2 also displayed enrichment at promoters when compared to GFI1B and FOS suggesting an important regulatory role (Figure 5.2b, c and d).

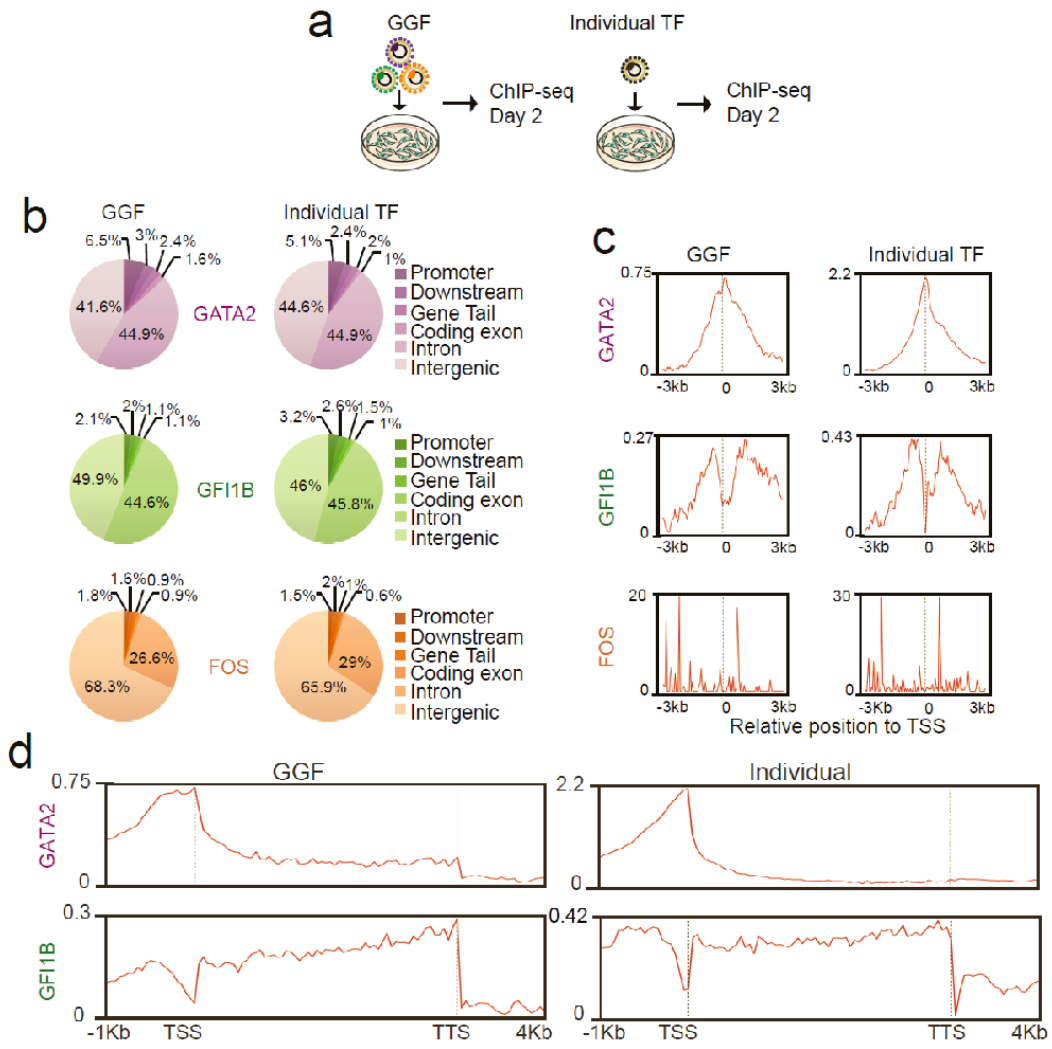



Figure 5.2 - **Genomic distribution of TF occupancy.** (**a**) Strategy for identifying GGF genomic binding sites. GGF factors were transduced in combination (left) or individually (right) and analyzed by ChIP-seq 2 days after adding Dox. (**b**) Genomic distribution of GGF peaks in transduced HDFs. (**c**) ChIP-seq read density relative to Transcription Start Site (TSS). The graph shows a plot of the average read coverage per million mapped reads centered to the TSS. (**d**) ChIP-seq signal distribution for GATA2 and GFI1B when GATA2, GFI1B and FOS (GGF) are expressed in combination (upper panels) or individually (lower panels). The plots display the average signal 1kb upstream the transcriptional start site (TSS), along gene bodies

and 4kb downstream the transcription termination site (TTS). Genic regions are represented as a 3kb long metagene.

Interestingly, GATA2 targets did not undergo major changes when the three TFs were expressed in combination or individually (Figure 5.3a). This result suggests that GATA2 displays independent targeting capacity as it binds to a similar set of targets with or without the presence of GFI1B and FOS (Figure 5.3a). In contrast, GFI1B depends on GATA2 to bind to most of its targets (Figure 5.3a). FOS showed a small number of targets suggesting that this TF does not have direct access to the chromatin during the initial stages of reprogramming (Figure 5.3a). GATA2 targets include the CD34 gene, the HSC homing molecule CD9 (Karlsson et al., 2013) and the EHT mediator GPR56 (Kartalaei et al., 2015) (Figure 5.3b).

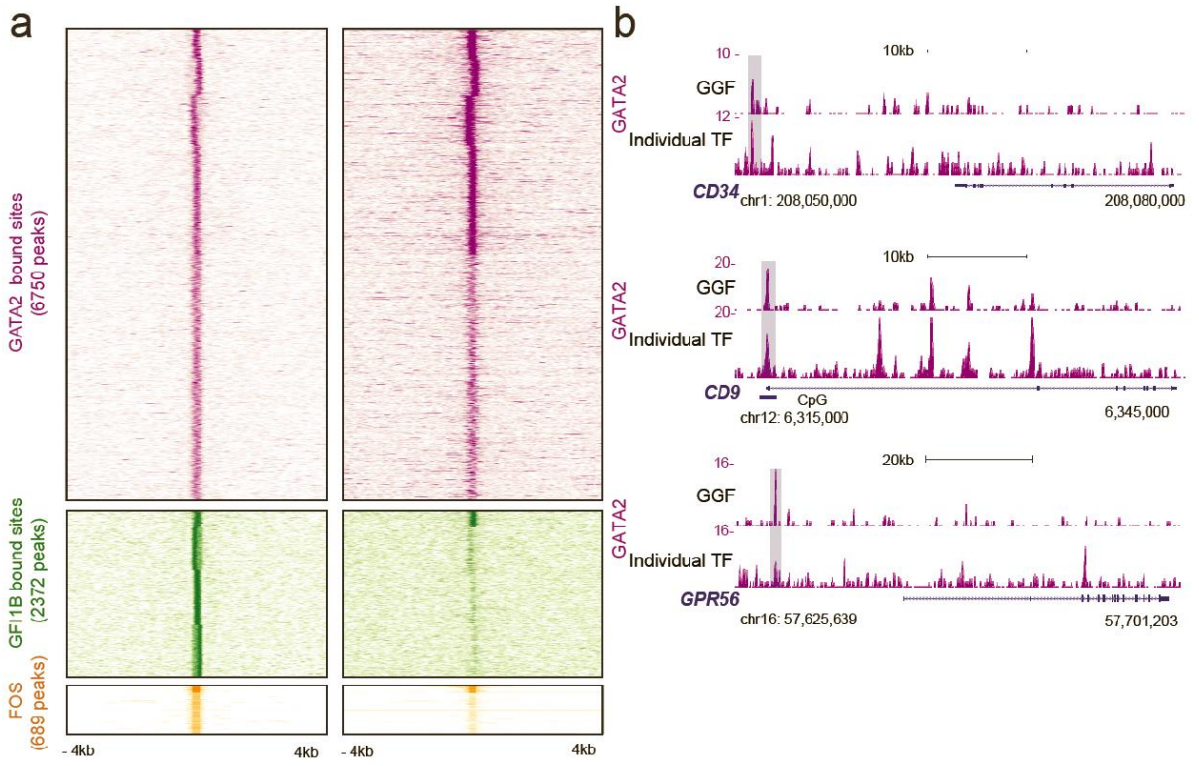
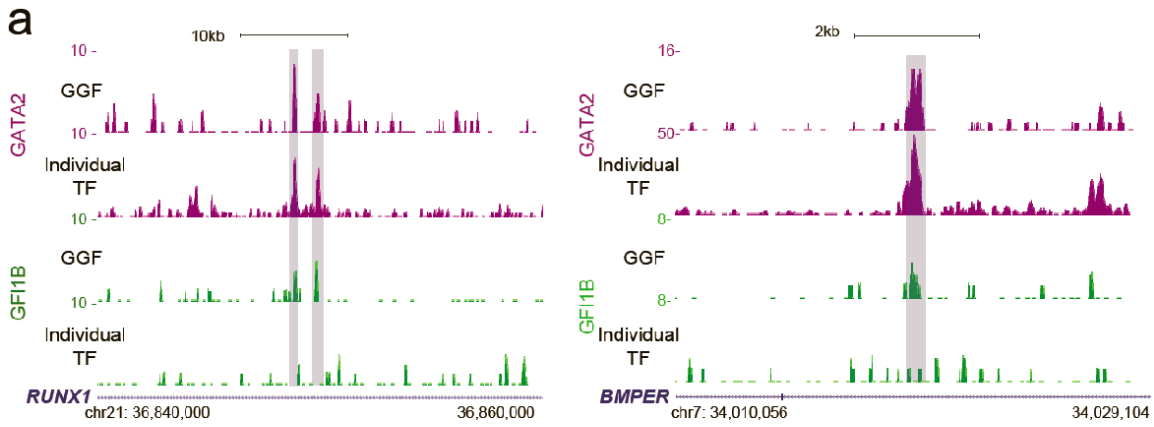



Figure 5.3 - **Analysis of TF occupancy reveals that GATA2 has both dominant and independent targeting capacity.** (**a**) Heatmaps representing genome-wide occupancy profile for GGF factors when expressed in combination or individually in HDFs. For each site the signal is displayed within 8kb window centered on individual peaks. (**b**) Genome browser profiles illustrating GATA2 binding sites at *CD34*, *CD9 and GPR56 loci*. The y-axis represents the total number of mapped reads. The boxes highlight shared peaks when GATA2 is expressed individually or in combination with GFI1B and FOS.

RUNX1 and BMPER (McGarvey et al., 2017) (Figure 5.4a), two important regulators of hematopoiesis, are targeted by both GATA2 and GFI1B in GGF induced fibroblasts. However, GFI1B binding was lost when GATA2 was not included, suggesting cooperative binding between GATA2 and GFI1B for the induction of RUNX1 and definitive hematopoiesis. Motif prediction for GATA2 in GGF induced fibroblasts showed that GATA motif was strongly enriched. This analysis also identified HIF1B, BORIS and ARID3A as regulators of hematopoietic reprogramming (Figure 5.4b, upper panel). For GFI1B targets, AP-1, HIF1B and GATA motifs were identified but the GFI motif was only strongly enriched when GFI1B was expressed individually (Figure 5.4b, lower panel). This suggests that GATA2 recruits GFI1B to its "natural" target sites. Overall these results provide insights as to how these factors engage the fibroblast genome and initiate hemogenic programming by targeting key hematopoietic regulators.




b

GGF

| Name | p-value | motif |
|---|---|---|
| **GATA** | 1e-928 | |
| HIF1b | 1e-771 | |
| BORIS | 1e-108 | |
| Arid3a | 1e-56 | |

Individual TF

| Name | p-value | motif |
|---|---|---|
| **GATA** | 1e-4458 | |
| BATF | 1e-4731 | |
| CEBP | 1e-681 | |
| TEAD | 1e-623 | |

GATA2

| Name | p-value | motif |
|---|---|---|
| **AP-1** | 1e-355 | |
| HIF1b | 1e-215 | |
| Jun **AP-1** | 1e-42 | |
| **GATA** | 1e-29 | |

| Name | p-value | motif |
|---|---|---|
| **GFI** | 1e-115 | |
| Jun **AP-1** | 1e-44 | |
| CEBPa | 1e-16 | |
| RUNX | 1e-13 | |

GFI1B

Figure 5.4 - **GATA2 binds to similar sites independently of the other TFs.** (**a**) GATA2 and GFI1B occupancy profile at the *RUNX1 locus.* The boxes highlight shared peaks between GATA2 and GFI1B only when expressed in combination. The genomic scale is in kilobases (kb). (**b**) *De novo* motif prediction for GATA2 and GFI1B target sites when expressed in combination or individually. The motifs for GATA, GFI and AP-1 factors are bolded.

To gain insight of the target genes contributions that are involved in the biological processes and pathways, we analyzed the gene ontology using DAVID GO analysis (Figure 5.5). We found that among the most enriched pathways were terms related with angiogenesis, vascular signaling and inflammation whereas the ontology for the biological processes highlighted cell motion, angiogenesis and blood vessel morphogenesis and development. The observation that many terms overlapped for GATA2 and for GFI1B lead us to hypothesized if those common terms could have a genome wide co-bound pattern.

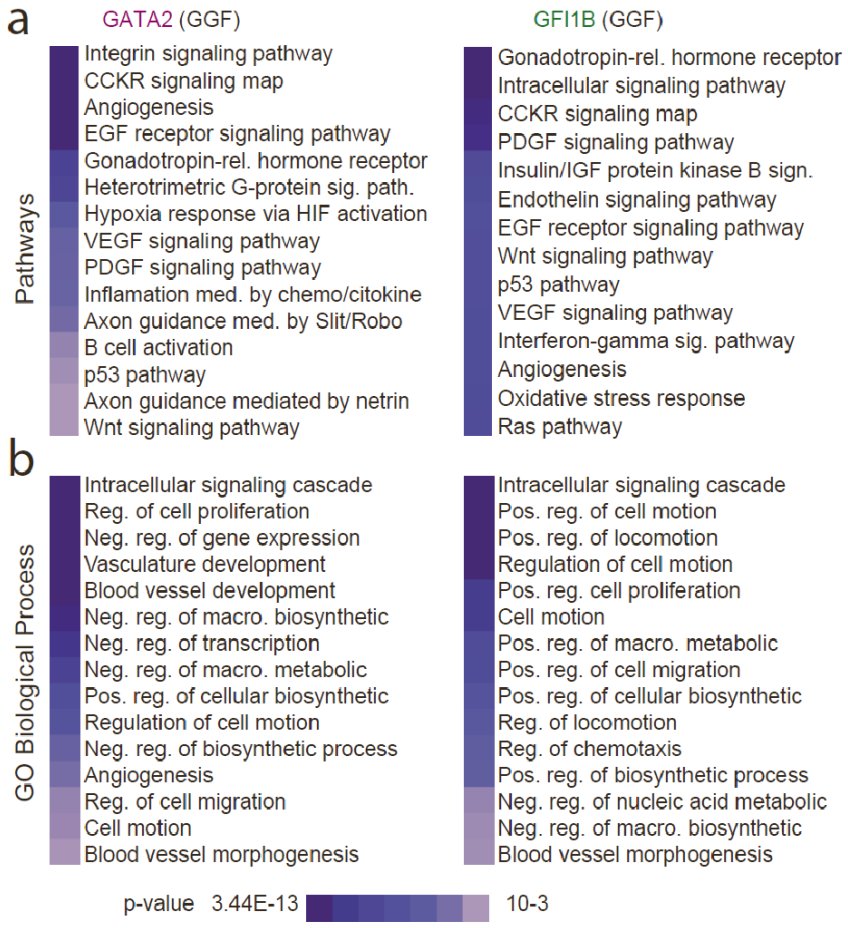



Figure 5.5 - **Gene ontology analysis of HDF transduced with GGF.** (**a**) Panther pathway enrichment analysis of genes bound by GATA2 (left panel) and GFI1B (right panel) in HDF transduced with GGF. (**b**) Biological processes GO terms enriched for GATA2 (left panel) and GFI1B (right panel) bound genes in HDF transduced with GGF.

## 5.3 **GATA2 and GFI1B interact and share a cohort of target sites**

We next investigated the extent of overlap between GATA2 and GFI1B genomic targets and whether they physically interact. By displaying GATA2 and GFI1B target sites we observed that 750 genomic positions were shared, representing 31.6% of total GFI1B targets (Figure 5.6a and appendix II). These include HSC and EHT regulators such as PRDM1 (Wang et al., 2008) and PODXL (Zhang et al., 2014) (Figure 5.6b). Motif comparison analysis showed significant similarity between GATA2 and GFI1B motifs (Jaccard similarity index = 0.1) (Vorontsov et al., 2013), supporting the interaction between the two TFs (Figure 5.6c). We then performed de novo motif prediction for the overlapping peaks. Interestingly, AP-1 motif was the most enriched followed by the GATA and GFI1 motifs, highlighting the cooperative action between GATA2, GFI1B and FOS during human hemogenic reprogramming (Figure 5.6d). Co-bound genes are part of pathways such as interferon-gamma signaling, inflammation and cytoskeletal regulation by Rho GTPases (Figure 5.6e), processes with demonstrated relevance for HSC emergence (Pereira et al., 2013, 2016). Gene ontology analysis of co-bound genes showed that cell motion and vasculature development were enriched terms (Figure 5.6f).

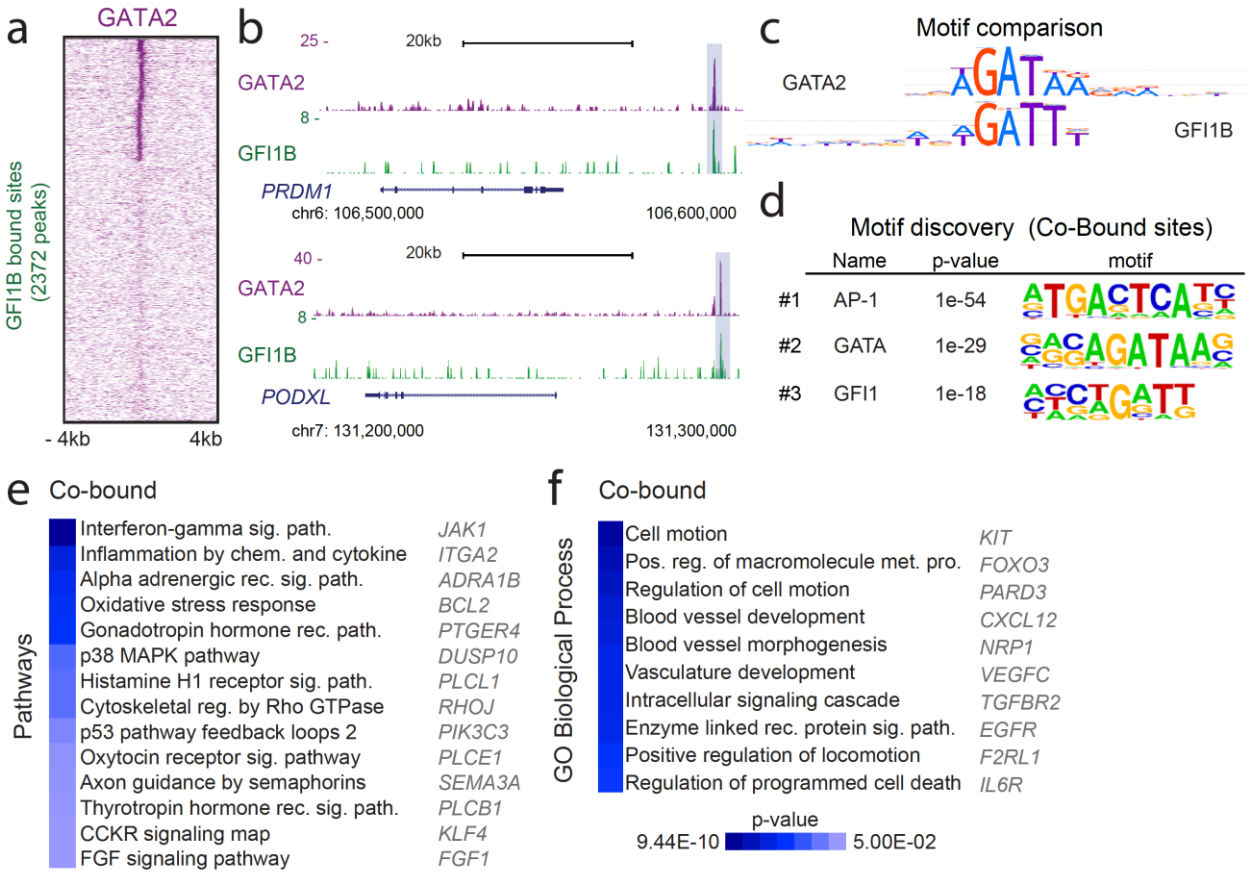



Figure 5.6 - **GATA2 and GFI1B interact and share a cohort of target sites.** (**a**) Heatmap representing genome-wide occupancy profiles GFI1B showing shared targets with GATA2 48h after induction of GGF with Dox. The signal at the corresponding genomic regions of TF binding is displayed across the other data set. For each site the signal is displayed within an 8kb window centered on individual peaks. (**b**) Genome browser profiles illustrating GATA2 and GFI1B binding sites at *PRDM1* and *PODXL loci*. The y-axis represents the total number of mapped reads. The light blue boxes highlight genomic positions co-occupied by GATA2 and GFI1B. (**c**) Motif comparison between GATA2 and GFI1B. Jaccard similarity coefficient = 0.1. (**d**) *De novo* motif discovery at co-bound sites. Top three most enriched motifs are shown along with p-values. (**e**) Panther pathway enrichment analysis of genes co-bound by GATA2 and GFI1B. Examples of co-bound genes are shown and heatmaps display p-values. (**f**) Biological processes GO terms enriched for co-bound genes. Examples of co-bound genes are shown and heatmaps display p-values.

We further interrogated our ChIP-seq data for the regulatory interactions between the three hemogenic TFs. Both GATA2 and GFI1B bind to their own locus suggesting auto- regulation as reported by May et al., 2013 and Anguita et al., 2010. In addition, GATA2 bind to a CpG island in the FOS locus and GFI1B binds to the GATA2 locus (Figure 5.7a). To confirm physical interaction, we have performed Co-Immunoprecipitation 48 hours after expression in fibroblasts. This analysis demonstrated an interaction between GATA2 and FOS and between GATA2 and

GFI1B (Figure 5.7b). This suggests that the interplay between GATA2, FOS and GFI1B is central for hemogenic reprogramming (Figure 5.7c).

