THE UNIVERSITY
of EDINBURGH

GENOME-WIDE TRANSCRIPTIONAL CHARACTERISATION AND INVESTIGATION OF THE MURINE NICHE FOR DEVELOPING HAEMATOPOIETIC STEM CELLS

Alison Clare McGarvey

Thesis submitted for the degree of Doctor of Philosophy
The University of Edinburgh
2016

Abstract

Haematopoietic stem cells (HSCs) are capable of differentiation into all mature haematopoietic lineages, as well as long-term self-renewal and are consequently able to sustain the adult haematopoietic system throughout life. Currently, in the mouse, HSCs are understood to first appear in the aorta-gonad-mesonephros (AGM) region at embryonic day 11 via a process of maturation from precursors (pre-HSCs). This maturation within the AGM region involves the complex interplay of signalling between cells of the niche and maturing precursor cell populations, but is relatively little understood at a molecular level. Recently our understanding of the AGM region has been refined, identifying the progression from E9.5 to E10.5 and the polarity along the dorso-ventral axis as clear demarcations of the supportive environment for HSC maturation. In this thesis, I investigated the molecular characteristics of these spatio-temporal transitions in the AGM region through the application of RNA-sequencing. This enabled the identification of molecular signatures which may underlie the supportive functionality of the niche. I further compared these expression signatures to the transcriptional profile of an independent cell type, also capable of supporting HSC maturation, the OP9 stromal cell line. By combining this transcriptional information with an *ex vivo* culture system, I screened a number of molecules for their ability to support HSC maturation from early precursors, leading to the discovery of a novel regulator of HSC maturation: BMPER. Further characterisation of this molecule enabled the identification of its specific cellular source and the proposal that through its action as an inhibitor of BMP signalling it facilitates the maturation of precursors into HSCs. These results lend further detail and support to the role of BMP signalling in the regulation of HSC maturation as well as demonstrating the potential of these transcriptional profiles to yield novel mechanistic insight.

Lay Summary

This project was an investigation of the earliest emergence of the stem cells which can produce the entire blood system. The time and place these blood stem cells first appear are known, but a full understanding of the critical environmental cues that regulate their formation is still lacking. Here we made a dataset which measures all of the genes present in the environment of emerging blood stem cells. By computational modelling we identified candidates from this data which are found in this region. Adding these to a culture of the precursors of blood stem cells showed one of these candidates improved the efficiency of blood stem cell formation. Further study revealed the specific cell types which produced this molecule and the mode of action that it works through. Hopefully this will provide some insight into how blood stem cells may be grown in the lab and also the availability of this dataset might aid future discovery of other factors which regulate this process.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Alison McGarvey
Edinburgh, 2016

Acknowledgements

Firstly, I would like to thank my two supervisors Professor Alexander Medvinsky and Dr Simon Tomlinson for giving me the chance to embark upon a project that required me to take time to learn new skills in several domains. Both Alexander and Simon have encouraged me with their passion for science, while providing critical feedback as well as scope for creative thinking.

Being a member of two research groups, I have been lucky to have contributions from many people throughout my project. I give particular thanks to Dr Stansislav Rybtsov and Dr Celine Souilhol for advising me scienficially and with many of the tricky experimental procedures involved in the field of developmental haematopoiesis. To Dr Florian Halbritter, Dr Duncan Godwin and Dr Jonathan Manning for helping me navigate the brave new world (for me) of bioinformatics and programming. To Anastasia Kousa for her friendship, much needed coffee breaks and moral support. To all current and previous members of both labs Dr Antoniana Batsivari, Anahi Binagui-Casas, Dr Jennifer Easterbrook, Kateryna Bilotkach, Dr Lucia Morgado-Palacin, Sara Tamagno, Edie Crosse, Dr Natalia Rybtsova, Dr Sabrina Gordon-Keylock, Dr Andrejs Ivanovs, Dr Anna Liakhovitskaia, Dr Boni Afouda, Dr Daria Paruzina, Dr David Hills, Dr Fiona Murphy, Heather Wilson, Dr Javier Gonzalez, Dr Jordi Senserrich, Michael Stockton, Dr Niamh Fanning, Dr Suling Zhao, Dr Yiding Zhao, Dr Vincent Frontera, James Ashmore, Dr Aidan McGlinchey and Will Bowring, for many scientific and personal contributions to my PhD experience.

Further thanks go to my PhD committee members Professor Val Wilson and Dr Tilo Kunath for insightful feedback and helping keep my project on track. To Drs Fiona Rossi, Claire Cryer, Bertrand Vernay for their technical support with flow cytometry and imaging, as well as Carol Manson, Andrew Dyer and all the staff who care for the mice used in the experiments here.

To the many PhD students and colleagues at the SCRM and University of Edinburgh with whom I have shared a lot of laughter, fun, travels and comradery in the ups and downs of PhD life. To my family, who give me strength. And to Luca, for being by my side on this journey.

“I was born not knowing and have had only a little time to change that here and there.”

Richard P. Feynman

Abbreviations

7AAD	7-Amino-Actinomycin D
β	Beta
µ	Micro
AGM	Aorta-Gonad-Mesonephros
Ao	Dorsal aorta
AoD	Dorsal aorta - Dorsal
AoV	Dorsal aorta - Ventral
bp	Base pair
BFU-E	Burst Forming Unit -Erythroid
Ca2+	Calcium
CSS	Cascading Style Sheets
CD	Cluster of differentiation
cDNA	Complementary Deoxyribonucleic Acid
CFU-C	Colony Forming Unit - Culture
CFU-GEMM	Colony Forming Unit - Granulocyte Erythroid Monocyte Megakaryocyte
CFU-GM	Colony Forming Unit - Granulocyte Macrophage
CFU-S	Colony Forming Unit - Spleen
DAPI	4,6-diamidino-2-phenylindole, dihydrochloride
DNA	Deoxyribonucleic acid
E	Embryonic day
EDTA	Ethyelediamine tetra-acetate
ee	Embryo equivalent
ES	Embryonic stem
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FDR	False discovery rate
g	Grams
HSC	Haematopoietic Stem Cell
HSPC	Haematopoietic Stem or Progenitor Cell
HTML	HyperText Markup Language
IMDM	Iscove's Modified Dulbecco's Medium
JSON	JavaScript Object Notation
Lin	Lineage (antigens B220, CD3e, Ter119, Gr1, CD11b)
LTR-HSC	Long-term repopulating definitive haematopoietic stem cell
m	Milli
M	Molar
Mg2+	Magnesium
mRNA	Messenger Ribonucleic Acid
n	Nano
PBS	Phosphate Buffered Saline

PCA	Principle component analysis
pre-HSC	Haematopoietic stem cell precursor
P/S	Penicillin/Streptomycin
P-Sp	Para-aortic splanchnopleura
qRT-PCR	Quantitative Reverse-Transcription Polymerase Chain Reaction
RIN	RNA integrity number
RPM	Reads per million
RPKM	Reads per kilobase per million
SAM	Sequence Alignment/Map
SCF	Stem cell factor
sp	Somite pair
SVG	Scalable vector graphics file
TAE	Tris-Acetate-EDTA
U	Units
UGR	Urogenital ridge
URL	Uniform Resource Locator
VC	VE-Cadherin
XML	EXtensible markup language
YS	Yolk Sac

Contents

LIST OF FIGURES

List of Tables

Chapter 1 Introduction

Haematopoietic stem cells (HSCs) are a subject of extensive research in terms of both their fundamental contribution to vertebrate physiology, and their versatile use as a therapeutic agent. Efforts to study their behaviour and molecular underpinnings are hampered by an inability to culture HSCs extensively *in vitro.* Understanding how HSCs could be generated from alternative sources, such as embryonic stem (ES) cells, *in vitro* is likely to greatly improve our ability to study their regulatory mechanisms, as well as provide insight into the aberrant processes that result in haematopoietic disease such as leukaemia. The formation of HSCs *de novo* during midgestation of embryonic development (Medvinsky and Dzierzak, 1996; Medvinsky et al., 1993; Müller et al., 1994) indicates that the generation of HSCs *in vitro* may be attainable through a deep understanding their *in vivo* developmental origins. Thus, the elucidation of the molecular events driving HSC maturation from their precursors in the embryo is an important step in bringing us closer to *in vitro* study of HSC biology, disease modelling and potential therapeutic advances.

In this chapter, I will introduce the concept of the haematopoietic stem cell (HSC) through its experimental definition and the role of HSCs in production and maintenance of the haematopoietic system. I will outline the studies that uncovered the embryonic origins of HSCs in the midgestation aorta-gonad-mesonephros (AGM) region, the embryonic precursors of HSCs and the remaining controversies in these definitions. Through a discussion of the niche, I aim to emphasise the importance of non-cell autonomous signalling on HSC maturation and the degree to which we currently understand its spatial and temporal demarcation, particularly during development. Finally, by reviewing our current understanding of molecular regulation of HSC development, and the recent explosion in transcriptional profiling technologies, I would like to highlight the context of this PhD project and the approach used to broaden our molecular understanding of the non-cell autonomous regulators of HSC development in the AGM region.

1.1 The haematopoietic hierarchy and the concept of haematopoietic stem cells

The haematopoietic system is a key multifunctional organ in all vertebrates, serving a highly active role, from the delivery of oxygen and nutrients to the defence against infection. Ultimately, such a multifaceted circulating system is one of the key features enabling multicellular organisms to reach a large size yet still maintain the needs of individual cells. Sustaining this activity requires a high turnover of cells, and consequently highly adaptive mechanism to maintain tissue homeostasis. As such the haematopoietic system has been an area of extensive study over the last 100 years.

1.1.1 Deriving the haematopoietic hierarchy model

The architecture of the haematopoietic system is a paradigm of tissue organisation (Orkin and Zon, 2008). The commonly accepted model depicts a single cell type, the haematopoietic stem cell, which produces the rest of the blood system through a hierarchy of gradually more fate-restricted progenitors. Any loss in terminally differentiated haematopoietic cells can therefore be quickly restored by a pool of upstream progenitors (Figure 1-1).

The elucidation of this hierarchy model arose in the 1950s and 1960s at the height of the "Atomic Era", when the effects of radiation damage to tissues were being studied extensively (Eaves, 2015). The potency of the haematopoietic organs was demonstrated by the recovery of damaged tissue in irradiated mice if transplanted with bone marrow or spleen of an untreated mouse (Jacobson et al., 1951; Lorenz et al., 1951). Cytological examination showed that this recovery was due to the replacement of the damaged tissue by the cells of the donor as opposed to a chemical (or 'humoral') stimulation for self-repair (Ford et al., 1956). These key experiments introduced the concept of cellular repopulation as a therapy, or indeed as it is commonly known today – an allogenic transplant. However, the nature of these therapeutic cells required further study.

The idea that the bone marrow hosts multipotent stem cells was in fact proposed much earlier than these transplantation experiments, through the histological observations of Alexander Maximow, who coined the "Unitarian Theory

of Haematopoiesis" (Maximow, 1924). However this was not shown empirically until Till and McCulloch's study of irradiation recovery with clonal experiments (Till and McCulloch, 1961). These experiments harnessed the observation that transplantation of irradiated mice results in the formation of nodules of mixed haematopoietic cell types in the spleen. Since the number of splenic nodules was linearly correlated to the number of cells transplanted, and by using mixtures of cytologically distinguishable donor cells, these nodules were established to be the clonal product of individual cells (Becker et al., 1963). Splenic colonies which include mixtures of all blood lineages including lymphoid, were later identified (Wu et al., 1967, 1968), and in some cases could form colonies in secondary recipients (Siminovitch et al., 1963), therefore exhibiting the hallmarks of a stem cell – the capacity to self-renew and differentiate. Furthermore, the majority of bone marrow cells which produce splenic colonies were in a relatively quiescent state (Becker et al., 1965; Bruce et al., 1966). Thus the model, which still prevails today, was established that bone marrow contains a quiescent pool of self-renewing haematopoietic stem cells (HSCs) which are capable of producing all adult blood lineages through differentiation.

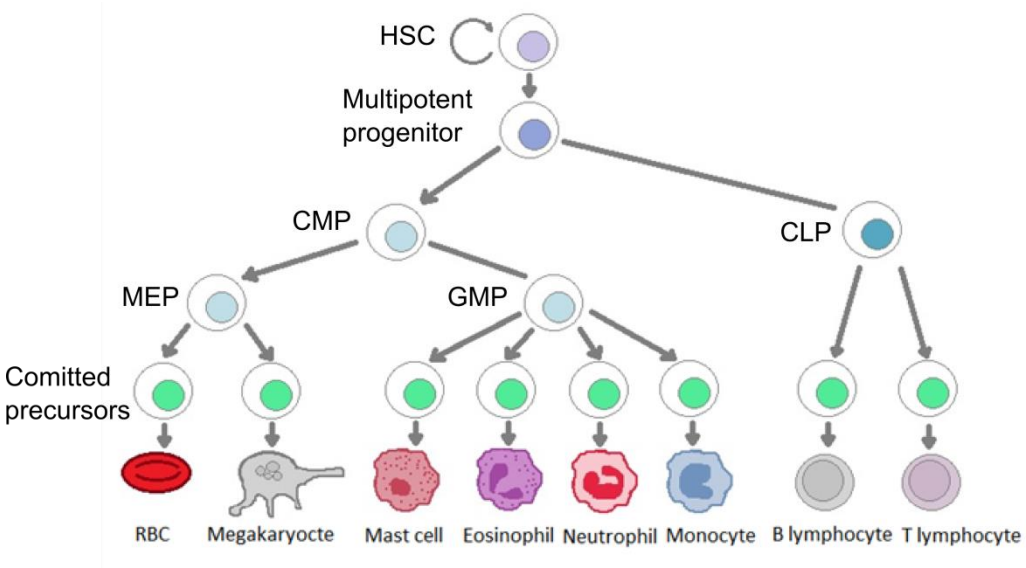


Figure 1-1 The hierarchical model of the haematopoietic system

General model of the haematopoietic system seeded by long-term repopulating haematopoietic stem cells (HSCs) which produce gradually more fate-restricted progenitors

including: CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte-erythrocyte progenitor; GMP, granulocyte-macrophage progenitor; and finally terminally differentiated blood cells. Based on the model in (Orkin and Zon, 2008).

1.1.2 Revising haematopoietic stem cell assays

While the concept of the haematopoietic stem cell devised by Till and McCulloch has held true, the interpretation of their experiments has been revised. Later studies of splenic colonies revealed them to be a transient population of progenitors (Hodgson and Bradley, 1979; Magli et al., 1982), which support the immediate recovery following irradiation, but are different to the true long term repopulating cells that sustain the recovery of the haematopoietic system (Jones et al., 1990). To truly measure single HSCs it was necessary to establish an assay that contains a mixture of transient repopulating cells, which are genetically distinct from a minimal number of long term repopulating cells (Szilvassy et al., 1990). This was titled the competitive repopulation assay. By providing these genetically distinct progenitors (or "carrier" cells), mice will recover from the irradiation even if the test cells (or "donor") are provided in low numbers. Repopulating HSCs can then be quantified by injecting donor cells at a limiting dilution, which ensures only a few of the transplanted mice will be reconstituted. Taking into account the number of mice which have been repopulated, and those which have not, the frequency of repopulating cells (HSCs) can be calculated using a Poisson model (Hu and Smyth, 2009).

An important goal in HSC research is to establish a small number of markers that uniquely distinguish HSCs and allow their isolation independently of transplantation studies. Presently, no single gene has been found to be expressed uniquely in HSCs, so combinations of markers are required for prospective isolation. Furthermore the fact that HSCs cannot be cultured as a clonal population, make it almost impossible to isolate a pure population HSC by phenotype alone. Although the transplantation assay is limited by its requirement for severe, unphysiological treatment and the inference of the HSC state from progeny, it is currently the only method that can reliably test a cell's capacity to produce all blood lineages and self-renew. Therefore in this thesis, the term HSC is only used for cells that have been identified in this way.

1.1.3 Reassessment of the haematopoietic hierarchy

Early models of the hierarchy of progenitors and fate restricted cells of the haematopoietic system came from histological observations of the consistent proximity of certain bone marrow cells (Maximow, 1924). As methods to culture haematopoietic cells *in vitro* improved, a fuller picture of the differentiation hierarchy was mapped using colony assays in semi-solid medium. An array of committed progenitors were defined through the identification of colonies consisting of restricted cell types: CFU-GM granulocyte/macrophage colonies (Bradley and Metcalf, 1966; Pluznik and Sachs, 1965), CFU-E erythroid colonies (Stephenson et al., 1971), CFU-M megakaryocyte colonies (Metcalf et al., 1975), and BFU-E burst forming colonies – earlier erythroid colonies (Axelrad, 1974).

While these colony experiments lay the foundations of the hierarchical model, the full set of progenitors and kinetics of differentiation has remained a topic of debate over the last 3 decades. A key challenge to the canonical hierarchy model is that phenotypically homogenous populations may still exhibit a functional heterogeneity, which may not be detectable in the assays used. Recent lineage tracing methods have proposed alternative differentiation routes (Sun et al., 2014). Tracing cell fates with a pulse of tagging, followed by a 3-4 month chase allowed assessment of progeny. Analysis of the progeny suggested that HSCs have minimal contribution to native haematopoiesis and that the main drivers of steady state haematopoiesis are lineage restricted and multipotent progenitors (Sun et al., 2014). In human, through subdivision of classical progenitor populations and assays of their properties, the existence of oligopotent progenitors has been questioned, with the suggestion that adult haematopoiesis is sustained by unipotent progenitors (Notta et al., 2016).

The haematopoietic hierarchy can alternatively be thought of and modelled mathematically as the function of its emergent properties (Whichard et al., 2010). This generally entails modelling the distribution of cell types in steady state or after perturbations as a function of birth/death or differentiation/self-renewal. A key subject of debate is to what extent the dynamics of the haematopoietic hierarchy and the fate changes of individual cells is a deterministic or stochastic process (Chang et al., 2008; M, 1999; Morrison and Weissman, 1994; Ogawa et al., 1983). An early simulation of the properties of colony forming cells through random birth (self-

renewal) or death (differentiation or cell death) suggested that the haematopoietic hierarchy could be described by stochastic models (Till et al., 1964). However the elucidation of key markers for prospective isolation, regulatory cytokines and lineage-specific transcription factors argues for a more deterministic view of haematopoiesis and its regulation (Morrison and Weissman, 1994; Rieger et al., 2009; Roeder and Glauche, 2006). Ultimately the most faithful model of cell behaviour may depend on the scale that the system is being modelled at as it is possible for properties of the system as a whole may appear deterministic, while the behaviour of individual cells or molecules may be better described as stochastic (Whichard et al., 2010). Such modelling may be useful in prediction of dynamics in response to perturbations such as malignancies and therapeutic interventions which attempt to restore the steady state system.

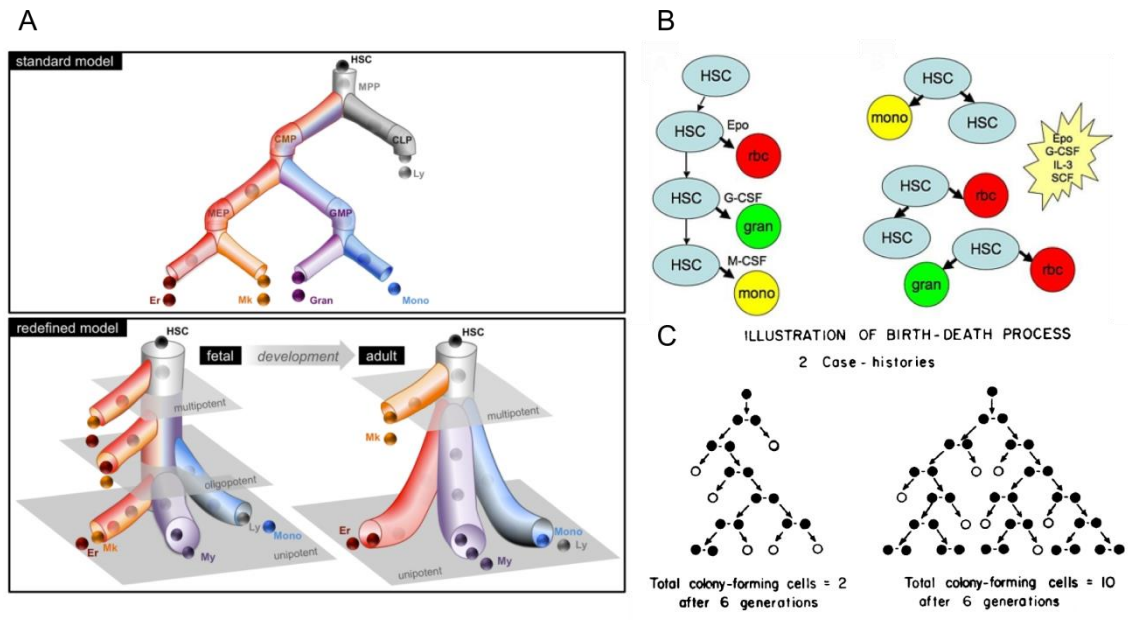


Figure 1-2 Revised models of the dynamics of haematopoietic differentiation

(A) From (Notta et al., 2016) showing revision of classical hierarchical model with less restricted oligopotent progenitors in the embryo, while in adult a lack of oligopotent progenitors. (B) From (Whichard et al., 2010) showing deterministic regulation of differentiation by cytokines (left) and stochastic regulation of differentiation where cytokines promote survival of cell types rather than differentiation (right). (C) From (Till et al., 1964) showing how simulation of differentiation based on a stochastic model can explain heterogeneity of colony assays, while at the level of the whole population the behaviour can be stable.

1.1.4 HSCs as a therapeutic agent

The transplantation of HSCs into patients is one of the best examples of an effective stem cell therapy. Bone marrow transplantation as a clinical therapy started being used in the mid-20th century (Thomas et al., 1957, 1959) in parallel with basic research into the effects of radiation damage and recovery by bone marrow transplant in mice. Interestingly, these first transplants preceded the understanding that stem cells are founders of the haematopoietic system, but applied the understanding that bone marrow is a potent haematopoietic organ. Since then, the elucidation of HSCs has improved the understanding that bone marrow transplantation provides stem cells and transient progenitors that can reconstitute the patient's haematopoietic system.

Bone marrow therapy is now used to treat a range of cancers of the blood or other haematopoietic malignancies (Table 1) (Copelan, 2006). The general principle

is to ablate the patient's own haematopoietic system, whereby the cancer or defective haematopoietic cells are removed, and replace it with a graft of HSCs and progenitors from a healthy donor (allogenic) or from the patient's own bone marrow (autologous) to reconstitute the system. It is now apparent that the grafted cells can also contribute to the removal of malignant cells in the patient as well as replacing them with normal cells (Copelan, 2006). The potential application to diseases that are not necessarily classed as haematopoietic disorders has been recently demonstrated in patients with multiple sclerosis. The aberrant autoimmune response, which manifests as a neurological disorder, could be relieved by essentially resetting the patient's haematopoietic system with an autologous HSC transplant (Atkins et al., 2016).

Although bone marrow transplantation is an effective therapy for a wide range of conditions, it holds two key limitations. Firstly, the ablation of the patient's intrinsic haematopoietic system can be highly toxic. The use of radiation has been replaced with a less toxic combination of Busulfan and Cyclophosphamide (Santos et al., 1983). However it remains one of the main causes of complications of transplantations and a major effort is to more selectively ablate the patient's immune system with antibody-based agents (Matthews et al., 1999; Subbiah et al., 2003). An alternative method is to shift the role of immune ablation to the graft rather than the pre-treatment whereby an immunosuppressive, but less severe, pre-treatment can be used (McSweeney et al., 2001). This, however, is less effective for patients with advanced malignancies (Levine et al., 2003).

The second limitation in allogenic transplantation is the possibility of graft versus host disease (GVHD), a severe immune reaction from the graft against the patient. Since the discovery of the role of the major histocompatibility complex and human leukocyte antigen (HLA) system on recognition of foreign antigens, it has been possible to screen potential donors to reduce the risk of GVHD (Thomas et al., 1977). However, in many cases it is not possible to find a suitable match. The use of bone marrow has gradually been replaced with CD34+ peripheral blood HSCs, which are easier to obtain and more rapidly reconstitute the haematopoietic system. However, this population hosts more T cells, and consequently poses a greater risk of GVHD. If a suitable match cannot be found, patients are often transplanted with

umbilical cord blood, which shows a lower risk of GVHD from lymphoid cells of the graft so requires less stringent HLA matching but reconstitutes more slowly, opening up a greater risk of infection in the patient (Wagner et al., 2002). Identifying a way of expanding cord blood cells is therefore a highly promising way of increasing availability donor cells for allogenic therapy (Jaroscak et al., 2003). Isolating HSCs from the earliest embryonic origin, the aorta-gonad-mesonephros (AGM) region (described in more detail in section 1.2.3), is potentially a more suitable source of HSCs for *ex vivo* expansion, as HSCs at this stage appear to show a greater capacity for self-renewal than adult HSCs (Ivanovs et al., 2011). A third goal in translating research in HSC development to a clinically useful result would be to direct differentiation of ES cells into haematopoietic stem cells. Combined with the ability to reprogram mature cells into an embryonic stem cell-like state termed induced pluripotent stem cells (iPSC), this could provide a suitable model for genetic modification or patient specific disease modelling as well as another route to autologous transplantation (Vo and Daley, 2015).

It is perhaps a fortunate consequence of the inherent mobility of the haematopoietic system that renders it highly amenable to ablation and reconstitution through transplant. These properties, and the effectiveness as a therapy, demonstrate that many aspects of haematopoietic stem cell research – including prospective enrichment with markers, following *in vivo* dynamics and the studying HSC ontogeny – have a real chance of improving the methods of HSC therapy, and ultimately improving outcomes for patients with a broad range of diseases.

Table 1. Diseases Commonly Treated with Hematopoietic Stem-Cell Transplantation.

Autologous transplantation*
Cancers
Multiple myeloma
Non-Hodgkin's lymphoma
Hodgkin's disease
Acute myeloid leukemia
Neuroblastoma
Ovarian cancer
Germ-cell tumors
Other diseases
Autoimmune disorders
Amyloidosis
Allogeneic transplantation†
Cancers
Acute myeloid leukemia
Acute lymphoblastic leukemia
Chronic myeloid leukemia
Myelodysplastic syndromes
Myeloproliferative disorders
Non-Hodgkin's lymphoma
Hodgkin's disease
Chronic lymphocytic leukemia
Multiple myeloma
Juvenile chronic myeloid leukemia
Other diseases
Aplastic anemia
Paroxysmal nocturnal hemoglobinuria
Fanconi's anemia
Blackfan–Diamond anemia
Thalassemia major
Sickle cell anemia
Severe combined immunodeficiency
Wiskott–Aldrich syndrome
Inborn errors of metabolism

* More than 30,000 autologous transplantations are performed annually worldwide, two thirds for multiple myeloma or non-Hodgkin's lymphoma.

† More than 15,000 allogeneic transplantations are performed annually worldwide, nearly half for acute leukemias. The vast majority are performed to treat lymphoid and hematologic cancers.

Table 1 Diseases that are currently treated with haematopoietic stem cell transplantation

From (Copelan, 2006), list of cancers and other diseases currently treated with allogenic and autologous haematopoietic stem cell transplantation.

1.2 Chasing HSCs: modelling the foundations of HSC ontogeny

The presence of a defined population of HSCs resident in the bone marrow prompts the question of how these cells are specified during embryogenesis, particularly since the appearance of circulating blood long precedes the formation of skeleton and bone marrow. This is not merely an esoteric question – understanding the formation of HSCs *in vivo* can help establish the fundamental requirements to

generate HSCs from alternative sources *in vitro*. Furthermore, during development HSCs are forming the growing blood population rather than just maintaining it (as in the adult). Consequently, the self-renewing embryonic HSC population is potentially a more potent therapeutic agent for reconstituting an ablated haematopoietic system than the relatively quiescent population of the adult.

A key challenge in the field of HSC research (to some extent still ongoing) has been to pinpoint the time and location in which HSCs first appear. The intrinsic mobility of haematopoietic cells makes this a more challenging feat than in most solid organ systems. Moreover, the reliance on measuring progeny as a proxy for the HSCs, rather than directly measuring HSCs, has produced seemingly conflicting findings that are still debated by some today.

1.2.1 The primitive wave of haematopoiesis

Haematopoietic cells are first detected in the mouse embryonic yolk sac at E7.5. Accordingly, the prevailing hypothesis for a long time was that HSCs originate from the yolk sac – a theory that can be easily reconciled with the trajectory of HSCs producing all other blood cell types. At this time point, nucleated erythrocytes are detectable in structures of the yolk sac called the blood islands (Maximow, 1924; Sabin, 1920). In sagittal sections, these appear to be a discrete compartment of undifferentiated cells surrounded by endothelial cells – ostensibly clonal structures. Following the derivation of *in vivo* and *in vitro* colony assays these structures were shown to host multi-lineage colony forming cells (CFU-S and CFU-C), bolstering the theory that long-term repopulating HSCs originate from the early yolk sac (Moore and Metcalf, 1970). This was reinforced by *in utero* transplantation of yolk sac cells into embryos (Toles et al., 1989; Weissman, I., 1978), although the long term contribution of these cells to adult haematopoiesis was low and irreproducible (Cumano et al., 2001; Medvinsky and Dzierzak, 1996; Medvinsky et al., 1993; Müller et al., 1994)

The yolk-sac theory of HSCs was challenged by fundamental chick-quail grafting experiments (Dieterlen-Lievre, 1975), and the discovery of limitations in the assays previously employed, such as the transient nature of *in vivo* splenic colonies (discussed further in 1.2.2). Moreover the yolk sac blood islands were in fact shown

to be an artefact of sagittal sectioning (Ferkowicz and Yoder, 2005). Indeed, when observed in whole mount these are actually a uniform band of cells surrounding the circumference of the yolk sac (Ferkowicz et al., 2003), which questions the theory of their clonality.

Although seemingly paradoxical given the trajectory of the haematopoietic hierarchy model, the now commonly accepted model for murine haematopoietic development, is that lineage restricted blood cells of the yolk-sac appear in a primitive wave of haematopoiesis which precedes the formation of HSCs (Dzierzak and Speck, 2008; Medvinsky et al., 2011; Orkin and Zon, 2008). This consists of erythrocytes, myeloid cells, and subsequent waves of progenitors (Bertrand et al., 2008; Palis et al., 1999), which are believed to be required to meet the initial energy/oxygen requirements of the developing embryo. Later this is superseded by a 'definitive' wave of haematopoiesis, in which the HSCs which populate the adult long-term are initiated (Figure 1-3; discussed in more detail below).

1.2.2 An embryonic source of definitive haematopoiesis

The intra-embryonic source of HSCs was first demonstrated by grafting experiments in the avian model. The premise of these experiments was the observation that previous studies of HSCs in yolk sac used day 7 embryos (Moore and Owen, 1967) in which the vasculature had already been established and therefore contribution from other organs could not be excluded from their results. Using chick and quail tissues, which could be distinguished by their nuclear morphology, embryonic tissue (from quail) was grafted onto extra-embryonic tissue (from chick blastoderm) at embryonic day 2, prior to vascularisation (Dieterlen-Lievre, 1975). The contribution of each of these tissues to adult haematopoietic organs could therefore be measured. Consistently, chick tissue was absent from the quail haematopoietic organs demonstrating that origins of haematopoietic cells must be intra-embryonic.

These grafts were unable to reach adulthood and therefore the long-term contribution to the adult haematopoietic system could not be determined. Importantly, the development of a method of distinguishing cells by B-antigens facilitated the use of allogenic chick-chick yolk sac grafts which could reach hatching stage (Lassila et al., 1978, 1982). In these chicks, yolk sac derived

erythrocytes were found transiently in the embryo, but after hatching only intra-embryonic cells were found in the peripheral blood. Thus, these experiments strongly suggested an embryonic and not extraembryonic source of cells that sustain the adult haematopoietic system long term.

1.2.3 The aorta-gonad-mesonephros as the primary origin of HSCs

The long-term haematopoietic contribution from the embryo body prompted the question of where the primary organ for long-term repopulating haematopoietic stem cells is located. Lineage tracing experiments in chick and Xenopus hinted that the source of adult haematopoiesis is the dorsal lateral plate mesoderm (Chen and Turpen, 1995; Dieterlen-Lièvre and Martin, 1981; Maéno et al., 1985). Comprehensive analysis of the mouse embryo from E8.0 (just as circulation begins), with the stringent assay for splenic colonies 11 days after transplantation (CFU-S_{11}) revealed that between embryonic day 9 and day 11, the embryo harbours cells of highly potent haematopoietic activity in a region encompassing the dorsal aorta, urogenital ridges, and mesonephros (AGM) region (Medvinsky et al., 1993). At the same stage, the yolk sac hosted 20 times fewer CFU-S_{11}. These experiments pointed to the AGM as the place of initiation of haematopoiesis, in agreement with studies in other species. However CFU-S_{11} assays didn't measure long term repopulation, moreover directed transplantation at these stages couldn't exclude the possibility that CFU-S_{11} are initiated elsewhere and migrate to the AGM region in the circulation. To address these points, studies of the long-term repopulating potential of the yolk sac and AGM were compared before and after culture (Medvinsky and Dzierzak, 1996; Müller et al., 1994). Using a newly development organ culture method it was possible to exclude the interference of migration in the embryo and demonstrate that the AGM region at E11 is the only site in with HSCs are autonomously initiated (Medvinsky and Dzierzak, 1996; Figure 1-3). While at E11.5 both yolk sac and AGM contained long-term repopulating HSCs, at E10.5 yolk sac had none and AGM had almost none (3/96). After organ culture of AGM region or yolk sac from E10.5, HSCs could only be detected from the AGM region. This suggested that HSCs are specified in the AGM region and subsequently migrate to other sites such as yolk sac.