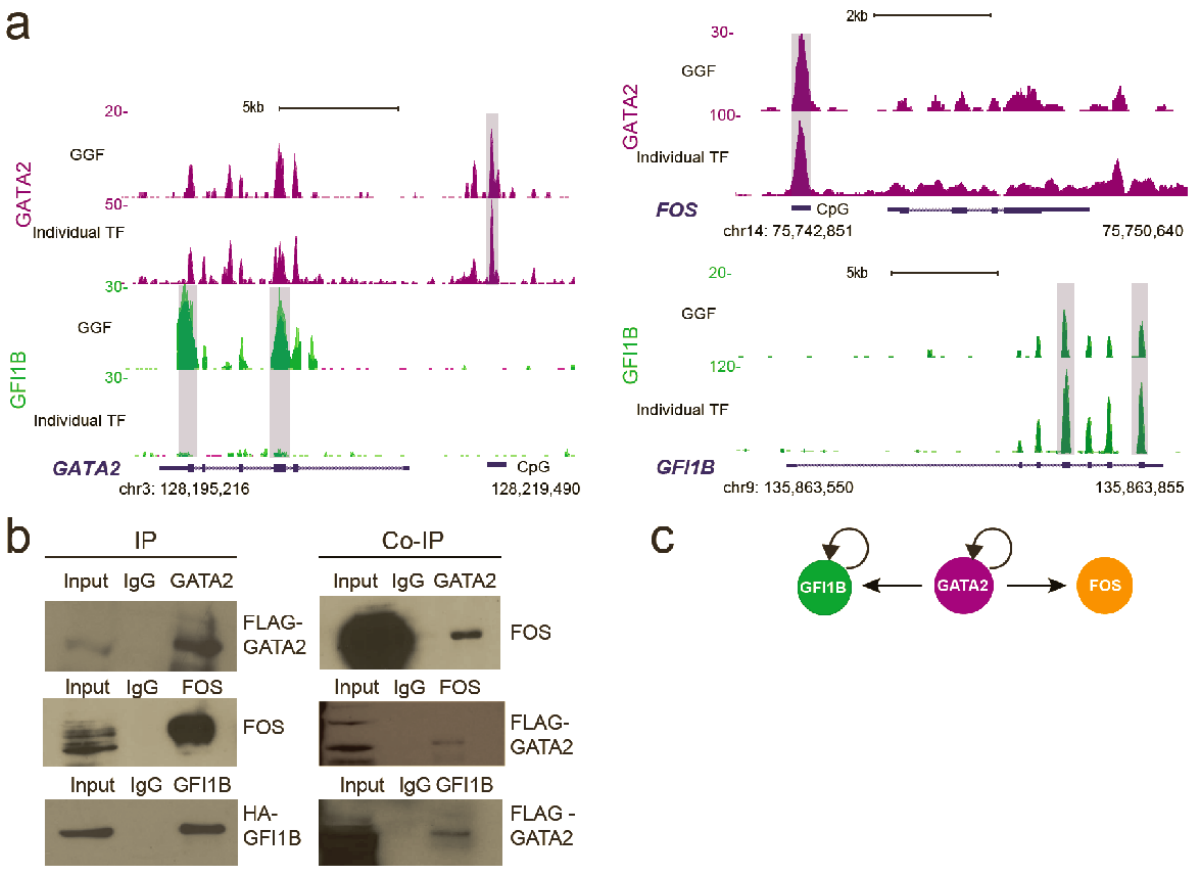



Figure 5.7 - **Interactions between the hemogenic TFs.** (**a**) GATA2 and GFI1B occupancy profile at the *GATA2*, *FOS* and *GFI1B* locus. The y-axis represents the total number of mapped reads. The light blue boxes highlight genomic positions occupied by GATA2 and GFI1B. (**b**) Immunoblots showing Immunoprecipitation (IP) of GATA2 (upper), FOS (middle) and GFI1B (lower) in HDFs 48h after induction of GGF (left panel). The right panel shows Co-IP detection for FOS (upper) and GATA2 (middle and lower). (**c**) The scheme summarizes the regulatory interactions between the three TFs. Both GATA2 and GFI1B are regulated by auto loop mechanisms and GATA2 binds to and physically interact with GFI1B and FOS.

## 5.4 **GATA2 and GFI1B engage open promoters and enhancers regions**

We next studied downstream mechanisms by correlating GATA2 and GFI1B binding sites with transcriptional changes (Figure 5.8a). We found that most GATA2 target genes are either upregulated, in agreement to its role as a transcriptional activator, or downregulated. We identified 1425 differentially expressed genes, which were bound by either GATA2 or GFI1B. Moreover 1186 genes were bound only by GATA2 and 182 were bound only by GFI1B. The 57 co-bound genes are targeting the cluster highly expressed in fibroblasts and a second sub group of genes that are upregulated

in induced CD49f$^{+}$CD34$^{+}$ cells (Figure 5.8a, Fisher t-test, $p< 10^{-10}$). To characterize the chromatin features associated with GATA2 and GFI1B engagement we used previously published ChIP-seq data sets for H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K9me3 and H3K36me3 in HDFs (Table 2.4). GATA2 and GFI1B bound sites in fibroblasts are enriched for marks associated with active promoters and enhancers such as H3K4me3, H3K27ac and H3K4me1 (Figure 5.8b). This result is consistent with the DNase I pattern in human dermal fibroblasts. GATA2 and GFI1B bind mostly to DNase I sensitive sites (Figure 5.8c). These results demonstrate that GATA2 and GFI1B preferentially bind to accessible chromatin mostly in promoters and enhancers regions. We further summarized the association between GATA2 and GFI1B binding and chromatin in fibroblasts using ChromHMM (Ernst and Kellis, 2012), a segmentation of the genome into 18 chromatin states based on the combinatorial patterns of chromatin marks. We confirmed the preference of GATA2 and GFI1B in active Transcriptional Start Sites (TSS), flanking upstream TSS and active enhancers (Figure 5.8d, blue). In addition, we have analyzed published data sets for histones marks in K562 cells and GATA2, GFI1B, and FOS ChIP-seq data in Hematopoietic Progenitor Cells (HPCs) (Table 2.4).In contrast to GATA2 and FOS, we observed a distinct pattern for GFI1B that is strongly enriched in bivalent/poised TSS (Figure 5.8d, orange). This dramatic shift in GFI1B targeting suggests that the cooperative interaction between GATA2 and GFI1B may be specific for the earlier stages of hematopoietic reprogramming and EHT that is lost in downstream hematopoietic progenitors.

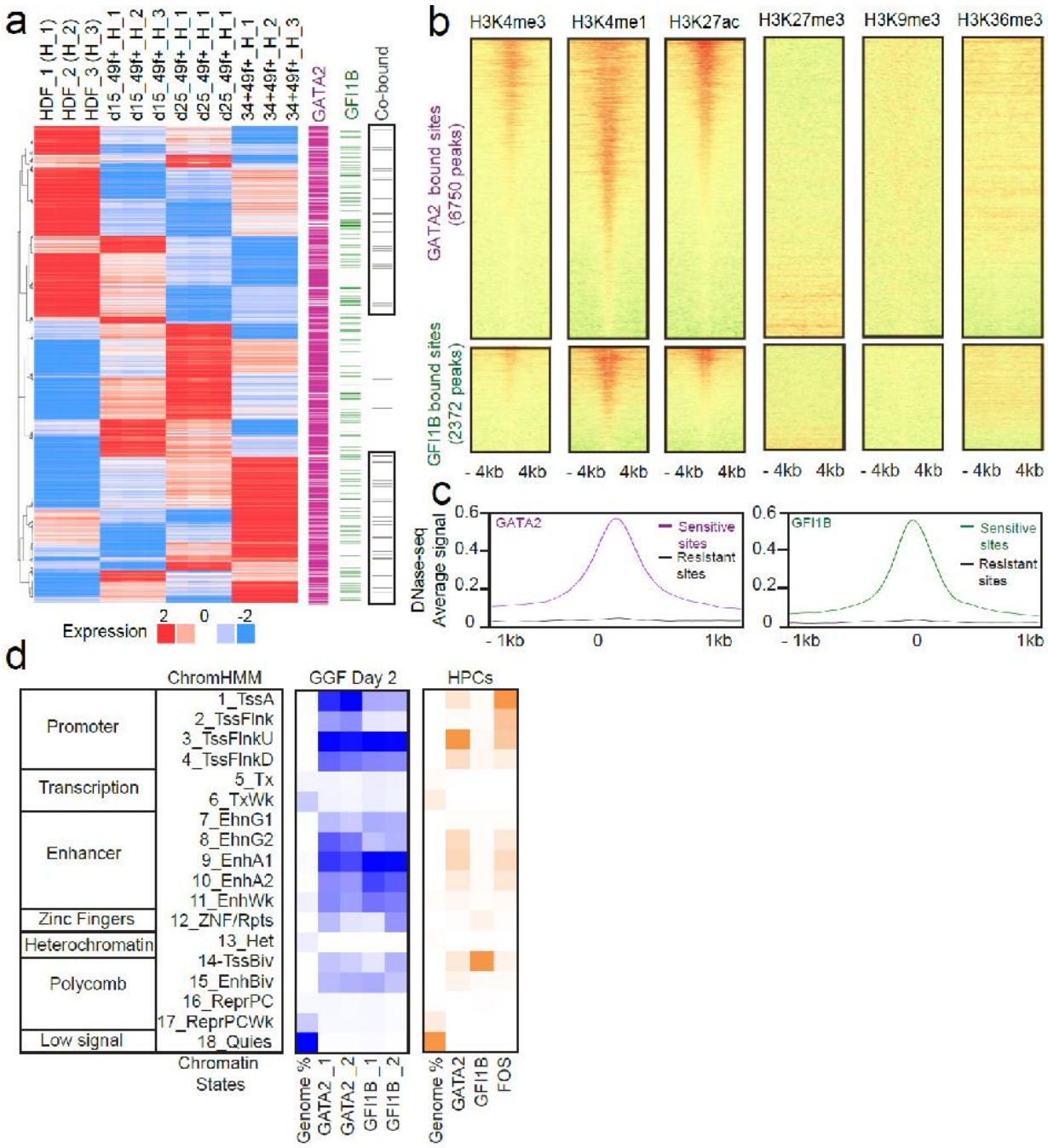



Figure 5.8 - **GATA2 and GFI1B engage open promoters and enhancers regions.** (**a**) Integration of RNA-seq and ChIP-seq datasets. The heatmap (left panel) shows1,425 genes with at least 1.5 fold expression changes across the dataset. Genes were identified as targets of GATA2 (purple), GFI1B (green) with co-bound targets in grey (right panels) based on binding peaks within the range of a 10kb window centered on the transcriptional start site. (**b**) Heatmaps of normalized tag densities representing HDF chromatin marks at GATA2 and GFI1B targets sites. The signal is displayed within a 8kb window centered on the binding sites. (**c**) Average DNase-seq signal of HDFs at GATA2 and GFI1B target sites. The signal is displayed within a 2kb window. (**d**) Heatmaps for chromatin-state functional enrichment. Rows represent chromatin states according to ChromHMM annotation. TssA (active promoters), TssFlnk (flanking promoters), TssFlnkU (flanking upstream promoters), TssFlnkD (flanking downstream promoters), Tx (strong transcription), TxWk (weak transcription), EnhG1/2 (Genic enhancers), EnhA1/2 (active enhances), EnhWk (weak enhancers), ZNF/Rpts (ZNF genes

& repeats), Het (Heterochromatin), TssBiv (Bivalent/Poised TSS), EnhBiv (Bivalent Enhancer), ReprPC (Repressed PolyComb), ReprPCWk (Weak Repressed PolyComb) and Quies (Quiescent). Blue panel shows the percentage of genome occupancy for GATA2 and GFI1B in GGF-transduced HDFs. Orange panel shows the percentage of genome occupancy for GATA2, GFI1B and FOS in human hematopoietic progenitors (HPCs).

## 5.5 **Discussion**

In chapter 4, we show data to support the hypothesis that $ACE^{+}CD34^{-}$ precursor cells later give rise to $ACE^{+}CD34^{+}$ cells contained in aortic clusters. In this chapter we found that *ACE*, *ITGA6* and *CD34* are GATA2 direct targets during the initial stages of hemogenic reprogramming, providing a direct mechanistic link between human hemogenic precursor phenotype and GATA2.

In addition to providing information for the identification of direct precursors of HSCs during development, our direct reprogramming strategy has identified pathways implicated in EHT and HSC specification. We have shown that, in both mouse and human, GATA2, FOS and GFI1B represent a conserved minimal TF network for hemogenic induction. In the mouse, we identified interferon signaling and inflammation as enriched pathways in hemogenic precursors (Pereira et al., 2013). Recently, multiple studies explored inflammation during EHT and HSC formation in mouse and zebrafish models (Espin-Palazon et al., 2017). Enrichment of this pathway during human hemogenic induction supports its importance for human EHT. Another example is the role of FOS and AP-1 in HSC specification. This was not revealed by classical genetic ablation, possibly due to compensatory effects. AP-1 is essential for human and mouse hemogenic induction (Lis et al., 2017; Pereira et al., 2013; Sandler et al., 2014) and has recently been implicated during EHT from mouse ESCs (Obier et al., 2016). Our studies corroborate that direct cellular reprogramming informs the process of HSC specification and offers tractable means to systematically interrogate pathways of human HSC formation. It will be interesting to address the role of other enriched pathways such as Cytoskeletal regulation by Rho GTPase and the Histamine H1 receptor pathway. Genes that are co-targeted during hematopoietic reprogramming by GATA2 and GFI1B either individually or in combination are additional candidates for further study e.g., *ITGA2*, *PLCL1*, *ADRA1B*, *RHOJ*, *SEMA3A*, and *PLCB1*.

Our findings show that GATA2 is the dominant TF and cooperates with GFI1B to engage open chromatin regions during the initial phases of reprogramming. In contrast to iPSC reprogramming, the mechanisms underlying direct hematopoietic reprogramming remain poorly understood. We have explored the regulatory mechanisms underlying hemogenic induction with Chip-sequencing at the initial stages of reprogramming. Two different models of action have been proposed for how TFs access chromatin to initiate reprogramming: 1) pioneer TFs, as a unique class of transcriptional regulators with the capacity to bind nucleosomal DNA (closed chromatin) and activate gene regulatory networks in target cells (Soufi et al., 2012; Zaret and Carroll, 2011) and 2) cooperative interaction among multiple TFs (Chronis et al., 2017) . In this recent study, iPSC reprogramming factors bind cooperatively to enhancer regions to direct somatic inactivation and pluripotent gene expression initiation (Chronis et al., 2017). Our data supports a model (Figure 5.9) whereby GATA2, GFI1B and FOS cooperate to silence fibroblast specific program and gradually impose the hemogenic program. GATA2 and GFI1B binding occurs at open chromatin, promoters, and enhancer regions supporting the cooperative binding model. GATA2 showed however independent targeting capacity and was crucial for the recruitment of GFI1B to target sites. Our finding mirrors the role of Gata2 during hematopoietic progenitor differentiation by recruiting Gata1 to erythroid genes (May et al., 2013). This suggests that GATA2 acts as a “pioneer” or a nucleation factor to initiate hematopoietic reprogramming at open chromatin regions. During mouse ESC differentiation, it was shown that AP-1 motifs were enriched in open chromatin regions and co-localized with TF-binding sites that were specific to HE cells (Goode et al., 2016). In human endothelial cells, AP-1 cooperates with GATA2 to induce key endothelial and inflammatory genes (Kawana et al., 1995; Linnemann et al., 2011). In contrast, Gfi1b is not a part of the heptad of TFs in mouse hematopoietic progenitors (Wilson et al., 2010). Indeed, we have confirmed that in human hematopoietic progenitors GFI1B has a very different binding pattern from GATA2 and FOS. We propose that GATA2 and GFI1B interaction is specific during hemogenic reprogramming and HSC specification. Recent Chip-seq data from mouse ESC-derived HE supports a similarity between Gata2 and Gfi1b target sites (Goode et al., 2016). Taken together, these data highlight the cooperative action between GATA2, FOS and GFI1B during human hematopoietic reprogramming, initially binding at

fibroblast-specific genes and then activating endothelial and hematopoietic gene signatures.

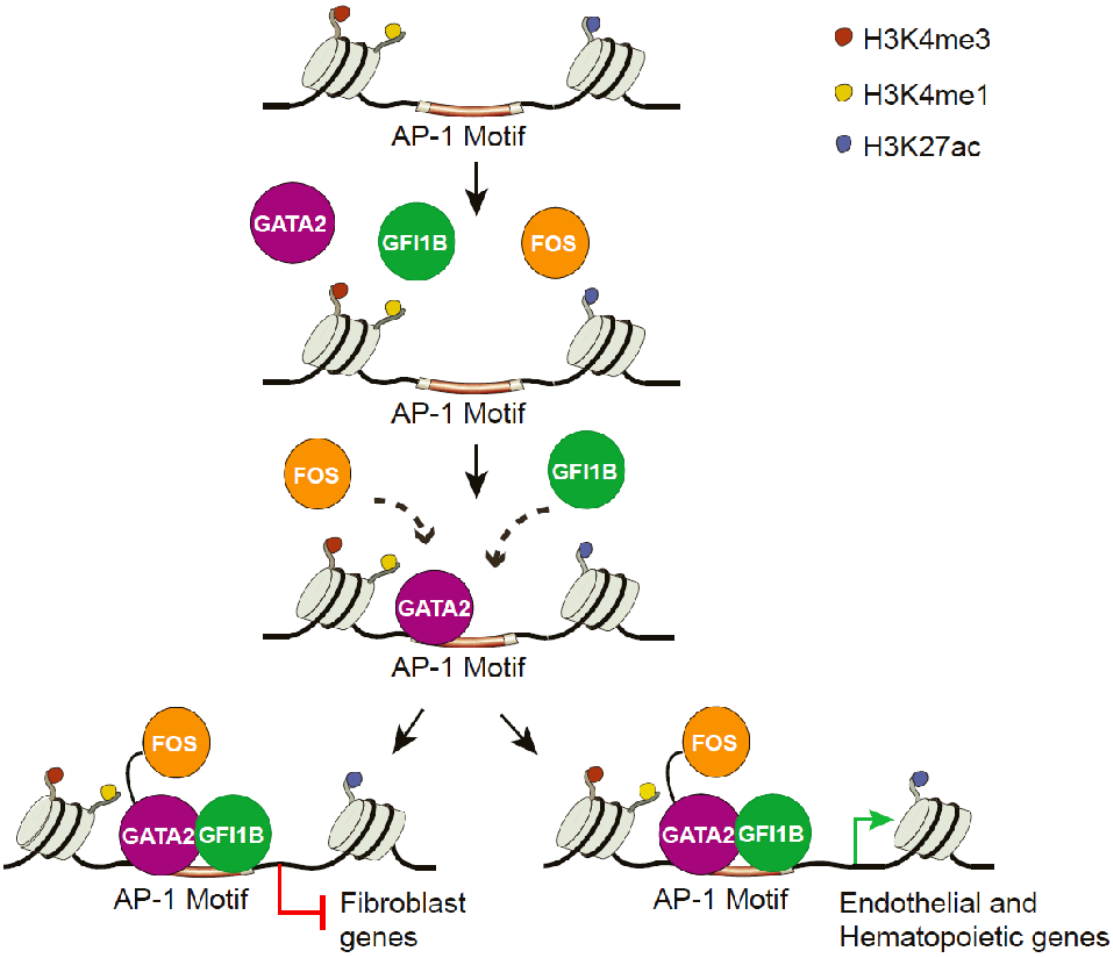



Figure 5.9 - **Model of the molecular mechanisms underlying the induction of hemogenesis in fibroblasts.** In induced cells, GATA2 binds open chromatin regions marked with H3K4me1, H3K4me3 and H3K27ac and recruits GFI1B and FOS into the AP-1 motifs to induce cell fate. Formation of the cooperative complex alters gene expression by silencing specific fibroblast genes and by activating endothelial and hematopoietic genes.

Collectively, our results suggest that HSC specification is controlled by the cooperative action of TFs and underscore the importance of GATA2-GFI1B interaction and initial engagement at open chromatin regions. In summary, we demonstrate that direct cellular reprogramming provides insights into the molecular mechanisms of human HSC specification. These studies also provide a platform to further the development of patient-specific HSPCs from easily accessible human dermal fibroblasts.

# Chapter 6 – General discussion and future directions

Establishing alternatives strategies to produce HSCs *in vitro* are needed. Over the last few years, significant attention has been given to the emergence of hematopoietic stem cells *in vivo* as a model to understand stem cell formation. Relevant to this work is the identification and isolation of a specific phenotype of hemogenic precursors in mid-gestation placentas from an *in vitro* system that can be matured to transplantable hematopoietic stem cells. Human fibroblasts can be reprogrammed into hemogenic cells by the same transcription factors previously reported. In addition, this thesis has explored the molecular mechanisms of the initial phases of reprogramming in order to develop an understanding of human endothelial-to-hematopoietic transition.