More recently, the AGM region has been verified as the first site of HSC initiation in the human embryo (Ivanovs et al., 2011). Long term repopulating HSCs can first be detected, through direct transplantation at Carnegie stage 14 and in higher numbers than at Carnegie stage 15. Only subsequently are they found in the yolk sac, liver and placenta. Taken together, these studies suggest a potentially universal process in vertebrates whereby the definitive HSCs that will feed the adult haematopoietic hierarchy are first initiated in dorsal aorta a few days after gastrulation. However, this initiation is quickly followed by migration to a variety of other embryonic sites as is discussed further in 1.2.5 and 1.4.2.

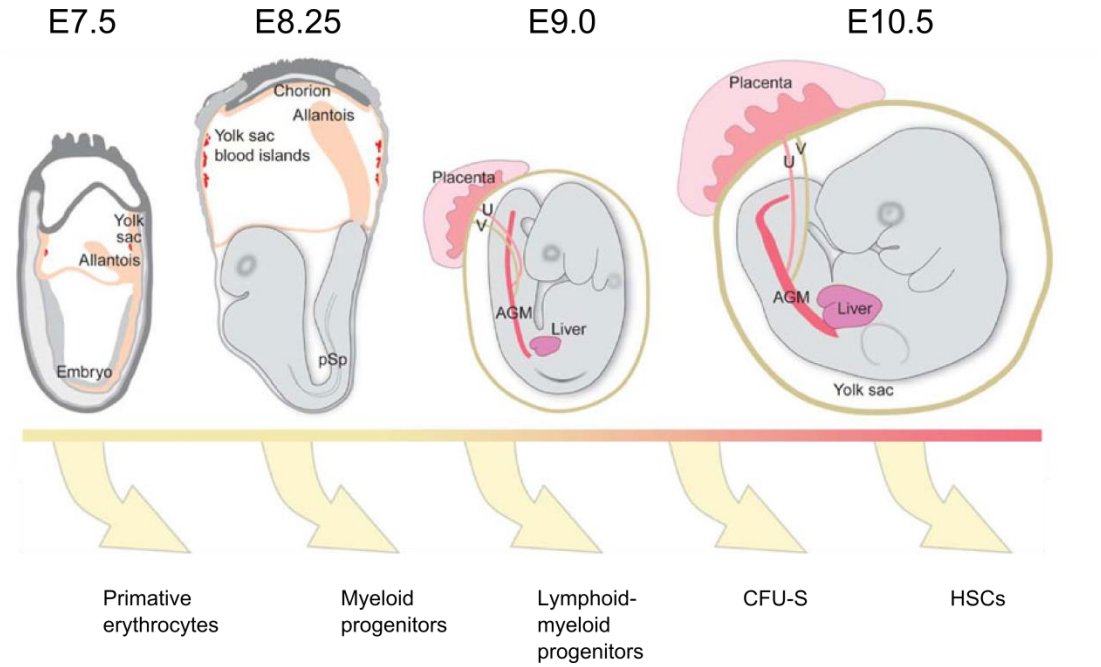


Figure 1-3 The ontogeny of haematopoiesis in the murine embryo

Adapted from (Dzierzak and Speck, 2008), shows the morphology of murine embryos from E7.5 to E10.5 and the sequential emergence of primitive haematopoietic progenitors at E7.5, followed by more mature progenitors, and then the first definitive haematopoietic stem cells (HSCs) at E10.5. P-Sp = para-aortic splanchnopleura, AGM = aorta-gonad-mesonephros, V = vitelline artery, U = umbilical artery.

1.2.4 Ongoing controversies on the earliest HSC precursors

While the AGM region as the earliest source of long-term repopulating cells is widely accepted, it is still debated. Some argue that the long-term repopulation assay is too stringent, and does not represent the physiological situation (Yoder and Hiatt, 1997). An alternative approach been proposed, whereby newborn mice are conditioned with myelosupressive busulfan – a milder treatment than irradiation, then transplanted with yolk sac cells. In these conditions, yolk sac cells gave rise to long-term reconstitution of the adult haematopoietic system (Yoder and Hiatt, 1997), but the same cells have never been shown to repopulate irradiated adult mice (Medvinsky and Dzierzak, 1996; Müller et al., 1994). Therefore, the nature of yolk sac cells in the development of the haematopoietic system remains unclear, and while it is not inconceivable that the yolk sac could harbour the earliest precursors of HSCs, it remains to be clearly shown.

1.2.5 Quantifying HSC emergence in the embryo

During the first few days of HSC appearance in the embryo, HSCs can be found at multiple locations. A striking transition is from the AGM region, in which HSCs are first detected (Medvinsky and Dzierzak, 1996), to the foetal liver, which harbours the largest number of haematopoietic stem cells during mid-gestation (Fleischman et al., 1982; Morrison et al., 1995). As early as E12 – one day after initiation in the AGM region – the foetal liver hosts approximately 50 HSCs and this pool rapidly expands in the next few days (Ema and Nakauchi, 2000). To explore the mechanisms, and potential reasons for such a dynamic migratory development, it is necessary to have a quantitative understanding of the embryonic locations that host HSCs during mid-gestation.

Quantification by direct transplantation in limiting dilution has demonstrated that the AGM region never hosts more than 1-2 HSCs (Gekas et al., 2005; Kumaravelu et al., 2002; Rybtsov et al., 2016). As well as the AGM region, approximately 1-2 HSCs have been found in the E11.5 yolk sac (Gekas et al., 2005; Kumaravelu et al., 2002; Müller et al., 1994), placenta (Gekas et al., 2005), circulating blood (Kumaravelu et al., 2002) and umbilical and vitelline arteries (de Bruijn et al., 2000; Gordon-Keylock et al., 2013). This could be an indication that

HSCs are initiated in multiple locations; however the presence of the circulation at this stage could alternatively facilitate the rapid migration to these sites (Figure 1-4). To distinguish between these hypotheses, explant culture systems have been used to autonomously expand HSCs from sites other than the AGM region, although the conclusions are not fully clear. The yolk sac has been shown to be capable of autonomously expanding HSCs at E12.5 but not E11.5 (Kumaravelu et al., 2002), however it remains unclear whether this is through initiation or expansion of the existing pool as HSCs were always present at the start of the culture. The placenta hosts an increasing number of HSCs between E11.5 and E13.5 (Gekas et al., 2005; Ottersbach and Dzierzak, 2005) but the capacity to autonomously initiate or expand HSCs has not been demonstrated. The extra-embryonic vessels have been shown to autonomously expand HSCs in culture, but in lower numbers than the AGM region (Gordon-Keylock et al., 2013). Importantly the foetal liver has never been shown to autonomously initiate long term repopulating HSCs in culture (Kumaravelu et al., 2002; Medvinsky and Dzierzak, 1996), except one study in which low level reconstitution of *Rag2-/-* mice was achieved (Kieusseian et al., 2012). These studies suggest that *de novo* HSC specification is largely from an arterial source, largely the AGM region, and to a lesser extent the extra-embryonic vessels.

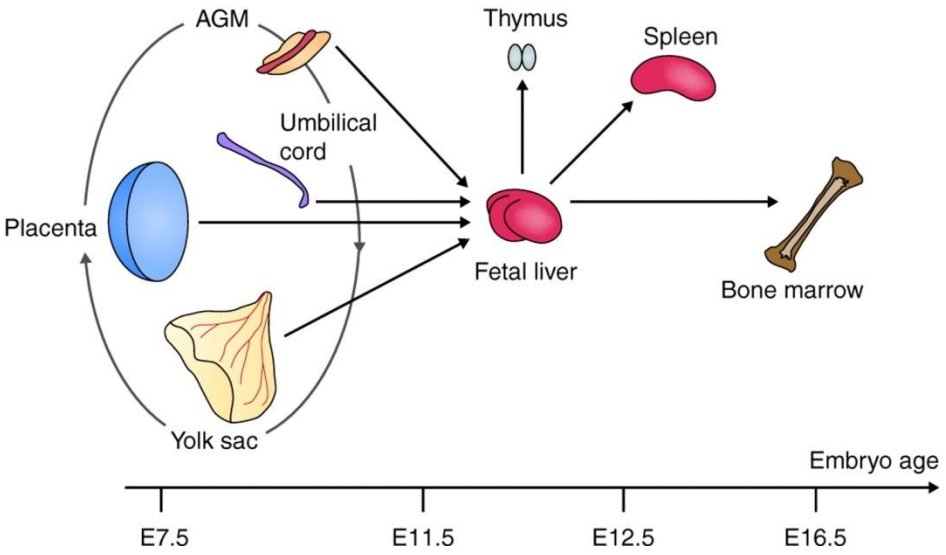


Figure 1-4 Multi-organ contribution to haematopoiesis during embryonic development

Adapted from (Medvinsky et al., 2011), shows the sites of appearance of haematopoietic progenitors and haematopoietic stem cells between E7.5 and E11.5 which potentially

migrate between different sites then colonise the foetal liver at E12.5. HSCs are expanded in the foetal liver and subsequently migrate to thymus, spleen and bone marrow.

Since the entire embryo contains only between 3 and 10 HSCs at E11.5, but by E12.5 has between 70 and 150 in the foetal liver, an important question is how HSC maturation or expansion can meet such a large population change. This could only be explained by either rapid expansion of the HSCs existing in the embryo at E11.5, or a maturation of precursor cells of HSCs. These two hypothesis were tested with a mixed culture of cells from the AGM region that had been labelled with carboxyfluorescein diacetate succinimidyl ester (CFSE) dye, enabling estimation of the number of cell divisions during the culture (Taoudi et al., 2008). The majority of repopulating HSCs in this culture were shown to undergo approximately four divisions, which could not explain the approximately 150-fold expansion in HSCs from the culture. This suggests that the predominant method of HSC expansion is through the maturation of a precursor cell type (pre-HSC) present in the AGM region. This quantitation of HSC dynamics in culture has recently been reconciled with the physiological process of HSC maturation in the murine embryo, through a new technique which quantifies precursors of pre-HSCs (Rybtsov et al., 2016). By taking early embryonic tissues and diluting them to limiting dilution prior to a period of culture, the number of pre-HSCs in the AGM region and the extra-embryonic vessels at late E11 has been quantified and shown to be equivalent to the numbers of HSCs in the foetal liver at E12. In contrast, the yolk sac and foetal liver have not been clearly shown to host pre-HSCs. Therefore, direct quantitative evidence suggests that the AGM region is capable of producing enough HSCs to seed the foetal liver, and it is therefore likely to be the predominant initiator of HSCs in the embryo (Figure 1-5).

Open questions remain as to exactly where HSC maturation from precursors mainly occurs. The likely explanations are either: HSCs are produced in the AGM region but immediately migrate after formation (therefore never more than one or two are detected); there is a short pulse of HSC maturation overnight in the AGM region which has not been captured experimentally; or that pre-HSCs gradually migrate to the extra-embryonic vessels and mature there, agreeing with the high

placental numbers (Gekas et al., 2005; Gordon-Keylock et al., 2013; Ottersbach and Dzierzak, 2005; Rybtsov et al., 2016). Or potentially a combination of the first and last hypothesis could be true: that pre-HSCs gradually migrate, but as soon as they mature, they enter circulated and migrate to foetal liver.

In summary, robust quantification of HSCs and pre-HSCs during development suggest that the AGM region is the most potent site of HSC initiation in the murine embryo, and hosts a large number of precursor cells that can mature into HSCs. To further our understanding of how HSCs are initiated, the AGM region must be examined with respect to both the lineage of cell types from which these are directly coming (which will be discussed in section 1.3) and the signals that direct these cells (which will be discussed in section 1.4 and 1.5).

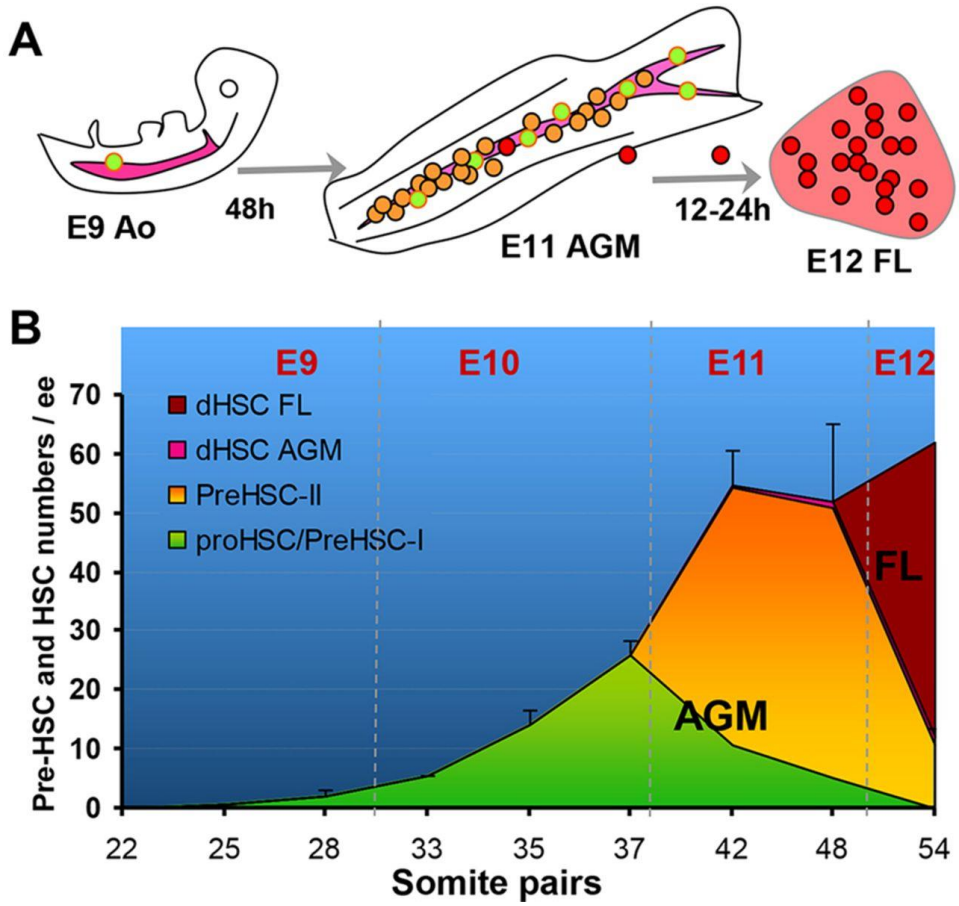


Figure 1-5 Quantification of the pre-HSC population supplying foetal liver HSCs

From (Rybtsov et al., 2016). Summarises quantitative experiments which support the model that a growing pool of precursor cells expands in number to reach sufficient levels to explain the rapid appearance of HSCs in the foetal liver at E12.

1.3 Tracing the genealogy of HSCs

The understanding of the first point in which haematopoietic stem cells appear in the embryo, and the demonstration of their autonomous initiation drives the question of which cells undergo this maturation to produce HSCs. Identifying the precursors of HSCs is essential in elucidation of the signalling mechanisms that enable HSC specification, and therefore how this could be replicated *in vitro*, for example to derive of HSCs from ES cells. Tracing rare HSCs from their embryonic ancestors *in vivo* is extremely challenging, and has required sophisticated combinations of imaging, markers and functional assays, which are outlined below.

1.3.1 The haemangioblast

The development of endothelium and blood has long been believed to be associated. As early as the 1920s the first haematopoietic cells forming in the yolk sac blood islands were observed to have a close and significant relationship with endothelial cells (Maximow, 1924; Sabin, 1920). This led to the hypothesis of a common precursor of haematopoietic cells and endothelium termed the haemangioblast.

Validation of the haemangioblast theory requires the proof that a single cell can produce both haematopoietic lineage and endothelial lineages, for example in a colony assay. A few studies have shown mixed endothelial and haematopoietic colonies from ES-blast derived colonies (Choi et al., 1998; Kennedy et al., 2007). Equivalent cells have been identified in the posterior primitive streak of the E7 embryo (Huber et al., 2004). However the physiological relevance of such cells has not been fully demonstrated as they could represent a multipotent mesodermal cell (Medvinsky et al., 2011). Therefore, the haemangioblast theory of HSC origins remains open to debate.

1.3.2 Haematogenic endothelium

A common theory of the early origins of HSCs is that they are initiated from endothelial cells. This theory is driven by observations of the morphology of the dorsal aorta, from which the emergence of luminal clusters of cells with endothelial and haematopoietic fate (termed "intra-aortic clusters") coincides with the first detection of HSCs (Dieterlen-Lièvre et al., 2006; Jaffredo et al., 1998). Thus, the precursors of HSCs are termed the "haematogenic endothelium", a specific subset of endothelial cells that can undergo a transformation into haematopoietic cells.

Support for the endothelial origin of HSCs has come from experiments using ES-derived cells in which a VC+CD45- population was shown to have preferential differentiation to myeloid and lymphoid fates over VC-CD45+ cells (Nishikawa et al., 1998a, 1998b). Moreover, live imaging has shown transitions from endothelial to haematopoietic cells in culture from ES-derived cells (Eilken et al., 2009), in the zebrafish aorta (Bertrand et al., 2010; Kissa and Herbomel, 2010) and from culture of AGM region slices (Boisset et al., 2010). These results reinforce the concept that endothelial cells are capable of producing haematopoietic cells.

Study of the endothelial to haematopoietic transition in mice has made use of the Sca1 (Ly6a) marker. Ly6a has been shown to aid enrichment of HSCs in the adult bone marrow to enable enrichment of HSCs (Spangrude et al., 1988); however not all embryonic HSCs could be enriched by this marker. A Ly6a-GFP transgenic model showed that Ly6a is indeed expressed in all HSCs of the AGM region and provides a > 100-fold enrichment in adult bone marrow (de Bruijn et al., 2002; Ma et al., 2002). Furthermore Ly6a-GFP+ cells have been visualised in the endothelium of the dorsal aorta (de Bruijn et al., 2002) and live imaging has shown that transition from Ly6a-GFP+ cells with endothelial morphology to rounded Ly6a-GFP+ cells with haematopoietic morphology (Boisset et al., 2010). Cumulatively these experiments have given evidence for an endothelial to haematopoietic transition underlying the specification of HSCs, and indeed has been used as the basis of genome-wide RNA-sequencing studies (Kartalaei et al., 2015). However, a remaining limitation of the hypothesis that haematogenic endothelium produces HSCs is that the experiments rely on phenotypic markers of HSCs. While correlated with the timing and location of HSC initiation, it still remains to be shown directly by a lineage tracing or maturation and long-term repopulation assay that the Ly6a-GFP+ endothelium is the direct precursor of HSCs in the AGM region.

1.3.3 The pre-HSC lineage

Tracing the immediate precursors of HSCs has been greatly aided by the development of a culture system that captures the maturation of HSCs, but is amenable to experimentation, due to a step in which cells are dissociated and can be reaggregated with cells of another source (Taoudi et al., 2008). This assay (referred to as the reaggregate culture hereafter), enabled the tracing of a VC+CD45+ population of cells, which are direct precursors of HSCs as the sorted population could be distinguished from the supportive niche cells by genetic marking with GFP. This protocol has been modified, by adding cells of the OP9 stromal cell line derived from osteopetrotic (op/op) newborn mouse calvaria (Nakano et al., 1994), to support the maturation of cells as early as E9.5 and thus the phenotype of the precursor lineage could be directly traced from this stage (Rybtsov et al., 2011, 2014). These experiments have therefore shown functionally that HSCs develop from a $VC^{+}CD45^{-}CD41^{lo}CD43^{-}$ (or pro-HSC) population present at E9.5 (Rybtsov et al., 2014), to a

$VC^{+}CD45^{-}CD41^{+}$ (or Type I pre-HSC) population present at E10.5 (Rybtsov et al., 2011) to a $VC^{+}CD45^{+}$ (or Type II pre-HSC) population at E11.5 (Taoudi et al., 2008) (Figure 1-6).

The importance of a VC as a marker of HSC precursors is in keeping with a potential endothelial ancestry. However, endothelial markers do not necessarily equate to endothelium, as morphology of the endothelium is the defining feature, and the direct spatial relationship to the aortic endothelium has not been definitively proven. The endothelial characteristics of HSC precursors can be rationalised by two main hypotheses: that precursors have an endothelial origin; or alternatively, endothelial markers may be transiently expressed in precursors as to aid their passage through the lining of the aortic endothelium *en route* from the mesenchyme to the circulatory system. Distinguishing these theories will require functional lineage tracing assays that can retain the *in vivo* spatial information.

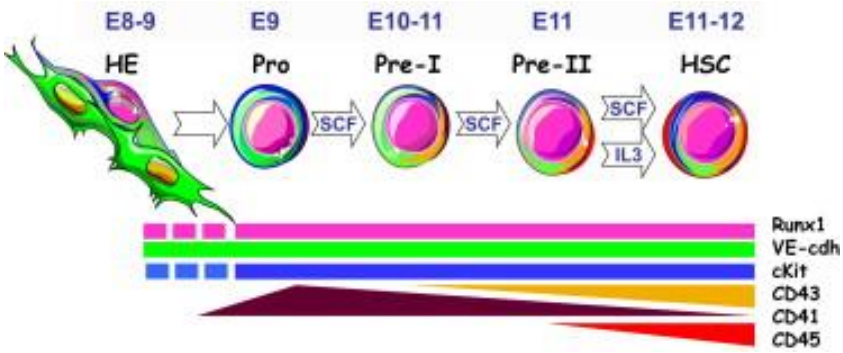


Figure 1-6 The precursor lineage producing HSCs in the AGM region

Diagram from (Rybtsov et al., 2014) shows the maturation of defined precursor populations in the AGM region between E9.5 to E11.5 can be captured through the sequential upregulation and downregulation of surface markers. The progression of precursors is dependent on SCF, and the final stage of maturation for Type II pre-HSC to HSC is also enhanced by IL3.

1.4 The haematopoietic stem cell niche through space and time

Like many developmental processes, the emergence and maintenance of HSCs requires a combination of inductive cues and cells competent to respond to

these cues. Such interchange of signalling may occur in a spatially restricted location, termed the niche. While a niche can be described as the specific tissue location in which a stem cell is normally detected, a more functional definition of a niche states that if transiently depleted of its stem cells, a niche should take up and maintain a newly introduced stem cell (Morrison and Spradling, 2008). The HSC environment meets both of these criteria: the position of HSCs is generally restricted to specific tissues, despite their ability to enter circulation; when depleted of HSCs through irradiation, adult haematopoietic organs are reconstituted by HSCs following transplant and the embryonic niche can take up developing HSCs after depletion (detailed in 1.4.2). Consequently, the haematopoietic system is regarded as an important paradigm of stem cell-niche interactions (Orkin and Zon, 2008). This section briefly discusses HSC niches in the adult and embryo. Specific focus is on how our understanding of the AGM region, as the niche in which HSCs first appear, has been refined to generate a spatio-temporal map of supportive regions, and the potential contribution of multiple cell types to this environment.

1.4.1 The adult bone marrow niche

In the adult, the bone marrow has long been understood to be a key haematopoietic location, and since the work of Till and McCulloch (Till and McCulloch, 1961), it has been shown to host haematopoietic stem cells. Furthermore, the reconstitution of bone marrow demonstrates its role is functional, not just circumstantial. HSCs can also be identified in spleen and liver, but generally only under stress conditions such as haematopoietic malignancies (Morrison and Spradling, 2008). As this is concomitant with a great expansion in numbers, it suggests that the haematopoietic system may use facultative niches to regulate different HSC functions.

The bone marrow is in fact a complex mixture of cell types, including endosteal surface, osteogenic cells, adipocytes and a vascular network of endothelial and perivascular cells. The exact cellular location of HSCs and role of each of these cell types in their regulation has remained a point of debate. As HSCs can only be defined functionally, identifying their exact spatial location is challenging. The establishment of refined sets of markers to enrich populations has aided this task, as

some studies use combinations of markers for imaging HSCs within the bone marrow niche (Kiel et al., 2005; Lo Celso et al., 2009; Nombela-Arrieta et al., 2013). Studies of endogenous HSC locations with the SLAM family of surface proteins show a large number of target cells adjacent to sinusoidal blood vessels, suggesting the importance of the vascular niche (Kiel et al., 2005). However the fact that these marker sets enrich HSCs to approximately 50% lends the question of which of the imaged cells in the niche are truly functional HSCs.

The osteoblastic niche has been proposed as a key regulator of HSC numbers (Calvi et al., 2003; Zhang et al., 2003). Genetic models in which osteoblast numbers could be stimulated to increase in number led to an increase in HSC numbers (Calvi et al., 2003; Zhang et al., 2003). The regulation of the osteoblastic niche is believed to be via adhesion molecules such as N-cadherin (Zhang et al., 2003) although this has been contradicted by depletion studies showing no effect on HSCs (Kiel et al., 2009), possibly implying regulation by redundant mechanisms. Ablation of critical regulatory cytokines in distinct population types has suggested that the HSCs mainly reside in the perivascular niche within bone marrow (Ding et al., 2012). However the exclusivity of conditional promotors has called into question whether different cell contributions can be exactly discriminated in this way (Hoggatt et al., 2016). Hence, the exact mechanisms of HSC regulation within the adult niche remains to be fully clarified, but is likely to be constituted by an assortment of cell types, whose roles can differentially regulate HSCs to meet their diverse functions.

1.4.2 The embryonic HSC niche – a migratory journey

Similarly to the adult, during embryonic development, different HSC behaviours correlate with different niches. During embryonic development, HSCs are found in a variety of locations at different times. The earliest HSCs are initiated in the AGM region, quickly followed by yolk sac, placenta and vitelline vessels but only ever in small numbers (Gekas et al., 2005; Kumaravelu et al., 2002; Medvinsky and Dzierzak, 1996; Müller et al., 1994; Rybtsov et al., 2016). Soon the majority of HSCs in the embryo are detected in the foetal liver, concomitant with a huge expansion in numbers (Ema and Nakauchi, 2000; Fleischman et al., 1982; Morrison et al., 1995). HSCs at this stage are found in a highly proliferative state compared to

adult (Bowie et al., 2006, 2007a). Subsequently HSCs are found in spleen and bone marrow (Christensen et al., 2004; Gekas et al., 2005) (Figure 1-4).

Thus, it can be inferred that the AGM region provides the niche that facilitates HSC specification but not maintenance, whilst foetal liver provides a niche for rapid HSC expansion, prior to relocation to bone marrow which maintains quiescent state for homeostatic roles. Indeed pre-HSCs of the AGM region cannot mature in to transplantable HSCs autonomously, suggesting *a priori* that they are dependent on external cues. Furthermore cells of the AGM region can be depleted of pre-HSCs, and provided external pre-HSCs which will mature into HSCs in reaggregate culture (Taoudi et al., 2008), demonstrating its functional role as a niche.

1.4.3 The spatial polarisation of the AGM region

To understand the niche cues that help direct HSC maturation, a precise understanding of the location of the niche is required. A lack of definitive markers of HSCs means that this has had to be determined through precise micro-dissection and functional experiments. Functional assays have shown that within the AGM region, the dorsal aorta is the earliest source of hematopoietic stem cells (de Bruijn et al., 2000). Within the aorta, the observation that intra-aortic clusters are preferentially found in the ventral wall has suggested that HSC activity may be ventrally polarised in the AGM region (Ciau-Uitz et al., 2000; Jaffredo et al., 1998; Smith and Glomski, 1982; Tavian et al., 1996). This has been shown in mouse by transplantation with AGM tissues that have been sub-dissected into dorsal (AoD) and ventral (AoV) parts and HSCs found to be approximately four times more frequent in the ventral side than dorsal side (Souilhol et al., 2016a; Taoudi and Medvinsky, 2007). Along the rostral-caudal axis, the dorsal aorta has been shown to harbour HSCs in a central region, close to the junction of the vitelline artery (Mascarenhas et al., 2009).

In addition to quantification of the number of HSCs in different aspects of the AGM region, the functional role of proximal tissues has been examined though the use of *ex vivo* reaggregate cultures. Co-culture of the AGM region with ventral or dorsal tissues (gut or notochord respectively), has indicated that ventral tissues promote HSC expansion from early E10 embryos, suggesting that HSC specification may be regulated by positional cues coming from surrounding tissues (Peeters et al.,

2009). Co-culture experiments combining dorsal and ventral tissues of E10.5 and E11.5 embryos has highlighted that the dorsal domain enhances the production of HSCs from E10.5 AoV, while at E11.5 the ventral domain enhances the production of HSCs from AoD (Souilhol et al., 2016a). Furthermore, the urogenital ridges (UGRs) enhance HSC production from AoV at E10.5 and E11.5. Thus, the position of HSCs and their precursors, as well as the cells/signals supporting their emergence, appears to be highly polarised in the AGM region (Figure 1-7). To what extent the polarisation of the environment influences the polarisation of HSC emergence, or the presence of HSCs determines the polarisation of supportive cues, remains an open question. Nonetheless, the refined functional understanding of spatial polarity of the AGM region provides an important platform for further exploration of molecular components regulating HSC emergence and is therefore one of the key foundations of this thesis (as discussed in section 1.7).

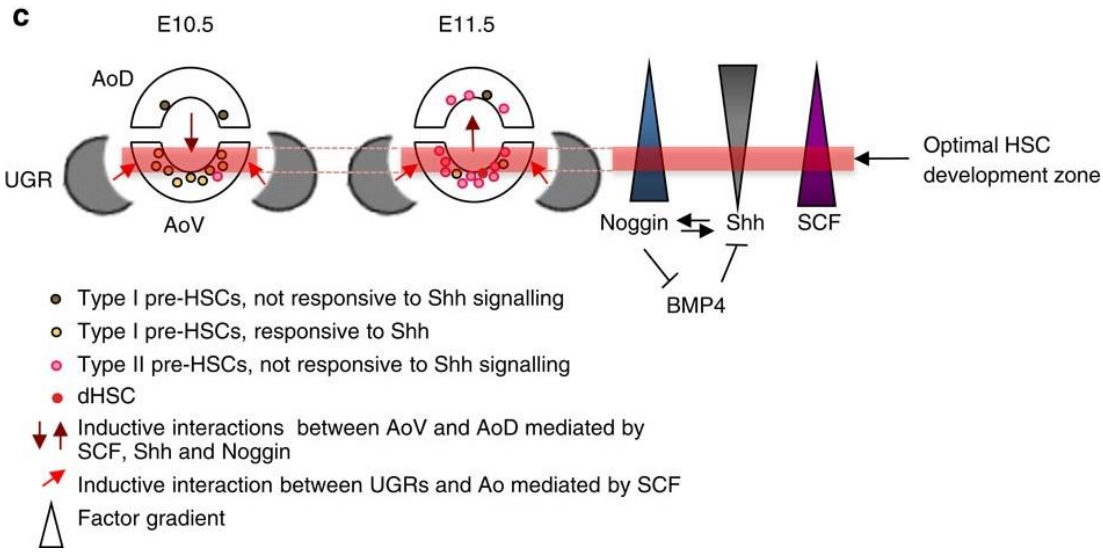


Figure 1-7 Dorso-ventral polarity of the AGM region

Diagram from (Souilhol et al., 2016a) summarises the polarised distribution of pre-HSCs and HSCs in the AGM region as well as the relative contribution of proximal tissues to the support of HSC maturation. Noggin, Shh and SCF are key secreted factors that appear to contribute to the polarisation of functional support from the niche.

1.4.4 Cellular constituents of the AGM region

The AGM region is a complex environment encompassing multiple cell types and developing organs (Figure 1-8). Although the emergence of HSCs appears to be associated with the floor of the dorsal aorta, the contribution of surrounding cell types is unclear, but likely to have an influential role through both short and long-range signalling. Therefore, understanding the complete picture of the cellular constituents of the niche should aid the elucidation of the full signalling environment of the AGM region.