## 6.1 **– Modeling hematopoiesis through cell fate reprogramming**

Several studies attempted to generate a hematopoietic stem cell from multiple sources via reprogramming such as embryonic stem cells (Kyba et al., 2002), induced pluripotent stem cells (Doulatov et al., 2013; Elcheva et al., 2014), fibroblasts (Batta et al., 2014; Pereira et al., 2013), lineage committed blood progenitors (Riddell et al., 2014) and endothelial cells (Sandler et al., 2014). However, a common question for direct or pluripotent reprogramming is to which extent are the converted cells truly similar to their *in vivo* counterparts. Using our direct hematopoietic reprogramming as a blue print of hemogenesis, in this thesis we identified the same phenotype previously reported *in vitro* (Pereira et al., 2013) and characterized the specific population expressing Sca1, Prom1, and CD34 (PS34) as hemogenic precursors. When analyzing the expression profile of PS34, we observed that specific markers for endothelial program (CD31, CD105, VeCad and Tie2) are detectable as early as E10.5. Along with those markers, we also display the gene expression of important transcription factors in EHT like *Sox17*, *Prom1*, *Podxl*, and *Kitl* that are highly expressed in the most immature PS34 population at E10.5 (Figure 3.6a). By PCA analysis and single cell pseudo-time reconstitution, it is possible to follow the developmental trajectory of PS34 cells. Therefore, this might indicate that PS34CD45$^{-}$ cells have already acquired hemogenic potential prior the EHT or during mesoderm induction. Recent studies reinforced this hypothesis by differentiating human pluripotent stem cells into CD34$^{+}$ HSCs-derived cells. They found that by inhibiting the activin-nodal signaling (Kennedy et al., 2012), by activating the WNT signaling

(Sturgeon et al., 2014) or by modulating both signaling pathways (Ng et al., 2016) at early days, it promotes the formation of different hematopoietic lineages with T lymphoid potential as a read out (Kennedy et al., 2012; Sturgeon et al., 2014) thus suggesting that the alteration of specific factors can modify the fate of hematopoietic cells at early stages of mesoderm development. Another study revealed that by inhibiting NOTCH signaling with GSI, in the first 3 days of culture coinciding with the beginning of EHT, reduced the $CD45^{+}$ population (Ditadi et al., 2015). Moreover, using a transgenic mouse model carrying the +23 *Runx1* enhancer, Swiers and colleagues identified at the clonal level that specification and determination of hemogenic endothelium cells to downstream hematopoietic lineages occurs at E9.5, up to 2 days earlier than expected (Swiers et al., 2013).

Similarly to the pluripotent reprogramming model, we also show that different cell types and cells from different species can be converted into hemogenic cells, in spite of the higher resistance to transformation by human cells (Hahn and Weinberg, 2002), demonstrating that this minimal network is evolutionarily conserved. In this study, we used the human ortholog versions of the already described TF combination for mouse fibroblasts – GATA2, GFI1B and FOS (Pereira et al., 2013) to efficiently convert human fibroblasts and by clustering gene expression of mouse and human datasets we find a robust transcriptional signature for hemogenic precursors cells thus suggesting a conserved process across evolutionary-diverged species (Figure 4.11c). Moreover, an independent study recently demonstrated the validity of our model by using the same TF combination to generate long-term HSCs from mouse induced pluripotent stem cells *in vivo* within teratomas (Tsukada et al., 2017). Conversely, switching across different species is not a straightforward event and in iPSCs, by comparing both human and mouse fibroblasts from previous studies, the authors revealed similar binding pattern of OSKM but divergent regulatory networks (Fu et al., 2018). Studies in cardiac reprogramming have failed to convert human fibroblasts with the TF combination of Gata4, Mef2 and Tbx5 described in mouse fibroblasts towards a cardiac fate (Fu et al., 2013; Nam et al., 2013).

Overall, the use of direct conversion provided a straightforward and an advantageous platform to isolate hemogenic precursors, to track their developmental path and to uncover molecular mechanisms while it alleviated technical issues like the scarcity of hemogenic cells and ethical issues relatively to human cells. In contrast to iPSCs, it

also eases the risk of tumorigenicity. The data presented in this thesis demonstrates that direct lineage conversion offers a versatile and powerful tool to shorten the gap between *in vivo* and *in vitro* HSC emergence by mimicking a complex process in a step wise manner.

Along with placenta, many other embryonic tissues have been documented with hemogenic potential such as the AGM region, the yolk sac and the endocardium (Nakano et al., 2013; Nishikawa et al., 1998; Rhodes et al., 2008; Swiers et al., 2013; Zovein et al., 2008). A recent study has suggested the embryonic head vasculature as a new site of *de novo* generation of HSCs (Li et al., 2012). Nevertheless, this location remains controversial (Iizuka et al., 2016), it would be of tremendous interest to investigate the presence of PS34 phenotype in other mouse embryonic tissues and in human placenta.

### 6.2 **– Functional maturation of induced cells**

In this thesis we have used clonogenic *in vitro* assays as a first measurement of stem and progenitor cell activity followed by transplantation of defined cell populations into immunocompromised mice. In chapter 3, we investigated how different stromal cells can affect the functional phenotype of PS34CD45$^{-}$ cells. Our data demonstrate that the clonogenic capacity is acquired when cells are co-cultured with OP9 cells and supplemented with specific cytokines. We also show that the engraftment made by PS34CD45$^{-}$ population was only achieved when the cells were co-cultured with a stromal cell line expressing the Notch ligand Delta 1 – OP9-DL1. This activation indicates an important function of Notch signaling for an efficient EHT and shows that PS34CD45$^{-}$ are upstream in the developmental path and rely on intimate interactions with other cells present in the microenvironment to complete maturation. Progress has been made in identifying and characterizing the *in vitro* conditions for long-term expansion and maintenance of pluripotent stem cells but the same has not been extensively done for HSCs. The contribution of a supportive environment *in vivo* has been demonstrated in 2013, when two studies put forward the proof of concept that hPSCs can generate HSCs using hPSCs-derived teratomas (Amabile et al., 2013; Suzuki et al., 2013). By co-injecting the iPSCs with OP9 cells these teratomas generated HSCs among other tissues with homing capacity to the mouse bone

marrow. These studies demonstrated that generation of HSCs via teratoma is variable and stochastic but it remains unclear the precise mechanism or factors that contributed to the generation of HSCs. Although this approach cannot meet the clinical application requirements, the production of HSCs in teratomas pave the way for the identification of specific factors given by the *in vivo* environment to be recapitulated in distinct *in vitro* systems.

The engraftment of immunodeficient mice by human cells has emerged as a pivotal experiment to predict human HSCs function. Despite many improvements, up to date no mouse model could fully support human HSCs (Beer and Eaves, 2015; Brehm et al., 2013). This fact is exemplified by the overall decline of human cell engraftment over time, the myeloid skewing over lymphoid lineage in recipient mice and the low efficiency in serial transplantations (Rongvaux et al., 2013; Willinger et al., 2011). In chapter 4, we show that upon transplantation of human $CD34^+CD49f^+$ cells into immunodeficient mice, human $CD45^+$ cells were detected in the peripheral blood up to 12 weeks. In addition, we detected B cells and T cells in both cell lines HDF and BJ overcoming an important caveat of this assay. However, these cells did not hold long-term reconstitution potential. It is also worth noting the fact that Tsukada and colleagues induced long-term engraftment *in vivo* by overexpressing fibroblasts with Gata2, Gfi1b and cFos within a teratoma (Tsukada et al., 2017). This gives us a clear indication that with a supportive niche, $CD34^+CD49f^+$ could also mature and possibly engraft secondary mice. However, further experiments are required to mature $CD34^+CD49f^+$ cells into fully functional HSCs. In this regard, two studies yielded functional hematopoietic stem cells by combining reprogramming with respective treatment in a supportive niche. Sugimura and colleagues used a combination of TFs (ERG, HOXA5, HOXA9, HOXA10, LCOR, RUNX1 and SPI1) in human iPSCs to convert them into immature HSCs that were further transplanted directly into the mice bone marrow to complete fully maturation and to gain long-term reconstitution potential (Sugimura et al., 2017). On the other hand, Lis and colleagues opted for a closer developmental cell source, adult endothelial cells, which were directly converted using specific TFs (Fosb, Gfi1, Runx1 and Spi1) and then co-cultured with an inductive vascular niche (Lis et al., 2017).

### 6.3 **– Mapping hemogenic transcription factor networks**

One of the first challenges in switching the reprogramming system from mouse to human cells is to establish the TF core. In this study, as previously discussed we have found that ectopic expression of GATA2, GFI1B and FOS is sufficient to efficiently reprogram human fibroblasts. In line with other studies, we confirm that GATA2 and FOS physically interact (Kawana et al., 1995) and withdrawal of either FOS or GATA2 from the cocktail abolishes the generation of $CD34^{+}CD49f^{+}$ cells (Figure 4.6) thus suggesting that both TF are an integral part of the hemogenic network. In endothelial cells, two studies showed that GATA2 co-occupied sites characteristically bound by JUN and FOS. Additionally, gene ontology analysis identified the co-targeted genes as inflammatory genes suggesting the existence of a GATA2/AP1 regulatory network to control inflammatory genes (Kanki et al., 2011; Linnemann et al., 2011). In our study we also have an enrichment of inflammatory pathways but since FOS had limited access binding the chromatin we do not see an overlap between GATA2 and FOS target genes. From single cell whole transcriptome analyses, data from E10.5 $PS34CD45^{-}$ to HSPCs placental cells indicates a reduction in the expression of *Fos* as the development progresses (Figure 3.11b). In addition, global gene expression profile of murine HSC compartments showed transient expression of both *Fos* and *Fosb*, they are upregulated in nascent HSCs from the YS and the AGM region but downregulated in highly proliferative FL HSCs (McKinney-Freeman et al., 2012).

An important mechanistic finding unveiled in this thesis is the GATA2 and GFI1B interaction. ChIP-seq reveals that GATA2 co-occupy significantly GFI1B target sites and their preferential motif is the AP-1 motif. Co-bound target gene list includes regulators of EHT and HSC emergence such as PODXL, PRDM1, BMPER and RUNX1. In fact, studies have proposed Gfi1 and Gfi1b as direct downstream targets of RUNX1 during EHT to repress endothelial markers (Lancrin et al., 2012), which is in agreement with our findings (Figure 4.6). In 2013, three independent studies showed occupation of *Gfi1b* locus by GATA2 in hematopoietic differentiated lineages (Guo et al., 2013; May et al., 2013; Moignard et al., 2013). Moignard and colleagues identified a regulatory network involving Gata2, Gfi1 and Gfi1b by single cell analysis, in which Gata2 activates Gfi1b and modulates Gfi1/Gfi1b cross-inhibition during the myelolymphoid lineage entry (Moignard et al., 2013). With a similar single-cell approach, Guo and colleagues showed that haploinsufficiency of Gata2 primes the

megakaryocytic and erythroid fate by downregulate *Gfi1b* (Guo et al., 2013). Furthermore, for erythroid differentiation to occur, Gata2 complexes with LSD1 a lysine-specific demethylase member of the CoREST repressive complex, to repress Gata1 in hematopoietic progenitor cells (Guo et al., 2016). Likewise, Gfi1 and Gfi1b also recruit LSD1 to epigenetically silence endothelial genes in hemogenic endothelium thus permitting the emergence of nascent HSCs (Thambyrajah et al., 2016). This suggests a putative interaction between Gata2 and Gfi1b mediated by Lsd1, therefore it is plausible to think that during direct hemogenic reprogramming, the silencing of the fibroblast program by GATA2 and GFI1B might be mediated through LSD1/CoREST complex. Further studies will be needed to test this hypothesis.

As described in the first chapter, several studies have been attempting to generate HSCs, frequently with TFs that are already implicated in hematopoiesis. An interesting fact is the limited TFs overlap in combinations across different studies, the exception is for members of RUNX1 family (RUNX1,Runx1t1 and Runx1c) (Batta et al., 2014; Cheng et al., 2016; Lis et al., 2017; Riddell et al., 2014; Sandler et al., 2014; Sugimura et al., 2017), Fos family (Fos and FOSB) (Pereira et al., 2013; Sandler et al., 2014) and Gfi1b family (Gfi1b and Gfi1) (Pereira et al., 2013; Sandler et al., 2014) as well as GATA2 (Batta et al., 2014; Cheng et al., 2016; Elcheva et al., 2014; Pereira et al., 2013; Vereide et al., 2014). But, is HSC emergence determined by only a limited number of TFs? The only studies showing long-term engraftment contain those exceptions in their combinations highlighting their importance and their role in HSC specification (Lis et al., 2017; Sugimura et al., 2017; Tsukada et al., 2017). In addition, RUNX1, Fos and Gfi1b seem to be substituted by their respective paralog gene in contrast to GATA2 suggesting that it cannot be replaced by another GATA family member.

### 6.4 - **Cooperative binding leads to induced hemogenic formation**

We have combined traditional molecular biology methods of co-immunoprecipitation followed by Western blotting, with high-throughput techniques based on next generation sequencing, to provide insight into hemogenic induction during the initiation phase of reprogramming. We interrogate how GGF shape hemogenic cell fate in fibroblasts and what are the regulatory mechanisms for the hemogenic

reprogramming. Hemogenic factors display significant differences in their binding pattern. GATA2 reveals consistency either individually or in combination with the hemogenic factors suggesting an independent and dominant target capacity and, we identify GATA2 as the TF that recruits GFI1B and FOS to regulatory regions. It has been suggested that GATA family members act as pioneer factors by binding closed chromatin and initiating gene regulatory networks (Cirillo et al., 2002; Takaku et al., 2016; Zaret and Carroll, 2011). Our data however support a cooperative action from the hemogenic TFs in agreement with findings described recently for iPSCs (Chronis et al., 2017). Furthermore, we reveal a specific chromatin signature of histone marks - H3K4me3, H3K27ac and H3K4me1- that predicts GATA2 and GFI1b occupancy. In a recent study, Wapinksi and colleagues studied neuronal reprogramming initial engagement of TF in short intervals for 12h, 24h, 36h and 48h and discovered that Ascl1 induces chromatin remodeling in fibroblasts as early as 12h (Wapinski et al., 2017). In hemogenic reprogramming, it would be interesting to learn if GATA2 opens the chromatin beforehand the 48h mark or if the GGF combination recruit remodelers.

In contrast, GFI1B and FOS do not show a dominant role during the earliest stages of hemogenic reprogramming. GFI1B depends on GATA2 to bind its targets while FOS showed a limited access to DNA. Despite being difficult to draw a clear conclusion about the mechanism by which FOS acts in cellular reprogramming, the data shown here support the importance of this TF in hemogenesis. Further studies will be useful to understand the precise mode of action of FOS during reprogramming and given that our ChIP-seq data is performed early in fibroblasts (48h post- transduction), it would be interesting to determine the binding sites of FOS in the $CD34^{+}CD49f^{+}$ hemogenic precursor population. ChIP-seq at such late time point would be of particular interest to understand if FOS have gained access to the chromatin during reprogramming, and it would be also useful to characterize its targets and to elucidate the regulatory network with GATA2.

Because of the technical challenge of having both appropriate number of transduced cells and high reprogramming efficiency, only few studies have explored factor binding in reprogrammed cells. Starting with different cell types and different combinations of TFs, the only shared mechanism is the repression of the pre-existing program and the activation of cell specific genes (Chronis et al., 2017; Rhee et al., 2017; Soufi et al., 2012; Wapinski et al., 2013). The use of RNA-seq and ChIP-seq integration provided

a unique glimpse into the mechanism of EHT during reprogramming. In agreement with the other studies, our data showed that both GATA2 and GFI1B bind directly fibroblast specific genes to repress them as well as they bind directly endothelial and hematopoietic genes to initiate transcription (Figure 5.8a).

As genome-wide data accumulates, integrative processes that combine information from multiple sources are key. In our study we employed ChromHMM to discover *de novo* re-occurring combinatorial and spatial patterns (Ernst and Kellis, 2012). We learned that GATA2 and GFI1B display a similar binding pattern in early hemogenic reprogramming and, in contrast GFI1B binding pattern diverge in hematopoietic progenitor cells. In mouse ESC-derived HE also demonstrated a similarity between Gata2 and Gfi1b target sites (Goode et al., 2016). However, ChIP-seq analysis from 10 key hematopoietic genes (Scl/Tal1, Lyl1, Lmo2, Gata2, Runx1, Meis1, Pu.1, Erg, Fli-1, and Gfi1b) in mouse hematopoietic progenitors showed few overlap between Gata2 and Gfi1b binding sites (Wilson et al., 2010). Overall, our studies suggest that GATA2 and GFI1B interaction is specific during hemogenic reprogramming and EHT and then lost in downstream hematopoietic progenitors (Figure 6.1).

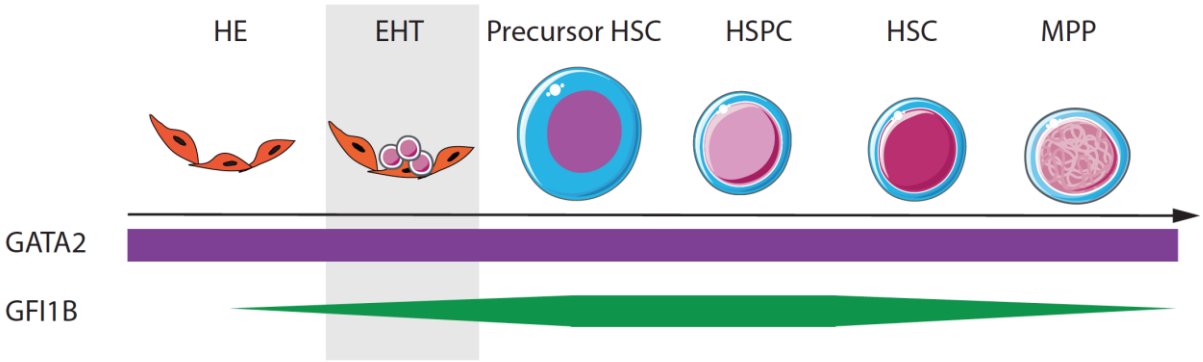



Figure 6.1 **- Schematic model representing the HSC specification.** Findings from our studies suggest a specific cooperation between GATA2 and GFI1B during endothelial-to-hematopoietic transition.

## 6.5 **- Concluding remarks**

In this thesis, I have presented evidence that hemogenic reprogramming faithfully recapitulates developmental hematopoiesis, which allowed the isolation of *in vivo* precursors of definitive mouse HSCs in mid-gestation placentas establishing a hemogenic reprogramming as a model to study hematopoiesis (Figure 6.2). I have also demonstrated the generation of hemogenic cells from human fibroblasts showing a conserved process and, I have revealed the molecular mechanism regulating HSC specification through GATA2 and GFI1B cooperation.

Collectively, the model described in this thesis moves the field to a better understanding of hemogenic fate. However, direct reprogramming is only in its childhood and the findings presented here represent just the first step towards clarity to a more complex developmental process. As the field continues to mature, direct hemogenic reprogramming will contribute to personalized medicine allowing patients to receive tailored treatments representing a new frontier in regenerative medicine (Figure 6.2).

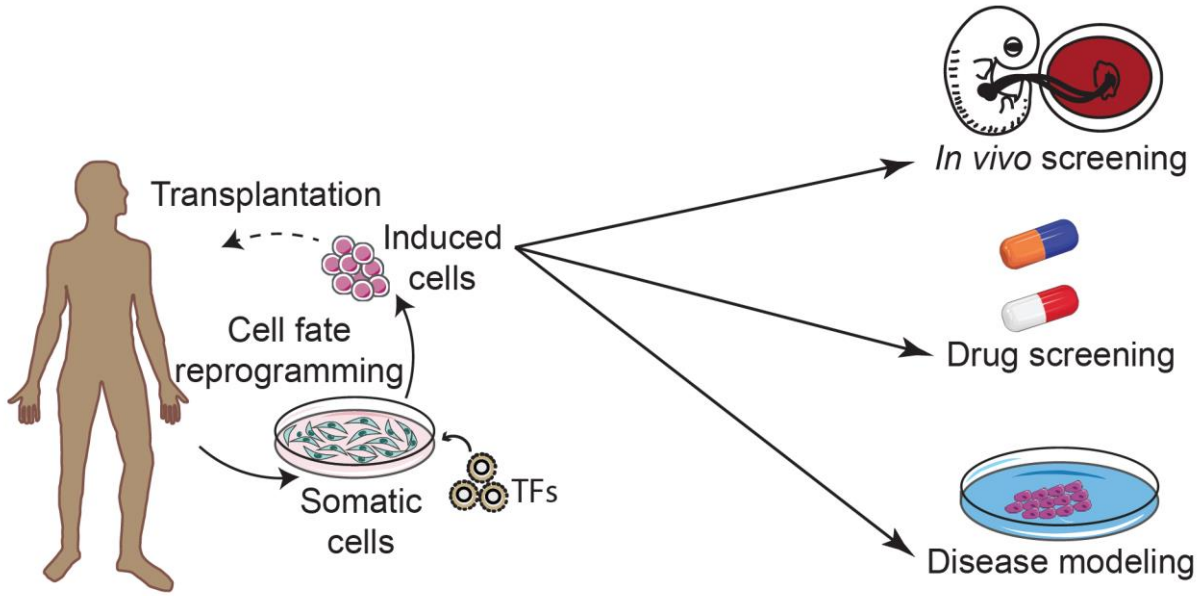



Figure 6.2 - **Schematic representation of the applicability of direct cell fate reprogramming.** Somatic cells such as fibroblasts can be reprogrammed with specific transcription factors and, the resulting induced cells can further be used to identify specific markers and phenotypes; can be used to test new pharmacological drugs and study *in vitro* disease models. In the future, induced cells with therapeutic potential could be implemented for clinical transplantation.

# References

Adamo, L., Naveiras, O., Wenzel, P.L., McKinney-Freeman, S., Mack, P.J., Gracia-Sancho, J., Suchy-Dicey, A., Yoshimoto, M., Lensch, M.W., Yoder, M.C., et al. (2009). Biomechanical forces promote embryonic haematopoiesis. Nature *459*, 1131–1135.

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. *44*, W3–W10.

Alvarez-Silva, M., Belo-Diabangouaya, P., Salaün, J., and Dieterlen-Lievre, F. (2003). Mouse placenta is a major hematopoietic organ. Development *130*, 5437–5444.

Amabile, G., Welner, R.S., Nombela-Arrieta, C., D'Alise, A.M., Di Ruscio, A., Ebralidze, A.K., Kraytsberg, Y., Ye, M., Kocher, O., Neuberg, D.S., et al. (2013). In vivo generation of transplantable human hematopoietic cells from induced pluripotent stem cells. Blood *121*, 1255–1264.

Ambrosini, G., Dreos, R., and Bucher, P. (2015). Principles of ChIP-seq data analysis illustrated with examples. Genomics Comput. Biol. *1*, 22.

Angel, P., and Karin, M. (1991). The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. Biochim. Biophys. Acta BBA - Rev. Cancer *1072*, 129–157.

Anguita, E., Villegas, A., Iborra, F., and Hernandez, A. (2010). GFI1B controls its own expression binding to multiple sites. Haematologica *95*, 36–46.

Arndt, K., Grinenko, T., Mende, N., Reichert, D., Portz, M., Ripich, T., Carmeliet, P., Corbeil, D., and Waskow, C. (2013). CD133 is a modifier of hematopoietic progenitor frequencies but is dispensable for the maintenance of mouse hematopoietic stem cells. Proc. Natl. Acad. Sci. U. S. A. *110*, 5582–5587.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Bates, D.L., Chen, Y., Kim, G., Guo, L., and Chen, L. (2008). Crystal Structures of Multiple GATA Zinc Fingers Bound to DNA Reveal New Insights into DNA Recognition and Self-Association by GATA. J. Mol. Biol. *381*, 1292–1306.

Batta, K., Florkowska, M., Kouskoff, V., and Lacaud, G. (2014). Direct Reprogramming of Murine Fibroblasts to Hematopoietic Progenitor Cells. Cell Rep. *9*, 1871–1884.

Beck, D., Thoms, J.A., Perera, D., Schütte, J., Unnikrishnan, A., Knezevic, K., Kinston, S.J., Wilson, N.K., O'Brien, T.A., Göttgens, B., et al. (2013). Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. Blood *122*, e12–e22.

Becker, A.J., McCulloch, E.A., and Till, J.E. (1963). Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells. Nature *197*, 452–454.

Beer, P.A., and Eaves, C.J. (2015). Modeling Normal and Disordered Human Hematopoiesis. Trends Cancer *1*, 199–210.

Bergiers, I., Andrews, T., Vargel Bölükbaşı, Ö., Buness, A., Janosz, E., Lopez-Anguita, N., Ganter, K., Kosim, K., Celen, C., Itır Perçin, G., et al. (2018). Single-cell transcriptomics reveals a new dynamical function of transcription factors during embryonic hematopoiesis. ELife *7*.

Bertrand, J.Y. (2005). Three pathways to mature macrophages in the early mouse yolk sac. Blood *106*, 3004–3011.

Bertrand, J.Y., Kim, A.D., Violette, E.P., Stachura, D.L., Cisson, J.L., and Traver, D. (2007). Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. Development *134*, 4147–4156.

Bertrand, J.Y., Cisson, J.L., Stachura, D.L., and Traver, D. (2010). Notch signaling distinguishes 2 waves of definitive hematopoiesis in the zebrafish embryo. Blood *115*, 2777–2783.

Bhatia, M., Bonnet, D., Kapp, U., Wang, J.C.Y., Murdoch, B., and Dick, J.E. (1997). Quantitative Analysis Reveals Expansion of Human Hematopoietic Repopulating Cells After Short-term Ex Vivo Culture. J. Exp. Med. *186*, 619–624.

Bhatia, M., Bonnet, D., Wu, D., Murdoch, B., Wrana, J., Gallacher, L., and Dick, J.E. (1999). Bone Morphogenetic Proteins Regulate the Developmental Program of Human Hematopoietic Stem Cells. J. Exp. Med. *189*, 1139–1148.

Bigas, A., Martin, D.I., and Bernstein, I.D. (1995). Generation of hematopoietic colony-forming cells from embryonic stem cells: synergy between a soluble factor from NIH-3T3 cells and hematopoietic growth factors. Blood *85*, 3127–3133.

Billerbeck, E., Barry, W.T., Mu, K., Dorner, M., Rice, C.M., and Ploss, A. (2011). Development of human CD4+FoxP3+ regulatory T cells in human stem cell factor-, granulocyte-macrophage colony-stimulating factor-, and interleukin-3-expressing NOD-SCID IL2R null humanized mice. Blood *117*, 3076–3086.

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

Boisset, J.-C., Clapes, T., Klaus, A., Papazian, N., Onderwater, J., Mommaas-Kienhuis, M., Cupedo, T., and Robin, C. (2015). Progressive maturation toward hematopoietic stem cells in the mouse embryo aorta. Blood *125*, 465–469.

Bonde, S., Dowden, A.M., Chan, K.-M., Tabayoyong, W.B., and Zavazava, N. (2008). HOXB4 But Not BMP4 Confers Self-Renewal Properties to ES-Derived Hematopoietic Progenitor Cells. Transplantation *86*, 1803–1809.

Bowles, K.M., Vallier, L., Smith, J.R., Alexander, M.R.J., and Pedersen, R.A. (2006). HOXB4 Overexpression Promotes Hematopoietic Development by Human Embryonic Stem Cells. STEM CELLS *24*, 1359–1369.

Brehm, M.A., Shultz, L.D., Luban, J., and Greiner, D.L. (2013). Overcoming Current Limitations in Humanized Mouse Research. J. Infect. Dis. *208*, S125–S130.

Briggs, R., and King, T.J. (1952). Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. Proc. Natl. Acad. Sci. U. S. A. *38*, 455–463.

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

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

Brunet, J.-P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proc. Natl. Acad. Sci. *101*, 4164–4169.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase. Cell *150*, 1209–1222.

Burns, C.E., Traver, D., Mayhall, E., Shepard, J.L., and Zon, L.I. (2005). Hematopoietic stem cell fate is established by the Notch–Runx pathway. Genes Dev. *19*, 2331–2342.

Burns, C.E., Galloway, J.L., Smith, A.C., Keefe, M.D., Cashman, T.J., Paik, E.J., Mayhall, E.A., Amsterdam, A.H., and Zon, L.I. (2009). A genetic screen in zebrafish defines a hierarchical network of pathways required for hematopoietic stem cell emergence. Blood *113*, 5776–5782.

Butko, E., Distel, M., Pouget, C., Weijts, B., Kobayashi, I., Ng, K., Mosimann, C., Poulain, F.E., McPherson, A., Ni, C.-W., et al. (2015). Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. Development *142*, 1050–1061.

Cai, Z., de Bruijn, M., Ma, X., Dortland, B., Luteijn, T., Downing, J.R., 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.

Campbell, K.H.S., McWhir, J., Ritchie, W.A., and Wilmut, I. (1996). Sheep cloned by nuclear ransfer from a cultured cell line. Nature *380*, 64–69.

Cao, X., Shores, E.W., Hu-Li, J., Anver, M.R., Kelsail, B.L., Russell, S.M., Drago, J., Noguchi, M., Grinberg, A., and Bloom, E.T. (1995). Defective lymphoid development in mice lacking expression of the common cytokine receptor γ chain. Immunity *2*, 223–238.

Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., Fahrig, M., Vandenhoeck, A., Harpal, K., Eberhardt, C., et al. (1996). Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. Nature *380*, 435–439.

Chadwick, K. (2003). Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. Blood *102*, 906–915.

Chan, K.-M., Bonde, S., Klump, H., and Zavazava, N. (2008). Hematopoiesis and immunity of HOXB4-transduced embryonic stem cell–derived hematopoietic progenitor cells. Blood *111*, 2953–2961.

Chanda, B., Ditadi, A., Iscove, N.N., and Keller, G. (2013). Retinoic Acid Signaling Is Essential for Embryonic Hematopoietic Stem Cell Development. Cell *155*, 215–227.

Chang, A.N., Cantor, A.B., Fujiwara, Y., Lodish, M.B., Droho, S., Crispino, J.D., and Orkin, S.H. (2002). GATA-factor dependence of the multitype zinc-finger protein FOG-1 for its essential role in megakaryopoiesis. Proc. Natl. Acad. Sci. *99*, 9237–9242.

Chen, M.J., Yokomizo, T., Zeigler, B.M., Dzierzak, E., and Speck, N.A. (2009). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. Nature *457*, 887–891.

Chen, M.J., Li, Y., De Obaldia, M.E., Yang, Q., Yzaguirre, A.D., Yamada-Inagawa, T., Vink, C.S., Bhandoola, A., Dzierzak, E., and Speck, N.A. (2011). Erythroid/Myeloid Progenitors and Hematopoietic Stem Cells Originate from Distinct Populations of Endothelial Cells. Cell Stem Cell *9*, 541–552.

Cheng, H., Ang, H.Y.-K., A. EL Farran, C., Li, P., Fang, H.T., Liu, T.M., Kong, S.L., Chin, M.L., Ling, W.Y., Lim, E.K.H., et al. (2016). Reprogramming mouse fibroblasts into engraftable myeloerythroid and lymphoid progenitors. Nat. Commun. *7*, 13396.

Chiu, C.-P., and Blau, H.M. (1985). 5-Azacytidine permits gene activation in a previously noninducible cell type. Cell *40*, 417–424.

Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J.C., and Keller, G. (1998). A common precursor for hematopoietic and endothelial cells. Development *125*, 725–732.

Choi, K.-D., Yu, J., Smuga-Otto, K., Salvagiotto, G., Rehrauer, W., Vodyanik, M., Thomson, J., and Slukvin, I. (2009). Hematopoietic and Endothelial Differentiation of Human Induced Pluripotent Stem Cells. Stem Cells *27*, 559–567.

Chotinantakul, K., and Leeanansaksiri, W. (2012). Hematopoietic Stem Cell Development, Niches, and Signaling Pathways. Bone Marrow Res. *2012*, 1–16.

Chowdhury, A.H., Ramroop, J.R., Upadhyay, G., Sengupta, A., Andrzejczyk, A., and Saleque, S. (2013). Differential Transcriptional Regulation of meis1 by Gfi1b and Its Co-Factors LSD1 and CoREST. PLoS ONE *8*, e53666.

Christensen, J.L., Wright, D.E., Wagers, A.J., and Weissman, I.L. (2004). Circulation and Chemotaxis of Fetal Hematopoietic Stem Cells. PLoS Biol. *2*, e75.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442–459.e20.

Ciau-Uitz, A., Pinheiro, P., Gupta, R., Enver, T., and Patient, R. (2010). Tel1/ETV6 Specifies Blood Stem Cells through the Agency of VEGF Signaling. Dev. Cell *18*, 569–578.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Clarke, R.L., Yzaguirre, A.D., Yashiro-Ohtani, Y., Bondue, A., Blanpain, C., Pear, W.S., Speck, N.A., and Keller, G. (2013). The expression of Sox17 identifies and regulates haemogenic endothelium. Nat. Cell Biol. *15*, 502–510.

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

Copelan, E.A. (2006). Hematopoietic stem-cell transplantation. N. Engl. J. Med. *354*, 1813–1826.

Cosgun, K.N., Rahmig, S., Mende, N., Reinke, S., Hauber, I., Schäfer, C., Petzold, A., Weisbach, H., Heidkamp, G., Purbojo, A., et al. (2014). Kit Regulates HSC Engraftment across the Human-Mouse Species Barrier. Cell Stem Cell *15*, 227–238.

Coulombel, L. (2004). Identification of hematopoietic stem/progenitor cells: strength and drawbacks of functional assays. Oncogene *23*, 7210.

Cox, B., Kotlyar, M., Evangelou, A.I., Ignatchenko, V., Ignatchenko, A., Whiteley, K., Jurisica, I., Adamson, S.L., Rossant, J., and Kislinger, T. (2009). Comparative systems biology of human and mouse as a tool to guide the modeling of human placental pathology. Mol. Syst. Biol. *5*.

Crisan, M., Kartalaei, P.S., Vink, C., Yamada-Inagawa, T., Bollerot, K., van IJcken, W., van der Linden, R., de Sousa Lopes, S.M.C., Monteiro, R., Mummery, C., et al. (2015). BMP signalling differentially regulates distinct haematopoietic stem cell types. Nat. Commun. *6*, 8040.

Cumano, A., and Godin, I. (2007). Ontogeny of the Hematopoietic System. Annu. Rev. Immunol. *25*, 745–785.

Cumano, A., Dieterlen-Lievre, F., and Godin, I. (1996). Lymphoid potential, probed before circulation in mouse, is restricted to caudal intraembryonic splanchnopleura. Cell *86*, 907–916.

Cumano, A., Ferraz, J.C., Klaine, M., Di Santo, J.P., and Godin, I. (2001). Intraembryonic, but not yolk sac hematopoietic precursors, isolated before circulation, provide long-term multilineage reconstitution. Immunity *15*, 477–485.

Dancis, J., Brown, G.F., Gorstein, F., and Balis, M.E. (1977). Treatment of hypoplastic anemia in mice with placental transplants. Blood *50*, 663–670.

David, L., and Polo, J.M. (2014). Phases of reprogramming. Stem Cell Res. *12*, 754–761.

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

Dick, J.E., Magli, M.C., Huszar, D., Phillips, R.A., and Bernstein, A. (1985). Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/Wv mice. Cell *42*, 71–79.

Dieterlen-Lièvre, F. (1975). On the origin of haemopoietic stem cells in the avian embryo: an experimental approach. Development *33*, 607–619.

Ditadi, A., Sturgeon, C.M., Tober, J., Awong, G., Kennedy, M., Yzaguirre, A.D., Azzola, L., Ng, E.S., Stanley, E.G., French, D.L., et al. (2015). Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. Nat. Cell Biol. *17*, 580–591.

Doan, L.L. (2004). Targeted transcriptional repression of Gfi1 by GFI1 and GFI1B in lymphoid cells. Nucleic Acids Res. *32*, 2508–2519.

Doré, L.C., Chlon, T.M., Brown, C.D., White, K.P., and Crispino, J.D. (2012). Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. Blood *119*, 3724–3733.

Doulatov, S., Notta, F., Laurenti, E., and Dick, J.E. (2012). Hematopoiesis: A Human Perspective. Cell Stem Cell *10*, 120–136.

Doulatov, S., Vo, L.T., Chou, S.S., Kim, P.G., Arora, N., Li, H., Hadland, B.K., Bernstein, I.D., Collins, J.J., Zon, L.I., et al. (2013). Induction of Multipotential Hematopoietic Progenitors from Human Pluripotent Stem Cells via Respecification of Lineage-Restricted Precursors. Cell Stem Cell *13*, 459–470.

Drake, C.J., Brandt, S.J., Trusk, T.C., and Little, C.D. (1997). TAL1/SCL is expressed in endothelial progenitor cells/angioblasts and defines a dorsal-to-ventral gradient of vasculogenesis. Dev. Biol. *192*, 17–30.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Eilken, H.M., Nishikawa, S.-I., and Schroeder, T. (2009). Continuous single-cell imaging of blood generation from haemogenic endothelium. Nature *457*, 896–900.

Elcheva, I., Brok-Volchanskaya, V., Kumar, A., Liu, P., Lee, J.-H., Tong, L., Vodyanik, M., Swanson, S., Stewart, R., Kyba, M., et al. (2014). Direct induction of haematoendothelial programs in human pluripotent stem cells by transcriptional regulators. Nat. Commun. *5*.

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

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Espín-Palazón, R., Stachura, D.L., Campbell, C.A., García-Moreno, D., Del Cid, N., Kim, A.D., Candel, S., Meseguer, J., Mulero, V., and Traver, D. (2014). Proinflammatory Signaling Regulates Hematopoietic Stem Cell Emergence. Cell *159*, 1070–1085.

Espin-Palazon, R., Weijts, B., Mulero, V., and Traver, D. (2017). Proinflammatory Signals as Fuel for the Fire of Hematopoietic Stem Cell Emergence. Trends Cell Biol.

Feng, B., Ng, J.-H., Heng, J.-C.D., and Ng, H.-H. (2009). Molecules that Promote or Enhance Reprogramming of Somatic Cells to Induced Pluripotent Stem Cells. Cell Stem Cell *4*, 301–312.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*, 1728–1740.

Feng, R., Desbordes, S.C., Xie, H., Tillo, E.S., Pixley, F., Stanley, E.R., and Graf, T. (2008). PU. 1 and C/EBPα/β convert fibroblasts into macrophage-like cells. Proc. Natl. Acad. Sci. *105*, 6057–6062.

Ferkowicz, M.J., Starr, M., Xie, X., Li, W., Johnson, S.A., Shelley, W.C., Morrison, P.R., and Yoder, M.C. (2003). CD41 expression defines the onset of primitive and definitive hematopoiesis in the murine embryo. Development *130*, 4393–4403.

Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K.S., Powell-Braxton, L., Hillan, K.J., and Moore, M.W. (1996). Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. Nature *380*, 439–442.

Foudi, A., Kramer, D.J., Qin, J., Ye, D., Behlich, A.-S., Mordecai, S., Preffer, F.I., Amzallag, A., Ramaswamy, S., Hochedlinger, K., et al. (2014). Distinct, strict requirements for Gfi-1b in adult bone marrow red cell and platelet generation. J. Exp. Med. *211*, 909–927.

Frame, J.M., McGrath, K.E., and Palis, J. (2013). Erythro-Myeloid Progenitors: "definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells. Blood Cells. Mol. Dis. *51*.

Fu, J.-D., Stone, N.R., Liu, L., Spencer, C.I., Qian, L., Hayashi, Y., Delgado-Olguin, P., Ding, S., Bruneau, B.G., and Srivastava, D. (2013). Direct Reprogramming of Human Fibroblasts toward a Cardiomyocyte-like State. Stem Cell Rep. *1*, 235–247.

Fu, K., Chronis, C., Soufi, A., Bonora, G., Edwards, M., Smale, S., Zaret, K.S., Plath, K., and Pellegrini, M. (2018). Comparisonof reprogramming factor targets reveals both species-specific and conserved mechanisms in early iPS cells. BioRxiv.

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics *31*, 3718–3720.

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

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

Gao, X., Wu, T., Johnson, K.D., Lahvic, J.L., Ranheim, E.A., Zon, L.I., and Bresnick, E.H. (2016). GATA Factor-G-Protein-Coupled Receptor Circuit Suppresses Hematopoiesis. Stem Cell Rep.

Garcon, L. (2005). Gfi-1B plays a critical role in terminal differentiation of normal and transformed erythroid progenitor cells. Blood *105*, 1448–1455.

Gekas, C., Dieterlen-Lièvre, F., Orkin, S.H., and Mikkola, H.K.A. (2005). The Placenta Is a Niche for Hematopoietic Stem Cells. Dev. Cell *8*, 365–375.

Gekas, C., Rhodes, K.E., Vanhandel, B., Chhabra, A., Ueno, M., and Mikkola, H.K.A. (2010). Hematopoietic stem cell development in the placenta. Int. J. Dev. Biol. *54*, 1089–1098.

Georgiades, P., Ferguson-Smith, A.C., and Burton, G.J. (2002). Comparative Developmental Anatomy of the Murine and Human Definitive Placentae. Placenta *23*, 3–19.

Gering, M., and Patient, R. (2005). Hedgehog Signaling Is Required for Adult Blood Stem Cell Formation in Zebrafish Embryos. Dev. Cell *8*, 389–400.

Gerri, C., Marass, M., Rossi, A., and Stainier, D.Y.R. (2018). Hif-1α and Hif-2α regulate hemogenic endothelium and hematopoietic stem cell formation in zebrafish. Blood *131*, 963–973.

Ghiaur, G., Ferkowicz, M.J., Milsom, M.D., Bailey, J., Witte, D., Cancelas, J.A., Yoder, M.C., and Williams, D.A. (2008). Rac1 is essential for intraembryonic hematopoiesis and for the initial seeding of fetal liver with definitive hematopoietic progenitor cells. Blood *111*, 3313–3321.

Gilmour, D.S., and Lis, J.T. (1984). Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. Proc. Natl. Acad. Sci. *81*, 4275–4279.

Godin, I.E., Garcia-Porrero, J.A., Coutinho, A., Dieterlen-Lievre, F., and Marcos, M.A.R. (1993). Para-aortic splanchnopleura from early mouse embryos contains B1a cell progenitors. Nature *364*, 67–70.

Goldie, L.C., Lucitti, J.L., Dickinson, M.E., and Hirschi, K.K. (2008). Cell signaling directing the formation and function of hemogenic endothelium during murine embryogenesis. Blood *112*, 3194–3204.

Goode, D.K., Obier, N., Vijayabaskar, M.S., Lie-A-Ling, M., Lilly, A.J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., et al. (2016). Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. Dev. Cell *36*, 572–587.

Gordon-Keylock, S., Sobiesiak, M., Rybtsov, S., Moore, K., and Medvinsky, A. (2013). Mouse extraembryonic arterial vessels harbor precursors capable of maturing into definitive HSCs. Blood *122*, 2338–2345.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Grainger, S., Richter, J., Palazón, R.E., Pouget, C., Lonquich, B., Wirth, S., Grassme, K.S., Herzog, W., Swift, M.R., Weinstein, B.M., et al. (2016). Wnt9a Is Required for the Aortic Amplification of Nascent Hematopoietic Stem Cells. Cell Rep. *17*, 1595–1606.

Grass, J.A., Jing, H., Kim, S.-I., Martowicz, M.L., Pal, S., Blobel, G.A., and Bresnick, E.H. (2006). Distinct Functions of Dispersed GATA Factor Complexes at an Endogenous Gene Locus. Mol. Cell. Biol. *26*, 7056–7067.

Grimes, H.L., Chan, T.O., Zweidler-McKay, P.A., Tong, B., and Tsichlis, P.N. (1996). The Gfi-1 proto-oncoprotein contains a novel transcriptional repressor domain, SNAG, and inhibits G1 arrest induced by interleukin-2 withdrawal. Mol. Cell. Biol. *16*, 6263–6272.

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

Guiu, J., Bergen, D.J.M., De Pater, E., Islam, A.B.M.M.K., Ayllón, V., Gama-Norton, L., Ruiz-Herguido, C., González, J., López-Bigas, N., Menendez, P., et al. (2014). Identification of Cdca7 as a novel Notch transcriptional target involved in hematopoietic stem cell emergence. J. Exp. Med. *211*, 2411–2423.

Guo, G., Luc, S., Marco, E., Lin, T.-W., Peng, C., Kerenyi, M.A., Beyaz, S., Kim, W., Xu, J., Das, P.P., et al. (2013). Mapping Cellular Hierarchy by Single-Cell Analysis of the Cell Surface Repertoire. Cell Stem Cell *13*, 492–505.

Guo, Y., Fu, X., Huo, B., Wang, Y., Sun, J., Meng, L., Hao, T., Zhao, Z.J., and Hu, X. (2016). GATA2 regulates GATA1 expression through LSD1-mediated histone modification. Am. J. Transl. Res. *8*, 2265–2274.

Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. Development *10*, 622–640.

Gurdon, J.B., and Byrne, J.A. (2003). The first half-century of nuclear transplantation. Proc. Natl. Acad. Sci. U. S. A. *100*, 8048–8052.