1.4.4.1 Endothelial cells

The endothelium – marked by its expression of markers such as VC (Lampugnani et al., 1992) and CD31 (Privratsky and Newman, 2014) – is one of the primary niche cell types (also a potential HSC precursor) during HSCs ontogeny, being one of the main constituents of the aortic wall. It is, however, potentially not a uniform cell type. Indeed through lineage tracing in chick, the ventral and dorsal aspect of the dorsal aorta have been show to distinct developmental origins with the ventral aspect originating from splanchnopleural mesoderm and the dorsal aspect from somatic mesoderm (Pardanaud et al., 1996). These diverse developmental origins could underlie their functional differences. Moreover, the understanding that HSCs, pre-HSCs and intra-aortic clusters only ever arise from arterial and not venous endothelium (de Bruijn et al., 2000; Gordon-Keylock et al., 2013; Taoudi and Medvinsky, 2007) suggests that the diversity of endothelial cells in the AGM region is an important consideration regarding their contribution to the niche environment and the HSC precursor lineage.

1.4.4.2 Perivascular cells

All vasculature endothelium is surrounded by a layer of non-endothelial, perivascular cells attached to the vessel wall (Armulik et al., 2011; Bianco et al., 2013; Jain, 2003). In the mouse bone marrow, these perivascular cells have been shown to have properties of skeletal stem cells such as clonal production of skeletal cells lineages, formation of bone-marrow ossicles that are engrafted by haematopoietic cells following transplantation, and self-renewal (Méndez-Ferrer et al., 2010; Sacchetti et al., 2007). These are interesting properties in the study of

haematopoiesis as they apparently can found the structure and regulate the vascular organisation necessary for a haematopoietic niche. However the generalisability of properties of perivascular cells is a subject of debate, as some claim that cells with similar properties of multipotency (indeed even the capacity to generate muscle, myocardium, endothelium) can be prospectively isolated from a range of tissues based on markers (Crisan et al., 2008; Meirelles et al., 2006). However such studies fail to show a robust, and consistent multipotent behaviour of these populations in clonal or *in vivo* assays, unless the cells are subjected to extended periods of culture (Bianco et al., 2013).

An alternative, hypothesis is that through a common mechanism of homophilic interaction with endothelial cells (mediated by CD146 and other genes), lineage restricted progenitors are recruited from the surrounding tissue to the endothelium and have different functional properties depending on the tissue context (Bianco et al., 2013; Díaz-Flores et al., 2009; Sacchetti et al., 2007). Although a layer of perivascular cells expressing CD146, PDGFRβ and NG-2 has been shown around the dorsal aorta (Mirshekar-Syahkal et al., 2014), little is known about the functional characteristics of these cells. The osteogenic, adipogenic and chondrogenic potential of the AGM after culture has been demonstrated (Mendes et al., 2005) but again the lack of clonal assays left the physiological relevance of these properties unclear. Given the skeletal potential and – even more interestingly – the vasculature modelling potential of cells perivascular to the bone marrow stroma (Sacchetti et al., 2007), the idea that perivascular cells may regulate the dorsal aorta and consequently HSC maturation in the AGM region is an attractive prospect but one which requires experimental support.

1.4.4.3 Gut

Ventral to the dorsal aorta is the developing gut. Co-culture experiments with the AGM region have shown that the gut provides some key inductive signalling to HSC development (Peeters et al., 2009). When cultured with an anti-hedgehog blocking antibody this inductive signalling from gut is abrogated, suggesting that the inductive signalling is via hedgehog signalling. However, the lack of specificity of the application of the blocking antibody obscures the conclusion that the gut itself provides hedgehog directly, as this result could be explained by a secondary role of

hedgehog signalling in the gut-AGM culture. Moreover a limitation to the conclusions about the role of the gut in this study is the presence of HSCs in the vitelline arteries which may contaminate the gut culture (Gordon-Keylock et al., 2013)

1.4.4.4 Primordial germ cells

The gut also contributes to the multi-cellular AGM environment as a source of primordial germ cells. Between the mouse embryonic stages E9.5 to E10.5, primordial germ cells migrate from the gut, through the ventral mesenchyme to the urogenital ridges cells (Medvinsky et al., 1993, 1996; Molyneaux et al., 2001; Wakayama et al., 2003; Yokomizo and Dzierzak, 2010). Whether they contribute to the development of HSCs is unknown – some studies have proposed that they could actually represent the early precursors of HSCs as they have multipotent characteristics, appear in the AGM region at the same time HSCs are specified, and can produce haematopoietic cells in culture (Rich, 1995). Transplantable HSCs have never been traced from a PGC cell of origin, however a dependency on common signalling such as SCF and Cxcl12 (Medvinsky et al., 1993, 1996; Molyneaux et al., 2003; Pesce et al., 1993), could imply a direct or indirect coregulatory interaction or perhaps a propensity of the AGM environment to provide key common signalling environments for developmental processes.

1.4.4.5 Mesonephros/mesonephric tubules

Another notable tissue in the AGM environment is the developing mesonephros, an intermediate structure during kidney development. The formation of this structure begins with the single epithelial layer mesonephric duct which develops in a cranial-caudal direction between E9.5 and E11.5 (Davidson, 2008; Dressler, 2006; Vize et al., 2003). As this duct progresses caudally, the mesenchyme condenses to produce mesonephric tubules. By E11.5 there are 11 pairs of tubules, which sit ventro-lateral to the dorsal aorta. In the mouse, the mesonephros is a transient structure that degenerates after E11.5 as its function is replaced by the metanephros, which forms the adult kidney. The metanephros is initiated at E10.5 from a distinct mesenchyme at the level of hindlimb called the metanephric mesenchyme. The influence of mesonephros on HSC development has not been

demonstrated, however (like PGCs) its appearance and degeneration coincides with the initiation of HSCs in the AGM region and it makes a significant contribution to the mesenchyme. Furthermore the expression of several genes associated with haematopoiesis has been shown in mesonephric tubules, such as *Ly-6E* (Miles et al., 1997), *Gata3* (Lakshmanan et al., 1999) and in the mesonephric duct: SCF (*Kitl*) (Souilhol et al., 2016a). Thus, it may be a potentially significant contributor to the signalling in the ventral mesenchyme.

1.4.4.6 Sympathetic nervous system

A role of the sympathetic nervous system has recently been implicated in regulation of HSCs of the AGM region (Fitch et al., 2012). In this study Gata3−/− embryos exhibited impaired sympathoadrenal differentiation and showed reduced functional and phenotypic HSCs compared with wild-type embryos. However, treatment with catecholamines was able to rescue HSC numbers in these embryos highlighting them as critical regulators. Moreover, AGM tissues from Th−/− embryos, which could not synthesize catecholamines, and wild-type AGM regions cultured as explants in the presence of a tyrosine hydroxylase inhibitor, also showed a reduction in HSC numbers.

1.4.4.7 Haematopoietic cells

Haematopoietic cells have a potential contribution to the AGM niche environment in many regards. Cells budding from the aortic lumen are exposed to the full blood circulation, which can carry signalling molecules from any tissue. The budding intra-aortic clusters may host a mixture of HSCs and progenitors, which may cross signal to each other. In bone marrow macrophages have been shown to regulate retention of HSCs in the bone marrow niche (Chow et al., 2011). The role of macrophages in the developmental niche remains unclear, although macrophage-colony-stimulating factor has been shown to reduce the number of haematopoietic progenitors in favour of endothelial cells in culture, and consequently OP9 cells, derived from M-CSF negative osteopetrotic (op/op) mice is a suitable stromal support for AGM culture (Minehata et al., 2002; Nakano et al., 1994). On the other

hand macrophages have been shown to mediate HSPC migration through the AGM stroma in zebrafish (Travnickova et al., 2015).

1.4.4.8 Osteogenic cells

In the adult bone marrow, osteoblast cells have been proposed to play and important role in regulating HSC numbers. In the AGM region, the expression of markers of osteogenic cells has been described (Mascarenhas et al., 2009; Mirshekar-Syahkal et al., 2014), however whether there is a functional basis to this is not known.

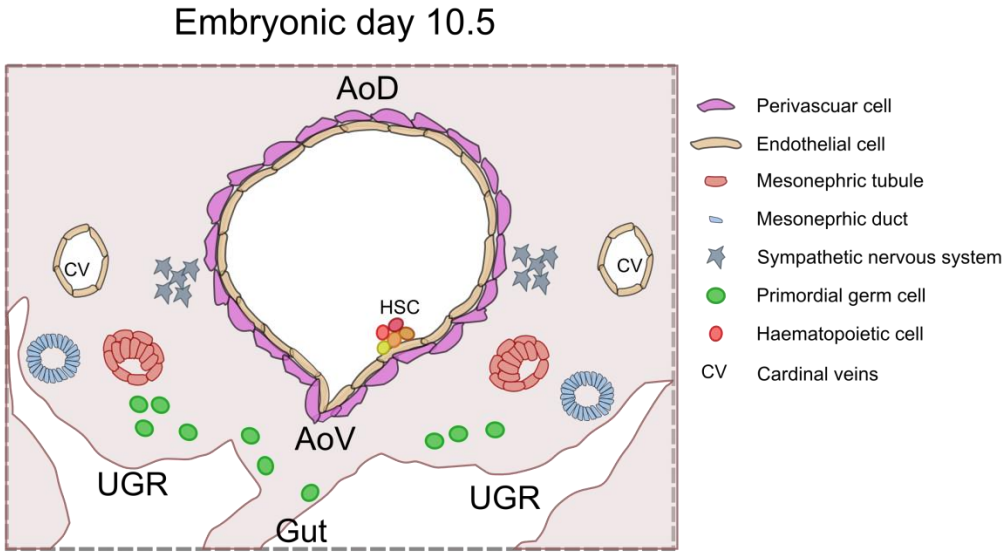


Figure 1-8 The cell composition of the AGM region

Schematic represents the different cell types constituting the mouse AGM region at E10.5. Representative symbols are described in the key and include perivascular cells, endothelial cells, mesonephric tubules, mesonephric ducts, sympathetic nervous system, primordial germ cells (PGCs) and haematopoietic cells.

1.5 The molecular regulation of HSC specification

The elucidation of the time, place, and lineage from which HSCs first emerge prompts the question of exactly how this transition is executed through a combination of molecular events within the developing cells and signalling cues from the environment. As HSCs cannot be cultured long term *in vitro*, this work has largely relied on *in vivo* studies of genetically modified mice. A number of regulatory molecules have been elucidated based on their knockout phenotypes, but

often the details of their action, or even whether they act in a cell-autonomous or non-cell autonomous fashion, remain elusive. In this section I will discuss the current understanding of the main molecular regulators of HSCs broadly divided into the transcription factors driving cell-autonomous regulation and the signalling pathways involved in non-cell autonomous regulation. Moreover I will briefly discuss how this understanding is being applied to seek methods for *de novo* HSC specification.

1.5.1 Transcription factors

Transcription factors are the key molecules defining cell identity. By regulating expression of a large number of genomic regions and, importantly, regulating each other, they can maintain a stable cell state or drive the transition into a new state. Therefore, understanding the essential transcription factors for establishment or maintenance of HSCs is an important area of research. A number of transcription factors, which are often aberrantly expressed in leukaemia, have been shown to have a regulatory role in haematopoiesis through *in vivo* studies of knockout mice.

The following discussion concerns three groups of regulators: key regulators of both primitive and definitive haematopoiesis, Scl and Lmo2; transcription factors that, interestingly, appear to act only on the maturation of HSCs (or endothelial to haematopoietic transition); and finally a group of regulators of HSC self-renewal HoxB4, HoxA9, PU.1, Erdr1, and cFos that expand existing HSCs.

1.5.1.1 Scl/Tal

The basic helix-loop-helix transcription factor Scl/Tal is a critical regulator of haematopoiesis. First discovered as a common aberrantly activated gene in T-cell leukaemias (Aplan et al., 1990; Brown et al., 1990) it has been shown to be essential for primitive and definitive haematopoiesis in the murine embryo. Double knockout mice die around E9.5, and show a complete lack of haematopoiesis in the yolk sac or embryo (Robb et al., 1995; Shivdasani et al., 1995). Furthermore, chimeric mice show that Scl knockout cells fail to contribute to definitive haematopoiesis later in development (Porcher et al., 1996) suggesting that Scl/Tal1 is a fundamental regulator of the haematopoietic cell lineage.

Conditional deletion in Mx1-Cre mice (when cre is expressed drives knockout after induction by interferon) studies suggest that Scl is indeed dispensable to HSC self-renewal, engraftment and differentiation (Mikkola et al., 2003). However a double knockout of Lyl1/Scl in adult abrogates their repopulating function and leads to apoptosis (Souroullas et al., 2009) suggesting that the role of Scl in steady state haematopoiesis is perhaps preserved through the action of a functionally redundant molecule.

1.5.1.2 Lmo2

The LIM-finger protein Lmo2, another aberrantly regulated gene in T-cell leukaemia, has been shown to abrogate haematopoiesis in knockout mice (Warren et al., 1994). These embryos fail to form blood islands in the yolk sac, do not form erythroid cells in the embryo or yolk sac and die at E10.5. In adult chimeric mice, cells derived from Lmo2-/- ES cells fail to contribute to any adult haematopoiesis, suggesting that, like Scl, Lmo2 is an essential regulator of both primitive and definitive haematopoiesis.

1.5.1.3 Runx1

Runx1 is believed to be a core regulator of definitive HSCs and has been extensively studied in the last 20 years. Also known as a member of the core binding factors (CBF) family or AML1, Runx1 was first cloned due to its association with the t(8;21) translocation which is frequently found in patients suffering from acute myeloid leukaemia (Miyoshi et al., 1991). Homozygous knockout of Runx1 leads to embryonic death between E12.5-E13.5 and is associated with haemorrhaging and a lack of mature haematopoietic cells in foetal liver (Sasaki et al., 1996; Wang et al., 1996a, 1996b) suggesting a specific effect on definitive haematopoiesis (in which HSCs are formed).

Subsequent analysis has shown that Runx1 is expressed in the endothelial and mesenchymal cells at all sites of HSC specification including the AGM region, vitelline and umbilical arteries and yolk sac (North et al., 1999). In Runx1 knockout mice, intra-aortic clusters fail to form, therefore combined with the haematogenic endothelium theory of HSC formation, this could suggest that Runx1 is expressed in precursors of HSCs and is necessary to facilitate the transition from HSC precursors

(or haematogenic endothelium) into mature HSCs (Medvinsky et al., 2011; Speck and Gilliland, 2002; Swiers et al., 2010; Yokomizo et al., 2001).

Whether Runx1 is associated with the specification or maintenance of HSCs is subject to debate. Establishing the exact time in which Runx1 knockout blocks HSC development requires careful *in vivo* genetic experiments to precisely time the knockout. Such experiments have suggested the Runx1 is necessary to promote the maturation from VE-cad+CD45−CD41+ (Type I pre-HSC) to VE-cad+CD45+CD41+ (Type II pre-HSC) but not earlier (Liakhovitskaia et al., 2014). While deletions of *Runx1* under a Vav1-Cre system indicate that Runx1 is dispensable in mature HSCs (Chen et al., 2009). Thus, Runx1 appears to have a key cell autonomous role in driving the maturation of the HSC precursor lineage.

1.5.1.4 GATA family

The Gata family of transcription factors has six members, which share a common zinc-finger DNA-binding motif. These are broadly divided into the class of haematopoietic regulators: Gata1, Gata2 and Gata3 (Orkin, 1995); and cardiac regulators: Gata4, Gata5, Gata6 (Charron and Nemer, 1999; Molkentin, 2000) although there may be some function outside of these tissues (Bresnick et al., 2012).

Of the three haematopoietic regulators, Gata2 has been strongly associated with HSC function specifically (Ling et al., 2004; Tsai and Orkin, 1997; Tsai et al., 1994). Gata2 homozygous knockout embryos die around E10.5 and exhibit severe anaemia. In chimeric mice, Gata2-/- cells are able to contribute to primitive erythrocytes but do not contribute to foetal liver or bone marrow, suggesting that Gata2 is specifically required for definitive haematopoiesis (Tsai et al., 1994). Mice with a dose deficiency of Gata2 show a reduced capacity to initiate or expand HSCs in culture of AGM region, but not yolk sac or foetal liver. Moreover bone marrow HSC numbers remain normal in Gata2+/- mice, but their capacity for secondary transplantation or competitive repopulation is markedly reduced (Ling et al., 2004). Thus, the function of Gata2 is potentially specific to the initiation and expansion of HSCs.

1.5.1.5 Gfi1

Gfi1 was first identified as a transcription factor capable of conferring growth-factor independent proliferation of T-cell lymphoma cell lines and Th2 cells (Akagi et al., 2004; Gilks et al., 1993; Zhu et al., 2002) but has also been shown to be expressed in HSCs (Phillips et al., 2000). Although knockout mice are not embryonic lethal, they have a reduced capacity to serially transplant mice, and are rapidly out-competed in competitive transplantation studies (Hock et al., 2004a). In concert with the increase proliferation rate in knockout mice, Gfi1 is believed to be a critical factor in restricting the proliferation rate of HSCs therefore retaining the pool of functionally integral HSCs.

Gfi1 has also be proposed as a regulator of endothelial to haematopoietic transition as overexpression was able to rescue the formation of haematopoietic colonies and CD41+ cells in Runx1-/- mice (Lancrin et al., 2012).

1.5.1.6 Other regulators of HSC self-renewal

Whilst the transcription factors described above appear to be key to establishment of the haematopoietic lineage, a number of other factors appear to regulate HSC functionality. Bmi1 (Lessard and Sauvageau, 2003; Park et al., 2003), Etv6 (Hock et al., 2004b) and PU.1 (Burda et al., 2010), knockouts result in adult HSC defects while Sox17 knockout has reduced foetal liver HSCs (Kim et al., 2007). Furthermore, HoxB4, HoxA9, PU.1, Erdr1, and cFos expand HSC activity when overexpressed *in vitro* (Deneault et al., 2009).

Many of the essential transcription factor regulators of HSCs have been identified in the last 25 years. This is often through the examination of a knockout mouse line which shows a haematopoietic deficiency, usually manifesting as lethality around E10.5, when definitive haematopoiesis is being initiated. However, establishing a deeper understanding of whether their role is in initiation, self-renewal, differentiation or engraftment in HSCs remains more challenging and relies on precisely regulated *in vivo* genetic experiments, as early lethality can obscure an essential role later on in development. Moreover, regulators of HSC functions such as self-renewal, or partial functional redundancy between transcription factors, may not be apparent from steady state numbers of HSCs *in vivo.* In these cases, competitive repopulation studies can be highly informative. Finally, since some of

these transcription factors are expressed in non-HSC cells (e.g. Runx1), exclusivity of these functions to the HSC lineage, or whether they act through a secondary mechanism (such as the niche cells) will be important to establish. As it stands, the full picture of transcriptional regulation of HSCs has not been elaborated, but all progress in this area has a major effect on the ability to model HSC behaviour, and potentially derive this *in vitro.*

1.5.2 Non-cell autonomous regulators of HSC emergence

Given the key role of the niche in defining HSC behaviour in the adult and during development, understanding the mediators of these niche regulators can give important insight into how HSCs can be controlled *in vitro*. Secreted molecules are important effectors of long and short-range signals between cells. A number of such molecules that have a significant impact on the regulation of HSCs have been elucidated and are discussed here, with particular focus on the regulation of HSC molecules that ontogeny.

1.5.2.1 Stem cell factor

Stem cell factor (SCF) or Kit-ligand (Kitl) is a secreted protein that acts through binding of cKit receptor. As cKit is found on all HSCs throughout development and adulthood (Gekas et al., 2005; Morrison et al., 1995; Ogawa et al., 1991; Rybtsov et al., 2014; Sánchez et al., 1996), and defects in the receptor have been shown to reduce Thy-1^{lo}Lin-Sca-1^{+} and CFU-S numbers (Ikuta and Weissman, 1992) as well as causing embryonic death due to anaemia (Broudy, 1997), SCF has been an important focus in the elucidation of secreted regulators of HSCs. In the adult bone marrow, perivascular cells have been shown to produce the primary source of SCF as conditional deletion with cre-recombinase under the regulation of leptin receptor expression leads to depletion of bone marrow HSCs (Ding et al., 2012).

Addition of SCF to cultures of foetal liver cells has been shown to increase the survival of HSCs *ex vivo* (Bowie et al., 2007b). SCF has also been added to cultures of HSC precursors from E11.5, E10.5 and E9.5 and has been shown to be necessary for the 150-fold expansion of HSCs from E11.5 AGM culture and sufficient to drive the maturation of pro-HSCs and Type I pre-HSCs (in the presence

of serum and other niche cells) (Rybtsov et al., 2011, 2014, 2016; Taoudi et al., 2008). SCF is expressed broadly in the ventral side of the AGM region (Souilhol et al., 2016a) and in the adult bone marrow its functional role is restricted to the endothelial and perivascular niche (Ding et al., 2012).

1.5.2.2 Interleukin 3

IL3 deficient mice have been shown to have normal blood counts, CFU-Cs and repopulating HSCs (Lantz et al., 1998; Mach et al., 1998) suggesting that IL3 does not affect steady state haematopoiesis. However during development, IL3 deficient mice have reduced HSC numbers in the AGM region, yolk sac and placenta, compared to wild-type controls, as well as in explant culture (Robin et al., 2006). Exogenous addition of IL3 has been shown to significantly contribute to the 150-fold expansion of HSCs in E11.5 reaggregate culture (Taoudi et al., 2008), but only with a role in progression of Type II pre-HSCs to HSCs suggesting its role is specific to stage (Rybtsov et al., 2014).

1.5.2.3 Flt3-ligand

Flt3-ligand is a soluble homodimeric protein expressed in bone marrow cells and myeloid and lymphoid haematopoietic cell lines which engages with Fms-tyrosine kinase 3 (Flt3) (Gilliland and Griffin, 2002). Intraperitoneal and subcutaneous injections of the ligand have been shown to increase the number of CFU-Cs and phenotypic HSCs (estimated by lin-Sca1+Kit+ cells) in the peripheral blood, suggesting that it drives the mobilisation of HSCs from the bone marrow into the peripheral blood (Brasel et al., 1996). *Flt3l* knockout mice show a mild reduction in colony forming cells, dendritic cells and natural killer cells (McKenna et al., 2000) but not HSCs in adult (Sitnicka et al., 2002), or foetal liver (Buza-Vidas et al., 2009). However, in *ex vivo* reaggregate culture of E11.5, FLT3L has been shown to act synergistically with SCF and IL3 to promote HSC expansion. As *Flt3l* knockout mice have not been examined at this embryonic stage, it is not clear whether this is a required or redundant function in the regulation of HSC maturation.

1.5.2.4 Notch signalling pathway

Notch signalling is a common regulator of developmental cell fate decisions. The Notch family of proteins (Notch1, Notch2, Notch3, Notch4) serve an unusual

function as both cell surface receptors and intra-cellular transcriptional regulators, and thus provide an important mechanism by which cells can directly regulate the gene expression of their neighbours (Milner and Bigas, 1999). A failure of *Notch1-/-* ES cells to contribute to the adult haematopoietic system in chimeric mice suggests that Notch signalling is essential for regulation of HSCs during development (Hadland et al., 2004). As well as *Notch1*, its ligands *Jag1* and *Jag2*, and its target *Hes1* are ventrally expressed in the mouse E9.5 and E10.5 dorsal aorta, implicating them in the time and place of HSC maturation (Robert-Moreno et al., 2005, 2008). However inducible expression of DNMAML, a dominant negative regulator of Notch signalling, and conditional deletion of *Rbpj* (the transcription factor regulated by the intracellular notch domain) demonstrate that Notch signalling is dispensable for adult haematopoietic stem cell function (Maillard et al., 2008).

Recently, use of a destabilised GFP reporter of the Notch target, Hes1, demonstrated that Notch signalling is active in HSC precursors as early as E10.5, while its activity is decreased in pre-HSCII at E11.5 (Souilhol et al., 2016b). Moreover the use of Notch inhibitors in reaggregate culture show that the dependency on Notch signalling is reduced during this time window, therefore confirming that Notch activity is important for the early stages of HSC specification, but that its role is dispensable once mature HSCs have formed.

1.5.2.5 Hedgehog signalling pathway

Conditional deletion of the hedgehog signalling target smoothened (SMO) using the Mx1-Cre mouse line, has demonstrated that hedgehog signalling is dispensable to adult HSC function (Gao et al., 2009). In zebrafish hedgehog (Hh) mutants, or Hh inhibitor treated cells have defects in adult but not embryonic haematopoiesis. This is via interference with migration of endothelial progenitors and arterial gene expression (Gering and Patient, 2005). In the mouse the expression of Hh in the gut ventral to the AGM region, and the reduction in HSC production after Hh inhibition has again implicated it as a regulator of HSCs/HSC emergence (Peeters et al., 2009).

Although the role of Hh has been proposed as a mechanism by which HSC emergence is inhibited from the dorsal side of the dorsal aorta, reaggregate culture studies have shown that the AoD in fact has a positive effect on HSC maturation

from the AoV and Shh may be one of the key mediators of this effect (Souilhol et al., 2016a).

1.5.2.6 BMP signalling pathway

The role of BMP signalling in the regulation of HSCs in the adult and in the developing embryo is subject to debate. A double knockout of *Smad1* and *Smad5* suggests that BMP signalling is dispensable in adult and foetal liver HSCs (Singbrant et al., 2010). On the other hand, in the adult bone marrow, block of BMP signalling can induce osteogenic niche cells, which correlates with an increase in HSC numbers (Zhang et al., 2003). Studies of perfusion of bone marrow with the BMP inhibitor Noggin, show high numbers of repopulated mice and suggest that the concomitant increase in *Cxcl12* expression may indicate that BMP inhibits homing to the bone marrow (Khurana et al., 2014). These results suggest that in the adult, BMP exerts a non-cell autonomous role of HSC regulation by potentially influencing the homing signals in the niche.

In the embryo, the role of BMP signalling is less clear. *Bmp4* expression is polarised to the ventral domain of the AGM region (Durand et al., 2007; Marshall et al., 2000; Wilkinson et al., 2009) where HSCs are preferentially identified. In zebrafish, it has been suggested that Bmp4 is responsible for the ventral polarisation of HSC emergence as a conditional ablation of the receptor reduced the number of *runx1+* cells (putative HSCs) (Wilkinson et al., 2009). High Bmp levels have also been shown to enhance survival of HSCs in *ex vivo* culture of placenta (Bhatia et al., 1999). From these observations it has been proposed that Bmp4 is positive regulator HSC specification during development.

In mouse, *Bmp4* has been shown to have higher expression in AGM stromal cell lines that support HSCs in culture relative to non-supportive lines (Durand et al., 2007). Subsequently, explant culture of E11 AGM in the presence of BMP4 led to an increase in repopulated mice from 58% to 77% of mice injected. Recently this result has been challenged with the finding that a reaggregate culture of E11.5 and E10.5 AGM generate fewer HSCs when BMP4 was present (Souilhol et al., 2016a). Reasons for this discrepancy may be due to the different culture conditions used. A culture system that supports expansion and maturation of HSCs from E11.5 AGM would expect to produce saturating levels of repopulation when mice are injected

with 1 embryo equivalent (ee) (Souilhol et al., 2016a). But in the first study (Durand et al., 2007) a relatively low level of repopulation was achieved in all conditions, suggesting that the culture system may be measuring HSC survival rather that maturation. The inhibitory effect of BMP4 on HSC maturation has been further supported by the rescue effect of HSC maturation in culture in the presence of the BMP inhibitor: Noggin (Souilhol et al., 2016a).

Further studies in mouse, have added confusion regarding the role of BMP in HSC development as, use of a reporter for BMP signalling has indicated that all repopulating HSCs in the E11.5 AGM region have been activated by BMP (Crisan et al., 2015). Although apparently contradictory to the inhibitory role of BMP signalling for HSC development (Souilhol et al., 2016a) these two findings could be reconciled by a more precise understanding of the timing of BMP signalling. (Crisan et al., 2015) use a mouse with a EGFP (enhanced green fluorescent protein) reporter fused to a BMP responsive element (BRE), a DNA sequence from the *Id1* promotor that is bound by phosphorylated-SMAD following engagement of the BMP receptor with a BMP ligand (Korchynskyi and Dijke, 2002; Monteiro et al., 2008). The details of this system are not stated in full detail, but a conventional EGFP has a half-life > 24h (Li et al., 1998). Consequently, BRE-GFP HSCs transplanted from E11 AGM region could have been activated by BMP earlier than E10 and retained the GFP marker, which could still fit the model that BMP signalling can inhibit HSC development later on at E11-E11.5.

Thus, the role of BMP in the regulation of HSCs remains unclear, but appears to be complex and context-dependent. To elucidate BMP's precise role, particularly during embryonic development, experimental interrogation must precisely define the timing at which is it required, and the processes it regulates (e.g. maturation or maintenance). Moreover, given the numerous ligands, receptors and modulators of the BMP signalling pathway, a broader characterisation of the pathway activation in the AGM region may help explain the mechanism of signalling.

1.5.2.7 Retinoic acid

Retinoic acid (RA), a small lipophilic ligand for nuclear RA receptors, is processed from vitamin A through the action of intracellular aldehyde

dehydrogenases (ALDHs). The activity of aldehydrogenases, and therefore capacity of these populations to synthesise retinoic acid, can be detected through a fluorescent metabolic product (aldefluor) that can be detected by FACS. HSCs in the bone marrow have been shown to be aldefluor positive (Aldh+) (Armstrong et al., 2004; Hess et al., 2004) suggesting that they the pathway is actively signalling in these cells. Moreover, RAR-γ has been show to maintain the HSC population in adult bone marrow (Purton et al., 2006; Walkley et al., 2007).

More recently, retinoic acid has recently been described as a regulator of HSC development from early precursors in the AGM region (Chanda et al., 2013). Both retinoic acid receptors, RAR-α and RAR-γ, were expressed in HSC precursors (AA4.1+/VEC+) of the E10.5 and E11.5 AGM region. All AGM HSCs were found in the Aldh+ population. Moreover deletion of retinal dehydrogenase 2 (*Raldh2*) abrogated repopulation from AGM HSCs while culture of E11.5 with RA signalling agonist increased the production of HSCs. Therefore retinoic acid signalling via RAR-α is an important regulator of HSC development.

1.5.2.8 Inflammatory signalling pathways

Inflammatory signalling is usually mediated by secreted cytokines which indicate cellular stress or damage or infection. In the context of a mature, developed haematopoietic system, these usually serve to activate or mobilise immune cells such as macrophages. Recently, adult HSCs have been shown to exit their relatively dormant state after induction the key inflammatory cytokines interferon alpha (IFN-α) and interferon gamma (IFN-γ) (Baldridge et al., 2010; Essers et al., 2009). Furthermore, studies in zebrafish (Sawamiphak et al., 2014) and murine models (Li et al., 2014) have implicated IFN-γ as a positive regulator of HSC development in the embryo. Zebrafish knockdowns (Sawamiphak et al., 2014) and mouse knockouts (Li et al., 2014) of IFN-γ and its cognate receptor, both showed a loss in HSPCs and HSCs respectively in the AGM region. Moreover, knockdowns of TNF receptors (most strikingly Tnf2) in zebrafish also resulted in loss of HSPCs in the aortic floor. In both (Sawamiphak et al., 2014) and (Espín-Palazón et al., 2014), these pro-inflammatory signals were proposed to converge on the Notch pathway to promote HSPC emergence from the endothelium. Whether this is the case in mouse remains

to be shown, particularly whether this model could reconciled with the transient requirement of Notch signalling for early pre-HSC maturation (previously discussed in section 1.5.2.4). IFN-α has also been suggested as a mechanism by which HSCs from the AGM region progress into a more mature state found in foetal liver (Kim et al., 2016). Therefore pro-inflammatory signalling appears to regulate HSC behaviour at all stages of their life span, although whether the same functions are affected in each stage remains unclear.

1.5.2.9 Wnt signalling pathway

Study of the role of Wnt/β-catenin in adult HSCs has produced some apparently contradictory results. Deficiency of *β-catenin* or *Wnt3a* is required for maintaining adult HSCs (Luis et al., 2009; Reya et al., 2003) while other studies have shown that in the absence *β/γ-catenin* genes, normal haematopoiesis was possible (Jeannet et al., 2008; Koch et al., 2008). On the other hand, over-expression or gain-of-function mutations of *β-catenin in vivo* appear to exhaust the pool of HSCs through excessive cycling (Kirstetter et al., 2006; Scheller et al., 2006). Thus, it appears a careful balance of Wnt signalling is required to maintain normal adult haematopoiesis.

During mouse embryonic development, Wnt signalling has been shown to be transiently required (Ruiz-Herguido et al., 2012) as explant culture of E10.5 AGM region with the Wnt inhibitor, PKF-115-584, resulted in a reduction of HSCs but culture of E11.5 AGM region showed no difference in HSC production when cultured with the inhibitor.

1.5.2.10 Others secreted regulators of HSC function

Additional secreted proteins such as thrombopoietin (TPO), Insulin-like growth factor 2 (IGF-2), fibroblast-growth factor 1 (FGF-1) and angiopoietin like 2 and 3 have been shown to expand HSC numbers *ex vivo* (Zhang et al., 2006), although their role during development hasn't been studied. Cxcl12 is essential for homing and engraftment to the bone marrow niche, as seen by failure of HSCs to engraft Cxcl12-/- mice (Ara et al., 2003). Non-murine models have also demonstrated the possible contribution of vascular endothelial growth factor (VEGF)

to HSC development. This ligand for FLK1 tyrosine kinase has been shown in zebrafish to be required for HSPC formation (Gering and Patient, 2005) but the role in mammals remains unclear.

Pathway	Role in adult HSC behaviour	Role in HSC formation during development	Dorso-ventral polarisation in the AGM region
SCF	cKit receptor expressed on all HSCs (Gekas et al., 2005; Morrison et al., 1995; Ogawa et al., 1991; Rybtsov et al., 2014; Sánchez et al., 1996)	Promotes maturation of pro-HSC to Type I pre-HSC (Rybtsov et al., 2014); promotes HSC survival in foetal liver (Bowie et al., 2007b)	SCF is ventrally polarised (Souilhol et al., 2016a)
Interleukin	IL3 deficient mice have normal steady state haematopoiesis (Lantz et al., 1998; Mach et al., 1998)	IL3 knockout mice have reduced HSC numbers in the AGM region, YS and placenta (Robin et al., 2006) ; contributes to the progression of Type I pre-HSCs to HSCs (Rybtsov et al., 2014)	Not known
Flt3	Subcutaneous injection drives HSC mobilisation into peripheral blood	Promotes HSC maturation in culture long with SCF and IL3 (Taoudi et al., 2008)	Not known
Notch	Dispensable in adult HSCs (Maillard et al., 2008)	Required for pre-HSC maturation but dispensable in mature HSCs (Hadland et al., 2004; Souilhol et al., 2016b)	Ventral expression of Jag1, Jag2, Notch1, Hes1 (Robert-Moreno et al., 2005, 2008)
Hedgehog	Dispensable in adult HSCs (Gao et al., 2009)	Mediates the enhancement of HSC maturation in AoV by AoD (Souilhol et al., 2016a)	Shh is dorsally polarised (Souilhol et al., 2016a)
BMP	Dispensable in adult HSCs (Singbrant et al., 2010); blocking BMP signalling may facilitate homing and engraftment (Khurana et al., 2014)	May promote HSC survival in explant culture (Durand et al., 2007); inhibits HSC maturation in reaggregate culture (Souilhol et al., 2016a)	BMP4 is ventrally polarised (Durand et al., 2007; Marshall et al., 2000; Wilkinson et al., 2009)
Retinoic acid	Bone marrow HSCs are Aldh+(Armstrong et al., 2004; Hess et al.,	Essential for embryonic HSC development (Chanda et al., 2013)	Homogenously distributed (Chanda et al., 2013)

	2004); RAR-γ regulates HSC maintenance (Purton et al., 2006; Walkley et al., 2007)		
Inflammatory signalling	Both positive (Baldridge et al., 2010) and negative (Snoeck et al., 1994) roles shown previously	Positive role of IFN-γ and TNF-α on HSC/HSPC development in the AGM region (Espín-Palazón et al., 2014; Li et al., 2014; Sawamiphak et al., 2014)	
Wnt signalling	Controversial, both positive and negative roles described	Required transiently for HSC maturation (Ruiz-Herguido et al., 2012)	Appears homogenous around aorta endothelium (Ruiz-Herguido et al., 2012)

1.5.3 Directed differentiation and *de novo* HSC formation *in vitro*

One of the fundamental goals of HSC biology is the ability to form HSCs *de novo in vitro* as a source of HSCs for molecular study, disease modelling, and potentially transplantation. Two main approaches aim to achieve this, both relying on lessons learned from molecular studies of the *in vivo* system. One approach is, analogous to reprogramming cells to a pluripotent state with "Yamanaka factors" (Takahashi and Yamanaka, 2006), to introduce the essential transcription factors expressed in HSCs to establish the transcriptional circuitry and therefore function of HSCs. The identification of transcription factors necessary for steady state haematopoiesis and development have been applied in a number of studies which were capable of producing haematopoietic cells from mature somatic cells, although not transplantable HSCs (Batta et al., 2014; Mitchell et al., 2014; Pereira et al., 2013; Sandler et al., 2014; Szabo et al., 2010). Refining the understanding of the initiators and maintainers of HSCs *in vivo* may help the production of transplantable HSCs in this manner.

Another approach for clinical research and potential therapeutic use would be to direct differentiation from ES cells or patient-derived induced pluripotent stem cells (iPSCs) using external regulatory cues. The advantages of this approach are that ES cells can be expanded indefinitely *in vitro* providing a rich (and genetically

tractable) supply; ES-like cells can be generated from any differentiated cell so can model a genetic abnormality of a patient, or could provide an autologous source of cells; and directed differentiation avoids the need for extensive genetic modification (e.g. retroviral transfection) before transplantation. The strongest efforts so far have produced ventral mesoderm by culturing embryoid bodies with BMP4 which increased the progenitor numbers and progenitor self-renewal (Chadwick et al., 2003). Others have cultured ES cells with bone marrow derived stromal cells (Kaufman et al., 2001; Vodyanik et al., 2005). However, these protocols have only produced cells capable of producing haematopoietic colonies *in vitro*. The failure of these protocols to produce transplantable cells has led to the hypothesis that the cells produced are more like primitive progenitors found in the yolk sac (Keller et al., 1993; Vo and Daley, 2015). Thus, an essential strategy to achieve differentiation of the definitive haematopoietic system will be to identify the key signalling which differs between primitive haematopoiesis and definitive haematopoiesis *in vivo*.

1.6 Sequencing technologies: an opportunity for large-scale characterisation of stem cell systems

Development of technology is closely associated with leaps in scientific insight. The field of genomics and transcriptomics is currently a key driver in the molecular characterisation of stem cell systems. The ability to quantify the expression of every gene in a cell has an enormous impact on the elucidation of the defining features of a cell type, the instructional changes that drive cells to differentiate, and the impact of environmental, or intracellular changes.

This section will briefly discuss how genome-wide technologies have been developed, and how recent applications of such technologies have provided novel insight into a number of aspects haematopoietic stem cell research.

1.6.1 The development of transcriptome expression profiling

Genomics and transcriptomics are built on the foundations of molecular biology: the structure of DNA (Watson and Crick, 1953), the nature of information encoding in genomes (Nirenberg and Matthaei, 1961), and the elucidation of this information transfer as a “central dogma” that DNA is transcribed as RNA and

translated into protein (Crick, 1970). Due to the inherent linearity of DNA and RNA, and their capacity for exponential amplification, the development of technologies to measure and quantify their information has greatly outstripped technologies that measure protein types. Consequently, measurement of gene expression is often used as a proxy for the characterisation of protein effectors of cell signalling.

These technologies have developed in two parallel, but highly interlinked streams: the physical development of sequencing machinery; and the methods of information processing and storage.

1.6.1.1 Transcriptomic technology development

Major developments in DNA sequencing technologies started with the invention of the Maxam Gilbert and Sanger sequencing methods (Maxam and Gilbert, 1977; Sanger et al., 1977). Due to its preferable radiation-free methodology, Sanger sequencing was widely adopted and, following developments in parallelisation and automation, eventually led to the landmark completion of the first human genome sequence (Consortium, 2004; Lander et al., 2001; Venter et al., 2001). This was quickly followed by the mouse genome (Chinwalla et al., 2002) and many other key model organisms.

Although elucidating the sequence of the full genome is important, to understand the dynamics of how this information is used in biological contexts, the expression products of the genome (i.e. RNA) must be identified and quantified. One of the earliest technologies that enabled this on a large-scale is the microarray. The principle of this is to fix a reference set of probes, representing known transcripts, to a chip and apply a sample of interest (Lockhart et al., 1996; Schena et al., 1995, 1996; Southern et al., 1992). Through hybridisation of probes with fluorescently labelled cDNA, the relative quantity of RNA species in the sample is inferred by the fluorescent signal on the chip. Following commercialisation of these microarrays, they have been widely applied in many fields (Hughes et al., 2001; Lipshutz et al., 1999; Morozova et al., 2009; Stoughton, 2005). However microarray technology faces a number of drawbacks: the dynamic range of the hybridisation technology does not appear to represent the dynamic range of gene expression, and there is an inherent bias in the detection of RNA species as the microarray probes are designed on prior information (Mortazavi et al., 2008; Wang et al., 2009). Although, the

development of tiling arrays – high density oligonucleotide arrays representing all sense and antisense strands of then genome – has reduced the prior assumptions about transcribed regions required, novel splice isoforms or non-mRNA molecules are more accurately identified by *de novo* identification i.e. sequencing (Agarwal et al., 2010).

Since the human genome project, there has been a major drive to reduce the costs and increase the throughput of sequencing technologies. This was enabled by the development of "next generation" or "massively parallel" sequencing methods which have been made available as commercial technologies such as Roche/454, Illumina Genome Analyser IIX, Applied Biosystems SOLiD, and more recently Illumina HiSeq, Pacific Biosciences RS platform (English et al., 2012) and Oxford Nanopore (Mikheyev and Tin, 2014; Stoddart et al., 2009). Many of these technologies achieve such high throughout sequencing through some common concepts. First, elimination of the cloning step during sample preparation and replacement with PCR based amplification of the fragmented sample (fixed to a surface, or in droplets of an emulsion) reduces time, cost, and species-specific biases. Secondly, using "sequencing-by-synthesis" enables massive parallelisation of the sequencing. Although different detection chemistries are used, the common principle is to label nucleotides (with fluorescence or chemiluminescence), so that as they are correctly incorporated to synthesised sequence they release a chemical signal (Margulies et al., 2005; Shendure et al., 2005). In the case of SOLiD parallel sequence detection is though rounds of labelled di-base primer ligation that are used to detect the sequence (Valouev et al., 2008). For PacBio single molecule real time sequencing uses zero-mode waveguides (English et al., 2012). Nanopore technologies measure a signal change (e.g. current) as a molecule translocates through a protein nanopore (Stoddart et al., 2009). In all cases, the number of fragments that can be sequenced in parallel is therefore in the order of millions as opposed to hundreds by capillary-based sanger sequencing (Bentley, 2006), so samples can be sequenced at a genome-wide or transcriptome-wide scale at reasonably low cost.

One remaining limitations of sequencing is the inability to sequence long reads accurately. In many contexts, where a reference genome is available, this can

be overcome computationally. However for some purposes, such as sequencing novel genomes or novel transcripts, this is important. A number of technologies such as Pacific Biosciences RS platform (English et al., 2012), and most prominently nanopore-based technologies, are currently in development aim to achieve this goal (Mikheyev and Tin, 2014; Stoddart et al., 2009).

Thus sequencing can be performed at a large scale to any DNA sample, whether from reverse-transcribed RNA (RNA-sequencing), whole genomes or DNA enriched by other methods. Many aspects of cell regulation can therefore be investigated from genome variations in populations (whole-genome sequencing), transcription dynamics (RNA-sequencing) to transcription factor binding and epigenetics (ChIP-seq, DNAse-seq, ATAC-seq, DamID-seq). Improvements in library preparation and sequencing sensitivity now facilitate applications of these techniques on a single-cell level, of which single-cell RNA-seq is perhaps currently the most prominent in the stem cell field (Grün and van Oudenaarden, 2015; Kolodziejczyk et al., 2015; Saliba et al., 2014). Indeed many more variations of sequencing technologies – largely based on modifications of the sample preparations – are applied to address different questions but for the scope of this thesis, RNA-sequencing will be the main focus.

1.6.1.2 Information processing for novel biological insight

The ability to produce data in the order of millions of sequence reads per sample presents a large computational challenge. Chiefly, how useful insight can be extracted, and how this can be made accessible to others in an interpretable format.

A principle challenge is the production of sequence information as short reads, which need to be assembled into a map of contiguous (for genome sequencing) or non-contiguous (for RNA-sequencing) sequences on a genome-wide scale. For RNA-sequencing data, two approaches are employed to reconstruct the transcriptome: mapping of transcribed sequence reads to coordinates on a reference genome and merging sequences whose alignment overlaps; or determining *de novo* whether reads should be assembled as contiguous transcripts.

Mapping reads to the genome and merging aligned reads that have significant overlap has traditionally been the most popular method for transcriptome reconstruction. This method has the advantage of being more sensitive, as genes

which are lowly expressed may not have sufficient coverage to be assembled *de novo*. One of the biggest computational challenges of RNA-seq read alignment is that reads may be non-contiguous as genes are broken up by introns. A number of computational tools have found fast, and memory efficient ways to model splice junctions and consequently reliably align non-contiguous reads such as Tophat, STAR and HISAT (Dobin et al., 2013; Kim et al., 2015; Trapnell et al., 2009). Subsequently, aligned reads with significant overlap are merged into putative transcripts and quantified. This requires the inference of which regions constitute a gene, and a way of coping with reads that show ambiguous alignment, or different isoforms. Some methods use prior information of annotated gene regions and simply quantify these such as HT-seq (Anders et al., 2015), while others model transcript isoforms independently of prior gene annotation such as Cufflinks (Trapnell et al., 2010).

Given the potential loss of information in mapping reads to a linear reference genome, which potentially doesn't represent the diversity of the transcriptome, *de novo* transcriptome assembly remains an important alternative method. Advantages of *de novo* assembly are apparent where there are significant differences between the reference and transcriptome, such as splice isoforms, gaps, missing genome information or transcriptome is significantly altered from genome e.g. in cancer studies. However, this method can suffer loss of accuracy when poor quality reads are present. The computational challenge of transcriptome alignment has been tackled by breaking down reads into k-mers (substrings of reads of length k) and constructing *de Bruijn* graphs whereby nodes represent k-mers and transcripts are defined by identifying a route through the edges which represent immediately overlapping k-mers by k-1 bases (Compeau et al., 2011; Grabherr et al., 2011). Matching reads to compatible transcripts creates a pseudo-alignment (differing from alignment in that the exact coordinates of the match are not retained), and quantified by expectation maximum procedures. This has recently been implemented as a fast and accurate tool called Kallisto (Bray et al., 2016).

After transcript reconstruction and quantification, insight is usually gained from an experiment by comparing the gene expression profiles and identifying the genes which show a significant change in association with a particular condition. To

distinguish between significant changes and those that are more likely a product of inherent variation in the measurements, an appropriate statistical model must be used. For RNA-seq data, this model should account for the differences in variance across the mean expression levels, the fact that RNA-seq data is discrete and that expression intensity measures are dependent on the levels of expression of other genes. An appropriate model for sequencing data is the Poisson model, which described the sampling of reads from a fixed pool of genes. A limitation of this model is the assumption that the mean and variance are equal, which is not the case for RNA-seq data, as the variation tends to be larger than the mean (termed over-dispersion) and also dependent on the mean (termed heteroscedasticity) (Marioni et al., 2008; Nagalakshmi et al., 2008; Rapaport et al., 2013). Tools such as EdgeR and DESeq therefore use the negative binomial distribution, which allows the incorporation of a dispersion factor:

$$v = \mu + \alpha\mu^2$$

for variance v, μ and dispersion α. Due to the generally low number of sample replication, these tools use borrowed information from other genes in the sample to estimate dispersion based on mean expression intensity (Anders and Huber, 2010; Robinson et al., 2010). On the other hand, limma voom, models expression intensity as a continuous variable and uses non-parametric regression to estimate the variance of expression intensity for incorporation into a linear model (Law et al., 2014). Currently, no one tool or model has been shown to hold a major advantage in sensitivity or accuracy in differential expression above all others, and DESeq, EdgeR and limma are all in commonly used for RNA-seq analysis (Rapaport et al., 2013; Seyednasrollah et al., 2015).

Finally, an essential advance in genomic and transcriptomic studies is the introduction of standard procedures to store and share data. Genome assemblies can be explored with genome browsers such as Ensembl (Yates et al., 2016), and UCSC (Kent et al., 2002). These are centralised repositories of whole-genome coordinates onto which individual annotations and experimental information can be mapped, providing a wealth of information accessible to any researcher. This information is

being continually updated and, most prominently, a highly significant update on the function of 80% of the genome was released by the Encyclopaedia of DNA Elements (ENCODE) project (The ENCODE Project Consortium, 2012) .

As the information produced from a sequencing experiment greatly exceeds the space allocated for publications, sequencing data is deposited in standardised repositories such as ArrayExpress and NCBI's Gene Expression Omnibus to support reproducible research (Barrett et al., 2013; Kolesnikov et al., 2015). Although the depth of information gained from each sequencing experiment may not be immediately utilised, it is becoming increasingly apparent that such data can yield further insight for researchers in the future where it may be put in the context of other data. As such, a number of initiatives aim to bring together sequencing data from many experiments to enable meta-analysis (Halbritter et al., 2012, 2014; Ruau et al., 2013). As well as accessing raw sequencing data, the final lists of genes from a bioinformatics analysis or experiment can enhance functional annotation of future studies if collated. This is facilitated by a number of databases such as the Gene Ontology database (Ashburner et al., 2000; Consortium, 2015), Kyoto Encyclopaedia of Genes and Genomes (Kanehisa et al., 2016), Biocarta (Nishimura, 2001), Protein Interaction Database (Schaefer et al., 2009) and Molecular Signatures Database (Subramanian et al., 2005).

Thus the last 10-15 years have produced an extraordinary leap in the capacity to sequence and, importantly, share insight from sequencing experiments. Although some challenges remain in the standardisation of experiments and information processing, this is an extremely powerful resource to the scientific community and is developing at such a speed that many aspects of sequencing and analysis have changed significantly during the course of this thesis.

1.6.2 The application of sequencing technologies in the study of HSCs and HSC ontogeny

The ability to identify the full cell transcriptome has many potential applications in the study of HSC ontogeny during development. A number of studies have applied transcriptional profiling methods to yield important insights, although

in many cases experimental systems still pose challenges that have not been fully met by current technologies.

1.6.2.1 Transcriptional profiling of the cell-autonomous regulators of HSCs

One on the first applications of transcriptional profiling was the use of microarrays to identify novel unique markers of HSCs (Kiel et al., 2005). Through the comparison of expression profiles of a functionally enriched HSC population and a multipotent progenitor population a number of surface receptors of the SLAM family, CD150, CD244, and CD48, were found to be differentially expressed and could be used (CD150+CD48−Sca-1+Lineage−c-kit+) to increase the enrichment of the repopulating HSCs from 20% to 47% of the population. More recently RNA-sequencing of single cells which were functionally enriched using different surface marker combinations was used to establish an optimal combination of markers $EPCR^{hi}SLAMSca^{hi}$, capable of enriching HSCs to approximately 67% (Wilson et al., 2015). Expression profiling is therefore a powerful tool in exploration of surface marker enrichment, and indeed single-cell profiling enables the heterogeneity of a population to be examined. The goal of 100% purity may now only be limited by the technical challenges of transplantation assays.

Recently the transition of HSC precursors into HSCs in the AGM region has been captured by several studies (Kartalaei et al., 2015; Li et al., 2014; Zhou et al., 2016). The first two of these compare the transcriptional profiles of Ly6a enrichment HSC/progenitor populations with endothelial (EC), or haematogenic endothelial cells (HEC) (Kartalaei et al., 2015; Li et al., 2014). Li and colleagues show that inflammatory signalling is enriched in the CD31+VEC+ESAM+Kit+Ly6aGFP+ population and, unexpectedly, Ifngr1-/- mice had a four-fold reduction in HSCs in the AGM region, umbilical cord and vitelline arteries, suggesting that interferon gamma signalling is required during development although in the adult affects only proliferation not absolute numbers of HSCs (Baldridge et al., 2010). Using a similar strategy to enrich EC, HEC and HSC populations with Ly6a-GFP reporter mice, Kartalaei and colleagues found that the core HSPC regulatory transcription factors (Wilson et al., 2010) were largely expressed in all these populations, with increasing levels in HSCs and HEC compared to EC. This data was used to identify a novel surface marker Gpr56, which is expressed in haematopoietic clusters, and when

knocked down in zebrafish reduced the number of *cmyb* positive cells (putative HSCs).

A third study compared endothelial cells (EC), Type I pre-HSCs, Type II pre-HSCs and HSCs at E12 and E14 by single cell RNA-seq (Zhou et al., 2016). More genes were expressed in pre-HSCs compared to ECs, but as pre-HSCs mature to HSCs the number of expressed genes decreases again. Most core HSPC regulatory transcription factors (Wilson et al., 2010) expressed in pre-HSCs and HSCs and a number of key transcription factors (notably Runx1 and Gfi1) were expressed in pre-HSCs and HSCs but not ECs, highlighting their key role in the HSC lineage. This could implicate them (as has already been discussed) as the necessary elements for trans-differentiation of progenitors to HSC phenotype. What remained unclear was the difference in transcriptional circuitry between pre-HSCs and HSCs i.e. what defines the stable HSC state compared to the precursor state.

The above studies show that regulation of the cell-autonomous signalling during maturation of precursors into HSCs can now be studied at a transcriptome-wide scale. However this still faces some technical limitations. Expression profiling of a small population or single cells requires extensive amplification and the subsequent noise in expression data can be difficult to distinguish from real biological variation (Kolodziejczyk et al., 2015; Stegle et al., 2015). Moreover, the limited ability to enriched 100% pure HSC populations to study introduces further limitations, which is likely to only be solved by studies at a single-cell level.

1.6.2.2 Modelling the transcriptional circuitry of HSCs

The lack of unique markers and assays to identify a pure HSC population have meant that expression profiling to identify the essential transcription factor networks determining HSC states have lagged behind those of other stem cell systems, for example embryonic stem cells where pure populations can be isolated and the core transcription factor network is known (Chen et al., 2008; Dunn et al., 2014; Festuccia et al., 2012; Hall et al., 2009; Li et al., 2005; Martello et al., 2012, 2013; Tai and Ying, 2013). Given the wide numbers of sequences that can generally be bound by individual transcription factors, stable cell states are believed to be largely the product of the combinatorial action of several transcription factors which

enables more specific target recognition. As well as regulating transcription of a broad range of effector genes, these regulate each other to maintain the system balance, but also respond to perturbations or cues in which the cell may need to change state. In HSPCs, the core transcription factor circuit is believed to be composed of a combination of transcription factors that, when deleted, have haematopoietic phenotypes (discussed in section 1.5.1). But more than just describe the transcription factors, efforts are being made to model the co-regulatory relationships to enable predicative modelling of cell behaviour.

Comparison of global binding of a number of known HSPC regulators by chromatin immunoprecipitation (ChIP-seq) has revealed pairwise co-occupancy, close proximity of binding, and combinatorial clustering of seven transcription factors (SCL, LYL1, LMO2, GATA2, RUNX1, FLI-1, and ERG) suggesting they represent a core transcriptional network classed as the "heptad transcription factors" (Wilson et al., 2010). Following on from this, mutagenesis and luciferase reporter approaches have been used to establish the nature of regulatory interactions between transcription factors (Schütte et al., 2016). Based on the expression levels and nature of interaction, the authors constructed a dynamic Bayesian network representing a Markov process whereby the probability of future states depends on the current state. With this model they demonstrate the steady state expression level of nine transcription factors is maintained over time. To validate the model, independently of the information used to build it, a comparison with known experimental phenotypes was made. *Ly1* and *Tal1* were computationally knocked down individually, as they are known individually not to effect adult haematopoiesis (Capron et al., 2006; Mikkola et al., 2003), and no effect on expression levels was seen in the rest of the network. However combinatorial knockdown, which causes loss of haematopoietic progenitors in adult (Souroullas et al., 2009), reduces transcription levels of *Runx1* and *Gata2* suggesting the model can at least partially capture functional effects.

In silico modelling is a highly promising avenue to enable prediction of HSC and other haematopoietic cell fate decisions. However a few remaining impediments exist. Defining transcription factor binding patterns on a genome wide scale through ChIP-seq currently requires large numbers of cells for input (approximately 1 million), which is not possible to attain from a HSC population. Moreover,

functional validation of the nature of interactions is difficult at high throughput in primary cells that cannot be cultured *in vitro*. Such studies will require developments in technology that enable genome-wide transcription factor binding to be assessed in smaller numbers of cells (which may soon be possible with DamID-seq or ATAC-seq technologies), and improvements in our ability to genetically manipulate primary cells.

1.6.2.3 Transcriptional profiling of the niche for developing HSCs

As well as profiling the cell-autonomous signalling in HSCs and the transition between precursors and HSCs, transcriptional profiling has enormous capacity to elucidate the niche signalling environment in which HSCs mature.

A few studies have examined the niche for emerging HSCs through transcriptional analysis with microarrays. The availability of stromal cells lines which have been shown to preferentially support HSCs in culture provides a tractable system for molecular profiling (Charbord et al., 2014; Durand et al., 2007). Through comparison of differentially expressed genes in supportive and non-supportive cell lines from multiple sources (including the AGM region, and foetal liver) a set of genes which potentially represented the supportive phenotype was identified and could be used as a predictor of independent supportive or non-supportive cell lines (Charbord et al., 2014). Validation of their functional effect in zebrafish identified Tgfbi, snai2, pax9, and ccdc80 morphants all had haematopoietic defects 36 hours post-fertilisation when HSPCs emerge, while the vasculature largely remained normal. However, although a number of these cell lines were derived from the AGM region, the lack of clear support for HSC maturation (rather than HSC maintenance) by these lines may limit the utility of this study in elucidating regulators of HSC emergence.

Another profiling study of the developing HSC niche focused on the *in vivo* AGM environment for developing HSCs (Mascarenhas et al., 2009). In this study, comparison of E9.5 and E11.5 embryos, as well as the spatial restriction toward the centre of the dorsal aorta compared to the rostral and caudal aspects was the basis of the separation of HSC supportive and non-supportive niches. Transcriptional profiling by microarray identified the cyclin-dependent kinase inhibitor p57Kip2, which is known to be expressed in adult HSCs, and whose expression was detected

in the E11 ventral mesenchyme. *p57Kip2+/−m* mice were shown to retain HSCs in the AGM region, suggesting it regulates HSC migration to the foetal liver, or proliferation in the AGM region. In this analysis Gata3 was also identified in the niche at the time and place HSCs are first detected. A follow up study identified an important role for Gata3 in regulating HSC maturation via the production of catecholamines (Fitch et al., 2012). Therefore this study has provided important insight into the regulators of HSC maturation. However, since this study was performed the differential characteristics of the AGM region along the dorso-ventral axis (Souilhol et al., 2016a), as well as the transition between be E9.5 and E10.5 to produce an autonomously supportive environment (Rybtsov et al., 2014) are better understood. Furthermore, the use of microarray may have limited the depth of differential gene expression changes that can be detected which, in such a heterogeneous tissue, may mask some important signalling changes. Thus, as elaborated in section 1.7, a comparison of the functional compartments of the AGM region remains to be explored in high depth.

1.7 Project aims and thesis structure

In summary, the last century of research has revealed that haematopoietic stem cells are detectable and quantifiable from the midgestation embryo to the adult haematopoietic organs (sections 1.1 and 1.2). The potency of these cells renders them an important therapeutic tool for haematopoietic malignancies and disorders (section 1.1.4). However, cultivating HSCs outside of the highly regulated *in vivo* environment remains extremely challenging, and is one of the most prohibitive aspects of research and clinical use. Study of the ontogeny of HSCs can tell us many things: how multipotent characteristics are constructed; how the requirements of the adult haematopoietic system are met by *de novo* specification and expansion of the HSC pool; and, importantly, how HSCs may be produced *in vitro* for experimental and therapeutic purposes. The embryonic origin of HSCs has therefore been an important area of research in the last 50 years and whilst the lineage of precursor cells that produce HSCs is considerably defined (section 1.2), the minimal molecular regulation that directs their fate remains elusive.

The objective of this thesis was to further our understanding of the molecular regulation of HSC ontogeny by investigating the signalling cues produced by the earliest embryonic niche: the AGM region. Recent insight into the spatio-temporal demarcation of the supportive AGM region subdomains (section 1.4), along with the accessibility of RNA-sequencing methods (section 1.6) presented the opportunity to characterise the full transcriptional changes underlying the AGM region's functionality at high depth. The additional support for HSC maturation provided by the OP9 cell line (section 1.3.3) gave a further basis for transcriptional profiling. Furthermore the *ex vivo* culture system which has enabled the AGM region's functionality to be determined, provided a tractable platform in which to functionally probe a number of molecules for their support of HSC development. Employing these approaches, I have conducted an investigation of the molecular mechanisms of the murine niche for developing HSCs, which I present here as three main bodies of work:

In **Chapter 3** I discuss the approach to transcriptional profiling of the AGM region and OP9 cells. I identify the transcriptional changes associated with the *in vivo* environment that supports HSC maturation, and the *in vitro* (OP9) environment. I identify a number of signalling pathways which appear to act together to effect these conditions. I compare these independent signatures and show that there is a degree of overlap in their supportive mechanisms. Finally, I discuss the development of a visualisation that permits access to this data resource.

Chapter 4 is a follow up from the transcriptional profiling where I describe the selection of candidates to test in a functional screen for novel effectors of HSC maturation from E9.5 precursors. I discuss the output of this screen in terms of production of long-term repopulating HSCs, as well as colony forming cells. I show that a gene identified from the RNA-sequencing data, *Bmper*, encodes a protein capable of increasing the efficiency of HSC maturation in this E9.5 embryo reaggregate culture, as well as maturation of precursors from E11.5 AoD.

In **Chapter 5,** I investigate the role of Bmper in the AGM environment in more detail. Using immunostaining to explore the BMPER protein distribution in the AGM region, I show that the protein appears to be present at particularly high levels in intra-aortic clusters and primordial germ cells. I further compare the expression of *Bmper* in different cell types within the AGM region and show that a non-endothelial CD146+ population encompassing perivascular cells, mesonephric tubules and metanephric mesenchyme express the transcript at highly enriched levels. By comparing the distribution of Bmper with nuclear pSMAD1/5/8 in the AGM region, I propose that Bmper functions as a BMP signalling inhibitor in the intra-aortic cluster environment where pre-HSCs most likely develop. Finally, I show that *Bmper* expression can be induced in the presence of BMP4, and draw the model that Bmper mediates a feedback-response mechanism to BMP signalling in the AGM region to precisely modulate the level of signal pre-HSCs experience as they mature into HSCs.

Chapter 2 Materials and Methods

2.1 General solutions

Dissection buffer:

Dulbecco's phosphate buffered saline (PBS) solution (with Mg2+ and Ca2+ ions; Sigma) containing 7% foetal calf serum (FCS) (Gibco, PAA, or Biosera) and 50 units/ml penicillin and streptomycin (P/S; Gibco).

Flow Cytometry buffer solution (FACS buffer):

Dulbecco's PBS (without Mg2+ and Ca2+) containing 7% FCS and 50 units/ml P/S.

OP9 culture medium:

Iscove's modified Dulbecco's medium (IMDM) (Invitrogen), 20% FCS, L-glutamine (4 mM), penicillin/streptomycin (50 U/ml), and 0.1mM β-mercaptoethanol.

Reaggregate culture medium:

IMDM (Invitrogen), 20% preselected heat-inactivated FCS (PAA/HyClone), L-glutamine (4 mM), penicillin/streptomycin (50 U/ml), 0.1mM β-mercaptoethanol, 100 ng/ml SCF, 100 ng/ml IL3 and 100 ng/ml Flt3l (termed 3GF), unless otherwise indicated (all purchased from Peprotech). In some cases, indicated in the text, 3GF or serum or both were not added to the culture medium.

2.2 Animals

Embryos were obtained from mating C57BL/6 (CD45.2/2) mice. Day 0.5 is determined on the morning of discovery of a vaginal plug. Embryo stage was determined more accurately by the number of somite pairs where E9.5 is attributed to embryos with 25–29 sp (forelimbs present), E10.5 to embryos with 35–39 sp (hindlimb and tail bud present but no eye pigmentation), E11.5 to embryos with 41–45 sp (anterior footplate and incomplete eye pigmentation). All experiments with animals were performed under a Project License granted by the Home Office (UK), University of Edinburgh Ethical Review Committee, and conducted in accordance

with local guidelines. All animals were housed within the University of Edinburgh following to the regulations of the Animals Scientific Procedures Act, UK, 1986. This involved provision of a constant supply of water and chow food, and a stable environment with a cycle of 14 hours light/10 hours. Litters remained with parents for three weeks postnatally then were weaned by separating parents and offspring. Mice used for matings were older than 6 weeks. In this project, C57BL/6 mice (Jackson Laboratories) were used.

2.3 Embryonic tissue isolation and preparation

Embryos were removed from the uterus following a schedule 1 method of cervical dislocation on a pregnant female. Embryos were then separated from extra-embryonic tissues including the yolk sac (YS) and amniotic sac. Embryonic tissues were dissected under a low power microscope in dissection buffer. To obtain the caudal part of E9.5 embryos, the anterior body above the head was removed. To obtain the AGM region in E10.5 and E11.5 embryos, first the head and anterior region above the heart was removed. Subsequently the neural tube, ventral tissue (including heart, liver and gut) and ribs were removed. Further subdissection of the AGM region included, removal of the urogenital ridges and separation of the dorsal aorta into ventral (AoV) and dorsal (AoD) parts.