Gurdon, J.B., Elsdale, T.R., and Fischberg, M. (1958). Sexually Mature Individuals of Xenopus laevis from the Transplantation of Single Somatic Nuclei. Nature *182*, 64–65.

Gutman, J.A., Ross, K., Smith, C., Myint, H., Lee, C.K., Salit, R., Milano, F., Delaney, C., Gao, D., and Pollyea, D.A. (2016). Chronic graft versus host disease burden and late transplant complications are lower following adult double cord blood versus matched unrelated donor peripheral blood transplantation. Bone Marrow Transplant. *51*, 1588.

Haar, J.L., and Ackerman, G.A. (1971). A Phase and Electron Microscopic Study of Vasculogenesis and Erythropoiesis in the Yolk Sac of the Mouse. Anat. Rec. *170*, 199–223.

Hadland, B.K. (2004). A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. Blood *104*, 3097–3105.

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

Hahn, W.C., and Weinberg, R.A. (2002). Rules for making human tumor cells. N. Engl. J. Med. *347*, 1593–1603.

Halder, G., Callaerts, P., and Gehring, W.J. (1995). Induction of Ectopic Eyes by Targeted Expression of the eyeless Gene in Drosophila. Science *267*, 1788–1792.

Hao, Q.-L., Shah, A.J., Thiemann, F.T., Smogorzewska, E.M., and Crooks, G.M. (1995). A functional comparison of CD34+ CD38-cells in cord blood and bone marrow. Blood *86*, 3745–3753.

Hara, T., Nakano, Y., Tanaka, M., Tamura, K., Sekiguchi, T., Minehata, K., Copeland, N.G., Jenkins, N.A., Okabe, M., Kogo, H., et al. (1999). Identification of podocalyxin-like protein 1 as a novel cell surface marker for hemangioblasts in the murine aorta-gonad-mesonephros region. Immunity *11*, 567–578.

Harris, J.E., Harris, T.H., Weninger, W., Wherry, E.J., Hunter, C.A., and Turka, L.A. (2012). A mouse model of vitiligo with focused epidermal depigmentation requires IFN-γ for autoreactive CD8+ T cell accumulation in the skin. J. Invest. Dermatol. *132*, 1869–1876.

Harris, J.M., Esain, V., Frechette, G.M., Harris, L.J., Cox, A.G., Cortes, M., Garnaas, M.K., Carroll, K.J., Cutting, C.C., and Khan, T. (2013). Glucose metabolism impacts the spatiotemporal onset and magnitude of HSC induction in vivo. Blood *121*, 2483–2493.

Hart, A., Melet, F., Grossfeld, P., Chien, K., Jones, C., Tunnacliffe, A., Favier, R., and Bernstein, A. (2000). Fli-1 is required for murine vascular and megakaryocytic development and is hemizygously deleted in patients with thrombocytopenia. Immunity *13*, 167–177.

He, Q., Zhang, C., Wang, L., Zhang, P., Ma, D., Lv, J., and Liu, F. (2015). Inflammatory signaling regulates hematopoietic stem and progenitor cell emergence in vertebrates. Blood *125*, 1098–1106.

Hochedlinger, K., and Jaenisch, R. (2002). Monoclonal mice generated by nuclear transfer from mature B and T donor cells. Nature *415*, 1035.

Hock, H., Hamblen, M.J., Rooke, H.M., Schindler, J.W., Saleque, S., Fujiwara, Y., and Orkin, S.H. (2004). Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. Nature *431*, 1002–1007.

Hole, N., Graham, G.J., Menzel, U., and Ansell, J.D. (1996). A limited temporal window for the derivation of multilineage repopulating hematopoietic progenitors during embryonal stem cell differentiation in vitro. Blood *88*, 1266–1276.

Huang, D.-Y. (2005). GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells. Nucleic Acids Res. *33*, 5331–5342.

Huang, D.-Y., Kuo, Y.-Y., Lai, J.-S., Suzuki, Y., Sugano, S., and Chang, Z.-F. (2004). GATA-1 and NF-Y cooperate to mediate erythroid-specific transcription of Gfi-1B gene. Nucleic Acids Res. *32*, 3935–3946.

Huber, T., Kouskoff, V., Fehling, J.H., Palis, J., and Keller, G. (2004). Haemangioblast commitment is initited in the primitive streak of the mouse mebryo. Nature *432*, 625–630.

Hwang, L.Y., Siegelman, M., Davis, L., Oppenheimer-Marks, N., and Baer, R. (1993). Expression of the TAL1 proto-oncogene in cultured endothelial cells and blood vessels of the spleen. Oncogene *8*, 3043–3046.

Ieda, M., Fu, J.-D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct Reprogramming of Fibroblasts into Functional Cardiomyocytes by Defined Factors. Cell *142*, 375–386.

Iizuka, K., Yokomizo, T., Watanabe, N., Tanaka, Y., Osato, M., Takaku, T., and Komatsu, N. (2016). Lack of Phenotypical and Morphological Evidences of Endothelial to Hematopoietic Transition in the Murine Embryonic Head during Hematopoietic Stem Cell Emergence. PLOS ONE *11*, e0156427.

Imanirad, P., Kartalaei, P.S., Crisan, M., Vink, C., Yamada-Inagawa, T., de Pater, E., Kurek, D., Kaimakis, P., van der Linden, R., Speck, N., et al. (2014). HIF1α is a regulator of hematopoietic progenitor and stem cell development in hypoxic sites of the mouse embryo. Stem Cell Res. *12*, 24–35.

Ito, M. (2002). NOD/SCID/gamma cnull mouse: an excellent recipient mouse model for engraftment of human cells. Blood *100*, 3175–3182.

Ivanovs, A., Rybtsov, S., Welch, L., Anderson, R.A., Turner, M.L., and Medvinsky, A. (2011). Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. J. Exp. Med. *208*, 2417–2427.

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

Jackson, M., Axton, R.A., Taylor, A.H., Wilson, J.A., Gordon-Keylock, S.A.M., Kokkaliaris, K.D., Brickman, J.M., Schulz, H., Hummel, O., Hubner, N., et al. (2012). HOXB4 Can Enhance the Differentiation of Embryonic Stem Cells by Modulating the Hematopoietic Niche. STEM CELLS *30*, 150–160.

Jackson, M., Ma, R., Taylor, A.H., Axton, R.A., Easterbrook, J., Kydonaki, M., Olivier, E., Marenah, L., Stanley, E.G., Elefanty, A.G., et al. (2016). Enforced Expression of HOXB4 in Human Embryonic Stem Cells Enhances the Production of Hematopoietic Progenitors but Has No Effect on the Maturation of Red Blood Cells. Stem Cells Transl. Med. *5*, 981–990.

Jaffredo, T., Gautier, R., Eichmann, A., and Dieterlen-Lièvre, F. (1998). Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. Development *125*, 4575–4583.

Jaffredo, T., Nottingham, W., Liddiard, K., Bollerot, K., Pouget, C., and Debruijn, M. (2005). From hemangioblast to hematopoietic stem cell: An endothelial connection? Exp. Hematol. *33*, 1029–1040.

Jain, J., Nalefski, E.A., McCaffrey, P.G., Johnson, R.S., Spiegelman, B.M., Papaioannou, V., and Rao, A. (1994). Normal peripheral T-cell function in c-Fos-deficient mice. Mol. Cell. Biol. *14*, 1566–1574.

Jang, I.H., Lu, Y.-F., Zhao, L., Wenzel, P.L., Kume, T., Datta, S.M., Arora, N., Guiu, J., Lagha, M., Kim, P.G., et al. (2015). Notch1 acts via Foxc2 to promote definitive hematopoiesis via effects on hemogenic endothelium. Blood *125*, 1418–1426.

Jay, K.E., Rouleau, A., Underhill, T.M., and Bhatia, M. (2004). Identification of a novel population of human cord blood cells with hematopoietic and chondrocytic potential. Cell Res. *14*, 268.

Jegalian, A.G., and Wu, H. (2002). Regulation of Socs Gene Expression by the Proto-oncoprotein GFI-1B: TWO ROUTES FOR STAT5 TARGET GENE INDUCTION BY ERYTHROPOIETIN. J. Biol. Chem. *277*, 2345–2352.

Jiang, P., and Singh, M. (2014). CCAT: Combinatorial Code Analysis Tool for transcriptional regulation. Nucleic Acids Res. *42*, 2833–2847.

Jing, L., Tamplin, O.J., Chen, M.J., Deng, Q., Patterson, S., Kim, P.G., Durand, E.M., McNeil, A., Green, J.M., Matsuura, S., et al. (2015). Adenosine signaling promotes hematopoietic stem and progenitor cell emergence. J. Exp. Med. *212*, 649–663.

Johnson, G.R., and Moore, M.A.S. (1975). Role of stem cell migration in initiation of mouse foetal liver haemopoiesis. Nature *258*, 726–728.

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

Johnson, R.S., Spiegelman, B.M., and Papaioannou, V. (1992). Pleiotropic effects of a null mutation in the c-fos proto-oncogene. Cell *71*, 577–586.

Jokubaitis, V.J., Sinka, L., Driessen, R., Whitty, G., Haylock, D.N., Bertoncello, I., Smith, I., Peault, B., Tavian, M., and Simmons, P.J. (2008). Angiotensin-converting enzyme (CD143) marks hematopoietic stem cells in human embryonic, fetal, and adult hematopoietic tissues. Blood *111*, 4055–4063.

Jordan, C.T., and Lemischka, I.R. (1990). Clonal and systemic analysis of long-term hematopoiesis in the mouse. Genes Dev. *4*, 220–232.

Kallianpur, A.R., Jordan, J.E., and Brandt, S.J. (1994). The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis. Blood *83*, 1200–1208.

Kamei, M., Brian Saunders, W., Bayless, K.J., Dye, L., Davis, G.E., and Weinstein, B.M. (2006). Endothelial tubes assemble from intracellular vacuoles in vivo. Nature *442*, 453–456.

Kandasamy, K., Mohan, S.S., Raju, R., Keerthikumar, S., Kumar, G.S.S., Venugopal, A.K., Telikicherla, D., Navarro, J.D., Mathivanan, S., and Pecquet, C. (2010). NetPath: a public resource of curated signal transduction pathways. Genome Biol. *11*, R3.

Kanki, Y., Kohro, T., Jiang, S., Tsutsumi, S., Mimura, I., Suehiro, J., Wada, Y., Ohta, Y., Ihara, S., Iwanari, H., et al. (2011). Epigenetically coordinated GATA2 binding is necessary for endothelium-specific endomucin expression. EMBO J. *30*, 2582–2595.

Karlsson, G., Rörby, E., Pina, C., Soneji, S., Reckzeh, K., Miharada, K., Karlsson, C., Guo, Y., Fugazza, C., Gupta, R., et al. (2013). The Tetraspanin CD9 Affords High-Purity Capture of All Murine Hematopoietic Stem Cells. Cell Rep. *4*, 642–648.

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

Kawana, M., Lee, M.E., Quertermous, E.E., and Quertermous, T. (1995). Cooperative interaction of GATA-2 and AP1 regulates transcription of the endothelin-1 gene. Mol. Cell. Biol. *15*, 4225–4231.

Keller, G., Kennedy, M., Papayannopoulou, T., and Wiles, M.V. (1993). Hematopoietic commitment during embryonic stem cell differentiation in culture. Mol. Cell. Biol. *13*, 473–486.

Kennedy, M., Firpo, M., Choi, K., Wall, C., Robertson, S., Kabrun, N., and Keller, G. (1997). A common precursor for primitive erythropoiesis and definitive haematopoiesis. Nature *386*, 488–493.

Kennedy, M., D'Souza, S.L., Lynch-Kattman, M., Schwantz, S., and Keller, G. (2007). Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. Blood *109*, 2679–2687.

Kennedy, M., Awong, G., Sturgeon, C.M., Ditadi, A., LaMotte-Mohs, R., Zúñiga-Pflücker, J.C., and Keller, G. (2012). T Lymphocyte Potential Marks the Emergence of Definitive Hematopoietic Progenitors in Human Pluripotent Stem Cell Differentiation Cultures. Cell Rep. *2*, 1722–1735.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Khandanpour, C., Sharif-Askari, E., Vassen, L., Gaudreau, M.-C., Zhu, J., Paul, W.E., Okayama, T., Kosan, C., and Möröy, T. (2010). Evidence that growth factor independence 1b regulates dormancy and peripheral blood mobilization of hematopoietic stem cells. Blood *116*, 5149–5161.

Kiel, M.J., Yilmaz, Ö.H., Iwashita, T., Yilmaz, O.H., Terhorst, C., and Morrison, S.J. (2005). SLAM Family Receptors Distinguish Hematopoietic Stem and Progenitor Cells and Reveal Endothelial Niches for Stem Cells. Cell *121*, 1109–1121.

Kim, C. (2015). iPSC technology-Powerful hand for disease modeling and therapeutic screen. BMB Rep. *48*, 256–265.

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

Kim, I., Yilmaz, Ö.H., and Morrison, S.J. (2005). CD144 (VE-cadherin) is transiently expressed by fetal liver hematopoietic stem cells. Blood *106*, 903–905.

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

Kitajima, K., Zheng, J., Yen, H., Sugiyama, D., and Nakano, T. (2006). Multipotential differentiation ability of GATA-1-null erythroid-committed cells. Genes Dev. *20*, 654–659.

Ko, L.J., and Engel, J.D. (1993). DNA-binding specificities of the GATA transcription factor family. Mol. Cell. Biol. *13*, 4011–4022.

Kobayashi, I., Kobayashi-Sun, J., Kim, A.D., Pouget, C., Fujita, N., Suda, T., and Traver, D. (2014). Jam1a–Jam2a interactions regulate haematopoietic stem cell fate through Notch signalling. Nature *512*, 319–323.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming Factor Expression Initiates Widespread Targeted Chromatin Remodeling. Cell Stem Cell *8*, 96–105.

Kolb, H.-J. (2008). Graft-versus-leukemia effects of transplantation and donor lymphocytes. Blood *112*, 4371–4383.

Konantz, M., Alghisi, E., Müller, J.S., Lenard, A., Esain, V., Carroll, K.J., Kanz, L., North, T.E., and Lengerke, C. (2016). Evi1 regulates Notch activation to induce zebrafish hematopoietic stem cell emergence. EMBO J. *35*, 2315–2331.

Koushik, S.V., Wang, J., Rogers, R., Moskophidis, D., Lambert, N.A., Creazzo, T.L., and Conway, S.J. (2001). Targeted inactivation of the sodium-calcium exchanger (Ncx1) results in the lack of a heartbeat and abnormal myofibrillar organization. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *15*, 1209–1211.

Krause, D.S., Fackler, M.J., Civin, C.I., and May, W.S. (1996). CD34: structure, biology, and clinical utility. Blood *87*, 1–13.

Krishnakumar, R., and Blelloch, R.H. (2013). Epigenetics of Cellular Reprogramming. Curr. Opin. Genet. Dev. *23*, 548–555.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Kulessa, H., Frampton, J., and Graf, T. (1995). GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. Genes Dev. *9*, 1250–1262.

Kumano, K., Chiba, S., Kunisato, A., Sata, M., Saito, T., Nakagami-Yamaguchi, E., Yamaguchi, T., Masuda, S., Shimizu, K., Takahashi, T., et al. (2003). Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells. Immunity *18*, 699–711.

Kuo, Y.-Y., and Chang, Z.-F. (2007). GATA-1 and Gfi-1B Interplay To Regulate Bcl-xL Transcription. Mol. Cell. Biol. *27*, 4261–4272.

Kyba, M., Perlingeiro, R.C., and Daley, G.Q. (2002). HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. Cell *109*, 29–37.

Labastie, M.-C., Cortés, F., Roméo, P.-H., Dulac, C., and Péault, B. (1998). Molecular identity of hematopoietic precursor cells emerging in the human embryo. Blood *92*, 3624–3635.

Lacaud, G., and Kouskoff, V. (2017). Hemangioblast, hemogenic endothelium, and primitive versus definitive hematopoiesis. Exp. Hematol. *49*, 19–24.

Lacaud, G., Gore, L., Kennedy, M., Kouskoff, V., Kingsley, P., Hogan, C., Carlsson, L., Speck, N.A., Palis, J., and Keller, G. (2002). Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro. Blood *100*, 458–466.

Lam, E.Y.N., Hall, C.J., Crosier, P.S., Crosier, K.E., and Flores, M.V. (2010). Live imaging of Runx1 expression in the dorsal aorta tracks the emergence of blood progenitors from endothelial cells. Blood *116*, 909–914.

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.

Lancrin, C., Mazan, M., Stefanska, M., Patel, R., Lichtinger, M., Costa, G., Vargel, O., Wilson, N.K., Moroy, T., Bonifer, C., et al. (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. Blood *120*, 314–322.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Larochelle, A., Savona, M., Wiggins, M., Anderson, S., Ichwan, B., Keyvanfar, K., Morrison, S.J., and Dunbar, C.E. (2011). Human and rhesus macaque hematopoietic stem cells cannot be purified based only on SLAM family markers. Blood *117*, 1550–1554.

Lassar, A.B., Paterson, B.M., and Weintraub, H. (1986). Transfection of a DNA locus that mediates the conversion of 10T12 fibroblasts to myoblasts. Cell *47*, 649–656.

Laurent, B., Randrianarison-Huetz, V., Frisan, E., Andrieu-Soler, C., Soler, E., Fontenay, M., Dusanter-Fourt, I., and Dumenil, D. (2012). A short Gfi-1B isoform controls erythroid differentiation by recruiting the LSD1-CoREST complex through the dimethylation of its SNAG domain. J. Cell Sci. *125*, 993–1002.

Lee, D., Park, C., Lee, H., Lugus, J.J., Kim, S.H., Arentson, E., Chung, Y.S., Gomez, G., Kyba, M., Lin, S., et al. (2008). ER71 Acts Downstream of BMP, Notch, and Wnt Signaling in Blood and Vessel Progenitor Specification. Cell Stem Cell *2*, 497–507.

Lee, S.-Y., Yoon, J., Lee, M.-H., Jung, S.K., Kim, D.J., Bode, A.M., Kim, J., and Dong, Z. (2012). The Role of Heterodimeric AP-1 Protein Comprised of JunD and c-Fos Proteins in Hematopoiesis. J. Biol. Chem. *287*, 31342–31348.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. *1*, 729–748.

Lee, Y., Manegold, J.E., Kim, A.D., Pouget, C., Stachura, D.L., Clements, W.K., and Traver, D. (2014). FGF signalling specifies haematopoietic stem cells through its regulation of somitic Notch signalling. Nat. Commun. *5*, 5583.

Lemischka, I.R. (2005). Stem Cell Biology: A View toward the Future. Ann. N. Y. Acad. Sci. *1044*, 132–138.

Lemischka, I.R., Raulet, D.H., and Mulligan, R.C. (1986). Developmental potential and dynamic behavior of hematopoietic stem cells. Cell *45*, 917–927.

Lentjes, M.H., Niessen, H.E., Akiyama, Y., de Bruïne, A.P., Melotte, V., and van Engeland, M. (2016). The emerging role of GATA transcription factors in development and disease. Expert Rev. Mol. Med. *18*.

Leung, A., Ciau-Uitz, A., Pinheiro, P., Monteiro, R., Zuo, J., Vyas, P., Patient, R., and Porcher, C. (2013). Uncoupling VEGFA Functions in Arteriogenesis and Hematopoietic Stem Cell Specification. Dev. Cell *24*, 144–158.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, P., Lahvic, J.L., Binder, V., Pugach, E.K., Riley, E.B., Tamplin, O.J., Panigrahy, D., Bowman, T.V., Barrett, F.G., Heffner, G.C., et al. (2015). Epoxyeicosatrienoic acids enhance embryonic haematopoiesis and adult marrow engraftment. Nature *523*, 468–471.

Li, Y., Esain, V., Teng, L., Xu, J., Kwan, W., Frost, I.M., Yzaguirre, A.D., Cai, X., Cortes, M., Maijenburg, M.W., et al. (2014). Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell production. Genes Dev. *28*, 2597–2612.

Li, Z., Lan, Y., He, W., Chen, D., Wang, J., Zhou, F., Wang, Y., Sun, H., Chen, X., Xu, C., et al. (2012). Mouse Embryonic Head as a Site for Hematopoietic Stem Cell Development. Cell Stem Cell *11*, 663–675.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lie, A.L.M., Marinopoulou, E., Li, Y., Patel, R., Stefanska, M., Bonifer, C., Miller, C., Kouskoff, V., and Lacaud, G. (2014). RUNX1 positively regulates a cell adhesion and migration program in murine hemogenic endothelium prior to blood emergence. Blood *124*, e11–e20.

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

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.

Linnemann, A.K., O'Geen, H., Keles, S., Farnham, P.J., and Bresnick, E.H. (2011). Genetic framework for GATA factor function in vascular biology. Proc. Natl. Acad. Sci. *108*, 13641–13646.

Lis, R., Karrasch, C.C., Poulos, M.G., Kunar, B., Redmond, D., Duran, J.G.B., Badwe, C.R., Schachterle, W., Ginsberg, M., Xiang, J., et al. (2017). Conversion of adult endothelium to immunocompetent haematopoietic stem cells. Nature *545*, 439–445.

Liu, F., Walmsley, M., Rodaway, A., and Patient, R. (2008). Fli1 Acts at the Top of the Transcriptional Network Driving Blood and Endothelial Development. Curr. Biol. *18*, 1234–1240.

Lux, C.T., Yoshimoto, M., McGrath, K., Conway, S.J., Palis, J., and Yoder, M.C. (2008). All primitive and definitive hematopoietic progenitor cells emerging before E10 in the mouse embryo are products of the yolk sac. Blood *111*, 3435–3438.

Majeti, R., Park, C.Y., and Weissman, I.L. (2007). Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood. Cell Stem Cell *1*, 635–645.

Mandal, L., Banerjee, U., and Hartenstein, V. (2004). Evidence for a fruit fly hemangioblast and similarities between lymph-gland hematopoiesis in fruit fly and mammal aorta-gonadal-mesonephros mesoderm. Nat. Genet. *36*, 1019–1023.

Marcelo, K.L., Goldie, L.C., and Hirschi, K.K. (2013). Regulation of Endothelial Cell Differentiation and Specification. Circ. Res. *112*, 1272–1287.