To dissociate tissues into single cell suspensions, tissues were transferred to 5 ml polystyrene tubes (BD Flacon) containing 900µl of dissection buffer with 100 µl of 10mg/ml collagenase dispase (Roche). The tissue and collagenase were incubated at 37°C for 40 min in a shaking water bath. The dissociation reaction was quenched with the addition of 4 ml FACS buffer followed by centrifugation at 350 x g for 5min at 4˚C and resuspension in FACS buffer. The volume of tissue cells was quantified as embryo equivalents (ee).

2.4 OP9 cell culture

2.4.1 Normal culture

Cell culture procedures were in class 2 laminar flow hoods (Nuaire and Bassaire) using aseptic technique. Incubation was at 37˚C in humidified incubators with 5% CO_2 (Sanyo). All media was stored for no longer than two weeks.

Additional reagents were prepared by the tissue culture facility. All solutions were stored at 4˚C but brought to reach room temperature prior to use.

OP9 cells were maintained in IMDM (Invitrogen), 20% foetal calf serum supplemented with L-glutamine (4 mM), penicillin/streptomycin (50 U/ml) on polystyrene culture dishes. Cells were passaged every 4-5 days, when confluent, using 4X trypsin to remove cells from plastic and dissociate cells, and then replated at a concentration between 1:5 and 1:10 of the confluent population.

2.4.2 Thawing and freezing cells

Vials of cells stored at -80 °C were quickly thawed in a 37 °C water bath. The cell suspension was transferred to 10 ml pre-warmed (37 °C) medium in a universal tube. The suspension was gently mixed then centrifuged for 5 min (350 x g). Cells were resuspended in media, transferred into the appropriate flask, incubated overnight and the next day, the medium was changed.

For freezing, cells were centrifuged at 350 x g for 5 min to form a pellet. Cells were resuspended in OP9 culture media, and 450 µl transferred to a labelled cryovial (Nunc). 50 µl of Dimethylsulfoxide (DMSO) (Sigma) (10%) was added drop-wise, then the vial was immediately placed on dry ice and stored at -80 °C for up to 6 months, or in liquid nitrogen for long-term storage.

2.4.3 Reaggregate culture

For reaggregate culture of OP9 cells, a single cell suspension was generated by adding trypsin and then centrifuged at 430 x g for 12 min in 200-µl pipette tips sealed with parafilm to form a pellet. Reaggregated cells were cultured at the liquid–gas interface on pre-soaked 0.8-µm nitrocellulose filters (Millipore) at 37 °C in 5% CO_2, for 48 h. Reaggregates were harvested and cells dissociated with collagenase/dispase.

2.4.4 Transfection of OP9 cells with a doxycycline inducible *Bmp4* overexpression plasmid

Bmp4 cDNA was cloned into a doxycycline inducible bicistronic expression vector pPBhCMV1-cHA-IRESVenuspA (gift from H. Niwa) by Dr David Hills. In this construct both Bmp4 and Venus were expressed upon induction with doxycycline. 100,000 OP9 cells were transfected with this construct by

electroporation using NEON transfecting system (Invitrogen). 24 h after electroporation, cells were cultured in the presence of 1µg/ml doxycycline (Clontech) and the Venus positive population was sorted. After a week, in the absence of doxycycline induction, the Venus negative population was sorted and maintained, thereby excluding cells with leaky expression. For induction of Bmp4 followed by RNA extraction, cells were cultured in OP9 media with 1µg/ml doxycycline for 24h or 48h (Figure 8-1).

2.5 RNA extraction for RNA-sequencing

Cell suspensions were centrifuged to form a pellet of cells, to which lysis buffer (RLT) containing β-mercaptoethanol (1:100) was added. The tissue was homogenised with a 20-guage (0.9 mm) needle. The remaining extraction procedure followed the protocol of Qiagen RNeasy minikit (QIAGEN) including DNase I treatment (QIAGEN). RNA was initially quantified by spectrophotometry with the NanoDrop (Thermo Fisher). More accurate assessment of RNA quality was measured with Agilent 2100 Bioanalyzer (Agilent) to ensure all samples had an RNA integrity number (RIN) > 9 (assisted by Eliane Salvo-Chirnside, University of Edinburgh Systems Biology Unit). The RNA integrity number is an estimation of the level of RNA degradation based on the ratio of 28S ribosomal RNA and 18S in the sample (Schroeder et al., 2006).

2.6 RNA-sequencing and analysis

2.6.1 Library preparation for RNA-sequencing

RNA sequencing libraries (unstranded) were prepared from total RNA at The Edinburgh Genomics facility in the University of Edinburgh, using a TruSeq RNA Library Preparation Kit (Illumina) which purifies mRNA by polyA enrichment. Samples were multiplexed so that they could be pooled and sequenced across several sequencing lanes.

2.6.2 RNA-sequencing

From these libraries 50 base single-end sequence reads were generated with Illumina HiSeq 2000/2500 (Illumina) across five sequencing lanes, yielding 50

million reads per sample. For sequencing OP9 cells, the same platform was used but all 10 samples were multiplexed and sequenced in one lane, yielding 12 million reads per sample. Sequencing depth was informed by a comparison of differential expression changes through *in silico* modelling of depth (detailed in section 3.2.4), and estimation of the impact of tissue heterogeneity. Sequencing was carried out in The Edinburgh Genomics facility in the University of Edinburgh.

2.6.3 Sequencing read alignment

Sequencing reads were de-multiplexed at The Edinburgh Genomics facility and supplied as FASTQ files. FASTQ files are a standard text based format including the nucleotide sequence of the read, and the measure of data quality for each nucleotide in the read. The quality score is a phred-like logarithmic score between 0 and 50, representing the probability that the nucleotide readout is an error (Ewing and Green, 1998). Reads were then processed using the RNA-seq analysis pipelines of GeneProf (Halbritter et al., 2012). Quality of reads was assessed with the Sequence Data Statistics module of GeneProf which enables visual inspection of library sizes, average read length, nucleotide composition per dataset, nucleotide position per cycle, average quality score per cycle, cumulative quality scores, average quality score and abundance of identical sequences. Sequencing reads were aligned to mouse genome NCBI37/mm9 with TopHat (2012-11-14) v1.2.0 (Trapnell et al., 2009) and Bowtie (2012-04-10) v0.12.3 (Langmead et al., 2009) using the default parameters including: a seed length of 28 bp; a maximum of 2 mismatches permitted per alignment; a maximum of 10 permitted alignments per read; a required intron length between 70 and 500,000 bp. The mRNA levels per gene were quantified using the reference genome annotation from the Ensembl database (Ensembl 58 Mouse Genes, NCBIM37). Briefly, all uniquely mapped reads overlapping with each annotated gene were summed. These preliminary counts were then used to assign ambiguously mapped reads (those with multiple matches in the genome) to their likely origin. To this end, a fraction of each read was assigned to each possible mapped gene where the fraction was proportional to the relative number of uniquely mapped genes in each possible gene assignment, according to the following equation:

$$count(g) = \sum_{r \in reads(g)} \frac{w(r)}{|align(r)|} \sum_{r \in reads(g)} \frac{w(r)}{\sum_{\hat{g} \in align(r)} \sum_{\hat{r} \in reads(\hat{g})} \frac{w(\hat{r})}{|align(\hat{r})|}}$$

where *count(g)* is the expression count for an arbitrary gene g, *reads(g)* are all reads aligning to gene g, *w(r)* is the weight of read r (usually 1:0) and *align(r)* are all possible alignments of read *r* (Halbritter, F., 2012). Alignment percentages were assessed with the Read Alignment Statistics module. Read counts were scaled as reads per million (RPM) according to:

$$rpm(g) = \frac{count(g) \times 10^6}{R}$$

where $rpm(g)$ represents the scale read counts for a gene locus, $count(g)$ represents the raw read counts for a particular gene locus and R represents the total number of aligned genes in a library (Halbritter, F., 2012).

For the purposes of this thesis, genes would be selected ultimately for a relatively low throughput screen of secreted factors in the AGM region. Given the relatively low throughput of the screen, steps were taken to reduce the chance that selected genes may be variable due to technical noise. Therefore, for downstream analysis, genes were selected here if their expression intensity values were greater that 0.5 RPM in at least one sample. Based on the previously reported distinction between lowly expressed genes (possible noise), and highly expressed genes (likely real expression), based on the expression distribution falling into two peaks (Hebenstreit et al., 2011; Figure 2-1), a threshold of RPM 0.5 was judged to be a reasonable balance between genes which may be indistinguishable from technical noise, and genes which may be lowly expressed.

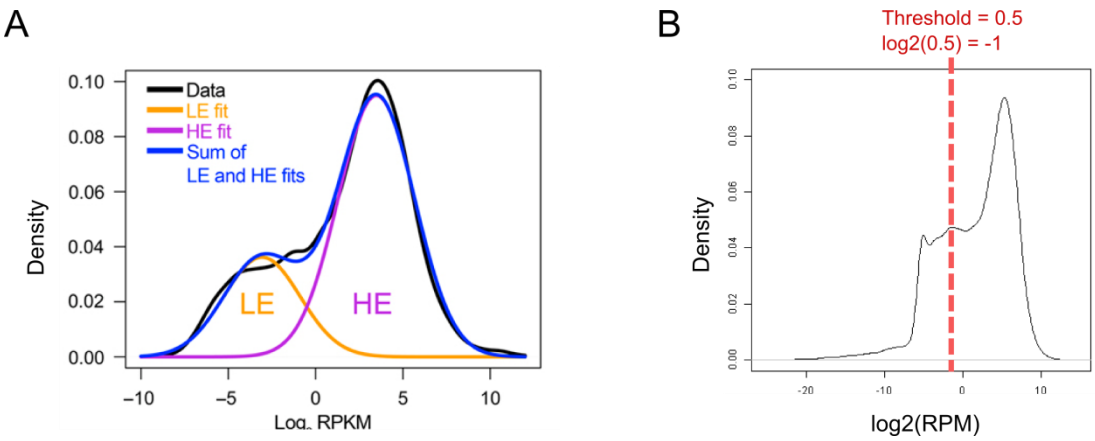


Figure 2-1 Selecting a threshold for gene expression

Plots show the kernel density estimates of expression level distribution from RNA-seq data. (A) Adapted from (Hebenstreit et al., 2011**)** shows the subdivision of the expression profile into lowly expression (LE) genes, which are probable noise; and highly expressed genes (HE), proposed to represent real expression of the sample. (B) Shows a representative profile from the AGM RNA-seq dataset produced in this thesis and the 0.5 RPM threshold, which aims to capture the main highly expressed genes.

2.6.4 R statistical programming environment

Unless stated otherwise, all downstream analysis of normalised read counts used the R statistical programming language and environment (R version 3.2.2 (2015-08-14), R Core Team, 2015) which includes a large number of the statistical techniques described below. Where the environment was extended through the use of external packages, they are cited in the text.

2.6.5 Sample clustering

2.6.5.1 Correlation heat maps

Comparison of sample expression profiles used all genes where RPM normalised read count values were greater that 0.5 RPM in at least one sample. Correlation was calculated between pairwise combinations of all samples with Spearman's rank correlation (due to its reduced sensitivity to outliers), where missing observations generate NA values ('cor' function from R stats).

2.6.5.2 Principle component analysis

Principle components of sample expression profiles (for genes with expression intensity greater that 0.5 RPM in at least one sample) were calculated ('prcomp' function from R stats). For principle component analysis, variables were zero-centred and scaled to unit variance. The proportion of total variance represented by each component was calculated as the square of the returned standard deviation values as a proportion of the sum of variances for all components. Samples were plotted on the basis of the first two principle components, which represent the most variation in the data and separate samples by categorical traits. For the AGM RNA-seq data, the association of each component with the known categorical trait of the samples was tested by one-way ANOVA. The data categorical traits were divided into "stage": whether the samples were from E9.5 or E10.5 embryos; or "position": whether samples were from the dorsal domain, or ventral domain (UGRs were included in ventral domain).

The methods for plotting principle components and testing association with categorical variables were implemented with assistance of code from Dr Florian Halbritter and Dr Jonathan Manning (Stem Cell Bioinformatics group). The script in the R programming language for plotting principle component analysis modified from a script from Dr Florian Halbritter (Halbritter et al., 2012):

```
#R code specifying pca plot function

pcaplot2 <- function (pca, grps, grp_names=levels(as.factor(grps)),
                      cex = 0.8, mar = c(5,5,1,1), ...) {

   #set plot parameters
   old_par <- par(no.readonly = TRUE);
   on.exit(par(old_par));
   par(cex = cex, mar = c(5, 3, 4, 1));

   #specifies layout of graphs
   layout(matrix(c(1,1,2,2,3), 2, 5, byrow = TRUE));

   #calculate proportion of total variance represented by each
   #principle component
   totalvariance <- sum(pca$sdev ^ 2);
   perc_of_var <- (pca$sdev ^ 2) / totalvariance;

   #plot the simple and cumulatve variance represented by each
   #principle component
   plot(1:length(pca$sdev),perc_of_var, type='b',lty=1, ylim=c(0,1),
        xlab='Number of Principal Components',
```

```
        ylab='Percent of Total Variance');
   cumulative_var <- rep(0,length(perc_of_var));
   for(i in 1:length(perc_of_var)) {
     for(j in 1:i) {
       cumulative_var[i] <- cumulative_var[i] + perc_of_var[j];
     };
   };

lines(1:length(pca$sdev),cumulative_var,col='red',type='b',lty=1);
   legend('topleft', c('simple','cumulative'), cex=0.8,lty=1,
            col=c('black','red'),bty='n');

   #specify sample colours
   grcols <- (c('goldenrod','goldenrod','goldenrod',
                'deepskyblue4','deepskyblue4','deepskyblue4',
                'hotpink4','hotpink4','hotpink4',
              'darkolivegreen4','darkolivegreen4','darkolivegreen4',
                'red3','red3','red3'))
   names(grcols) <- grp_names;

   #matrix specifying which principle components to plot
   pcs <- matrix(c(1,2),ncol=1);
   for(i in 1:ncol(pcs)) { xx <- pca$x[, pcs[1,i]];
                           yy <- pca$x[, pcs[2,i]];
                           rrange <- max(xx)-min(xx);
                           xxlim <- c(min(xx)-rrange*0.1,
                                      max(xx)+rrange*0.1);
                           rrange <- max(yy)-min(yy);
                           yylim <- c(min(yy)-rrange*0.1,
                                      max(yy)+rrange*0.1);
                           plot(xx,
                                yy,
                                type = 'p',
                                pch = 21,
                                xlab = paste('PC',pcs[1,i],sep=''),
                                ylab = paste('PC',pcs[2,i],sep=''),
                                xlim=xxlim,
                                ylim=yylim,
                                col="white",
                                bg=grcols[grps],
                                cex=3, ...);
                           abline(h = 0, col = 'gray', lty = 2);
                           abline(v = 0, col = 'gray', lty = 2); };

#initiate plot
   plot.new(); legend('right',grp_names,pch=16,col=grcols,bty='n');
}


 #calculate principle components and execute pcaplot function
 #where expression_matrix is a matrix of the normalised expression
 #values with column names representing the labels to be shown in
the plot
   pcaplot(
        prcomp(
            as.matrix(t(expression_matrix)),
            scale=T,
            center=T
            ),
```

```
        colnames(expression_matrix)
        )


 #calculate principle coponents and execute pcaplot function
  pcaplot(
            prcomp(
                    as.matrix(
                                t(DESeq.allFCRPM1_0.3)),
                                scale=T,
                                center=T
                                )
                    ,colnames(DESeq.allFCRPM1_0.3)
            )
```

The Script in the R programming language to calculate the association of each principle component with categorical traits of the samples: embryonic stage and position (i.e. AGM subdomain). This code was generated by Jonathan Manning and modified by me to run the AGM region RNA-seq data:

```
#R code to calculate association of each component with categorical
#traits by ANOVA

#calculates the percentage of variance represented by each principle
#component
 fraction_explained <- round(
                            (pca$sdev[1:15])^2 /
                            sum(pca$sdev[1:15]^2), 3
                            )*100

#defines the catgorical traits of samples in keeping with the column
#order
 experiment<-data.frame(
                    Stage=factor(
                                c(
                                    rep("E9.5",6),
                                    rep("E10.5",9)
                                    )
                                ),
                    Position=factor(
                                    c(
                                        rep("Dorsal",3),
                                        rep("Ventral",3),
                                        rep("Dorsal",3),
                                        rep("Ventral",6)
                                        )
                                    )
                            )

#initiate the p values matrix
 pvals <- matrix(
                    data=NA,
                    nrow=ncol(experiment),
                    ncol=15,
```

```
                    dimnames=list(
                                colnames(experiment),
                                paste(
                                        paste('PC', 1:15, sep=''),
                                        ' (',
                                        fraction_explained,
                                        '%)',
                                        sep='')
                                    )
                    )

#loops through the categorical trait definitions and calculates an
#association with each
# principle component by ANOVA
 for (i in 1:ncol(experiment)){
   for (j in 1:15){
       fit <- aov(pca$x[,j] ~ experiment[,i])
       pvals[i,j] <-  summary(fit)[[1]][["Pr(>F)"]][[1]]
   }
```

2.6.6 Gene-wise clustering to identify tissue signatures

2.6.6.1 Preparation of input genes for clustering

The most variant genes, with expression greater than 0.5 RPM in at least one sample, were selected as input for gene-wise clustering (the choice of input genes is detailed further in section 3.2.2, Figure 3-4). Variant genes were selected by calculating coefficient of variation with R package ‘genefilter’ (Gentleman R, Carey V, Huber W and Hahne F, 2016). Variant genes were then centred by subtracting the medial expression for every gene.

2.6.6.2 Clustering with ConsensusClusterPlus

Gene clusters were calculated using the ConsensusClusterPlus R package version 1.22.0 (Wilkerson and Hayes, 2010). To ensure the main determinant of clustering was the degree of differential expression between samples, Pearson correlation of median centred gene expression values was used as the distance measure. Average linkage hierarchical clustering was used as the clustering algorithm. Consensus values, which are the number of times two items occupied the same cluster as a proportion of the number of times they occurred in the same subsample, were calculated from 50 iterations of the clustering. Iterated clustering

and calculation of consensus values was calculated for all values of K (number of clusters) between 2 and 10.

2.6.6.3 Selection of input genes and value of K numbers of clusters

The number input genes and number of clusters (K) used to partition genes, were selected through assessment of the stability of clusters i.e. in which clusters had the highest consensus values. This was assessed by calculation and inspection of the cumulative distribution function and delta are under curve for each value of K (Figure 2-2). The cumulative distribution function (CDF) calculates the cumulative proportion of the population that has consensus values over the range [0,1] :

$$CDF(c) = \frac{\sum_{i<j} 1\{M(i,j) \leq c\}}{N(N-1)/2}$$

where *1{. . .}* denotes the indicator function, *M(i, j)* denotes entry *(i, j)* of the consensus matrix *M*, and *N* is the number of rows (and columns) of *M* (Monti et al., 2003). Delta Area Under Curve is the change in area under the CDF curves for each increase in value of K (Monti et al., 2003). A low Delta Area Under Curve was used as an indication that further increase in K did not significantly change the distribution of consensus values and would likely result in overfitting.

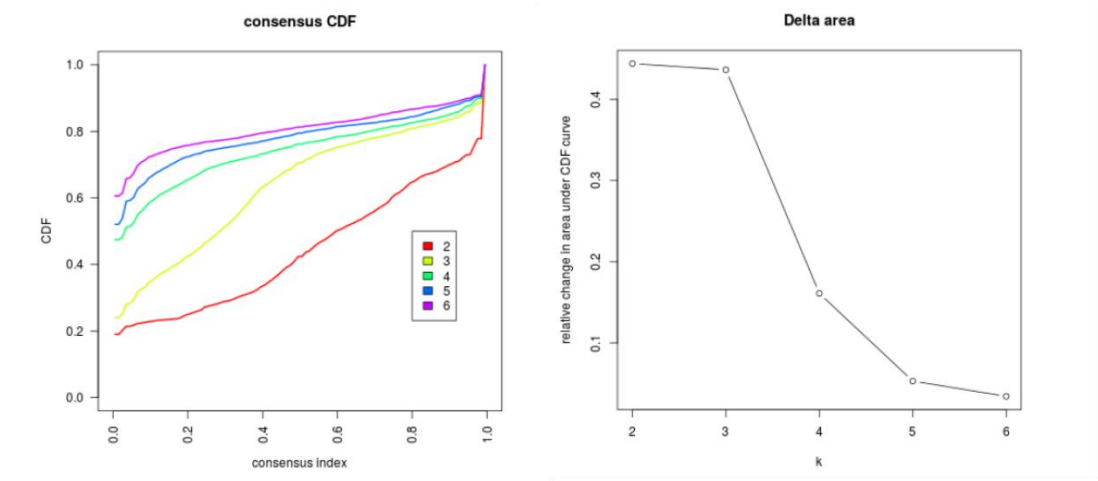


Figure 2-2 Example of consensus distribution function for selecting value of K clusters

Adapted from (Wilkerson and Hayes, 2010) ConsensusClusterPlus tutorial[1]. Left shows the consensus distribution for consensus values over range [0, 1] for values of K between 2 and 6. Right shows the relative change in the area under the CDF curve providing an informative guide to the minimum value of K producing a stable partitioning of clusters.

2.6.6.4 Assessment of cluster-trait association

The association of gene clusters with phenotypes was determined through calculation of the mean expression level of genes belonging to each cluster and association with categorical values "E10.5 AoV", "E10.5 AoD", "E10.5 UGR", "E9.5 AoV" and "E9.5 AoD" calculated by one-way ANOVA.

2.6.6.5 Functional enrichment of clusters

Enrichment of gene clusters with groups of genes associated with particular traits was determined by calculation of enrichment of Gene Ontology terms with the R package TopGO (Alexa A and Rahnenfuhrer J, 2016). Terms were selected from the Biological Processes Ontology. Enrichment of Gene Ontology terms was with weighted Fisher's statistic using the 3000 dynamically expressed genes as background and the genes of a particular cluster as the gene set of interest. Weighting accounts for the topology of the GO graph structure (Alexa et al., 2006). On the basis that 'child' GO terms represent more specific biological information

1 http://www.bioconductor.org/packages//2.7/bioc/vignettes/ConsensusClusterPlus/inst/doc/ConsensusClusterPlus.pdf

than ‘parent’ GO terms, weighed Fisher’s enrichment assesses terms where a child has higher enrichment than the parent, and in such cases genes annotated in the child term are given a lower weighting in the parent. Multiple hypothesis correction was not applied, as it has been considered overly conservative and non-representative of the assumptions and procedures used for weighted Fisher’s exact test of Gene Ontologies, risking loss of information (Alexa A and Rahnenfuhrer J, 2016). Instead, enriched ontologies were ranked by p-value and, for ease of interpretation, the top ten enriched terms are displayed for each gene cluster.

2.6.7 Differential expression analysis

For comparisons between conditions in both the AGM RNA-seq datasets and OP9 datasets, differentially expressed genes were calculated by applying the negative binomial distribution with R/Bioconductor package DESeq (2011-03-15) (Anders and Huber, 2010) and selecting those with an absolute fold change > 2. P-values were adjusted for multiple hypothesis correction with the Benjamini and Hochberg procedure (Benjamini and Hochberg, 1995) to produce false discovery rates (FDR). Genes were selected where the FDR was less than 0.05.

2.6.8 Pairwise comparative gene set enrichment analysis

Relative gene set enrichment in pairwise comparisons between whole sample expression profiles was calculated using the Limma package in R/Bioconductor (Ritchie et al., 2015). Briefly, trimmed mean of M values (TMM) scale normalization method (Robinson and Oshlack, 2010) is applied to read counts followed by mean-variance modelling at the observational level (voom) using an experiment design matrix (Law et al., 2014). Comparative gene set enrichment was calculated with ROAST (rotation gene set testing) which uses a Monte Carlo simulation technology instead of permutation to reduce the rate of false positives in gene set enrichment. This removes the need for gene-wise permutation, which doesn’t account for inter-gene correlations, and allow a permutation across complex experimental design with low replication. Pathways were obtained from the Molecular Signatures Database (Subramanian et al., 2005), and mapped from human to mouse with a homology file from MGI HOM_MouseHumanSequence.rpt (retrieved 2015-10-20). Significant pathways were determined as those with FDR <

0.2. Genes contributing to pathway enrichment were determined by as differentially expression genes (using the same linear model applied with Limma ROAST) and uncorrected with a p-value < 0.2 due to the contribution of multiple genes to the pathway enrichment scores.

2.6.9 Hypergeometric test

Significant overlaps between differentially expressed genes in OP9 samples, and AGM gene clusters, were calculated by hypergeometric distribution (R phyper), where the number of genes in the intersection represents *k* successes in *n* draws (where n is the number of genes in a cluster), from a population *N* (where *N* represents the total number of genes expressed in OP9 cells), that contains *K* successes (where *K* represents the number of significantly up-regulated genes in reaggregated OP9 cells).

2.6.10 *In silico* sub-sampling of sequencing depth

Modelling the RNA-seq expression data at lower sequencing depth used R package ShortRead version 1.24.0 (Morgan et al., 2009). The function FastqSampler selects a defined number of reads randomly from the sample FASTQ files and writes these to new FASTQ files. Following subsampling of 12.5 million and 25 million reads, FASTQ files were aligned and processed as described above.

2.6.11 Meta-analysis and correlation of *Bmp4* and *Bmper* in external datasets

Meta-analysis was undertaken with datasets obtained from the sequence read archive (SRA) or the European Nucleotide Archive (ENA) (Leinonen et al., 2011; European Nucleotide Archive, EMBL-EBI) (SRP033554, SRP049826, SRP045264, SRP036025, SRP023312, SRP026702, ERP001549, ERP001606) through the automated import tool in GeneProf (Halbritter et al., 2012), and processed with GeneProf as described above (section 2.6.3). Read counts for each experiment were merged in R and normalised with the updated DEseq2 version 1.8.2 which enabled use of treatment condition as a factor for variance stabilising transformation, to reduce the dependence of variance on the mean without discarding experimental conditions as noise (Love et al., 2014). Correlation between genes was calculated

with Spearman's rank correlation to reduce the impact of variance due to technical variation of different studies.

2.7 Construction of an interactive visualisation and database

The web-based interactive visualisation for exploration of the AGM RNA-seq and OP9 RNA-seq data was built on an infrastructure developed by Dr Duncan Godwin, in which all scripts and references to server locations where data is processed are packaged into a few lines of code, which can be included in any HTML page.

Pre-processed expression data is stored in a TXT file, with "Gene_name", "Ensembl_ID", "Entrez_ID"; followed by the RPM expression values of each sample such as: "E9AoD_RPM", "E9AoV_RPM", "E10AoD_RPM", "E10AoV_RPM", "E10UGR_RPM"; the fold change values between samples such as: " FC_E10AoVvsE9AoV", "FC_E10AoVvsE10AoD"; then annotation information for each gene defined by gene ontology categorisation such as :

"Secreted" , "Transmembrane", "Transcription_factor"; and finally whether the gene was a member of pre-computed clusters (clustering method described in section 2.6.5).

The image that represents each sample and the experimental set-up is provided as scalable vector graphics (SVG) file drawn in Inkscape[2]. This is an XML-based vector image format which essentially contains the specifications of a two dimensional image in a text file that is then rendered and displayed by a web browser. Because the file is text-based, specific elements of the image can be selected and modified by JavaScript.

The TXT and SVG files are retrieved locally using JAVA (code to do this was written by Dr Duncan Godwin). The JAVA objects created here are converted to JavaScript Object Notation (JSON), producing a string that can be processed with JavaScript. The main interactions between objects, and components reactive to user input are specified in jQuery[3], a JavaScript library. For example, aspects of the SVG

[2] https://inkscape.org/en/
[3] https://jquery.com/

images, which react to user interactions, were given specific element labels that are retrieved by jQuery enabling changes to their CSS (Cascading Style Sheet) attributes, as well as selection of columns of the table to filter. The expression data was displayed with the DataTables plugin from jQuery[4], and sliders and check boxes were from jQuery UI[5]. The main structure of the visualisation is defined in HTML, and jQuery was used to append interaction to this structure. The links to Ensembl (Yates et al., 2016) and the Mouse Atlas Project (Richardson et al., 2013) were generated by pasting the selected gene ID into a fixed URL pattern.

2.8 HSC maturation *ex vivo*

For *ex vivo* reaggregate cultures, caudal parts were dissected from E9.5 embryos and the AGM region was dissected then subdissected into AoV and AoD from E11.5 embryos (as described in 2.3). Dissected embryonic tissues were dissociated by collagenase/dispase and then either self-reaggregated or co-aggregated with OP9 stromal cells. For self-reaggregation, AGM cell suspensions were centrifuged at 430 x g for 12 min in 200 µl pipette tips sealed with parafilm to form a pellet. For co-aggregation with OP9, cell suspensions of 1 ee of embryo cells were mixed with 10^5 OP9 cells prior to centrifugation. Cell aggregates or explants were cultured at the liquid–gas interface on 0.8-µm nitrocellulose filters (Millipore) at 37 °C in 5% CO_2 (which were placed on media at least 1 h before cells added), for either 5 days (with E11.5 cells) or 7 days (with E9.5 cells) in 5 ml reaggregate culture media. For E9.5 reaggregate culture, 2 ml of culture media was added for the first 24 h, then this was replaced with 5 ml fresh media for the rest of the culture period. Additional recombinant proteins were GDF-3, CHRDL2, BMPER, INHBB, IBSP, IGFBP3, NELL1 and WNT2b (R&D Systems); CXCL10 and CCl4 (Peprotech), added at concentrations specified below in the results. After the stated culture period, the whole membrane was immersed collagenase/dispase (Roche) for 40 min at 37 °C, to remove reaggregates and dissociate them into a single cell suspension.

[4] https://datatables.net/
[5] https://jqueryui.com/

2.9 Colony forming assay

Methylcellulose based MethoCult medium containing erythropoietin, IL-3, IL-6, and SCF (M3434, Stem Cell Technologies) was thawed at room temperature and treated with 50 units/ml P/S. Cultures were processed and plated in MethoCult according to the manufacturer's instructions (www.stemcell.com) at a concentration between 0.005 ee and 0.05 ee (indicated in figure legends in the results) depending on whether cytokines had been added to the culture. Haematopoietic colonies were counted and scored after 7-9 days of differentiation. The colonies were scored according to described standard criteria (Medvinsky et al., 2007).

2.10 Long-Term Repopulation Assay

Prior to transplantations, recipient mice (C57BL/6 CD45.1/2) were irradiated with a total dose of 9.5 Gy, split into two doses that were separated by at least 3 hours and delivered by sealed Cs source at a rate of 21.6rad/min. Donor cells were injected intravenously into recipient mice along with 20,000 bone marrow carrier cells from C57BL/6 CD45.1/1 mice. Carrier cells were obtained by flushing the tibias and femurs with cell suspension buffer using 26-gauge syringe needle (BD Microlance), followed by mechanical dispersion and filtering through a 40μm cell strainer (BD Falcon). The injected dose of donor cells is stated in the text per embryo equivalent unit i.e. unit of cells equivalent to the number present in one embryo. The transplanted dose was adjusted depending on the culture system so small numbers of HSCs would be detected, but to ensure controls did not reach a saturating level of repopulation. 200μl of cell suspension was used for each injection, which was delivered using a 30-gauge syringe needle (BD Plastipak). All injections were into the lateral tail vein according to the procedures described in the project license and the regulations of the Animals Scientific Procedures Act, UK, 1986.

To detect long-term haematopoietic repopulation, peripheral blood was collected 16 weeks after transplantation by bleeding the tail vein into 500 μl of 5 mM EDTA/PBS. Erythrocytes were depleted using PharM Lyse (BD) and cells were stained with anti-CD16/32 (Fc block), anti-CD45.1-APC, and anti-CD45.2-PE, monoclonal antibodies at concentrations listed in Table 2, according to the protocol

below (section 2.11.1). The percentage of donor CD45.2 cells was analysed using FACSCalibur and flowJo software (TreeStar).

Multilineage contribution was detected from peripheral blood between 12-16 weeks after transplantation. Briefly, contribution to all blood lineages was detected by exclusion of recipient CD45.1+ cells and staining with lineage-specific monoclonal antibodies for Mac1, CD3e, Gr1, and B220 at concentrations listed in Table 2. Cells were analysed using analysed with BD LSR Fortessa analyser (5 laser or 4 laser) (BD Bioscience) and flowJo software (TreeStar).

2.11 Flow cytometry

Antigen	Clone	Isotype	Working Conc.	Conjugate	Supplier
CD16/CD32 (Fc block)	93	rIgG2b	1µg/ml	none	eBioscience
CD45	30F-11	rIgG2b	2µg/ml	V450, BV650, V500	BD Horizon
Mac1/CD11b	M1/70	rIgG2b	2.5µg/ml	Biotin, PE, PerCP_Cy5.5, FITC	eBioscience
CD3e	145-2C11	hIgG	2µg/ml	Biotin, PE, PerCP_Cy5.5, APC	eBioscience
CD41	MWReg30	rIgG1	10µg/ml	BV421	Biolegend
Gr1	RB6-8C5	rIgG2b	1.5µg/ml	Biotin, PE, PerCP_Cy5.5, PE_Cy7	eBioscience
B220/CD45R	RA3-6B2	rIgG2a	2.5µg/ml	Biotin, PE	eBioscience
Ter119	TER-119	rIgG2b	2µg/ml	Biotin, PE, PerCP_Cy5.5, FITC	eBioscience
CD45.1	A20	mIgG2a	2µg/ml	APC	eBioscience
CD45.2	104	mIgG2a	2µg/ml	PE	eBioscience
VC/CD144	eBioBV13	rIgG1	2µg/ml	ef660	eBioscience
CD146	ME-9F1	IgG2a	0.5µg/ml	PE	Biolengend
CD43	eBioR/60	rIgM	1µg/ml	PE, Biotin	eBioscience
Streptavidin	N/A	N/A	0.1µg/ml	BV560	Biolegend
7AAD	N/A	N/A	0.5µg/ml	N/A	Mol. Probes

Table 2 Antibodies used for flow cytometry

Details the antigen, clone, antibody type and source, working concentration, conjugates and suppliers used for flow cytometry experiments detailed below.

2.11.1 Staining cells for flow cytometry

Single cell suspensions were obtained as described in section 2.3 (embryo tissues) and 2.10 (adult blood cells). Antibody staining was in 5ml polystyrene tubes (BD Falcon) or in U bottomed 96-well plates (Sterilin). The antibodies and the concentration used are described in Table 2, and are specified in the results. Cells were spun down at 350 x g for 5min at 4˚C then resuspended in the primary antibody cocktail together with Fc block. Primary antibodies were incubated with the cell suspensions for 30 min at 4˚C or on ice in the dark. In parallel, additional control samples were stained with FMO (fluorescence minus one), where all antibodies were added except for one. As well as this single stains for each antibody were prepared using OneComp or UltraComp eBeads (eBiosciences) or cells. After incubation, cells were washed in FACS buffer then centrifuged at 350 x g for 5min at 4˚C. If any of the primary antibodies were conjugated with Biotin, a further stain containing the appropriate fluorochrome-conjugated streptavidin was added to the cells and incubated for 30 min at 4˚C or on ice. Cells were then washed again, spun down at 350 x g for 5 min at 4˚C and resuspended in cell suspension solution with 7AAD and analysed by flow cytometry.

For intracellular staining with anti-BMPER for flow cytometry, cells were first incubated with antibodies against extracellular antigens as above. Cells were then fixed in 4% paraformaldehyde at 4 °C for 10 min. This was followed by a wash with PBS, then incubation with PBS/0.1% Triton X-100 for 10 min at 4 °C. This solution was replaced with a blocking solution PBS/7% FCS/0.1% Triton X-100 (FACS/0.1% Triton) for 30 min. Anti-BMPER was diluted in FACS/0.1% Triton at the concentrations described below (Table 3) and applied to cells for 1 h. Cells were then washed with FACS/0.1% Triton and then the secondary antibody, anti-rabbit-AF488, was applied for 45 min. Subsequently samples were prepared for flow cytometry as described above.

2.11.2 Flow cytometry analysis

Flow cytometry analysis was performed using a dual laser FACScalibur (BD Bioscience) or BD LSR Fortessa analyser (5 laser or 4 laser) (BD Bioscience). Compensation adjustments were made based on appropriate single stains and gates were defined by FMO controls. Viable cells were defined and selected as those with

a low 7AAD intake. Cell size and granularity were assessed by forward and side scatter profiles respectively. Data acquisition was performed using BD CellQuestTM software (BD Bioscience). Data analysis and flow cytometric statistical analysis was performed using FlowJo software (Tree Star, Inc).

2.11.3 Fluorescence-activated cell sorting of embryonic tissues

For fluorescence-activated cell sorting of embryonic tissues, single cell suspensions were obtained as described in section 2.3) and stained with appropriate antibodies (Table 2) as described in section 2.11.1, including FMO controls and single stain compensation controls. Cell sorting was with the assistance of Dr Fiona Rossi and Dr Claire Cryer with the FACSAriaII and analysis with FACSDiva software (BD Bioscience). Viable cells were selected based on a low uptake of 7AAD and mature blood lineage cells were depleted by taking only cells negative for Ter119, Gr1, CD3e and CD11b. The remaining populations were defined by gates set using FMO controls and are detailed in the results.

2.12 Immunostaining and microscopy

Antigen	Clone	Isotype	Working Conc.	Conjugate	Supplier
CD45	AF114	Goat	2µg/ml	none	R&D Systems
CD31	MEC13.3	Rat	2.5µg/ml	none	Pharmingen
BMPER	ab73900	Rabbit	1µg/ml	none	Abcam
SSEA-1	MC-480	IgM	0.5µg/ml	Biotin	eBioscience
CD146	AF6106	Sheep	2µg/ml	none	R&D Systems
Phospho-SMAD1/5/8	D5B10	Rabbit	1µg/ml	none	Cell Signalling Technology
PAX2	Poly19010	Rabbit	5µg/ml	none	BioLegend
Goat IgG	NL001	donkey	10µg/ml	NL557	R&D Systems
Rabbit IgG	N/A	donkey	20µg/ml	AF647	Invitrogen
Rat IgG	N/A	donkey	20µg/ml	AF488	Invitrogen
Rabbit IgG	N/A	donkey	20µg/ml	AF488	Invitrogen
Streptavidin	N/A	N/A	0.5µg/ml	PE	Pharmingen
DAPI	N/A	N/A	0.5µg/ml	N/A	Biotium

Table 3 Antibodies used for immunohistochemistry

Details the antigen, antibody type and source, working concentration, conjugates and suppliers used for immunostaining experiments detailed below.

2.12.1 Sample fixing, embedding and sectioning

Embryos were fixed in 4 % paraformaldehyde at 4 °C overnight then embedded in gelatin by first incubating in 15 % sucrose for 2h at 4°C, followed by PBS/15 % sucrose/ 7 % gelatin at 37°C followed by flash freezing in liquid nitrogen. Transverse sections of 7 μm were cut with CM1900 Cryostat (Leica). Sections were permeabilised with PBS/0.5% Triton X-100 for 10 min then blocked for 30 min with PBS/10% foetal calf heat-inactivated serum (FCS).

2.12.2 Immunostaining extracellular markers

Antibodies were diluted in PBS/2% FCS. Primary staining was with rat anti-CD31 rabbit anti-BMPER, goat anti-CD45, anti-SSEA-1-biotin or sheep anti-CD146, overnight followed by incubation with anti-goat NL577, anti-rat Alexa Fluor 488, anti-rabbit Alexa Fluor 647, or streptavidin-PE for 2 h (antibodies and concentrations listed in Table 3), followed by counterstaining with DAPI. Images were acquired with an inverted confocal microscope (SP8, Leica) and processed using ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2015).

2.12.3 Immunostaining intracellular markers

For pSMAD1/5/8 and PAX2 staining, blocking was with PBS/10% serum/1% BSA/0.1% Triton X-100 and antibodies were diluted in PBS/5% serum/1% BSA/0.1% Triton X-100.

2.13 Expression profiling by qRT-PCR

2.13.1 Preparation of cDNA from bulk populations of cells

For quantitative real-time PCR (qRT-PCR) from bulk populations of cells (> 50,000), RNA extraction from a pellet of cells was as described above with Qiagen

RNeasy microkit or minikit (Qiagen), except in some cases the pellet either lysed immediately or was snap frozen and stored at -20°C before lysis and RNA extraction. cDNA was prepared with SuperScript® III Reverse Transcriptase (Invitrogen), random hexamer primers (Invitrogen) and 10 mM dNTP Mix (Invitrogen). Primers, RNA and dNTPs were heated at 65°C for 5 minutes and incubated on ice for 1 minute. The remaining buffers and reverse transcriptase were added to samples and incubated first at 25°C for 5 minutes then 50°C for 1 h for extension, followed by inactivation at 70°C for 15 minutes

2.13.2 Preparation of cDNA from 200 cell

For detection in small populations of cells, up to 200 cells were directly sorted by FACS into 10 µl of × 2 Reaction Mix (CellsDirect, Invitrogen) and 0.2 µl RNase inhibitor (SUPERase-In Ambion AM2694). Superscript III/Taq mix (CellsDirect) and gene-specific primers (10 µM each) were added to the cell lysate to directly reverse transcribe and amplify cDNA (PCR programme: 50 °C for 15 min; 95 °C for 2 min; 18 cycles of: 95 °C for 15 s; 60 °C for 4 min), with a protocol optimised by Kathy O'Neill (Professor Clare Blackburn Lab).

2.13.3 qRT-PCR

Where RNA was extracted from bulk populations of cells (> 50,000), qRT-PCR was carried out with using LightCycler 480 SYBR Green I MasterMix (Roche) for detection. Where RNA was extracted from 200 cells, Lightcycler 480 Probes Master Kit (Roche) was used for qRT–PCR detection on diluted cDNA. All qRT-PCR used two or more biological replicates and expression was measured relative to TATA-binding protein (Tbp). Control samples that underwent the Taq amplification in the absence of SuperScript III reverse transcriptase, (or had not undergone reverse transcription), were used to assess contamination or amplification of genomic DNA.

Statistical analysis of qRT-PCR data used a two-tailed Student's t-test to compare values that had been normalised to Tbp. P-values are stated in the figure legends. To plot the relative change in expression, all values were then normalised to one sample.

2.13.4 Primers for qRT-PCR

Primers for qRT-PCR were designed with Universal Probe Library Assay Design Center (Roche) and are listed below. Primers were designed which span introns and have a melting temperature between 59-60°C. The presence of interference due to primer dimers or contamination was assessed through the denaturation and reannealing of primers as a melt curve analysis with SYBR Green I MasterMix reagents. The efficiency of primers within a range of 1.8 to 2.1 duplications per cycle was calculated by serial dilution of cDNA at 4 concentrations.

MGI symbol	5' primer seq	3' primer seq	Amplicon size	UPL probe #
Bmp4	gaggagtttccatcacgaaga	Gctctgccgaggagatca	122	89
Bmper	tgtgcaagttcggtagcaag	tgcagttgactgaggaccac	60	58
Tbp	ggggagctgtgatgtgaagt	ccaggaaataattctggctca	82	97

Chapter 3 Transcriptional profiling of the niche for developing HSCs

3.1 Introduction

The AGM region, the niche in which haematopoietic stem cells mature, has been functionally mapped through comparison of the autonomous capacity of tissues to support HSC maturation and the degree to which this can be enhanced by external factors (Rybtsov et al., 2014; Souilhol et al., 2016a; Taoudi and Medvinsky, 2007). These studies have addressed temporal and spatial aspects of the niche respectively. While reaggregation and culture of E10.5 AGM tissues for 4 days will generate transplantable HSCs, the E9.5 AGM region can only generate haematopoietic stem cells after 7 days in culture with OP9 stromal cells (Rybtsov et al., 2014). This can, in part, be explained by the lower number of HSC precursor cells at this earlier E9.5 stage (Rybtsov et al., 2016). But the additional requirement for OP9 cells and the extensive time-lag in development suggests that there is also a lack of sufficient signalling cues from the primary niche cells of this culture system, to fully support the maturation process. It has also been shown, through separate cultures of the ventral (AoV) and dorsal (AoD) domains of the dorsal aorta, that the generation of HSCs is spatially polarised towards the ventral region(Souilhol et al., 2016a; Taoudi and Medvinsky, 2007). Again this can partially be explained by a ventral polarisation of precursor cells (pre-HSCs), but co-cultures of labelled AoV, AoD and UGRs show that the presence AoD, UGR and AoV each increase support for HSC maturation at E10.5 and E11.5 respectively. This suggests that AGM region subdomains each host important inductive signalling cues (Souilhol et al., 2016a).

Some of the signalling factors underlying these inductive functions of different AGM domains have been identified, such as the highly ventralised expression of SCF, Noggin and the dorsalised expression of sonic hedgehog (Shh) which have all been shown to support HSC maturation (Mukouyama et al., 1998; Peeters et al., 2009; Rybtsov et al., 2014; Souilhol et al., 2016a; Taoudi et al., 2008).

However the investigation of this inductive signalling has not yet been addressed on the genome-wide scale. Screening the full transcriptome of a cell type/tissue can elucidate the gene regulatory networks underlying these phenotypes, but can also facilitate the identification of protein level interactions such as surface markers and secreted molecules (although a potentially significant protein level regulation should be kept in mind). In this context, given the well-characterised function of the AGM region as a niche, it was thought that a transcriptional approach would be an effective strategy to characterise the molecular effectors of this function.

A number of studies have implemented genome-wide expression profiling technologies to gain important insight into the processes underlying HSC maturation (Kartalaei et al., 2015; Li et al., 2014; McKinney-Freeman et al., 2012) however they have largely focused on enriched populations (using different marker combinations) of HSCs, progenitors and precursors. Studies of the niche so far have focused on stromal cell lines (Charbord et al., 2014), or have lacked the important spatial resolution along the dorsal-ventral axis (Mascarenhas et al., 2009). Therefore, obtaining a deep transcription profile based on the most up-to-date understanding of the functional demarcation of the AGM region was considered the best way to gain an understanding of the non-cell autonomous regulation of HSC maturation.

Here I will describe the production of RNA-seq datasets from AGM regions of E9.5 and E10.5 embryos, subdissected into AoD, AoV and, at E10.5, UGR; and from OP9 cells cultured in submersed or reaggregated conditions. I will discuss alignment and quality control of the data; comparison of samples based on their expression profiles; identification of *de novo* gene sets representing tissue-specific gene sets through clustering; analysis of functionally enriched gene sets through two methods; and how comparison of the independent AGM region and OP9 transcriptional profiles shows a significant overlap in genes expressed in E10 AoV and reaggregated OP9. Ultimately, I will emphasise the potential of these datasets to inform functional studies of the molecular regulation of HSC maturation. Finally, through the production of an interactive visualisation I will demonstrate the utility an interface into these datasets to enable future exploration.

3.2 Results

Aspects of this work were conducted with the assistance of Dr Stanislav Rybtsov, Dr Celine Souilhol, Dr Florian Halbritter, Dr Jonathan Manning and Dr Duncan Godwin. Dr Stanislav Rybtsov and Dr Celine Souilhol sub-dissected the AGM regions and assisted in extraction of the RNA for sequencing. Dr Florian Halbritter provided part of the code for the display of principle component analysis. Dr Jonathan Manning provided part of the code for testing association of principle components with categorical traits, and differential gene set analysis. Dr Duncan Godwin provided assistance with the JAVA coding portion of the interactive data visualisation.

3.2.1 RNA-sequencing of AGM subdomains reveals temporal and spatial transcriptional changes in the embryonic HSC niche

In order to interrogate the expression changes which underlie the differences in functionality of AGM subdomains and developmental changes, the AGM regions of E9.5 and E10.5 embryos were sub-dissected into dorsal (AoD) and ventral (AoV) domains of the dorsal aorta, as well as the urogenital ridges (UGR) at E10.5 (Figure 3-1). This is in accordance with regions which have previously been functionally characterised in reaggregate cultures (Rybtsov et al., 2014, 2016; Souilhol et al., 2016; Taoudi and Medvinsky, 2007). Pools of these tissues from between 15 to 34 embryos yielded more than 5 μg of RNA from each sample (Table 4), which is sufficient to generate libraries for RNA-sequencing without the need for pre-amplification (dissection and RNA extraction was with the assistance of Dr Stanislav Rybtsov and Dr Celine Souilhol).

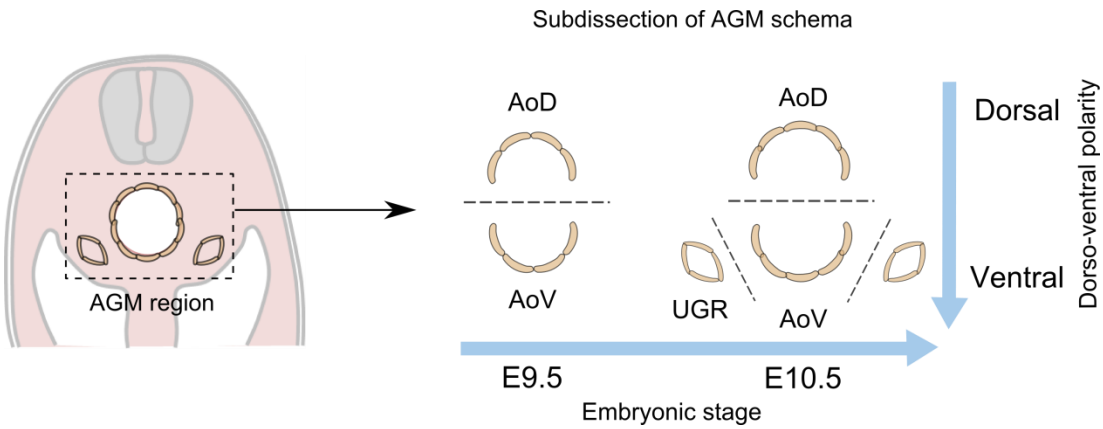


Figure 3-1: Aorta Gonad Mesonephros dissection strategy for RNA-sequencing

Schematic for subdissection of the aorta-gonad-mesonephros (AGM) region for RNA-extraction and sequencing. The dorsal aorta and urogenital ridges were isolated from E9.5 and E10.5 embryos. At E10.5 the urogenital ridges (UGRs) were separated from the dorsal aorta, while at E9.5 they are not sufficiently distinct to allow separation. The dorsal aorta was dissected into dorsal (AoD) and ventral (AoV) sides. These subdomains were pooled from multiple embryos for RNA extraction and sequencing.

Sample Name	Embryo stage	Tissue	Number of embryos	Nanodrop RNA conc (ng/µl)	Bioanalyser RNA conc (ng/µl)	RIN	Total volume (µl)	Total RNA (µg)
E10D1	E10.5	AoD	25	145	152	9.9	45.00	6.84
E10D2	E10.5	AoD	34	250	360	9.6	46.00	16.56
E10D3	E10.5	AoD	32	122	152	9.5	46.00	6.99
E10V1	E10.5	AoV	25	184	174	10	45.00	7.83
E10V2	E10.5	AoV	34	220	298	9.9	46.00	13.71
E10V3	E10.5	AoV	32	135	242	10	46.00	11.13
E10U1	E10.5	UGR	25	245	430	10	45.00	19.35
E10U2	E10.5	UGR	34	300	412	9.5	46.00	18.95
E10U3	E10.5	UGR	32	208	298	9.3	46.00	13.71
E9D1	E9.5	AoD	14	230	772	9.7	45.00	34.74
E9D2	E9.5	AoD	16	224	606	10	47.00	28.48
E9D3	E9.5	AoD	15	738	1026	10	47.00	48.22
E9V1	E9.5	AoV	14	250	480	9.6	45.00	21.60
E9V2	E9.5	AoV	16	286	753	10	43.00	32.38
E9V3	E9.5	AoV	15	203	666	10	47.00	31.30

Table 4 AGM region sample information for RNA-sequencing

Columns indicate the abbreviated sample names (used in subsequent figures); embryonic stage; subdissected tissue: dorsal half of the dorsal aorta (AoD), ventral side of the dorsal aorta (AoV) and urogenital ridges (UGR); “Number of embryos” indicates the number of embryos of the same stage from which tissues were dissected and pooled; “NanoDrop RNA conc” is the concentration of RNA estimated by spectrophotometry; “Bioanalyser RNA conc” is the concentration of RNA measured by the Agilent Bioanalyser; “RIN” is the RNA

integrity number; the total amount of RNA was calculated from the total volume multiplied the RNA concertation measured by Bioanalyser.

The average quality score per cycle was plotted for each sample (Figure 3-2A). The quality score (or phred score) is the probability that a base was called accurately during the sequencing process (Ewing and Green, 1998). Low scores can indicate errors in the sequencing process, for example degradation of the sequencing chemistry towards the end of the run, or a short drop in quality could indicate a bubble or some sort of transient problem in the sequencer. Scores above 20 throughout the reads indicated that the bases have a high probability of being called accurately (> 99%). Nucleotide composition per dataset displays the nucleotide distribution per library. The number of nucleotides of A, T, C, G would be expected to be equal in all libraries, so an imbalance can indicate a contaminant either with overrepresented sequences (for example adapter dimers) or another species, or a problem with the sequencing machinery. The AGM RNA-seq data showed an equal proportion of each nucleotide (Figure 3-2B) suggesting that there was no significant contamination.

Sequenced reads were aligned using the RNA-sequencing pipeline in GeneProf with Bowtie (Halbritter et al., 2012; Langmead et al., 2009) (detailed in section 2.6.3). The percentages of reads aligning to a single position, multiple positons, or no positions in the genome demonstrate that the majority of reads in all samples can be unambiguously aligned to one position in the mouse genome (Figure 3-2C). Of these aligned reads, approximately 50% of alignments are to a unique position in the genome (Figure 3-2D), suggesting they are not duplicated sequences arising during the PCR amplification stage of library preparation. In sum, the sequencing data was determined to be of good quality for further analysis.

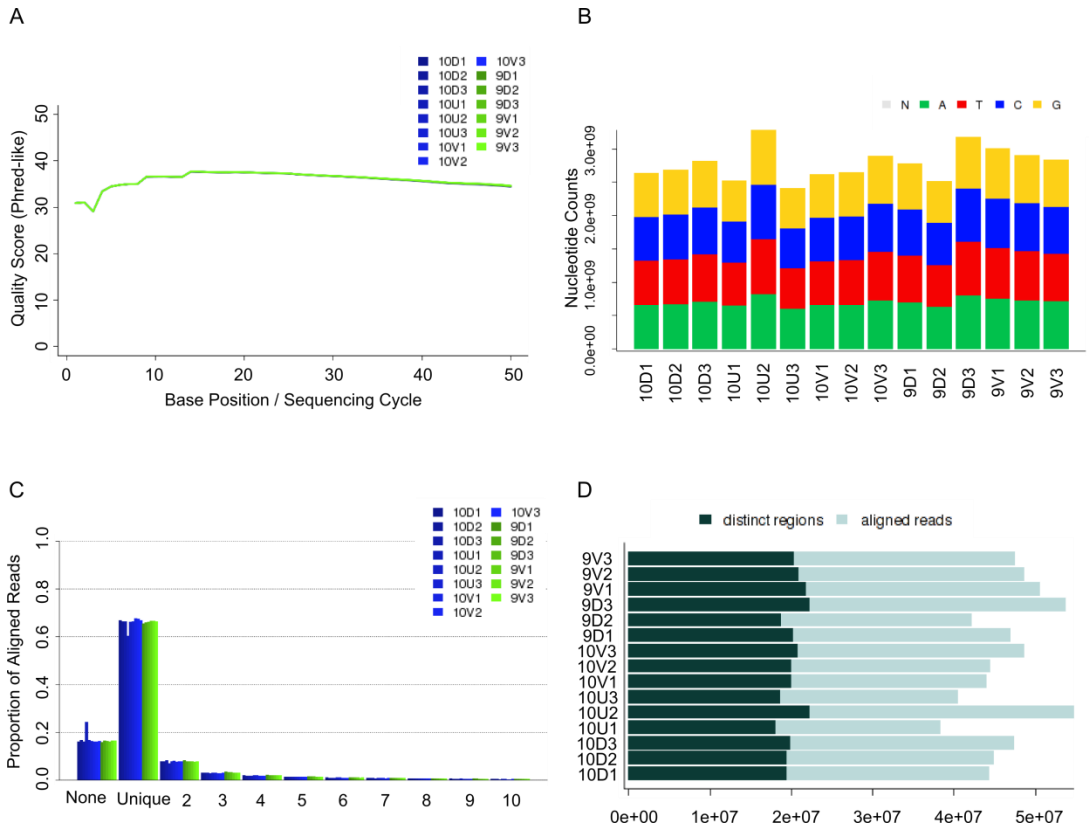


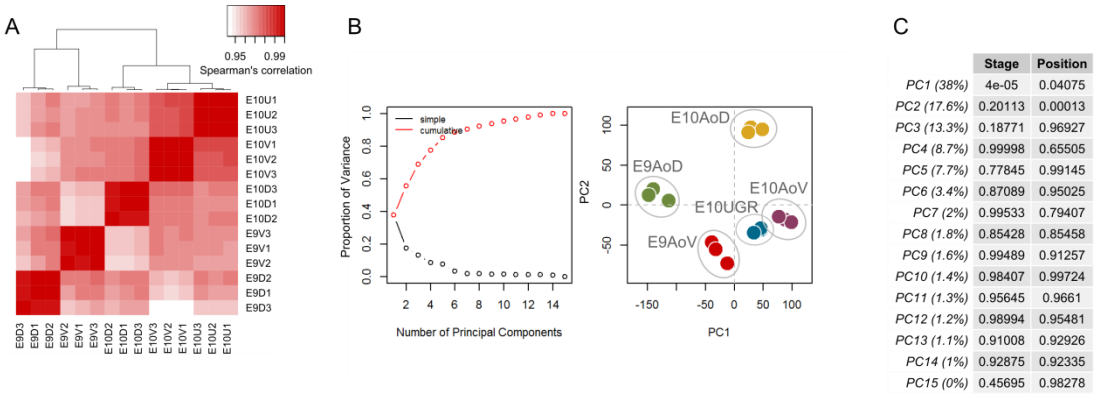
Figure 3-2 Alignment and quality control of RNA-sequencing data from the AGM-region

These plots are based on Sequence Data and Statistics plots from GeneProf. (A) Plots the mean quality score of all reads in each sample for each base position. (B) Shows the number of counts of each nucleotide (A, T, C, G) in each sample as a measure of distribution of nucleotides. (C) Shows the proportion of reads which do not align to the genome, align to one place, or ambiguously align at more than one place in the genome. (D) Represents the number of distinct genomic regions to which a read aligns compared to the total number of aligned reads.

3.2.2 Tissue specific gene expression signatures reflect the differential anatomy of the AGM region and highlight transitions in pathway regulation

To assess variation between samples based on their expression profiles, the expression of all detected genes in the samples were compared by correlation analysis (Figure 3-3A). Biological replicates showed the strongest correlation followed by samples from the same developmental stages. This suggests that the known functional and phenotypic differences between these samples are indeed

associated with strong transcriptional differences. Another method of determining differences between sample expression profiles is to plot the expression datasets based on their principle components. On the basis of the first two principle components (which account for 38% and 17.6% of variance in the data respectively), samples are separated on the plot first by stage then by dorsal-ventral polarity (Figure 3-3B). An analysis of variance (ANOVA) shows that the first two principle components are significantly associated with sample stage and dorso-ventral polarity respectively (Figure 3-3C), suggesting that the developmental progression between E9.5 and E10.5 elicits the strongest transcriptional shift between these samples, whilst the dorso-ventral polarity has a lesser, but still significant impact on the transcriptional differences.



	Stage	Position
PC1 (38%)	4e-05	0.04075
PC2 (17.6%)	0.20113	0.00013
PC3 (13.3%)	0.18771	0.96927
PC4 (8.7%)	0.99998	0.65505
PC5 (7.7%)	0.77845	0.99145
PC6 (3.4%)	0.87089	0.95025
PC7 (2%)	0.99533	0.79407
PC8 (1.8%)	0.85428	0.85458
PC9 (1.6%)	0.99489	0.91257
PC10 (1.4%)	0.98407	0.99724
PC11 (1.3%)	0.95645	0.9661
PC12 (1.2%)	0.98994	0.95481
PC13 (1.1%)	0.91008	0.92926
PC14 (1%)	0.92875	0.92335
PC15 (0%)	0.45695	0.98278

Figure 3-3 Comparing the transcriptional output of AGM region subdomains through clustering and principle component analysis

(A) Clustering of each AGM region RNA-seq sample with distance based on Spearman's ranked correlation. Heat map colouring displays the correlation level between each sample. (B) Principle component analysis of AGM samples based on normalised expression values including the proportion of total variance constituted by each component, and the distance between each sample based on the first two components. Green = E9AoD, Red = E9AoV, Yellow = E10AoD, Purple = E10AoV, Blue = E10UGR. (C) P-values from ANOVA testing association of categorical terms, stage and dorsal-ventral position, with first two principle components.

Given the strong transcriptional differences between samples it was possible to identify tissue-specific expression signatures by a relatively unsupervised clustering method. Using the most variant genes amongst the samples, genes were

partitioned into groups which share similar profiles across the data using ConsensusCluster iterative clustering methodology (Wilkerson and Hayes, 2010). Genes were median-centred, the differences in expression profile of each gene compared by Pearson correlation, then K clusters of genes with highly correlated expression profiles determined using hierarchical clustering. This clustering was iterated 50 times, and a consensus score assigned to every gene pair, signifying the proportion of iterations in which the two genes fall into the same cluster. The stability of cluster partitioning was determined with the cumulative distribution function, which presents the distribution of consensus values for each value of K (Figure 3-4). A high area under curve in the CDF plot, which plateaus at about 0.8, suggests that the majority of consensus values are either 1 or 0, meaning that clusters are non-random. A suitable number of genes, which would produce stable clusters was chosen by repeating this clustering approach for inputs between 1000 and 6000 most variant genes, and examining the resultant CDF plots (Figure 3-4). As the number of input genes increases, the plateau is less marked, suggesting that a large number of genes are not stably assigned to clusters. To strike a balance between the production of stable clusters, and comprehensive representation of the landscape of transcription, and input of the 3000 most variant genes was selected for clustering.