Marconcini, L., Marchiò, S., Morbidelli, L., Cartocci, E., Albini, A., Ziche, M., Bussolino, F., and Oliviero, S. (1999). c-fos-induced growth factor/vascular endothelial growth factor D induces angiogenesis in vivo and in vitro. Proc. Natl. Acad. Sci. U. S. A. *96*, 9671–9676.

Martin, C., Beaupain, D., and Dieterlen-Lievre, F. (1978). Developmental relationships between vitelline and intra-embryonic haemopoiesis studied in avian 'yolk sac chimaeras.' Cell Differ. *7*, 115–130.

Martowicz, M.L., Grass, J.A., Boyer, M.E., Guend, H., and Bresnick, E.H. (2005). Dynamic GATA Factor Interplay at a Multicomponent Regulatory Region of the *GATA-2* Locus. J. Biol. Chem. *280*, 1724–1732.

Mascarenhas, M.I., Parker, A., Dzierzak, E., and Ottersbach, K. (2009). Identification of novel regulators of hematopoietic stem cell development through refinement of stem cell localization and expression profiling. Blood *114*, 4645–4653.

Matsumoto, K., Isagawa, T., Nishimura, T., Ogaeri, T., Eto, K., Miyazaki, S., Miyazaki, J., Aburatani, H., Nakauchi, H., and Ema, H. (2009). Stepwise Development of Hematopoietic Stem Cells from Embryonic Stem Cells. PLoS ONE *4*, e4820.

May, G., Soneji, S., Tipping, A.J., Teles, J., McGowan, S.J., Wu, M., Guo, Y., Fugazza, C., Brown, J., Karlsson, G., et al. (2013). Dynamic Analysis of Gene Expression and Genome-wide Transcription Factor Binding during Lineage Specification of Multipotent Progenitors. Cell Stem Cell *13*, 754–768.

Mayani, H., Dragowska, W., and Lansdorp, P.M. (1993). Characterization of functionally distinct subpopulations of CD34+ cord blood cells in serum-free long-term cultures supplemented with hematopoietic cytokines. Blood *82*, 2664–2672.

McDermott, S.P., Eppert, K., Lechman, E.R., Doedens, M., and Dick, J.E. (2010). Comparison of human cord blood engraftment between immunocompromised mouse strains. Blood *116*, 193–200.

McGarvey, A.C., Rybtsov, S., Souilhol, C., Tamagno, S., Rice, R., Hills, D., Godwin, D., Rice, D., Tomlinson, S.R., and Medvinsky, A. (2017). A molecular roadmap of the AGM region reveals BMPER as a novel regulator of HSC maturation. J. Exp. Med. jem.20162012.

McGrath, K.E., Frame, J.M., Fegan, K.H., Bowen, J.R., Conway, S.J., Catherman, S.C., Kingsley, P.D., Koniski, A.D., and Palis, J. (2015). Distinct sources of hematopoietic progenitors emerge before HSCs and provide functional blood cells in the mammalian embryo. Cell Rep. *11*, 1892–1904.

McKinney-Freeman, S., Cahan, P., Li, H., Lacadie, S.A., Huang, H.-T., Curran, M., Loewer, S., Naveiras, O., Kathrein, K.L., Konantz, M., et al. (2012). The Transcriptional Landscape of Hematopoietic Stem Cell Ontogeny. Cell Stem Cell *11*, 701–714.

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

Medvinsky, A., Rybtsov, S., and Taoudi, S. (2011). Embryonic origin of the adult hematopoietic system: advances and questions. Development *138*, 1017–1031.

Medvinsky, A.L., Samoylina, N.L., Müller, A.M., and Dzierzak, E. (1993). An early pre-liver intraembryonic source of CFU-S in the developing mouse. Nature *364*, 64–67.

Melchers, F. (1979). Murine embryonic B lymphocyte development in the placenta. Nature *277*, 219–221.

Migliaccio, G., Migliaccio, A.R., Petti, S., Mavilio, F., Russo, G., Lazzaro, D., Testa, U., Marinucci, M., and Peschle, C. (1986). Human embryonic hemopoiesis. Kinetics of progenitors and precursors underlying the yolk sac----liver transition. J. Clin. Invest. *78*, 51–60.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55.

Mikkola, H.K.A., and Orkin, S.H. (2006). The journey of developing hematopoietic stem cells. Development *133*, 3733–3744.

Minegishi, N., Suzuki, N., Yokomizo, T., Pan, X., Fujimoto, T., Takahashi, S., Hara, T., Miyajima, A., Nishikawa, S.-I., and Yamamoto, M. (2003). Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos. Blood *102*, 896–905.

Mitchell, R., Szabo, E., Shapovalova, Z., Aslostovar, L., Makondo, K., and Bhatia, M. (2014a). Molecular Evidence for OCT4-Induced Plasticity in Adult Human Fibroblasts Required for Direct Cell Fate Conversion to Lineage Specific Progenitors. STEM CELLS *32*, 2178–2187.

Mitchell, R.R., Szabo, E., Benoit, Y.D., Case, D.T., Mechael, R., Alamilla, J., Lee, J.H., Fiebig-Comyn, A., Gillespie, D.C., and Bhatia, M. (2014b). Activation of Neural Cell Fate Programs Toward Direct Conversion of Adult Human Fibroblasts into Tri-Potent Neural Progenitors Using OCT-4. Stem Cells Dev. *23*, 1937–1946.

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

Monteferrario, D., Bolar, N.A., Marneth, A.E., Hebeda, K.M., Bergevoet, S.M., Veenstra, H., Laros-van Gorkom, B.A.P., MacKenzie, M.A., Khandanpour, C., Botezatu, L., et al. (2014). A Dominant-Negative *GFI1B* Mutation in the Gray Platelet Syndrome. N. Engl. J. Med. *370*, 245–253.

Moore, M.A.S., and Metcalf, D. (1970). Ontogeny of the Haemopoietic System: Yolk Sac Origin of In Vivo and In Vitro Colony Forming Cells in the Developing Mouse Embryo. Br. J. Haematol. *18*, 279–296.

Moore, M.A.S., and Owen, J.J.T. (1967). Chromosome Marker Studies in the Irradiated Chick Embryo. Nature *215*, 1081–1982.

Moore, K.A., Ema, H., and Lemischka, I.R. (1997). In vitro maintenance of highly purified, transplantable hematopoietic stem cells. Blood *89*, 4337–4347.

Morrison, S.J., Uchida, N., and Weissman, I.L. (1995). The biology of hematopoietic stem cells. Annu. Rev. Cell Dev. Biol. *11*, 35–71.

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

Motabi, I.H., and DiPersio, J.F. (2012). Advances in stem cell mobilization. Blood Rev. *26*, 267–278.

Muench, M.O., Kapidzic, M., Gormley, M., Gutierrez, A.G., Ponder, K.L., Fomin, M.E., Beyer, A.I., Stolp, H., Qi, Z., Fisher, S.J., et al. (2017). The human chorion contains definitive hematopoietic stem cells from the 15 $^{th}$ week of gestation. Development dev.138438.

Mukouyama, Y., Chiba, N., Hara, T., Okada, H., Ito, Y., Kanamaru, R., Miyajima, A., Satake, M., and Watanabe, T. (2000). The AML1 Transcription Factor Functions to Develop and Maintain Hematogenic Precursor Cells in the Embryonic Aorta–Gonad–Mesonephros Region. Dev. Biol. *220*, 27–36.

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

Murray, P.D.F. (1932). The Development in vitro of the Blood of the Early Chick Embryo. Proc. R. Soc. B *111*, 497–521.

Nakajima-Takagi, Y., Osawa, M., Oshima, M., Takagi, H., Miyagi, S., Endoh, M., Endo, T.A., Takayama, N., Eto, K., Toyoda, T., et al. (2013). Role of SOX17 in hematopoietic development from human embryonic stem cells. Blood *121*, 447–458.

Nakamura-Ishizu, A., and Suda, T. (2013). Hematopoietic stem cell niche: An interplay among a repertoire of multiple functional niches. Biochim. Biophys. Acta BBA - Gen. Subj. *1830*, 2404–2409.

Nakano, H., Liu, X., Arshi, A., Nakashima, Y., van Handel, B., Sasidharan, R., Harmon, A.W., Shin, J.-H., Schwartz, R.J., Conway, S.J., et al. (2013). Haemogenic endocardium contributes to transient definitive haematopoiesis. Nat. Commun. *4*, 1564.

Nam, Y.-J., Song, K., Luo, X., Daniel, E., Lambeth, K., West, K., Hill, J.A., DiMaio, J.M., Baker, L.A., Bassel-Duby, R., et al. (2013). Reprogramming of human fibroblasts toward a cardiac fate. Proc. Natl. Acad. Sci. U. S. A. *110*, 5588–5593.

Natoli, G. (2010). NF-κB: no longer an island, but a piece of a continent. EMBO Rep. *11*, 246–248.

Ng, C.E.L., Yokomizo, T., Yamashita, N., Cirovic, B., Jin, H., Wen, Z., Ito, Y., and Osato, M. (2010). A Runx1 Intronic Enhancer Marks Hemogenic Endothelial Cells and Hematopoietic Stem Cells. STEM CELLS *28*, 1869–1881.

Ng, E.S., Azzola, L., Bruveris, F.F., Calvanese, V., Phipson, B., Vlahos, K., Hirst, C., Jokubaitis, V.J., Yu, Q.C., Maksimovic, J., et al. (2016). Differentiation of human embryonic stem cells to HOXA+ hemogenic vasculature that resembles the aorta-gonad-mesonephros. Nat. Biotechnol. *34*, 1168–1179.

Nicolini, F.E., Cashman, J.D., Hogge, D.E., Humphries, R.K., and Eaves, C.J. (2004). NOD/SCID mice engineered to express human IL-3, GM-CSF and Steel factor constitutively mobilize engrafted human progenitors and compromise human stem cell regeneration. Leukemia *18*, 341–347.

Nie, B., Wang, H., Laurent, T., and Ding, S. (2012). Cellular reprogramming: a small molecule perspective. Curr. Opin. Cell Biol. *24*, 784–792.

Nishikawa, S.-I., Nishikawa, S., Kawamoto, H., Yoshida, H., Kizumoto, M., Kataoka, H., and Katsura, Y. (1998). In vitro generation of lymphohematopoietic cells from endothelial cells purified from murine embryos. Immunity *8*, 761–769.

Nobuhisa, I., Osawa, M., Uemura, M., Kishikawa, Y., Anani, M., Harada, K., Takagi, H., Saito, K., Kanai-Azuma, M., Kanai, Y., et al. (2014). Sox17-Mediated Maintenance of Fetal Intra-Aortic Hematopoietic Cell Clusters. Mol. Cell. Biol. *34*, 1976–1990.

Norkin, M., Lazarus, H., and Wingard, J. (2013). Umbilical cord blood graft enhancement strategies: has the time come to move these into the clinic? Bone Marrow Transplant. *48*, 884–889.

North, T., Gu, T.-L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marín-Padilla, M., and Speck, N.A. (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. Development *126*, 2563–2575.

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

Notta, F., Doulatov, S., Laurenti, E., Poeppl, A., Jurisica, I., and Dick, J.E. (2011). Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. Science *333*, 218–221.

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., et al. (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. Blood *110*, 4188–4197.

Oberlin, E., Tavian, M., Blazsek, I., and Péault, B. (2002). Blood-forming potential of vascular endothelium in the human embryo. Development *129*, 4147–4157.

Obier, N., Cauchy, P., Assi, S.A., Gilmour, J., Lie-A-Ling, M., Lichtinger, M., Hoogenkamp, M., Noailles, L., Cockerill, P.N., Lacaud, G., et al. (2016). Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate. Development dev.139857.

Ogawa, M., Tajima, F., Ito, T., Sato, T., Laver, J.H., and Deguchi, T. (2001). CD34 expression by murine hematopoietic stem cells. Ann. N. Y. Acad. Sci. *938*, 139–145.

Ohbo, K., Suda, T., Hashiyama, M., Mantani, A., Ikebe, M., Miyakawa, K., Moriyama, M., Nakamura, M., Katsuki, M., and Takahashi, K. (1996). Modulation of hematopoiesis in mice with a truncated mutant of the interleukin-2 receptor gamma chain. Blood *87*, 956–967.

Oka, C., Nakano, T., Wakeham, A., de la Pompa, J.L., Mori, C., Sakai, T., Okazaki, S., Kawaichi, M., Shiota, K., and Mak, T.W. (1995). Disruption of the mouse RBP-J kappa gene results in early embryonic death. Development *121*, 3291–3301.

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.

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

Ottersbach, K., and Dzierzak, E. (2005). The Murine Placenta Contains Hematopoietic Stem Cells within the Vascular Labyrinth Region. Dev. Cell *8*, 377–387.

Ottersbach, K., Smith, A., Wood, A., and Göttgens, B. (2010). Ontogeny of haematopoiesis: recent advances and open questions. Br. J. Haematol. *148*, 343–355.

Ozawa, Y., Towatari, M., Tsuzuki, S., Hayakawa, F., Maeda, T., Miyata, Y., Tanimoto, M., and Saito, H. (2001). Histone deacetylase 3 associates with and represses the transcription factor GATA-2. Blood *98*, 2116–2123.

Palis, J., and Yoder, M.C. (2001). Yolk-sac hematopoiesis: the first blood cells of mouse and man. Exp. Hematol. *29*, 927–936.

Palis, J., Robertson, S., Kennedy, M., Wall, C., and Keller, G. (1999). Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. Development *126*, 5073–5084.

Park, C.Y., Majeti, R., and Weissman, I.L. (2008). In vivo evaluation of human hematopoiesis through xenotransplantation of purified hematopoietic stem cells from umbilical cord blood. Nat. Protoc. *3*, 1932–1940.

Passweg, J.R., Baldomero, H., Bader, P., Bonini, C., Cesaro, S., Dreger, P., Duarte, R.F., Dufour, C., Kuball, J., Farge-Bancel, D., et al. (2016). Hematopoietic stem cell transplantation in Europe 2014: more than 40 000 transplants annually. Bone Marrow Transplant. *51*, 786–792.

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

Patient, R.K., and McGhee, J.D. (2002). The GATA family (vertebrates and invertebrates). Curr. Opin. Genet. Dev. *12*, 416–422.

Pereira, C.F., Terranova, R., Ryan, N.K., Santos, J., Morris, K.J., Cui, W., Merkenschlager, M., and Fisher, A.G. (2008). Heterokaryon-Based Reprogramming of Human B Lymphocytes for Pluripotency Requires Oct4 but Not Sox2. PLoS Genet. *4*, e1000170.

Pereira, C.F., Piccolo, F.M., Tsubouchi, T., Sauer, S., Ryan, N.K., Bruno, L., Landeira, D., Santos, J., Banito, A., Gil, J., et al. (2010). ESCs Require PRC2 to Direct the Successful Reprogramming of Differentiated Cells toward Pluripotency. Cell Stem Cell *6*, 547–556.

Pereira, C.-F., Lemischka, I.R., and Moore, K. (2012). Reprogramming cell fates: insights from combinatorial approaches: Reprogramming factors to instruct cell fate. Ann. N. Y. Acad. Sci. *1266*, 7–17.

Pereira, C.-F., Chang, B., Qiu, J., Niu, X., Papatsenko, D., Hendry, C.E., Clark, N.R., Nomura-Kitabayashi, A., Kovacic, J.C., Ma'ayan, A., et al. (2013). Induction of a Hemogenic Program in Mouse Fibroblasts. Cell Stem Cell *13*, 205–218.

Pereira, C.-F., Chang, B., Gomes, A., Bernitz, J., Papatsenko, D., Niu, X., Swiers, G., Azzoni, E., de Bruijn, M.F.T.R., Schaniel, C., et al. (2016). Hematopoietic Reprogramming In Vitro Informs In Vivo Identification of Hemogenic Precursors to Definitive Hematopoietic Stem Cells. Dev. Cell *36*, 525–539.

Perlin, J.R., Sporrij, A., and Zon, L.I. (2017). Blood on the tracks: hematopoietic stem cell-endothelial cell interactions in homing and engraftment. J. Mol. Med. *95*, 809–819.

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

Petrenko, O., Beavis, A., Klaine, M., Kittappa, R., Godin, I., and Lemischka, I.R. (1999). The molecular characterization of the fetal stem cell marker AA4. Immunity *10*, 691–700.

Pinello, L., Xu, J., Orkin, S.H., and Yuan, G.-C. (2014). Analysis of chromatin-state plasticity identifies cell-type-specific regulators of H3K27me3 patterns. Proc. Natl. Acad. Sci. *111*, E344–E353.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. Cell *151*, 1617–1632.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., and Orkin, S.H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell *86*, 47–57.

Prashad, S.L., Calvanese, V., Yao, C.Y., Kaiser, J., Wang, Y., Sasidharan, R., Crooks, G., Magnusson, M., and Mikkola, H.K.A. (2015). GPI-80 Defines Self-Renewal Ability in Hematopoietic Stem Cells during Human Development. Cell Stem Cell *16*, 80–87.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rahmig, S., Kronstein-Wiedemann, R., Fohgrub, J., Kronstein, N., Nevmerzhitskaya, A., Bornhäuser, M., Gassmann, M., Platz, A., Ordemann, R., Tonn, T., et al. (2016). Improved Human Erythropoiesis and Platelet Formation in Humanized NSGW41 Mice. Stem Cell Rep. *7*, 591–601.

Ramalho-Santos, M., and Willenbring, H. (2007). On the Origin of the Term “Stem Cell.” Cell Stem Cell *1*, 35–38.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nat. Genet. *38*, 500–501.

Rhee, C., Lee, B.-K., Beck, S., LeBlanc, L., Tucker, H.O., and Kim, J. (2017). Mechanisms of transcription factor-mediated direct reprogramming of mouse embryonic stem cells to trophoblast stem-like cells. Nucleic Acids Res. *45*, 10103–10114.

Rhodes, K.E., Gekas, C., Wang, Y., Lux, C.T., Francis, C.S., Chan, D.N., Conway, S., Orkin, S.H., Yoder, M.C., and Mikkola, H.K.A. (2008). The Emergence of Hematopoietic Stem Cells Is Initiated in the Placental Vasculature in the Absence of Circulation. Cell Stem Cell *2*, 252–263.

Riddell, J., Gazit, R., Garrison, B.S., Guo, G., Saadatpour, A., Mandal, P.K., Ebina, W., Volchkov, P., Yuan, G.-C., Orkin, S.H., et al. (2014). Reprogramming Committed Murine Blood Cells to Induced Hematopoietic Stem Cells with Defined Factors. Cell *157*, 549–564.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47.

Robert-Moreno, A. (2005). RBPj -dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells. Development *132*, 1117–1126.

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

Robert-Moreno, À., Guiu, J., Ruiz-Herguido, C., López, M.E., Inglés-Esteve, J., Riera, L., Tipping, A., Enver, T., Dzierzak, E., Gridley, T., et al. (2008). Impaired embryonic haematopoiesis yet normal arterial development in the absence of the Notch ligand Jagged1. EMBO J. *27*, 1886–1895.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., and Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. *12*, R22.

Robin, C., Ottersbach, K., Durand, C., Peeters, M., Vanes, L., Tybulewicz, V., and Dzierzak, E. (2006). An Unexpected Role for IL-3 in the Embryonic Development of Hematopoietic Stem Cells. Dev. Cell *11*, 171–180.

Robin, C., Bollerot, K., Mendes, S., Haak, E., Crisan, M., Cerisoli, F., Lauw, I., Kaimakis, P., Jorna, R., Vermeulen, M., et al. (2009). Human Placenta Is a Potent Hematopoietic Niche Containing Hematopoietic Stem and Progenitor Cells throughout Development. Cell Stem Cell *5*, 385–395.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Rodrigues, N.P. (2005). Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. Blood *106*, 477–484.

Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K.E., Guyot, B., Heck, A.J., Vyas, P., de Boer, E., Grosveld, F., and Strouboulis, J. (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. EMBO J. *24*, 2354–2366.

Rongvaux, A., Takizawa, H., Strowig, T., Willinger, T., Eynon, E.E., Flavell, R.A., and Manz, M.G. (2013). Human Hemato-Lymphoid System Mice: Current Use and Future Potential for Medicine. Annu. Rev. Immunol. *31*, 635–674.

Rossant, J., and Cross, J.C. (2001). Placental development: lessons from mouse mutants. Nat. Rev. Genet. *2*, 538.

Rowlinson, J.M., and Gering, M. (2010). Hey2 acts upstream of Notch in hematopoietic stem cell specification in zebrafish embryos. Blood *116*, 2046–2056.

Ruiz-Herguido, C., Guiu, J., D'Altri, T., Inglés-Esteve, J., Dzierzak, E., Espinosa, L., and Bigas, A. (2012). Hematopoietic stem cell development requires transient Wnt/β-catenin activity. J. Exp. Med. *209*, 1457–1468.

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

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O., and Medvinsky, A. (2014). Tracing the Origin of the HSC Hierarchy Reveals an SCF-Dependent, IL-3-Independent CD43− Embryonic Precursor. Stem Cell Rep. *3*, 489–501.

Sabin, F.R. (1921). STUDIES ON BLOOD: THE VITALLY STAINABLE GRANULES AS A SPECIFIC CRITERION FOR ERYTHRO--- BLASTS AND THE DIFFERENTIATION OF ,THE THREE STRAINS OF THE WHITE BLOOD---CELLS AS SEEN IN THE LIVING CHICK'S YOLK-SAC. John Hopkins Hosp. Bull. *368*, 314–321.

Saleque, S., Cameron, S., and Orkin, S.H. (2002). The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. Genes Dev. *16*, 301–306.

Saleque, S., Kim, J., Rooke, H.M., and Orkin, S.H. (2007). Epigenetic Regulation of Hematopoietic Differentiation by Gfi-1 and Gfi-1b Is Mediated by the Cofactors CoREST and LSD1. Mol. Cell *27*, 562–572.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming. Cell Stem Cell *7*, 64–77.

Samokhvalov, I.M., Samokhvalova, N.I., and Nishikawa, S. (2007). Cell tracing shows the contribution of the yolk sac to adult haematopoiesis. Nature *446*, 1056–1061.