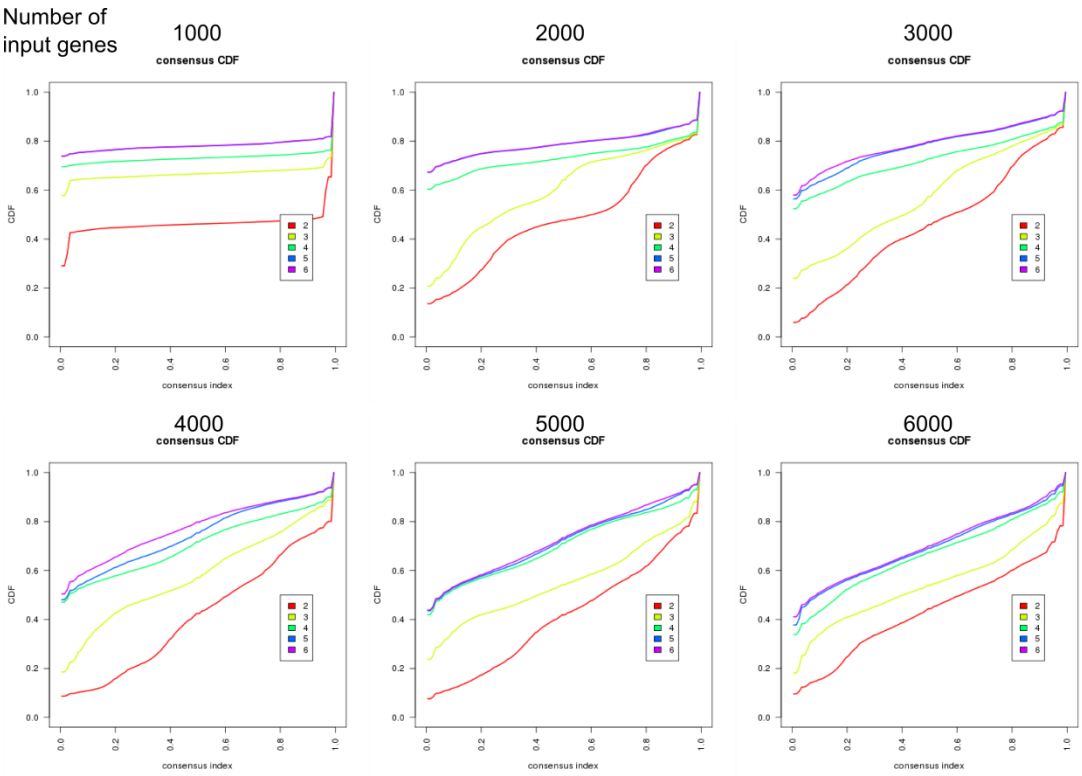
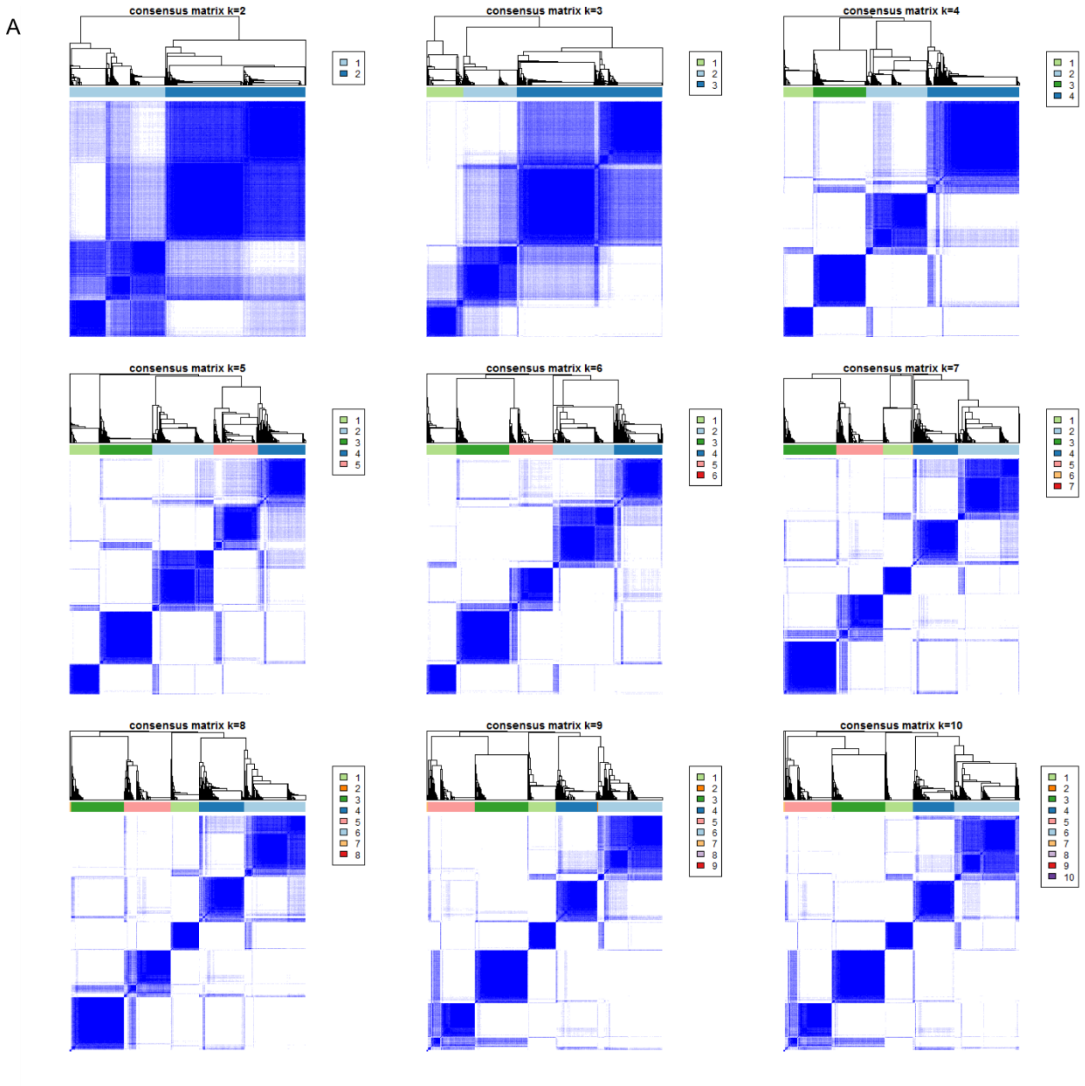


Figure 3-4 Selecting the number of input genes for clustering

Plots display the cumulative distribution function calculated for the number of genes displayed above each plot (between 1000 and 6000). Input genes are ranked as the genes with the highest coefficient of variation across the samples. Each clustering is iterated 50 times, and performed for K clusters between 2 and 6.

To determine the appropriate value of K to generate stable clusters, this clustering was repeated for values of K between 2 and 10 (Figure 3-5A). The relative change in stability of cluster partitioning for each value of K is assessed by the delta area under curve (Figure 3-5B and C), which measures the change in distribution of consensus values as the value of K increases was evaluated. A low are under the curve as K increases, suggests that there is no longer an increase in the proportion of consensus values that are 0 or 1 and therefore further division into clusters are likely to be arbitrary. Therefore in this clustering, the value of K=5 was selected as this is the value at which there were the highest proportion of consensus values equal to either 0 or 1, while values of K above 5 do not further increase the area under curve (Figure 3-5B and C).



D

Cluster mean-trait ANOVA

Phenotype	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
E9AoD	2.074e-08	0.1846391938	0.39591550817	0.03239630564	0.18156455637
E9AoV	0.56151425276	0.07573289664	1e-11	0.87301681336	0.38609987268
E10AoD	0.66339921718	1.38501e-06	0.21270901452	0.35187078796	0.12252566908
E10AoV	0.14499340665	0.70698537093	0.65066019743	4.7232e-07	0.75537957842
E10UGR	0.31746372672	0.49749069351	0.31108359545	0.79367840296	3.3854e-07

Figure 3-5 Gene-wise clustering to identify tissue specific gene expression signatures

(A) Heat maps represent the consensus values and cluster divisions for K clusters (K=2, 3,…,10). Each column and row represents one of the 3000 genes input. The heat map colour gradient represents consensus probability that two genes fall into the same cluster, while colour labelling above columns represents the assignment of genes into K clusters. (B) The cumulative distribution function (CDF) represents the cumulative proportions of the total possible gene-gene interactions that have consensus index values from 0 to 1. This distribution is displayed for each value of K clusters. (C) Displays the relative change in area under the CDF plot for each value of K from 2 to 10. (D) Results of one-way ANOVA comparing the mean gene expression profile of each clusters versus the categorical trait of each sample phenotype.

To assess whether there was a relationship between gene-clusters (which had been generated in an unsupervised manner, independently of known phenotypic information) and the sample phenotypes, the gene-cluster expression means were compared between samples and significance assessed by ANOVA. Each sample showed a significant association with a unique AGM subdomain (Figure 3-5D), as indeed each cluster had a higher expression mean in one sample relative to all others (Figure 3-6) suggesting that the five gene-clusters can be ascribed as representative expression signatures of each AGM subdomain.

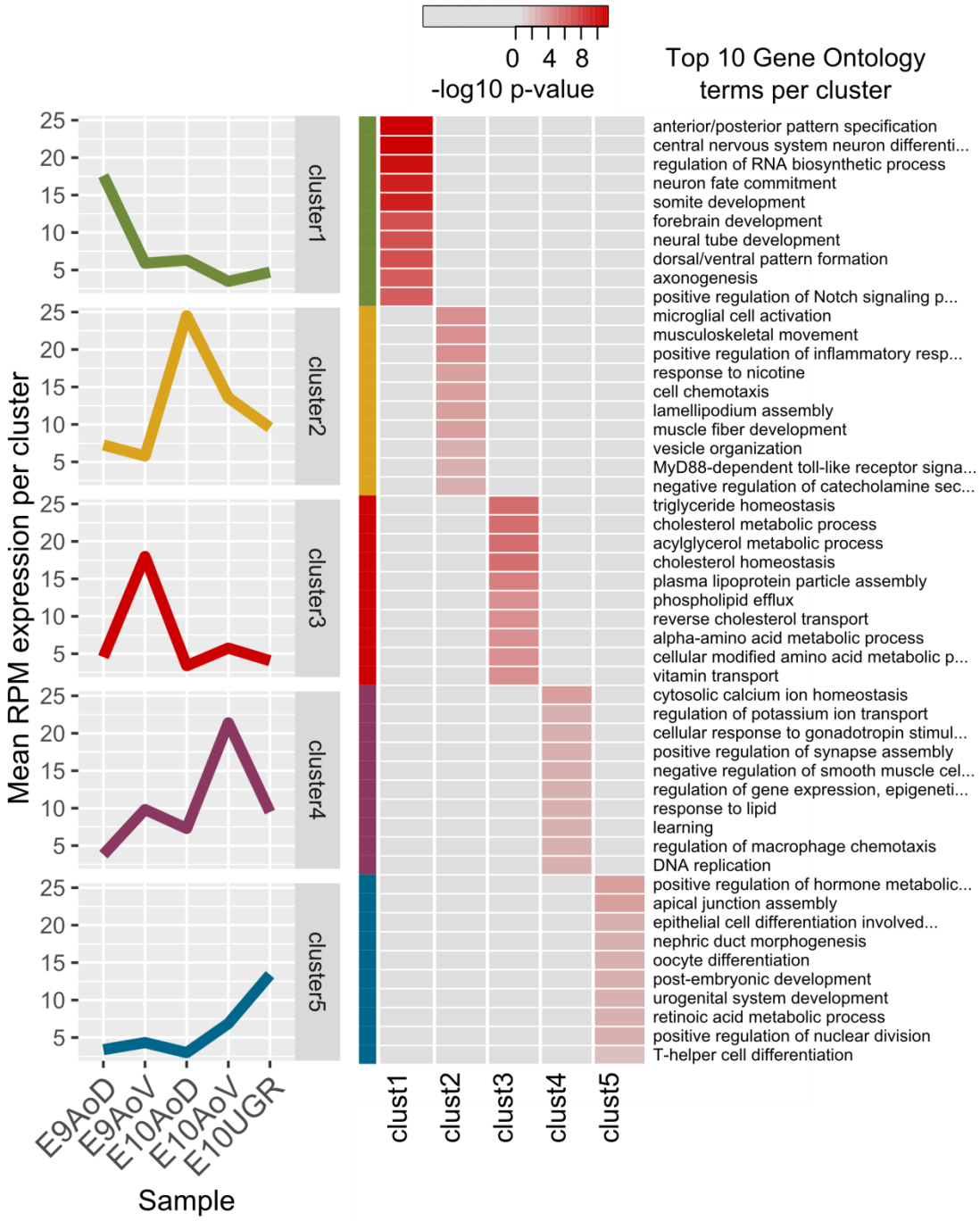


Figure 3-6 Exploration of tissue signatures through gene ontology enrichment

Plots on the left display the profile of the mean expression of each gene cluster. On the right the top 10 most significantly enriched gene ontology terms in each cluster, ranked from lowest to highest p-values. The heat map colours represent the p-value transformed by -log10 from weighted Fisher's exact test.

The broad functional categories encompassed by genes of each cluster were determined through enrichment of gene ontologies (GO biological processes). Significant enrichment is evaluated by Fisher's exact test which calculates the chance that the members of a GO term are found in a particular cluster of genes, given the background of the 3000 total genes represented by all 5 clusters. The top 10 enriched terms recapitulate the known anatomy (Figure 3-6) with terms associated with "somite development", "neuron fate commitment", "forebrain development" in the dorsal tissues; "oocyte differentiation", "urogenital system development", "nephric duct morphogenesis" in cluster 5, and haematopoietic signatures such as "positive regulation of inflammatory response", "cell chemotaxis" and "regulation of macrophage chemotaxis" in cluster 2 and 4 (i.e. E10.5 dorsal aorta tissues).

Selected gene ontologies enriched in cluster 4

GO.ID	Term	Annotated genes	p-value (weightFisher)	Genes contributing to enrichment
GO:0060638	mesenchymal-epithelial cell signaling	5	0.00537	Fgf10,Hgf,Tnc,Wnt2b
GO:0001706	endoderm formation	14	0.00829	Dusp1,Gata4,Gata6,Inhba,Nanog,Nog,Pou5f1
GO:0045766	positive regulation of angiogenesis	32	0.01032	Angpt2,C5ar1,Ccbe1,Ctsh,Gata4,Gata6,Hgf,Isl1,Prkcb,Ptk2b,Runx1,Srpx2
GO:0043534	blood vessel endothelial cell migration	13	0.02281	Angpt2,Egr3,Klf4,Nr4a1,Ptk2b,Srpx2
GO:0010758	regulation of macrophage chemotaxis	7	0.00355	C5ar1,Ccl2,Mmp28,Ptk2b,Rarres2
GO:0098542	defense response to other organism	62	0.00776	C5ar1,Cxcl1,Ddx60,Epx,Fgr,Gbp2,Iigp1,Il27ra,Irf5,Isg15,Lyz1,Mpo,Mx2,Naip6,Nlrc4,Oasl2,Oprk1,Spn,Spon2,Syk
GO:0032649	regulation of interferon-gamma productio...	10	0.02583	Fzd5,Il27ra,Isg15,Isl1,Runx1
GO:0071371	cellular response to gonadotropin stimul...	9	0.00222	Egr2,Egr3,Egr4,Gata4,Gata6,Inhba
GO:0019827	stem cell maintenance	30	0.01664	Dppa2,Esrrb,Fgf10,Klf4,Nanog,Nog,Phf19,Piwil2,Pou5f1,Prdm14,Tcl1
GO:0008584	male gonad development	26	0.02506	Esr2,Fgf9,Gata4,Inhba,Lhx9,Nupr1,Rxfp2,Tcf21,Tex19.1,Zfpm2
GO:0030510	regulation of BMP signaling pathway	30	0.01664	Bmper,Cav1,Chrdl2,Crb2,Fstl3,Gata4,Gata6,Gdf3,Kcp,Nanog,Nog

Table 5 Gene ontology enrichment of E10.5 AoV

Selected gene ontologies that are significantly enriched in cluster 4 ($p < 0.05$, Fisher's exact test) with particular relevance to haematopoiesis, and the genes that contribute to this enrichment.

As E10.5 AoV is the first site of HSC emergence in the mouse embryo, the pathways enriched in its associated gene cluster (Cluster 4) were focused on in more detail (Table 5). In this group, a number of terms associated with endothelial cells such as "positive regulation of angiogenesis" and "blood vessel endothelial cell migration" were enriched. This may be associated with a requirement for

remodelling of the endothelium as intra-aortic clusters form (Zovein et al., 2010). A number of pro-inflammatory signatures such as "regulation of macrophage chemotaxis", "defence response to other organism" and "regulation of interferon gamma-production" were significantly enriched in accordance with recent elucidation of interferon-gamma and pro-inflammatory signalling as important regulators of HSC development (Espín-Palazón et al., 2014; Li et al., 2014). As well as this, a number of terms associated with pluripotency/germ cells were enriched such as "cellular response to gonadotropin stimulation", "stem cell maintenance" and "male gonad development". These terms are likely to be the gene signature of primordial germ cells which are known to migrate through the ventral mesenchyme at E10.5 *en route* to the urogenital ridges (Medvinsky et al., 1993; Molyneaux et al., 2001; Yokomizo and Dzierzak, 2010). Finally a significant pathway of interest is "regulation of BMP signalling pathway". This is in support of recent studies of the role of BMP signalling and regulation by Noggin in HSC emergence (Souilhol et al., 2016). Thus the agreement of these signatures with published information about the regulation HSC development suggests that the expression signatures capture physiologically and functionally relevant information and therefore harnessing their high molecular detail is likely to enable elucidation of novel regulators of HSC development.

3.2.3 Gene set enrichment analysis captures the upregulation of TGF-β superfamily and pro-inflammatory signalling in AoV at E10.5

To gain a more focused understanding of the haematopoietic signalling changes during the E9.5 to E10.5 and dorsal-ventral transitions, comparative enrichment analysis of canonical signalling pathways was performed using Limma ROAST (Wu et al., 2010). In contrast to the pathway enrichment compared to a background (section 3.2.2), this method determines the probability that a set of genes is more highly expressed in one sample relative to another. The advantage of this method being that it is less sensitive to false positives as it does not assume independence between genes in terms of their expression; however it is limited to only pairwise comparisons between samples (Wu et al., 2010). The pathways tested have been previously implicated in the regulation of haematopoiesis and include:

inflammation, interferon and TNF (Espín-Palazón et al., 2014; Li et al., 2014), chemokines, cytokines, JAK/STAT, interleukins (Robin et al., 2006; Taoudi et al., 2008), insulin/insulin-like (Mascarenhas et al., 2009), NF-κB (González-Murillo et al., 2015), VEGF (Burns et al., 2009; Kabrun et al., 1997; Lugus et al., 2009), PDGF (Chhabra et al., 2012; Levéen et al., 1994; Soriano, 1994), Hedgehog (Durand et al., 2007; Peeters et al., 2009; Wilkinson et al., 2009), Notch (Gama-Norton et al., 2015; Hadland et al., 2015), SCF/Kit (Ding et al., 2012; Rybtsov et al., 2014), Wnt (Ruiz-Herguido et al., 2012; Sturgeon et al., 2014; Trompouki et al., 2011), TGF-β (Blank and Karlsson, 2015), BMP (Crisan et al., 2015; Durand et al., 2007; Souilhol et al., 2016), Nodal (Nostro et al., 2008; Pina et al., 2015; Sturgeon et al., 2014). All pathways were obtained from the Molecular Signature Database (Subramanian et al., 2005) (Table 6) and to ensure the search was comprehensive, all available annotations of these pathways were included from different databases such as KEGG (Kanehisa et al., 2016), BioCarta (Nishimura, 2001), PID (Schaefer et al., 2009), and Reactome (Croft et al., 2014; Fabregat et al., 2016).

Pathways
KEGG_NOTCH_SIGNALING_PATHWAY
KEGG_TGF_BETA_SIGNALING_PATHWAY
KEGG_VEGF_SIGNALING_PATHWAY
KEGG_JAK_STAT_SIGNALING_PATHWAY
BIOCARTA_ACE2_PATHWAY
BIOCARTA_CCR3_PATHWAY
BIOCARTA_CXCR4_PATHWAY
BIOCARTA_IGF1_PATHWAY
BIOCARTA_IL2_PATHWAY
BIOCARTA_IL3_PATHWAY
BIOCARTA_IL4_PATHWAY
BIOCARTA_IL5_PATHWAY
BIOCARTA_IL6_PATHWAY
BIOCARTA_IL7_PATHWAY
BIOCARTA_INSULIN_PATHWAY
BIOCARTA_NFKB_PATHWAY
BIOCARTA_PDGF_PATHWAY
BIOCARTA_CCR5_PATHWAY
BIOCARTA_SHH_PATHWAY
BIOCARTA_TNFR1_PATHWAY
BIOCARTA_TNFR2_PATHWAY
BIOCARTA_VEGF_PATHWAY
BIOCARTA_WNT_PATHWAY
PID_NOTCH_PATHWAY
PID_WNT_SIGNALING_PATHWAY
PID_BMP_PATHWAY
PID_HEDGEHOG_2PATHWAY
PID_HEDGEHOG_GLI_PATHWAY
PID_KIT_PATHWAY
PID_TGFBR_PATHWAY
REACTOME_SIGNALING_BY_WNT
REACTOME_SIGNALING_BY_SCF_KIT
REACTOME_SIGNALING_BY_NODAL
REACTOME_REGULATION_OF_KIT_SIGNALING
REACTOME_SIGNALING_BY_BMP
REACTOME_INTERFERON_GAMMA_SIGNALING
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
REACTOME_INTERFERON_SIGNALING
REACTOME_SIGNALING_BY_NOTCH
REACTOME_SIGNALING_BY_TGF_BETA_RECEPTOR_COMPLEX
IMMUNE_RESPONSE

Table 6 Gene sets selected for comparative enrichment in E10.5 AoV vs E9.5 AoV or E10.5 AoD

Canonical pathways from Molecular Signatures Database selected based on previous association with HSC development in literature for testing enrichment in E10.5 AoV compared to E9.5 AoV or E10.5 AoD.

Focusing on E10.5 AoV, a core subset of these pathways was identified which significantly increase (FDR < 0.2) in E10.5 AoV with respect to both the spatial and temporal transitions of the niche: SCF/Kit, Jak/Stat, BMP, VEGF, NF-κB, TGF-β, Hedgehog (Figure 3-7). Notably the enrichment of SCF/Kit in E10.5 AoV, is consistent with the understanding that SCF (Kitl) is an essential signalling factor in the niche (Rybtsov et al., 2014) and suggests that key regulators of the niche are captured in this data. To gain more detailed insight into the mode of signalling active in these tissues, the genes contributing to the enrichment of each pathway were identified as those which significantly change in the same direction as the enrichment (Figure 3-7B). A key contributing gene to SCF/Kit enrichment in E10.5 AoV is Kitl, which has previously been shown to be a supportive regulator of HSC development in E10.5 AoV (Souilhol et al., 2016a). The activation of hedgehog, is largely associated with the signal responsive elements of the pathway (i.e. Gli3), and notably Shh itself is highly expressed in AoD, agreeing with previous findings that Shh may be responsible for the inductive capacity of AoD in co-culture with AoV (Souilhol et al., 2016a).

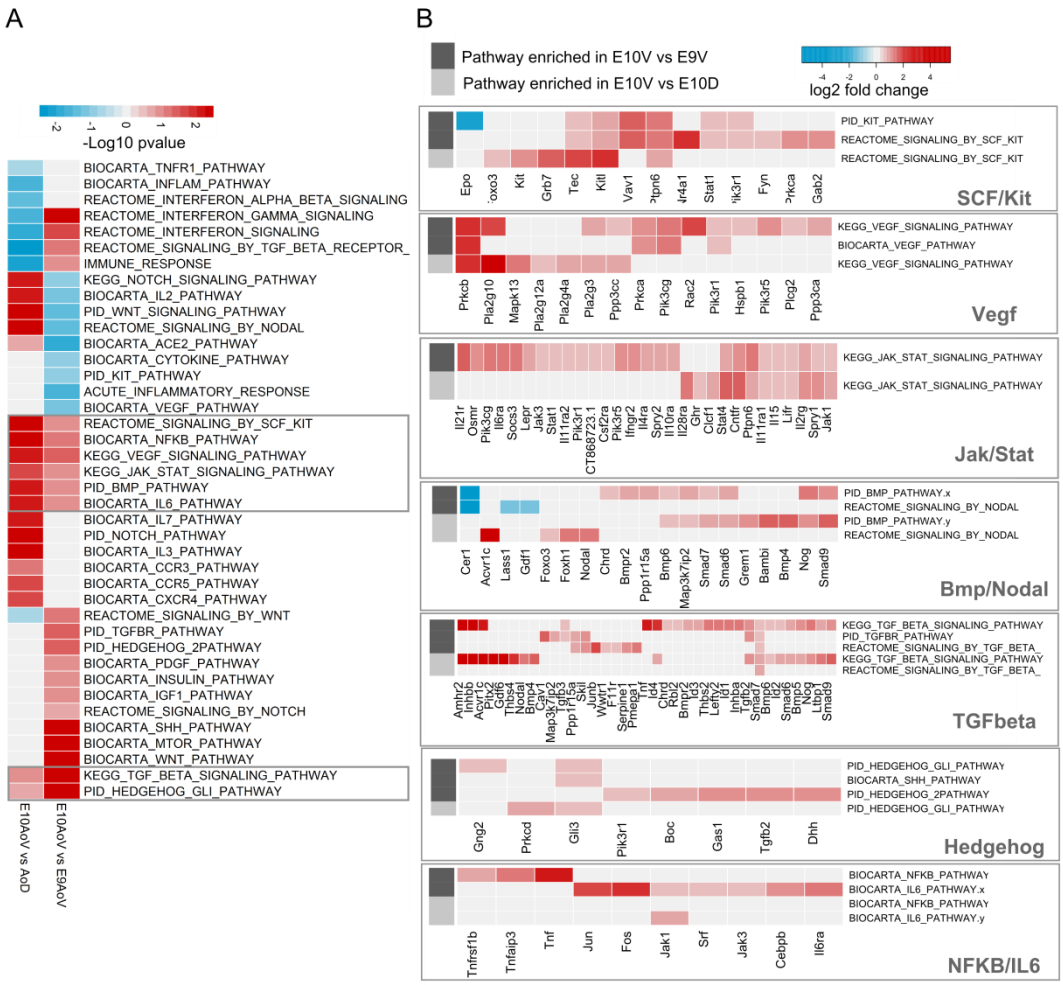


Figure 3-7 Comparative enrichment of haematopoiesis associated canonical pathways in E10.5 AoV vs E9.5 AoV or E10.5 AoD

(A) Enrichment of haematopoiesis-associated canonical pathways from the Molecular Signatures Database during the E9.5 to E10.5 and AoD to AoV transition. Colour scale represents -log10 multiple hypothesis corrected p-values. Boxes highlight the pathways significantly enriched in both E9.5 to E10.5 and AoD to AoV transition. (B) Genes contributing to enrichment of signalling pathways which are upregulated in both E10AoV vs E9AoV and E10AoV vs E10AoD. Heat map colour indicates either the fold change expression in E10.5 vs E9.5 (rows indicated by dark grey) or fold change expression in E10AoV vs E10AoD (rows indicated by light grey).

BMP and TGF-β superfamily signalling, which have previously been implicated in the regulation of HSC expansion and maturation (Crisan et al., 2015; Durand et al., 2007; Souilhol et al., 2016a), are significantly enriched, as was also

seen in Gene Ontology enrichment. A range of ligands such as Bmp5, Bmp6, Inhbb, Inhba, and Tgfb2 are enriched in E10AoV, suggesting that several branches of the TGF-β superfamily elicit effects in the HSC niche in addition to the canonical BMP4 pathway. Moreover the upregulation of Smad6, Smad7, Smad9 and Noggin in E10AoV vs E9AoV and E10AoV vs E10AoD as well as Chordin in E10AoV vs E9AoV, are indicative of potential site specific inhibitory state of BMP signalling as the AGM region develops.

The upregulated Jak/Stat, IL6 and NFKB pathways (Figure 3-7B) as well as enrichment for GO terms "macrophage chemotaxis" and "regulation of interferon-gamma production" (Figure 3-6B) suggest a significant pro-inflammatory signature within the ventral side of the AGM region. Jak/Stat signalling is a key axis in a number of cytokine pathways such as interferon signalling in macrophage activation (Schindler et al., 2007; Shuai and Liu, 2003; Stark and Darnell Jr., 2012), while IL6 and TNF are pro-inflammatory cytokines directly produced by macrophages (Aggarwal et al., 2012; Snick, 1990). This highlights pro-inflammatory signalling, potentially through activated macrophages, as a key activity within the niche and given the recently demonstrated effect of inflammatory signalling on HSC development (Espín-Palazón et al., 2014; Li et al., 2014; Sawamiphak et al., 2014) suggests that genes associated with these pathways may be involved in regulation of HSC maturation.

Thus comparative pathway enrichment provides additional detail to the gene ontology enrichment analysis in the elucidation of SCF, Jak/Stat, IL6 and NFKB pathways. But a common signature in both of these methods is the consistent enrichment of pro-inflammatory signalling and BMP/regulation of BMP signalling, suggesting that members of these pathways may be important to pursue for functional testing.

3.2.4 Evaluating the effect of sequencing depth on differential expression to guide further RNA-sequencing studies

The RNA from AGM tissues was sequenced at relatively high depth (50 million reads per sample) to ensure the transcriptional information was as complete

as possible. However it has been suggested that for differential expression analysis, such depth may not be necessary and indeed the stronger effect on sensitivity of identifying differentially expressed genes is the number of biological replicates (Liu et al., 2013). To explore this further, a lower depth of sequencing was simulated *in silico* by randomly sampling reads from the AGM RNA-seq dataset.

A

Number of differentially expressed gene in each comparison for depth 50M reads

vs	10V	10D	10U	9V	9D
10V	-	2646	2135	3110	4427
10D		-	2895	3443	3372
10U			-	3090	3753
9V				-	3059
9D					-

Number of differentially expressed gene in each comparison for depth 25M reads

vs	10V	10D	10U	9V	9D
10V	-	2331	2331	2748	4005
10D		-	2594	3002	3019
10U			-	2687	3334
9V				-	2734
9D					-

Number of differentially expressed gene in each comparison for depth 12.5M reads

vs	10V	10D	10U	9V	9D
10V	-	2037	1615	2359	3636
10D		-	2267	2581	2685
10U			-	2235	2866
9V				-	2355
9D					-

B

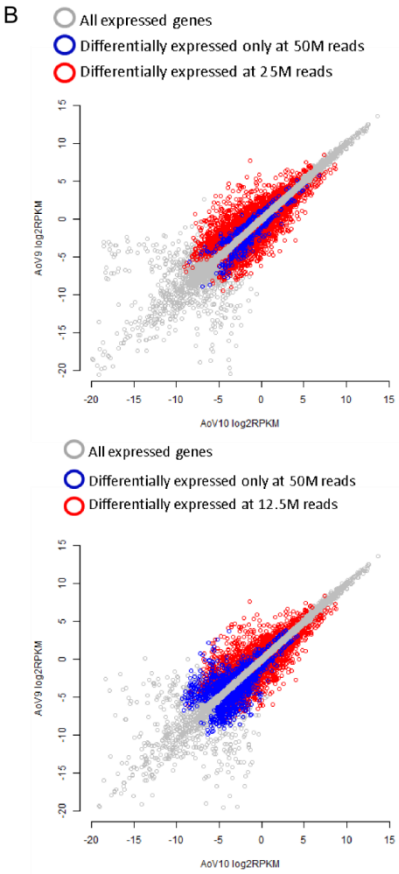


Figure 3-8 Comparison of differentially expressed genes in simulated sequencing depths of 50 million reads, 25 million reads and 12.5 million reads per sample

(A**)** Tables summarise the numbers of differentially expressed genes (absolute fold change ≥ 2 and FDR ≤ 0.2) in each pairwise comparison of tissues sequenced to a depth of 50 million reads, 25 million reads or 12.5 million reads. (B) Plots show E9AoV versus E10AoV log2 RPKM values overlaying all detected genes (grey), differentially expressed genes at 50 M reads/sample (blue), and differentially expressed genes in subsamples of 25 million reads/sample or 12.5 million reads per sample respectively (red).

A sequencing depth of 25 million reads/sample and 12.5 million reads/sample were simulated and aligned to the genome as with the full 50 million reads/sample. Differentially expressed genes were identified in pairwise comparisons of all samples

using an absolute fold change greater than 2 and FDR less than 0.2 as threshold (Figure 3-8A). Comparison of these sequencing depths demonstrates that the number of differentially expressed genes is not directly proportional to the depth of sequencing as even with a quarter of the sequencing depth approximately three quarters of the differentially expressed genes can be identified (Figure 3-8A). Notably the differentially expressed genes that are missed at lower sequencing depth are those expressed at a lower level, and those with a smaller differential expression change (Figure 3-8B).

A

Number of differentially expressed genes in E9V vs E10V

		Sequencing depth (million reads per sample)		
		50	**25**	**12.5**
Number of replicates	**3**	3163	2747	2359
	2	2643	2209	1833
	1	513	439	392

B

% loss of DE genes with loss of replicates

		Sequencing depth (million reads per sample)			
		50	**25**	**12.5**	**MEAN**
Change in replicate no.	**3 to 2**	16.4	19.6	22.3	19.4
	2 to 1	80.6	80.1	78.6	79.8

C

% loss of DE genes with loss of sequencing depth

		Number of replicates			
		3	**2**	**1**	**MEAN**
Change in seq depth	**50M to 25M**	13.2	16.4	14.4	14.7
	25M to 12M	14.1	17.0	10.7	13.9

Figure 3-9 Comparison of the impact of sequencing depth and biological replication of DE genes

(A) Table shows the number of differentially expressed (DE) genes between E9.5 AoV and E10.5 AoV which have an absolute fold change ≥ 2 and FDR ≤ 0.2 for simulated sequencing depths. The number was calculated for three simulated sequencing depths and different numbers of biological replicates (value shows the mean number of DE genes for all combinations of biological replicate comparisons). (B) Table shows, for each sequencing depth, the loss of DE genes, as a percentage of the total, as the number of replicates is reduced. (C) Shows for each number of replicates, the loss of DE genes, as a percentage of the total, when sequencing depth is halved.

A comparison was then made of the impact of sequencing depth, compared to the impact of the number of biological replicates on the number of statistically significantly differentially expressed genes. In all cases the loss of one biological replicate resulted in fewer differentially expressed genes than the loss of half of the sequencing depth. For example, when the number of biological replicates was reduced from 3 to 2, 19.4% of the genes were lost as opposed to only 14.7% lost when the sequencing depth was halved. Furthermore there was a dramatic loss of

80% of the differentially expressed genes when the number of biological replicates was reduced from 2 to 1.[6] Thus the power of differential expression analysis does not show a linear relationship with the number of biological replicates or the sequencing depth and this has been explored further in (Liu et al., 2013). Consequently, the conclusion drawn from these results was that to reduce costs in future RNA-sequencing experiments, a reduction in depth would likely be less detrimental to statistical power than a reduction in replication consistent with the findings of (Liu et al., 2013).

3.2.5 An RNA-seq dataset from OP9 cells reveals a major transcription change in reaggregate culture conditions compared to submersed flat culture

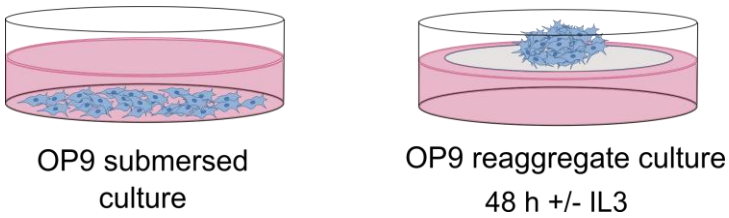


Figure 3-10 Generation of RNA-seq data from OP9 cells in different culture conditions

Transcriptome data was generated from OP9 cells cultured in two conditions: left shows flat submersed culture on plastic; right shows OP9 cells cultured at the air-liquid interface for 48h after reaggregation. Cells cultured in reaggregate conditions were cultured either with or without IL3.