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

Sandler, V.M., Lis, R., Liu, Y., Kedem, A., James, D., Elemento, O., Butler, J.M., Scandura, J.M., and Rafii, S. (2014). Reprogramming human endothelial cells to haematopoietic cells requires vascular induction. Nature *511*, 312–318.

Sasaki, K., Yagi, H., Bronson, R.T., Tominaga, K., Matsunashi, T., Deguchi, K., Tani, Y., Kishimoto, T., and Komori, T. (1996). Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor beta. Proc. Natl. Acad. Sci. *93*, 12359–12363.

Sassone-Corsi, P., Sisson, J.C., and Verma, I.M. (1988). Transcriptional autoregulation of the proto-oncogene fos. Nature *334*, 314–319.

Sawamiphak, S., Kontarakis, Z., and Stainier, D.Y.R. (2014). Interferon Gamma Signaling Positively Regulates Hematopoietic Stem Cell Emergence. Dev. Cell *31*, 640–653.

Schmitt, C.E., Lizama, C.O., and Zovein, A.C. (2014). From transplantation to transgenics: Mouse models of developmental hematopoiesis. Exp. Hematol. *42*, 707–716.

Schneuwly, S., Klemenz, R., and Gehring, W.J. (1987). Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. Nature *325*, 816–818.

Schulz, D., Vassen, L., Chow, K.T., McWhirter, S.M., Amin, R.H., Möröy, T., and Schlissel, M.S. (2012). Gfi1b negatively regulates Rag expression directly and via the repression of FoxO1. J. Exp. Med. *209*, 187–199.

Sekiya, S., and Suzuki, A. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. Nature *475*, 390–393.

Shalaby, F., Ho, J., Stanford, W.L., Fisher, K.-D., Schuh, A.C., Schwartz, L., Bernstein, A., and Rossant, J. (1997). A Requirement for Flk1 in Primitive and Definitive Hematopoiesis and Vasculogenesis. Cell *89*, 981–990.

Shaulian, E., and Karin, M. (2001). AP-1 in cell proliferation and survival. Oncogene *20*.

Shu, J., Zhang, K., Zhang, M., Yao, A., Shao, S., Du, F., Yang, C., Chen, W., Wu, C., Yang, W., et al. (2015). GATA family members as inducers for cellular reprogramming to pluripotency. Cell Res.

Shultz, L.D., Ishikawa, F., and Greiner, D.L. (2007). Humanized mice in translational biomedical research. Nat. Rev. Immunol. *7*, 118–130.

Silva, J., Chambers, I., Pollard, S., and Smith, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. Nature *441*, 997–1001.

Sinka, L., Biasch, K., Khazaal, I., Peault, B., and Tavian, M. (2012). Angiotensin-converting enzyme (CD143) specifies emerging lympho-hematopoietic progenitors in the human embryo. Blood *119*, 3712–3723.

Sitnicka, E. (2003). Human CD34+ hematopoietic stem cells capable of multilineage engrafting NOD/SCID mice express flt3: distinct flt3 and c-kit expression and response patterns on mouse and candidate human hematopoietic stem cells. Blood *102*, 881–886.

Snow, J.W., Trowbridge, J.J., Johnson, K.D., Fujiwara, T., Emambokus, N.E., Grass, J.A., Orkin, S.H., and Bresnick, E.H. (2011). Context-dependent function of “GATA switch” sites in vivo. Blood *117*, 4769–4772.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and Impediments of the Pluripotency Reprogramming Factors’ Initial Engagement with the Genome. Cell *151*, 994–1004.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. Cell *161*, 555–568.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining Molecular Cornerstones during Fibroblast to iPS Cell Reprogramming in Mouse. Cell Stem Cell *2*, 230–240.

Stevenson, W.S., Morel-Kopp, M.-C., Chen, Q., Liang, H.P., Bromhead, C.J., Wright, S., Turakulov, R., Ng, A.P., Roberts, A.W., Bahlo, M., et al. (2013). GFI1B mutation causes a bleeding disorder with abnormal platelet function. J. Thromb. Haemost. *11*, 2039–2047.

Sturgeon, C.M., Ditadi, A., Awong, G., Kennedy, M., and Keller, G. (2014). Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. Nat. Biotechnol. *32*, 554–561.

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

Sugimura, R., Jha, D.K., Han, A., Soria-Valles, C., da Rocha, E.L., Lu, Y.-F., Goettel, J.A., Serrao, E., Rowe, R.G., Malleshaiah, M., et al. (2017). Haematopoietic stem and progenitor cells from human pluripotent stem cells. Nature *545*, 432–438.

Sugiyama, D., Ogawa, M., Hirose, I., Jaffredo, T., Arai, K., and Tsuji, K. (2003). Erythropoiesis from acetyl LDL incorporating endothelial cells at the preliver stage. Blood *101*, 4733–4738.

Sumanas, S., Gomez, G., Zhao, Y., Park, C., Choi, K., and Lin, S. (2008). Interplay among Etsrp/ER71, Scl, and Alk8 signaling controls endothelial and myeloid cell formation. Blood *111*, 4500–4510.

Suzuki, N., Yamazaki, S., Yamaguchi, T., Okabe, M., Masaki, H., Takaki, S., Otsu, M., and Nakauchi, H. (2013). Generation of Engraftable Hematopoietic Stem Cells From Induced Pluripotent Stem Cells by Way of Teratoma Formation. Mol. Ther. *21*, 1424–1431.

Swiers, G., De Bruijn, M., and Speck, N.A. (2010). Hematopoietic stem cell emergence in the conceptus and the role of Runx1. Int. J. Dev. Biol. *54*, 1151–1163.

Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K.D., et al. (2013). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. Nat. Commun. *4*.

Szabo, E., Rampalli, S., Risueño, R.M., Schnerch, A., Mitchell, R., Fiebig-Comyn, A., Levadoux-Martin, M., and Bhatia, M. (2010). Direct conversion of human fibroblasts to multilineage blood progenitors. Nature *468*, 521–526.

Szalóki, N., Krieger, J.W., Komáromi, I., Tóth, K., and Vámosi, G. (2015). Evidence for Homodimerization of the c-Fos Transcription Factor in Live Cells Revealed by Fluorescence Microscopy and Computer Modeling. Mol. Cell. Biol. *35*, 3785–3798.

Tachibana, M., Amato, P., Sparman, M., Gutierrez, N.M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.-S., Sritanaudomchai, H., et al. (2013). Human Embryonic Stem Cells Derived by Somatic Cell Nuclear Transfer. Cell *153*, 1228–1238.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr. Biol. *11*, 1553–1558.

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.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861–872.

Takaku, M., Grimm, S.A., Shimbo, T., Perera, L., Menafra, R., Stunnenberg, H.G., Archer, T.K., Machida, S., Kurumizaka, H., and Wade, P.A. (2016). GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. Genome Biol. *17*.

Takayama, N., Nishikii, H., Usui, J., Tsukui, H., Sawaguchi, A., Hiroyama, T., Eto, K., and Nakauchi, H. (2008). Generation of functional platelets from human embryonic stem cells in vitro via ES-sacs, VEGF-promoted structures that concentrate hematopoietic progenitors. Blood *111*, 5298–5306.

Tan, Y.-T., Ye, L., Xie, F., Beyer, A.I., Muench, M.O., Wang, J., Chen, Z., Liu, H., Chen, S.-J., and Kan, Y.W. (2018). Respecifying human iPSC-derived blood cells into highly engraftable hematopoietic stem and progenitor cells with a single factor. Proc. Natl. Acad. Sci. *115*, 2180–2185.

Taoudi, S., and Medvinsky, A. (2007). Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta. Proc. Natl. Acad. Sci. *104*, 9399–9403.

Taoudi, S., Gonneau, C., Moore, K., Sheridan, J.M., Blackburn, C.C., Taylor, E., and Medvinsky, A. (2008). Extensive Hematopoietic Stem Cell Generation in the AGM Region via Maturation of VE-Cadherin+CD45+ Pre-Definitive HSCs. Cell Stem Cell *3*, 99–108.

Tashiro, K., Kawabata, K., Omori, M., Yamaguchi, T., Sakurai, F., Katayama, K., Hayakawa, T., and Mizuguchi, H. (2012). Promotion of hematopoietic differentiation from mouse induced pluripotent stem cells by transient HoxB4 transduction. Stem Cell Res. *8*, 300–311.

Tavian, M., and Peault, B. (2005). Embryonic development of the human hematopoietic system. Int. J. Dev. Biol. *49*, 243–250.

Tavian, M., Coulombel, L., Luton, D., Clemente, H.S., Dieterlen-Lievre, F., and Peault, B. (1996). Aorta-associated CD34+ hematopoietic cells in the early human embryo. Blood *87*, 67–72.

Tavian, M., Hallais, M.-F., and Péault, B. (1999). Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. Development *126*, 793–803.

Thambyrajah, R., Mazan, M., Patel, R., Moignard, V., Stefanska, M., Marinopoulou, E., Li, Y., Lancrin, C., Clapes, T., Möröy, T., et al. (2016). GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. Nat. Cell Biol. *18*, 21–32.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Till, J.E., and McCulloch, E.A. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat. Res. *14*, 213–222.

Timmermans, F., Velghe, I., Vanwalleghem, L., De Smedt, M., Van Coppernolle, S., Taghon, T., Moore, H.D., Leclercq, G., Langerak, A.W., Kerre, T., et al. (2009). Generation of T Cells from Human Embryonic Stem Cell-Derived Hematopoietic Zones. J. Immunol. *182*, 6879–6888.

Tipping, A.J., Pina, C., Castor, A., Hong, D., Rodrigues, N.P., Lazzari, L., May, G.E., Jacobsen, S.E.W., 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.

Tong, B., Grimes, H.L., Yang, T.-Y., Bear, S.E., Qin, Z., Du, K., El-Deiry, W.S., and Tsichlis, P.N. (1998). The Gfi-1B Proto-Oncoprotein Represses $p21^{WAF1}$ and Inhibits Myeloid Cell Differentiation. Mol. Cell. Biol. *18*, 2462–2473.

Tong, Q., Dalgin, G., Xu, H., Ting, C.-N., Leiden, J.M., and Hotamisligil, G.S. (2000). Function of GATA Transcription Factors in Preadipocyte-Adipocyte Transition. Science *290*, 134–138.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Tsai, F.-Y., and Orkin, S.H. (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*, 3636–3643.

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.

Tsukada, M., Ota, Y., Wilkinson, A.C., Becker, H.J., Osato, M., Nakauchi, H., and Yamazaki, S. (2017). In Vivo Generation of Engraftable Murine Hematopoietic Stem Cells by Gfi1b, c-Fos, and Gata2 Overexpression within Teratoma. Stem Cell Rep. *9*, 1024–1033.

Turner, R., and Tijan, R. (1989). Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers. Science *243*, 1689–1694.

Unger, C., Kärner, E., Treschow, A., Stellan, B., Felldin, U., Concha, H., Wendel, M., Hovatta, O., Aints, A., Ährlund-Richter, L., et al. (2008). Lentiviral-Mediated HoxB4 Expression in Human Embryonic Stem Cells Initiates Early Hematopoiesis in a Dose-Dependent Manner but Does Not Promote Myeloid Differentiation. STEM CELLS *26*, 2455–2466.

Vassen, L., Fiolka, K., Mahlmann, S., and Möröy, T. (2005). Direct transcriptional repression of the genes encoding the zinc-finger proteins Gfi1b and Gfi1 by Gfi1b. Nucleic Acids Res. *33*, 987–998.

Vassen, L., Fiolka, K., and Moroy, T. (2006). Gfi1b alters histone methylation at target gene promoters and sites of gama-satalite containing heterochromatin. EMBO J. *25*, 2409.

Vassen, L., Okayama, T., and Moroy, T. (2007). Gfi1b:green fluorescent protein knock-in mice reveal a dynamic expression pattern of Gfi1b during hematopoiesis that is largely complementary to Gfi1. Blood *109*, 2356–2364.

Vassen, L., Beauchemin, H., Lemsaddek, W., Krongold, J., Trudel, M., and Möröy, T. (2014). Growth Factor Independence 1b (Gfi1b) Is Important for the Maturation of Erythroid Cells and the Regulation of Embryonic Globin Expression. PLoS ONE *9*, e96636.

Vereide, D.T., Vickerman, V., Swanson, S.A., Chu, L.-F., McIntosh, B.E., and Thomson, J.A. (2014). An Expandable, Inducible Hemangioblast State Regulated by Fibroblast Growth Factor. Stem Cell Rep. *3*, 1043–1057.

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

Vierbuchen, T., and Wernig, M. (2012). Molecular Roadblocks for Cellular Reprogramming. Mol. Cell *47*, 827–838.

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.

Visvader, J., and Adams, J.M. (1993). Megakaryocytic differentiation induced in 416B myeloid cells by GATA-2 and GATA-3 transgenes or 5-azacytidine is tightly coupled to GATA-1 expression. Blood *82*, 1493–1501.

Vogeli, K.M., Jin, S.-W., Martin, G.R., and Stainier, D.Y.R. (2006). A common progenitor for haematopoietic and endothelial lineages in the zebrafish gastrula. Nature *443*, 337–339.

Vorontsov, I.E., Kulakovskiy, I.V., and Makeev, V.J. (2013). Jaccard index based similarity measure to compare transcription factor binding site models. Algorithms Mol. Biol. *8*, 23.

Voyta, J.C., Kelsen, D.P., Butterfield, C.E., and Zetter, B.R. (1984). Identification and isolation of endothelial cells based on their increased uptake of acetylated-low density lipoprotein. J. Cell Biol. *99*, 2034–2040.

Walsh, N.C., Kenney, L.L., Jangalwe, S., Aryee, K.-E., Greiner, D.L., Brehm, M.A., and Shultz, L.D. (2017). Humanized Mouse Models of Clinical Disease. Annu. Rev. Pathol. Mech. Dis. *12*, 187–215.

Wang, D., Zhuang, L., Gao, B., Shi, C.-X., Cheung, J., Liu, M., Jin, T., and Wen, X.-Y. (2008). The Blimp-1 gene regulatory region directs EGFP expression in multiple hematopoietic lineages and testis in mice. Transgenic Res. *17*, 193–203.

Wang, L., Menendez, P., Shojaei, F., Li, L., Mazurier, F., Dick, J.E., Cerdan, C., Levac, K., and Bhatia, M. (2005). Generation of hematopoietic repopulating cells from human embryonic stem cells independent of ectopic *HOXB4* expression. J. Exp. Med. *201*, 1603–1614.

Wang, L.C., Kuo, F., Fujiwara, Y., Gilliland, D.G., Golub, T.R., and Orkin, S.H. (1997). Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. EMBO J. *16*, 4374–4383.

Wang, Q., Stacy, T., Miller, J.D., Lewis, A.F., Gu, T.-L., Huang, X., Bushweller, J.H., Bories, J.-C., Alt, F.W., Ryan, G., et al. (1996). The CBFβ Subunit Is Essential for CBFα2 (AML1) Function In Vivo. Cell *87*, 697–708.

Wang, Z.-Q., Ovitt, C., Grigoriadis, A.E., Möhle-Steinlein, U., Rüther, U., and Wagner, E.F. (1992). Bone and haematopoietic defects in mice lacking c-fos. Nature *360*, 741–745.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical Mechanisms for Direct Reprogramming of Fibroblasts to Neurons. Cell *155*, 621–635.

Wapinski, O.L., Lee, Q.Y., Chen, A.C., Li, R., Corces, M.R., Ang, C.E., Treutlein, B., Xiang, C., Baubet, V., Suchy, F.P., et al. (2017). Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. Cell Rep. *20*, 3236–3247.

Wardle, F.C., and Tan, H. (2015). A ChIP on the shoulder? Chromatin immunoprecipitation and validation strategies for ChIP antibodies. F1000Research *4*.

Wei, Y., Ma, D., Gao, Y., Zhang, C., Wang, L., and Liu, F. (2014). Ncor2 is required for hematopoietic stem cell emergence by inhibiting Fos signaling in zebrafish. Blood *124*, 1578–1585.

Weissman, I.L. (2000). Stem cells: units of development, units of regeneration, and units in evolution. Cell *200*, 157–168.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324.

Wilkinson, R.N., Pouget, C., Gering, M., Russell, A.J., Davies, S.G., Kimelman, D., and Patient, R. (2009). Hedgehog and Bmp Polarize Hematopoietic Stem Cell Emergence in the Zebrafish Dorsal Aorta. Dev. Cell *16*, 909–916.

Willinger, T., Rongvaux, A., Strowig, T., Manz, M.G., and Flavell, R.A. (2011). Improving human hemato-lymphoid-system mice by cytokine knock-in gene replacement. Trends Immunol. *32*, 321–327.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H.S. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810–813.

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

Wilson, A., Oser, G.M., Jaworski, M., Blanco-Bose, W.E., Laurenti, E., Adolphe, C., Essers, M.A., Macdonald, H.R., and Trumpp, A. (2007). Dormant and Self-Renewing Hematopoietic Stem Cells and Their Niches. Ann. N. Y. Acad. Sci. *1106*, 64–75.

Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E., et al. (2010). Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. Cell Stem Cell *7*, 532–544.

Wong, P.M., Chung, S.-W., Reicheld, S.M., and Chui, D.H. (1986). Hemoglobin switching during murine embryonic development: evidence for two populations of embryonic erythropoietic progenitor cells. Blood *67*, 716–721.

Wu, C., Jin, X., Tsueng, G., Afrasiabi, C., and Su, A.I. (2016). BioGPS: building your own mash-up of gene annotations and expression profiles. Nucleic Acids Res. *44*, D313–D316.

Wu, X., Lensch, M.W., Wylie-Sears, J., Daley, G.Q., and Bischoff, J. (2007). Hemogenic Endothelial Progenitor Cells Isolated from Human Umbilical Cord Blood. STEM CELLS *25*, 2770–2776.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663–676.

Xu, W., and Kee, B.L. (2007). Growth factor independent 1B (Gfi1b) is an E2A target gene that modulates Gata3 in T-cell lymphomas. Blood *109*, 4406–4414.

Xu, M., Tsuji, K., Ueda, T., Mukouyama, Y., Hara, T., Yang, F.-C., Ebihara, Y., Matsuoka, S., Manabe, A., and Kikuchi, A. (1998). Stimulation of mouse and human primitive hematopoiesis by murine embryonic aorta-gonad-mesonephros–derived stromal cell lines. Blood *92*, 2032–2040.

Xue, Y., Gao, X., Lindsell, C.E., Norton, C.R., Chang, B., Hicks, C., Gendron-Maguire, M., Rand, E.B., Weinmaster, G., and Gridley, T. (1999). Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. Hum. Mol. Genet. *8*, 723–730.

Yamada, Y., Warren, A.J., Dobson, C., Forster, A., Pannell, R., and Rabbitts, T.H. (1998). The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. Proc. Natl. Acad. Sci. U. S. A. *95*, 3890–3895.

Yamanaka, S., and Blau, H.M. (2010). Nuclear reprogramming to a pluripotent state by three approaches. Nature *465*, 704–712.

Yoder, M.C., Hiatt, K., Dutt, P., Mukherjee, P., Bodine, D.M., and Orlic, D. (1997). Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac. Immunity *7*, 335–344.

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

Yoshimoto, M., Montecino-Rodriguez, E., Ferkowicz, M.J., Porayette, P., Shelley, W.C., Conway, S.J., Dorshkind, K., and Yoder, M.C. (2011). Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential. Proc. Natl. Acad. Sci. *108*, 1468–1473.

Young, P.E., Baumhueter, S., and Lasky, L.A. (1995). The sialomucin CD34 is expressed on hematopoietic cells and blood vessels during murine development. Blood *85*, 96–105.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science *318*, 1917–1920.

Yue, R., Li, H., Liu, H., Li, Y., Wei, B., Gao, G., Jin, Y., Liu, T., Wei, L., Du, J., et al. (2012). Thrombin Receptor Regulates Hematopoiesis and Endothelial-to-Hematopoietic Transition. Dev. Cell *22*, 1092–1100.

Yurino, A., Takenaka, K., Yamauchi, T., Nunomura, T., Uehara, Y., Jinnouchi, F., Miyawaki, K., Kikushige, Y., Kato, K., and Miyamoto, T. (2016). Enhanced Reconstitution of Human Erythropoiesis and Thrombopoiesis in an Immunodeficient Mouse Model with KitWv Mutations. Stem Cell Rep. *7*, 425–438.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zhang, C.-Y., Yin, H.-M., Wang, H., Su, D., Xia, Y., Yan, L.-F., Fang, B., Liu, W., Wang, Y.-M., Gu, A.-H., et al. (2017). Transforming growth factor-β1 regulates the nascent hematopoietic stem cell niche by promoting gluconeogenesis. Leukemia *32*, 479–491.

Zhang, H., Nieves, J.L., Fraser, S.T., Isern, J., Douvaras, P., Papatsenko, D., D'Souza, S.L., Lemischka, I.R., Dyer, M.A., and Baron, M.H. (2014). Expression of Podocalyxin separates the hematopoietic and vascular potentials of mouse ES cell-derived mesoderm. Stem Cells Dayt. Ohio *32*, 191–203.

Zhang, P., Behre, G., Pan, J., Iwama, A., Wara-aswapati, N., Radomska, H.S., Auron, P.E., Tenen, D.G., and Sun, Z. (1999). Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proc. Natl. Acad. Sci. U. S. A. *96*, 8705–8710.

Zhang, P., He, Q., Chen, D., Liu, W., Wang, L., Zhang, C., Ma, D., Li, W., Liu, B., and Liu, F. (2015). G protein-coupled receptor 183 facilitates endothelial-to-hematopoietic transition via Notch1 inhibition. Cell Res. *25*, 1093–1107.

Zhen, F., Lan, Y., Yan, B., Zhang, W., and Wen, Z. (2013). Hemogenic endothelium specification and hematopoietic stem cell maintenance employ distinct Scl isoforms. Development *140*, 3977–3985.

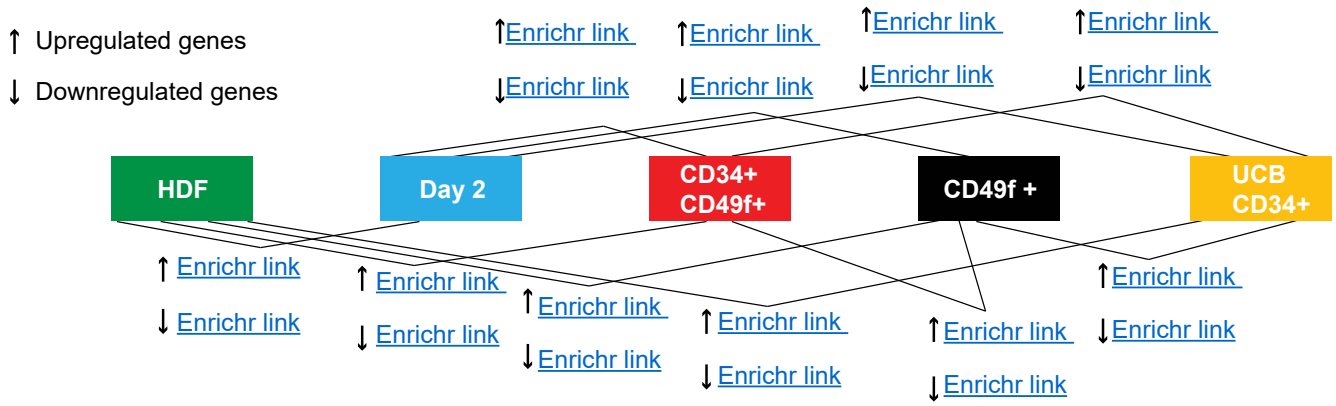
Zovein, A.C., Hofmann, J.J., Lynch, M., French, W.J., Turlo, K.A., Yang, Y., Becker, M.S., Zanetta, L., Dejana, E., Gasson, J.C., et al. (2008). Fate Tracing Reveals the Endothelial Origin of Hematopoietic Stem Cells. Cell Stem Cell *3*, 625–636.

Zovein, A.C., Turlo, K.A., Ponec, R.M., Lynch, M.R., Chen, K.C., Hofmann, J.J., Cox, T.C., Gasson, J.C., and Iruela-Arispe, M.L. (2010). Vascular remodeling of the vitelline artery initiates extravascular emergence of hematopoietic clusters. Blood *116*, 3435–3444.

Zweidler-McKay, P.A., Grimes, H.L., Flubacher, M.M., and Tsichlis, P.N. (1996). Gfi-1 encodes a nuclear zinc finger protein that binds DNA and functions as a transcriptional repressor. Mol. Cell. Biol. *16*, 4024–4034.

# Appendices

a Single Cell RNA-seq Gene list Enrichment Analysis




b

| Differential expression | Enrichr links |
|---|---|
| up HDF vs HDF Day 2 | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce39 |
| down HDF vs HDF Day 2 | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3a |
| up HDF vs CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3b |
| down HDF vs CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3c |
| up HDF vs CD34+CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3d |
| down HDF vs CD34+CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3e |
| up HDF vs UCB CD34+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3f |
| down HDF vs UCB CD34+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3g |
| up HDF Day 2 vs CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3h |
| down HDF Day 2 vs CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3i |
| up HDF Day 2 vs CD34+CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3j |
| down HDF Day 2 vs CD34+CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3k |
| up HDF Day 2 vs UCB CD34+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3l |
| down HDF Day 2 vs UCB CD34+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3m |
| up CD49f +vs CD34+CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3n |
| down CD49f+ vs.CD34+CD49f+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3o |
| up CD49f+ vs UCB CD34+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3p |
| down CD49f+ vs UCB CD34+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3q |
| up CD34+CD49f+vs UCB CD34+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3r |
| down CD34+CD49f+ vs UCB CD34+ | http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=1ce3s |

Appendix I - **Gene list enrichment analysis based on single-cell RNA-Seq data.** (**a**)The 500 most differentially expressed genes between HDF, day 2, day 15 CD49f+, day 25 CD49f+CD34+ and CD34+ UCB were used from 286 single cells. The most significantly differential expressed genes were identified in one population relative to their expression in other populations. Upregulated and downregulated genes are enriched in the population on the right hand side; these can be displayed by Enrichr hyperlinks (**b**) for functional enrichment of gene sets.

## Appendix II -List of genes co-bound by GATA2 and GFI1B

| | | | | | |
|---|---|---|---|---|---|
| RBFOX2 | MATN1-AS1 | PABPC1P2 | MIR5702 | GLIS3 | DRAM1 |
| FNBP1 | ATF6 | RSU1 | DMTF1 | MMP24 | ADAMTS14 |
| TYW5 | PKIG | RIPK2 | MIR1343 | EFEMP1 | CCDC85A |
| CACNA2D1 | LINC00544 | ATXN1 | ROR1 | C6ORF223 | ME3 |
| LNPEP | PID1 | DKFZP686O1327 | UTRN | DBC1 | TBX2 |
| C12ORF75 | ITSN1 | LOC100130357 | LOC256021 | ZNF648 | LINC00595 |
| COL1A2 | PDGFC | CGNL1 | NSMAF | PIM1 | KCNH7 |
| EML6 | GATA6 | LINC00710 | SLA | ZNF644 | MYO10 |
| SCGB1A1 | NR2E3 | ZNF462 | MIR708 | TXNDC8 | KIAA1549L |
| FEZ2 | ALKBH8 | NANOS1 | LOC100289473 | PPP1R12A | PPM1B |
| AGTR1 | GATA2 | STAT1 | NTPCR | TMCO5A | C1D |
| RPL22L1 | LIPA | ANKRD30BL | PLOD2 | GFPT1 | SUCLA2 |
| ZFYVE9 | MYLK | F3 | EGFR | KSR1 | MYO1B |
| NCAPH | MIR569 | LOC644838 | FOXO3 | TNFRSF19 | RPL21P44 |
| IER5L | C12ORF66 | MIR3938 | ZBTB20 | ITGA2 | SLC25A13 |
| LOC728392 | FAM110B | DAB2IP | SLC9A9-AS1 | C10ORF11 | SMIM21 |
| C8orf87 | FAM134B | FABP5 | LOC100132735 | EIF2AK3 | TRIM33 |
| TMEM72 | EPB41L2 | SLC39A10 | SLN | NRG1 | SLC25A15 |
| ZEB1-AS1 | LTA4H | SNX16 | HAS2 | RAD23B | LINC00589 |
| LURAP1L | LMCD1-AS1 | FLJ22447 | SH3BP4 | NPC1 | LOC728724 |
| DPYSL3 | DUSP5 | PELI2 | SLC16A7 | PARD3B | LINC00340 |
| BBX | ZEB2-AS1 | C11ORF34 | LINC00597 | MCC | SH2D4A |
| SNX27 | DUSP1 | CDKL1 | LINC00113 | QKI | PTEN |
| SPOCK1 | KCNB1 | C11ORF53 | MAP3K8 | ARL4C | FAM208B |
| VTI1A | EMP1 | LRRC17 | PITX2 | SP3 | LOC100271702 |
| MIR548AR | LOC643723 | EFCAB5 | ATP9A | NT5C1B | SQRDL |
| KLF12 | SAMD4A | TMEM53 | CERS5 | LINC00327 | LOC151171 |
| NEBL | TMCO4 | TMEM260 | C11ORF87 | NAF1 | ZNF608 |
| ABCA5 | MMP10 | FSTL5 | FST | PIK3C3 | LOC100130539 |
| NR2F2 | CENPE | EEA1 | LRRC40 | KCNK2 | OVERLAP |
| PBX3 | TMEFF2 | PIN1P1 | FOXN2 | LOC728743 | PGM3 |
| EXOC6B | TTLL7 | ADAM28 | FRMD5 | ARL4A | BEND3 |
| ARHGAP26-AS1 | MMP16 | EFCAB1 | FRMD6 | MIR548A2 | LRRC8C |
| ABCA9 | C4ORF6 | RICTOR | RGCC | LINC00320 | PCBD1 |
| MIR548AX | DPYD | JAK1 | STK24 | KHDRBS3 | KCNJ2 |
| MIR558 | BHLHE40 | APPL2 | CDC42EP3 | LOC100505718 | CTHRC1 |
| MIR557 | LOC441009 | FBXW7 | TSNAX-DISC1 | MIR1260B | DUT |
| IL22RA2 | RAPGEF6 | C11ORF65 | LOC100505702 | LOC100289673 | ACTN4 |
| MACC1-AS1 | LOC285627 | FBXW8 | ST6GALNAC5 | NREP | PLEKHA2 |
| MIR563 | PET112 | PRRX1 | INO80D | ATP5J | LRRN3 |
| MIR561 | KCNC4 | HSD52 | FTO | BICD1 | ENTPD6 |
| PRKAR1A | POU5F1P4 | CORO2B | GRIA1 | ADAMTS12 | JAKMIP2-AS1 |
| KCNMA1 | MRPL33 | FOXP1 | KCNE4 | LRRC16A | HFE2 |
| MIR548AC | C12ORF36 | CXCL12 | LOC284998 | LOC644649 | DCLK2 |

| | | | | | |
|---|---|---|---|---|---|
| MIR548AD | RRAS2 | SLC9A9 | ELFN2 | GOLIM4 | SFRP2 |
| LOC645434 | TSPAN5 | LOC100499183 | TMEM244 | NAV2 | RERE |
| ULK4 | HIVEP2 | SMYD2 | ASAP1 | NAV3 | LAMA2 |
| ITGBL1 | PIEZO2 | KIF1B | FAM196B | SGOL1 | NR3C1 |
| ALPK2 | INHBA-AS1 | SNX9 | HSD17B2 | ETV4 | CBL |
| ALDH8A1 | ARHGEF3-AS1 | SOCS5 | KIF5B | PJA2 | FAM154A |
| EPHA4 | CCDC93 | SNX6 | SLIT2 | C4ORF45 | CAND1 |
| IFLTD1 | FOXD1 | CD34 | FGF20 | AHI1 | ZIC1 |
| EPAS1 | LINC00052 | SOCS6 | NDUFV2 | TMEM214 | SFTA1P |
| SATB2 | LOC400027 | MME | CCDC140 | ACOX3 | SUMO1P1 |
| CYB5D2 | VEGFC | MRPS24 | RUNX1T1 | IL8 | PRELID2 |
| HERC4 | NBPF1 | UBE2E3 | ARHGEF12 | SGK223 | KIAA1462 |
| IFI16 | RGMB | MRPS22 | CCDC132 | ZMYND11 | ARMC9 |
| DHX35 | ENAH | UBE2E1 | PRNP | LOC440311 | REV1 |
| E2F7 | PTPRE | MN1 | NABP1 | TMEM45B | DKK1 |
| LOC401177 | NR5A2 | ACTL8 | INHBA | HS3ST3A1 | DKK2 |
| WWTR1 | RAP2B | EPGN | LOC152742 | RAP1GDS1 | DKK3 |
| EGR1 | TOX2 | GRM5-AS1 | SSH1 | LAMC1 | NUP93 |
| FAM9C | KITLG | KIT | KCNU1 | HMGB1 | LOC730101 |
| XRCC5 | KCNS2 | KIF26B | ZNF92 | C4ORF32 | MARCKS |
| RASSF9 | TAX1BP1 | TPST1 | GCLC | GHR | POLR3B |
| ANK3 | PABPC1 | SGSM3 | TMEM236 | FUT8 | CD9 |
| PRG4 | PDZRN3 | CD44 | CARD18 | MGC34034 | ASB7 |
| ANK2 | DEPDC5 | ERRFI1 | COL21A1 | MIR205HG | CDK17 |
| IFFO2 | LOC100506776 | NRP1 | DNAL1 | CPXM2 | TES |
| IFI44 | PCSK1 | NRP2 | TCF4 | ECT2 | LINC00640 |
| MACROD2-AS1 | SNORA14A | POPDC2 | DPYD-AS1 | S100G | COL12A1 |
| PTK2 | BARHL1 | DIRAS3 | GCLM | ATF7IP | XYLT1 |
| EXT1 | SNORA14B | SH3RF1 | VCL | EED | ADCY8 |
| PIK3R1 | TNFRSF11A | AKAP13 | DYRK2 | DAPK2 | MSANTD1 |
| MIR4799 | KIAA1324L | AKAP12 | COX18 | EFTUD1P1 | MYC |
| SDC2 | TMPRSS5 | RSRC1 | FLJ37505 | DDX10 | FLRT2 |
| MIR1288 | KIF14 | ENC1 | HTRA1 | SOD3 | PLCE1 |
| NEGR1-IT1 | SLC7A14 | THSD7A | ETS1 | MIR1915 | CYP1B1 |
| RCOR3 | PTPRK | XDH | PDS5A | TBX18 | ANKS1B |
| MIR1284 | GRIK2 | CCT5 | CALD1 | GREM2 | NHEJ1 |
| LPGAT1 | TFB1M | CD55 | CAMTA1 | EEPD1 | AFAP1L1 |
| ECI2 | SLC7A11 | SPINK4 | ZNF385D | PTHLH | FONG |
| LDLRAD3 | PTPRJ | OSR2 | RGS9 | SNTB1 | CUBN |
| LPAR3 | C8B | CACHD1 | IL6R | TBX15 | STARD13 |
| PALM2-AKAP2 | KALRN | CCDC158 | LOC338758 | CRY1 | FZD8 |
| MTMR9LP | SNX4 | ST7-AS2 | BCKDHB | RBM43 | PRKCE |
| C4ORF51 | PPP3CA | TMX3 | ZFHX3 | DAAM1 | PLK2 |
| MRPL3 | SOCS1 | SCN8A | MKL1 | NACAP1 | INO80 |
| IMMP2L | PPP3CC | OTUD4 | HMGA2 | SPATA31D1 | PARD3 |

| | | | | | |
|---|---|---|---|---|---|
| MICAL2 | CTSO | ASAP2 | SSTR1 | TRIB1 | ELL2 |
| 42339 | ADM | TLE3 | MIR4714 | MIR1252 | MIR2277 |
| ZFAT-AS1 | AEBP2 | ANKRD28 | REEP3 | SKOR1 | GSAP |
| GPR137C | MIR875 | AGK | C3ORF67 | IRX3 | MIR4451 |
| LPHN2 | MIR630 | CEP112 | PGPEP1L | LINC00474 | MIR4693 |
| HECW2 | C14ORF182 | YIPF5 | TOX | CKAP2L | APP |
| PRKD1 | LINC00837 | VRK2 | EIF4G3 | C1ORF140 | TTLL11 |
| LPHN3 | TIMP3 | BTBD9 | FAM25A | NSMCE2 | WDSUB1 |
| TSPYL5 | JARID2 | SLC8A1-AS1 | CALCOCO1 | NGF | NUDT4P1 |
| LOC100500773 | LOC440028 | CNOT8 | IRS1 | KLF4 | AKAP7 |
| TRIO | FAM160B1 | RAB38 | IRS2 | ENOX1 | CBLB |
| TMEM200B | TMEM14E | ABI1 | CUL2 | KLF2 | HSF2BP |
| NEDD4L | YTHDF3 | SLCO2A1 | AK5 | PKM | HS6ST1 |
| BCL10 | RNLS | ITGA11 | FGF1 | KLF7 | ANTXR2 |
| WDR41 | CDKN2B | LOC401321 | MIR4700 | NFASC | APCDD1 |
| IPO11 | GPC6-AS2 | GPHB5 | ABHD10 | VAPA | ARNTL |
| SLC6A4 | TMOD3 | TNIK | IGFBP7 | MIR1246 | AKAP2 |
| GJA1 | CD180 | CNIH3 | ZNF521 | PPAP2B | C3ORF17 |
| DSTYK | KIDINS220 | SNRPE | KCNIP1 | MIR3668 | PDCD10 |
| CCRN4L | SULF2 | PHLDB2 | RFC3 | MIR4757 | NIPAL2 |
| LOC100130298 | FAM104A | SEMA5A | LINC00691 | EVL | CHEK2 |
| KIAA0319L | MKLN1 | LINC00486 | C15ORF54 | CSNK1G1 | HABP4 |
| POU5F1B | PLCB4 | CLEC2L | CDK7 | CSNK1G3 | HMCN1 |
| BARD1 | TMEM158 | FIGN | BCL2 | POM121L2 | TBL1X |
| ZNF366 | EBNA1BP2 | C1ORF110 | MIR1265 | SEMA3A | MIR4687 |
| GSDMC | COL4A2 | AHR | MTFR2 | NCKAP5 | APBB2 |
| ANXA2 | MSRB3 | CLU | CPO | WIPF1 | MIR4681 |
| PHF11 | LINC00626 | TANK | BNC1 | SEMA3D | SIAH1 |
| TCF12 | PLCB1 | CEP128 | MAPKAP1 | LOC400456 | SLC6A15 |
| TACR1 | ZNF330 | PAPSS2 | GADL1 | LINC00460 | HAS2-AS1 |
| CDC7 | SPEF2 | AP3M2 | TARS | SEMA3C | NMBR |
| C9ORF84 | LOC339894 | RAB20 | UBR2 | KIAA1432 | C8ORF4 |
| C1ORF21 | LOC285768 | NSD1 | WDFY3-AS2 | PRR5L | NUAK1 |
| SCARNA27 | FMN2 | RNF19B | ADRA1B | AMBRA1 | MIR4437 |
| MIR620 | GLI2 | CFL2 | PTBP2 | PIK3C2A | MIR4439 |
| NCOR2 | LOC285762 | MACROD2 | FGGY | THSD4 | MIR4677 |
| LINC00607 | SLC8A1 | RBBP8 | DUSP10 | ADAMTS6 | MAGEB2 |
| GPR128 | GLI3 | KIRREL3-AS2 | METTL15 | SGCE | VGLL4 |
| EXOC4 | GRAMD3 | ARHGAP5-AS1 | HMBOX1 | HECTD2 | TSHZ2 |
| PHF17 | TMEM144 | TRPC4 | TNPO1 | PSD3 | DCUN1D5 |
| PAM | UBL3 | OSBPL3 | MROH9 | CACNG1 | PITPNA |
| CYB561 | ROPN1L | GTF3A | C3ORF55 | LINC00457 | TMTC3 |
| LNX1-AS2 | ANKRD55 | TRAPPC9 | CCDC34 | IL1R1 | SCARNA3 |
| GAS7 | UVRAG | PGD | MIR4522 | ACSL3 | TWIST1 |
| GPR126 | TLE4 | RCAN2 | MPDZ | RUNX1 | TMTC1 |

| | | | | |
|---|---|---|---|---|
| ADK | PHLDA1 | SYNJ2 | C3ORF58 | MIR3129 |
| PDK4 | MPP7 | AZI2 | RBMS3 | NARS2 |
| SOX9 | NOV | REPS2 | B4GALT1 | RNU6-19 |
| MIR4660 | F2RL1 | POSTN | TSEN15 | GNA13 |
| TNS3 | APBA1 | CDYL | NVL | PALMD |
| EDN1 | FAM69A | ESR1 | MCTP2 | LSAMP-AS3 |
| SIGLEC15 | CEACAM22P | SMAD6 | USP35 | LRIG1 |
| SLC35E3 | PRIM2 | ARHGAP12 | TRPS1 | LRIG3 |
| TSC22D2 | AMIGO2 | DLG2 | ST8SIA6-AS1 | COLGALT2 |
| PXDNL | PTGER4 | MDFIC | GTDC1 | MAPK6 |
| LIF | LYZL1 | TMEM117 | CLEC14A | LEKR1 |
| LSAMP | FAIM2 | CDH11 | STK4 | METTL4 |
| AP3B1 | TRAK1 | COL6A3 | DPP4 | ZDHHC17 |
| LHFPL2 | ROBO2 | DDAH1 | HK2 | RPS4X |
| RHOB | CLDN23 | CDH13 | SCFD2 | SPRED2 |
| RAD51B | AMN1 | DYNLRB2 | SLC7A11-AS1 | DOCK2 |
| PRPF18 | KIF13A | CASQ2 | NRIP1 | BMPR1A |
| FLJ12334 | MIR4466 | MIR4262 | CASP4 | SPTSSB |
| KBTBD3 | MYOZ2 | DOCK4 | LRRTM2 | GHSR |
| CC2D1B | LOC100506422 | CTTNBP2 | PDE4B | MIR4422 |
| TOP2B | THRB-AS1 | PXN | OPCML | PLSCR5 |
| TRAM1 | PTCD2 | SH3KBP1 | CXADR | MYO6 |
| CLIC4 | ALYREF | NTM | CCBE1 | MAPKAPK2 |
| ZCCHC11 | SHISA9 | COPB1 | PTGIS | TACSTD2 |
| CITED2 | TMEM135 | ARHGAP18 | UNQ6975 | CORO1C |
| SLC35F4 | TBL1XR1 | LPP | TMEM30C | MAPK13 |
| LINC00161 | SCG2 | PODXL | PDE4D | C9ORF92 |
| SLC44A3 | STAM2 | MIR4268 | LOC153910 | PLCL1 |
| STC2 | PDE7B | PMAIP1 | ALOX5AP | MIR1208 |
| SLC35F2 | MPHOSPH6 | MIR3175 | LCA5 | FNIP2 |
| HP07349 | GALNT15 | LOC100507557 | DCK | LOC286189 |
| CTAGE11P | ZCCHC7 | CMTM8 | ARID1B | CMYA5 |
| RNU6-79 | ZCCHC6 | TIPARP | DCN | MIR4789 |
| AUH | SUCNR1 | ABCC2 | ALDH1A3 | RAP1B |
| MAP4 | MIR4280 | NRG1-IT3 | PDE5A | PRDM1 |
| PITRM1 | RGS17 | RHBDD2 | SPRY1 | ETV1 |
| LOC100506403 | ZFP36L1 | SMURF1 | CAMK2D | LOC100287944 |
| MIR3157 | ARHGAP35 | PPP2R5C | NXN | PRKCA |
| TINAG | FBLN5 | DTNBP1 | LOC100131060 | |
| MIR3152 | NT5E | TGFBR2 | TM2D3 | |
| ARNT2 | ARHGAP42 | MEIS2 | TM2D1 | |
| SVIL | BASP1 | COQ2 | AQP11 | |
| NFYB | CWC22 | MIR585 | THBS1 | |
| DNAJC5B | BMPER | RASA2 | KANSL1 | |
| FAM46A | CLASP1 | AKNAD1 | SLC4A7 | |