[6] Without replicates most statistical models cannot judge the significance of a change. DESeq can produce an FDR value with single replicate, using the mean for genes with comparable expression levels to estimate dispersion, however this is likely to have limited reliability.

Sample name	Cell type	Passage number	Conc. of IL3 (μg/μl)	Culture method	RNA volume (μl)	RNA conc. with nanodrop (μg/ul)	Total RNA (μg)
R1	OP9	p9	-	Reaggregate	50	0.38	19
R2	OP9	p23	-	Reaggregate	50	0.17	8.5
R3	OP9	p14	-	Reaggregate	50	0.73	36.5
R1L	OP9	p9	200	Reaggregate	50	0.408	20.4
R2L	OP9	p23	200	Reaggregate	50	0.148	7.4
R3L	OP9	p14	200	Reaggregate	50	0.44	22
S1	OP9	p9	-	Submersed	60	1.36	81.6
S2	OP9	p14	-	Submersed	30	2	60
SM1	OP9 (de Bruijn lab)	p10	-	Submersed	30	0.3	9
SM2	OP9 (de Bruijn lab)	p10	-	Submersed	30	0.4	12

Table 7 Preparation of RNA from OP9 cells for RNA sequencing

Columns indicate: sample abbreviations, cell type (stated when from de Bruijn lab), passage number used, concentration of IL3 ligand if added, culture method either "Reaggregate" meaning aggregation of cells by centrifugation and then culture on a membrane at the air-liquid interface for 48h, RNA volume, the estimated concentration by spectrophotometry and the estimated total weight of RNA produced for input in sequencing library preparation.

The OP9 stromal cell line has been shown to support maturation of HSCs *ex vivo* when co-cultured with either sorted pre-HSCs or whole AGM tissues containing early precursors (Rybtsov et al., 2011, 2014; Souilhol et al., 2016). This cell line is derived from calvaria of new-born op/op mice which lack macrophage colony stimulating factor (M-CSF) (Kobayashi et al., 1996; Nakano et al., 1994). It was hypothesised that transcriptional profiling of this cell line may yield further key molecular signatures that confer support of HSC maturation. Importantly, support of HSC maturation is greatly enhanced when cells are cultured as a reaggregates on a membrane at the air-liquid interface of the culture media (Rybtsov et al., 2011; Taoudi et al., 2008), compared to submersed flat culture. Therefore to elucidate the difference between these conditions, as well as generate a transcriptional profile of OP9 cells, RNA was extracted and sequenced from OP9 cells that had been grown in flat, submersed conditions and after reaggregation followed by culture on a membrane for 48h. As well as this IL3, which has been shown to regulate HSC development, was added to some of the reaggregated cells to try and elucidate its stromal effect. Finally, due to known variations in cell lines between labs, a comparison was made between cells from Medvinsky lab stocks which are capable of supporting HSC maturation from pro-HSCs, Type I and Type II pre-HSCs

(Rybtsov et al., 2011, 2014) and from OP9 cells from the lab of Professor Marella de Bruijn (Figure 3-10 and Table 7).

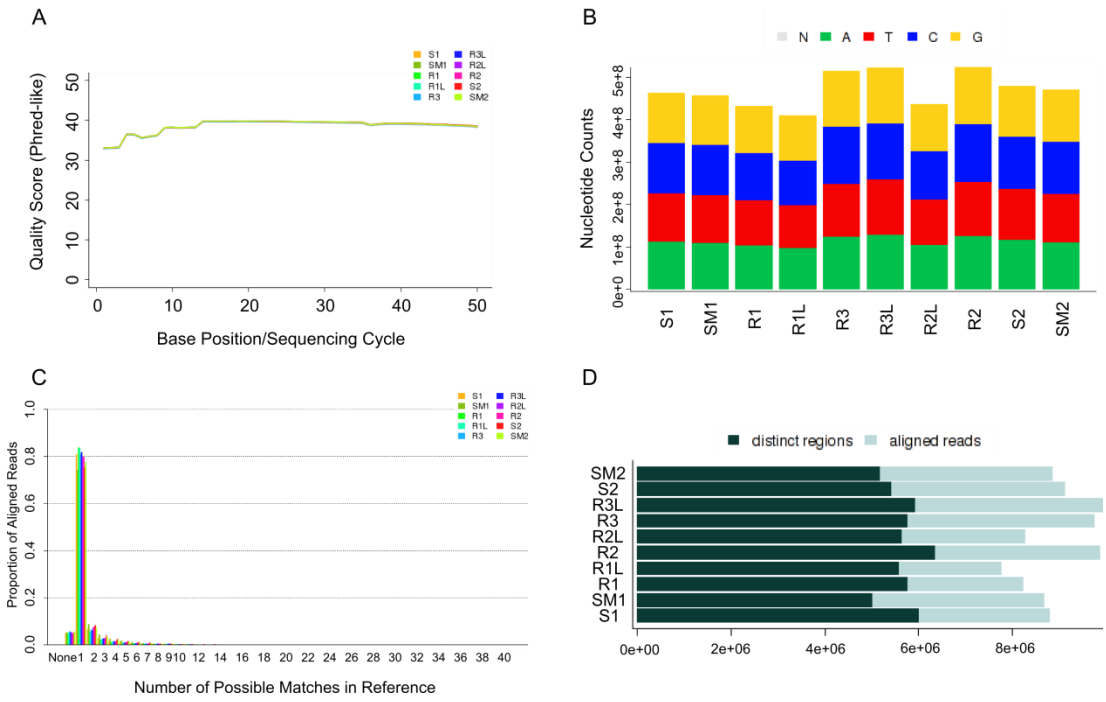


Figure 3-11 OP9 RNA-seq quality control and alignment results

As in (Figure 3-2) these plots are based on Sequence Data and Statistics plots from GeneProf. (A) Plots the mean quality score of all reads in each samples for each base position. (B) Shows the number of counts of each nucleotide (A, T, C, G) in each sample as a measure of distribution of nucleotides. (C) Shows the proportion of reads which do not align to the genome, align to one place, or ambiguously align at more than one place in the genome. (D) Represents the number of distinct genomic regions to which a read aligns compared to the total number of aligned reads.

Sequenced reads were aligned as described above with the RNA-sequencing pipeline in GeneProf (Halbritter et al., 2012). The nucleotide sequence quality scores > 20, balanced nucleotide composition and read alignment ~ 80% (Figure 3-11) suggest that sequencing was of good quality and ready for downstream analysis according the parameters described above (section 3.2.1). Comparison of the full transcriptional profiles of each sample by correlation and PCA (Figure 3-12) suggest that culture method (i.e. whether cells were cultured in flat submersed conditions or at the air liquid interface after reaggregation) confers the major transcriptional

difference between samples. With the exception of one of the flat OP9 samples, hierarchical clustering based on correlation distance partitions samples into two major clusters of either flat or reaggregated samples (Figure 3-12A). Similarly, the first principle component, which accounts for > 40% of the variation in the data, separates samples by these different culture conditions (Figure 3-12B). PCA suggests that there also notable differences between OP9 cells from the Medvinsky lab and from the de Bruijn lab as the second principle component (~20% of the variance in the data) separates the submersed OP9 samples by lab (Figure 3-12B). Contrastingly, the treatment with IL3 conferred no discernible difference in the transcriptional profiles.

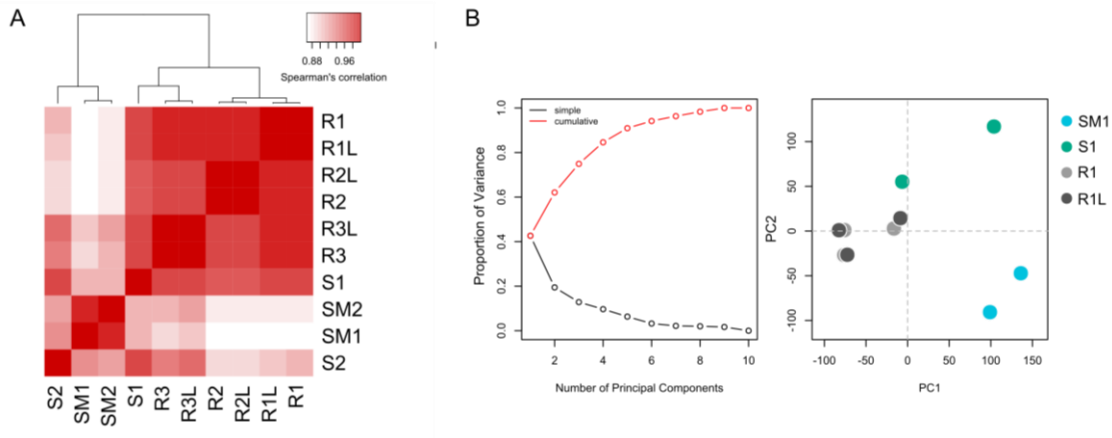
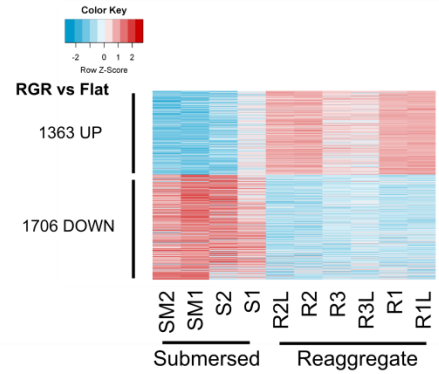


Figure 3-12 Comparison of transcriptional profiles of OP9 cells in different culture conditions and from different sources

(A) Clustering of each OP9 RNA-seq sample with distance based on Spearman's ranked correlation. Heat map colouring displays the correlation level between each sample. (B) Principle component analysis of OP9 samples based on normalised expression values including the proportion of total variance constituted by each component, and the distance between each sample based on the first two components. SM = OP9 cells from de Bruijn lab cultured in submersed conditions; S = OP9 cells from Medvinsky lab cultured in submersed conditions; R = OP9 cells from Medvinsky lab cultured in reaggregate conditions for 48 h; RL = OP9 cells from Medvinsky lab cultured in reaggregate conditions for 48 h and treated with IL3.

Focusing on the condition that gives the major functional and transcriptional difference, a comparison of differentially expressed genes in flat culture and reaggregate cultured OP9 cells was performed with DESeq with the threshold for absolute fold change greater than 2 and FDR less than 0.05. A large number of genes were significantly up-regulated (1363) and down-regulated (1706) between flat and reaggregate culture conditions (Figure 3-13A). The up-regulated genes show notable enrichment for Notch signalling and Integrin canonical pathways (Figure 3-13B). Although the cell-autonomous role of Notch is shown to be essential for early stages of pre-HSC maturation but is down-regulated in later precursors (Hadland et al., 2004; Souilhol et al., 2016b), the signalling enrichment in OP9 cells indicates a potential role for Notch in inducing a supportive transcriptional program in the stroma. Downregulated genes encompass a number of cell cycle genes (Figure 3-13C) which is in line with observed arrest of cell division in reaggregates in comparison to flat OP9 culture (data not shown).

A



B

Pathway	NGenes	PropDown	PropUp	Direction	PValue	FDR
KEGG_LYSOSOME	112	0.1070	0.589	Up	0.001	0.00627
PID_INTEGRIN3_PATHWAY	38	0.1050	0.447	Up	0.001	0.00627
REACTOME_IRON_UPTAKE_AND_TRANSPORT	32	0.1560	0.469	Up	0.001	0.00627
REACTOME_PIP3_ACTIVATES_AKT_SIGNALING	25	0.1600	0.520	Up	0.001	0.00627
REACTOME_SIGNALING_BY_NOTCH3	10	0.0000	0.800	Up	0.001	0.00627
KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_GLOBO_SERIES	14	0.0714	0.286	Up	0.001	0.00627
KEGG_RIBOFLAVIN_METABOLISM	13	0.1540	0.385	Up	0.001	0.00627
REACTOME_P130CAS_LINKAGE_TO_MAPK_SIGNALING_FOR_INT	14	0.0714	0.429	Up	0.001	0.00627
BIOCARTA_PLATELETAPP_PATHWAY	13	0.0000	0.462	Up	0.001	0.00627
REACTOME_GLYCOSPHINGOLIPID_METABOLISM	30	0.2000	0.400	Up	0.002	0.01450
REACTOME_KERATAN_SULFATE_DEGRADATION	10	0.2000	0.600	Up	0.002	0.01450
REACTOME_NRIF_SIGNALS_CELL_DEATH_FROM_THE_NUCLEUS	10	0.1000	0.700	Up	0.002	0.01450
REACTOME_A_TETRASACCHARIDE_LINKER_SEQUENCE_IS_REQU	22	0.0455	0.545	Up	0.002	0.01450
REACTOME_HS_GAG_DEGRADATION	19	0.0000	0.421	Up	0.002	0.01450
REACTOME_SIGNALING_BY_NOTCH2	10	0.0000	0.800	Up	0.002	0.01450
REACTOME_RECEPTOR_LIGAND_BINDING_INITIATES_THE_SEC	9	0.0000	0.889	Up	0.002	0.01450
REACTOME_SIGNALING_BY_NOTCH4	10	0.0000	0.800	Up	0.002	0.01450
BIOCARTA_GABA_PATHWAY	9	0.0000	0.222	Up	0.002	0.01450
REACTOME_BMAL1_CLOCK_NPAS2_ACTIVATES_CIRCADIAN_EXP	28	0.0000	0.679	Up	0.003	0.02090

C

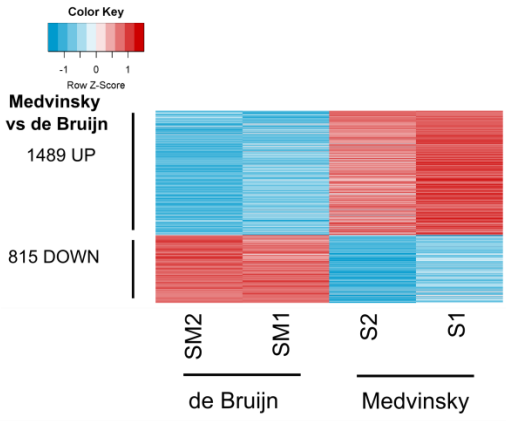
Pathway	NGenes	PropDown	PropUp	Direction	PValue	FDR
REACTOME_METABOLISM_OF_PROTEINS	300	0.450	0.1770	Down	0.001	0.00627
REACTOME_CELL_CYCLE_MITOTIC	272	0.647	0.0919	Down	0.001	0.00627
REACTOME_MITOTIC_M_M_G1_PHASES	151	0.702	0.0596	Down	0.001	0.00627
REACTOME_CELL_CYCLE_CHECKPOINTS	105	0.705	0.0571	Down	0.001	0.00627
REACTOME_G1_S_TRANSITION	97	0.742	0.0412	Down	0.001	0.00627
REACTOME_SYNTHESIS_OF_DNA	86	0.779	0.0465	Down	0.001	0.00627
KEGG_PYRIMIDINE_METABOLISM	79	0.544	0.1270	Down	0.001	0.00627
REACTOME_MITOTIC_PROMETAPHASE	73	0.616	0.0959	Down	0.001	0.00627
REACTOME_METABOLISM_OF_NUCLEOTIDES	58	0.534	0.0862	Down	0.001	0.00627
KEGG_GLYCOLYSIS_GLUCONEOGENESIS	47	0.383	0.1060	Down	0.001	0.00627
KEGG_ARGININE_AND_PROLINE_METABOLISM	47	0.340	0.1060	Down	0.001	0.00627
REACTOME_PROTEIN_FOLDING	45	0.578	0.2220	Down	0.001	0.00627
PID_FANCONI_PATHWAY	44	0.614	0.1360	Down	0.001	0.00627
PID_PLK1_PATHWAY	43	0.628	0.1400	Down	0.001	0.00627
REACTOME_TELOMERE_MAINTENANCE	41	0.683	0.0976	Down	0.001	0.00627
REACTOME_G2_M_CHECKPOINTS	41	0.805	0.0000	Down	0.001	0.00627
PID_ATR_PATHWAY	39	0.590	0.1030	Down	0.001	0.00627
PID_FOXM1_PATHWAY	38	0.500	0.2890	Down	0.001	0.00627
REACTOME_MITOCHONDRIAL_PROTEIN_IMPORT	38	0.632	0.1320	Down	0.001	0.00627
REACTOME_TRNA_AMINOACYLATION	36	0.639	0.0278	Down	0.001	0.00627

Figure 3-13 Differential expression analysis between OP9 culture conditions

(A) Shows the differentially expressed (DE) genes between OP9 cells cultured in submersed conditions and cultured in reaggregate conditions for 48h. DE genes are defined as having fold change ≥ 2 and FDR ≤ 0.05. (B) The top 20 significantly enriched pathways (calculated by Limma ROAST) in the genes up-regulated after reaggregation and (C) down-regulated after reaggregation. Pathways are from the entire Molecular Signatures Database canonical pathways dataset.

To explore the transcriptional differences between cell lines from different labs, the genes differentially expressed between OP9 cells (of two different passages) from Medvinsky lab and from de Bruijn lab were compared (Figure 3-14A). A number of genes show significantly increased expression (1489 genes ≥ 2-fold, FDR ≤ 0.05) and decreased (815 genes < 0.5-fold, FDR < 0.05). The significantly differentially expressed genes between Medvinsky and de Bruijn labs include the VEGF pathway, although the enrichment analysis as limited by the low number of replicates.

A



B

Pathway	NGenes	PropDown	PropUp	Direction	PValue	FDR
PID_VEGF_VEGFR_PATHWAY	10	0.10	0.400	Up	0.001	0.166
REACTOME_BINDING_AND_ENTRY_OF_HIV_VIRION	2	0.00	0.500	Up	0.001	0.166
KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450	46	0.13	0.348	Up	0.001	0.166
REACTOME_HORMONE_SENSITIVE_LIPASE_HSL_MEDIATED_TRI	10	0.30	0.000	Down	0.001	0.166

Figure 3-14 Differentially expressed genes in OP9 cell lines from different labs

(A) Shows the differentially expressed (DE) genes between OP9 cells stocks from the Medvinsky lab compared to stocks from the de Bruijn lab. DE genes are defined as having fold change ≥ 2 and FDR ≤ 0.05. (B) The significantly enriched pathways in comparisons of the two OP9 cell lines. Pathways are from the Molecular Signatures Database canonical pathways (Subramanian et al., 2005).

3.2.6 Comparison of the expression profiles of the *in vitro* and *in vivo* niches supportive of HSC development identifies a common molecular signature

As OP9 cells provide a supportive niche for HSC maturation as well as the primary cells of the AGM region, the relationship between these two niches was

explored in further detail. Given the greater support for HSC maturation from reaggregated OP9 cells compared to flat cultured OP9 cells, this specific transcriptional profile (i.e. the 1363 significantly up-regulated genes) was compared to the gene-clusters representing each AGM subdomain (in section 3.2.2, Figure 3-2). A hypergeometric test of the overlap between these signatures, demonstrated that only Cluster 4 (which had shown significant association with E10.5 AoV) shows a significant overlap ($p < 0.05$) with the transcriptional profile of reaggregated OP9 cells (Figure 3-15A). This therefore suggests that reaggregated OP9 cells more closely resemble the supportive *in vivo* environment for HSC maturation at E10.5 AoV, rather than the earlier E9.5 or E10.5 AoD environment. Thus these two transcriptome datasets are able to link two independent niche environments for HSC maturation. Moreover, this overlap signifies a common molecular program between OP9 cells and E10AoV (Figure 3-15B, Table 8) which includes the key haematopoietic transcriptional regulator, Runx1; several other molecules associated with lymphoid regulation: Mme (also known as acute lymphocytic leukaemia antigen, CD10), Egr2, Egr3 (Li et al., 2012); a number of structural proteins: Col12a1, Elastin, Matn2, Thbs2; signalling molecules associated with neural development: Ntrk3, protocadherins, Ror1; and a modulator of BMP signalling: Bmper. The expression of Runx1 in OP9 cells is in-line with previous findings that Runx1 is found in non-haematopoietic cells layers of the AGM region (North et al., 1999) and highlights a potential non-cell autonomous role in the regulation of HSC development. The upregulation of genes of the protocadherin-beta cluster (Pcdhb 6, 7, 8, 15, 16, 21) was notable in this signature. Protocadherins have previously been shown to engage in homophilic interactions and regulate neuronal development (Chen and Maniatis, 2013), but their role in OP9 cells or the AGM region remains unclear.

A

Cluster	Cluster size	Overlap	Overlap proportion	p-value
cluster1	379	22	0.058	0.9420
cluster2	777	71	0.091	0.1150
cluster3	676	34	0.050	0.9990
cluster4	597	63	0.106	0.0127
cluster5	566	43	0.076	0.6200

B

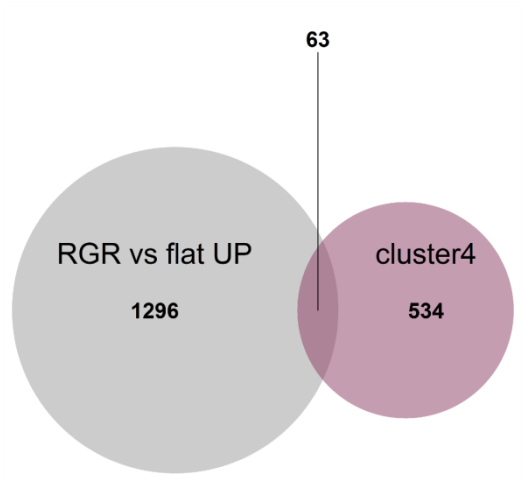


Figure 3-15 Comparison of OP9 and AGM region expression profiles

(A) Overlap of gene members of each AGM cluster, with the 1363 up-regulated genes in OP9 after reaggregation. Rows highlighted in pink indicate significant overlaps, with p-values < 0.05 as determined by hypergeometric test. 'Cluster' refers to the AGM gene clusters determined in section 3.2.2, Figure 3-2; 'Cluster size' is the number of genes belonging to each cluster; 'Overlap' is the number of genes in the overlap between each cluster and the 1363 upregulated genes in RGR OP9; 'Overlap proportion' refers to the proportion of each gene cluster in overlap; 'p-value' is determined by hypergeometric test. (B) Overlap between AGM cluster 4 and the 1363 genes up-regulated in OP9 after reaggregation.

Ensembl.Gene.ID	Name	log2FC 10Vvs9V	log2FC RGRvsFlat
ENSMUSG00000052026	Slc6a7	3.91	2.04
ENSMUSG00000053675	Tgm5	3.76	1.23
ENSMUSG00000059146	Ntrk3	3.61	2.63
ENSMUSG00000031963	Bmper	3.59	2.41
ENSMUSG00000050587	Lrrc4c	3.41	1.55
ENSMUSG00000051678	Pcdhb6	2.97	5.1
ENSMUSG00000033730	Egr3	2.93	3.44
ENSMUSG00000033715	Akr1c14	2.81	1.31
ENSMUSG00000027875	Hmgcs2	2.76	3.19
ENSMUSG00000032332	Col12a1	2.7	1.59
ENSMUSG00000027820	Mme	2.52	3.18
ENSMUSG00000028565	Nfia	2.44	1.14
ENSMUSG00000047033	Pcdhb15	2.33	1.44
ENSMUSG00000063687	Pcdhb5	2.13	3.01
ENSMUSG00000026628	Atf3	2.11	3.2
ENSMUSG00000047910	Pcdhb16	2.1	1.79
ENSMUSG00000045876	Pcdhb8	2.02	5.03
ENSMUSG00000023034	Nr4a1	2.01	3.28
ENSMUSG00000045062	Pcdhb7	1.96	2.46
ENSMUSG00000029135	Fosl2	1.94	1.1
ENSMUSG00000038354	Ankrd35	1.83	2.4
ENSMUSG00000037868	Egr2	1.83	1.81
ENSMUSG00000051242	Pcdhb9	1.81	3.03
ENSMUSG00000070867	RP23-309K20.1	1.77	1.31
ENSMUSG00000067283	Hspa1a	1.76	2.33
ENSMUSG00000043313	Pcdhb19	1.74	2.84
ENSMUSG00000051486	Pcdhb11	1.7	2.59
ENSMUSG00000044043	Pcdhb14	1.69	1.58
ENSMUSG00000035305	Ror1	1.68	2.75
ENSMUSG00000046387	Pcdhb17	1.63	1.45
ENSMUSG00000029675	Eln	1.62	6.93
ENSMUSG00000027827	Kcnab1	1.61	1.36
ENSMUSG00000022324	Matn2	1.55	2.28
ENSMUSG00000022952	Runx1	1.46	1.45
ENSMUSG00000044022	Pcdhb21	1.46	1.48
ENSMUSG00000045903	Npas4	1.46	1.58
ENSMUSG00000023885	Thbs2	1.43	1.12
ENSMUSG00000030711	Sult1a1	1.41	2.2
ENSMUSG00000055435	Maf	1.41	2.41
ENSMUSG00000075047	AC129601.1	1.34	2.72
ENSMUSG00000051041	Olfml1	1.09	5.79
ENSMUSG00000032179	Bmp5	1.06	4.86
ENSMUSG00000060224	Pyroxd2	1.04	1.73
ENSMUSG00000035184	AC154516.1	1.03	2.44
ENSMUSG00000086514	AL591075.2	0.977	1.16
ENSMUSG00000028214	Gem	0.976	2.27
ENSMUSG00000020122	Egfr	0.863	1.85
ENSMUSG00000078816	Prkcc	0.852	1.93
ENSMUSG00000067780	Pi15	0.775	1.9
ENSMUSG00000036545	Adamts2	0.757	1.45
ENSMUSG00000059921	Unc5c	0.749	1.51
ENSMUSG00000054555	Adam12	0.74	2
ENSMUSG00000000126	Wnt9a	0.716	1.94
ENSMUSG00000029217	Tec	0.704	2.92
ENSMUSG00000028487	Bnc2	0.679	1.79
ENSMUSG00000069920	B3gnt9	0.632	1.2
ENSMUSG00000006586	Runx1t1	0.313	2.95
ENSMUSG00000012428	Steap4	0.313	4.35
ENSMUSG00000031125	3830403N18Rik	0.0857	1.33
ENSMUSG00000046318	Ccbe1	0.0331	1.83
ENSMUSG00000023341	Mx2	-0.0296	2.31
ENSMUSG00000030717	Nupr1	-0.069	1.57
ENSMUSG00000079484	Phyhd1	-0.284	1.18

Table 8 Intersection between Cluster 4 (E10AoV) and genes up-regulated in reaggregate conditions

All genes in intersection between Cluster 4 (E10AoV) and genes up-regulated in reaggregate conditions. Ranked by log2 fold-change expression in E10 AoV vs E9 AoV, and including the log2 fold-change expression in OP9 between RGR and flat culture conditions.

3.2.7 An interactive visualisation enables convenient browsing of the AGM and OP9 data resources

3.2.7.1 The need for simple interface to RNA-seq data

RNA-seq experiments produce vast amounts of data – as was the case for the data generated in this thesis – which can be challenging to handle for a variety of reasons. Firstly, the large volume of data has heavy computational requirements, and even a relatively simple action such as opening the file containing normalised read counts can be difficult on a low power computer. Secondly, the amount of data likely exceeds the functional validation that can be performed in a single project. Therefore it may be useful to be able to refer back to the data at a later date, and even perform basic comparisons based on new hypotheses. For example, in this thesis the main focus was selection of genes which increase in the AGM region between E9.5 and E10.5 (see section 4.2.2). However others may be interested in the genes highly expressed in E10.5 AoD compared to E10.5 AoV. If researchers with little bioinformatics experience and insufficient computational power wish to return to the data this can therefore be challenging if they are only given raw data files of FASTQ, SAM (Sequence Alignment/Map) or even normalised read count files. To ensure this data would be openly available for anyone interested in the biological system in the future, an interactive visualisation of the data was produced which could be accessed via any web browser.

3.2.7.2 An interactive visualisation of samples, expression data and tools for data-browsing

A web-based visualisation was generated to enable researchers with little bioinformatics experience and potentially low computational power to easily access the RNA-seq data generated in this thesis. The overall design includes an image of the sample origins providing a quick prompt as to the experimental design; a data-table including the normalised expression counts (RPM) values for every gene detected in every sample and some pre-computed fold change values between sample groups; a panel on the right displays the tools to enable filtering of this data table (Figure 3-16). The user can interact with the image, selecting one or more samples, and then apply actions such as a minimum expression level, selection of genes of specific functional characteristics, or alternatively select a pre-computed cluster of genes; all of which will result in a filtering of the data table. The final filtered table can be copied, or downloaded as an excel file. The data table can also be searched to find expression information on specific genes the data table was linked to two databases, Ensembl (Yates et al., 2016) and the Mouse Atlas Project (Richardson et al., 2013) enable further gene-specific information to be quickly (Figure 3-17).

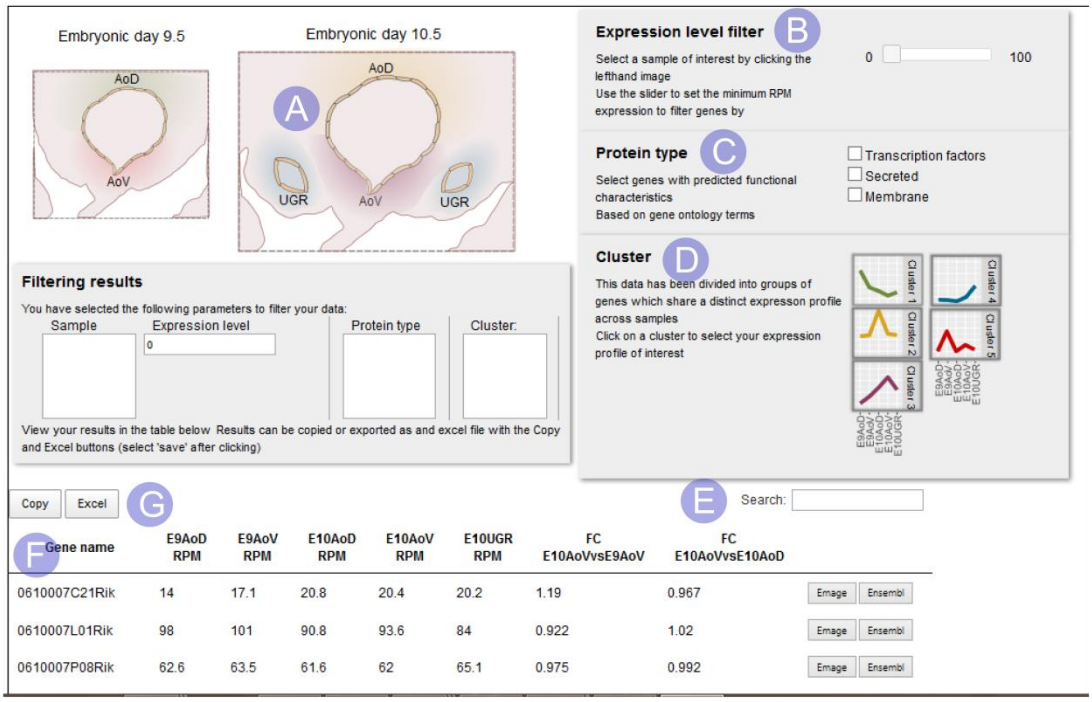


Figure 3-16 Interactive visualisation for exploring transcriptional data

(A) An image representing the experimental set-up which can be clicked to select one or more samples. (B) After a sample has been selected, the minimum expression level in that sample can be selected with a filter. (C) Genes in the selected sample can be selected by the type of protein encoded. (D) A pre-computed cluster of genes can be selected. (E) Typing the gene MGI symbol or Ensembl Gene ID to identify the specific gene. (F) The output of all selections/filtering will be displayed in the table, which can be ordered by each row. (G) After filtering the table can be copied to the clipboard, or exported as an excel file.

3.2.7.3 Example use case

The utility of this resource can be shown through a hypothetical use case. For example, a researcher may ask the question: which transcription factors which are highly expressed in E10.5 AoD compared to E10.5 AoV? They first click on sample AoD and it will become highlighted in yellow. They then set a minimum expression level with the slider to select all genes with RPM > 50. To select transcription factors they use the "Transcription factors" tick-box under "protein type". The data table has now been filtered, but results are displayed in alphabetical order. To rank the results based on the differential expression between E10.5 AoD and E10.5 AoV they click on the column title "FC E10AoVvsE10AoD". The first result is Pax1. Clicking on

the links to Ensembl and EMAGE quickly takes the user to more information on this gene and its previously described pattern of expression in the mouse embryo. Finally this smaller subset of approximately 200 genes can be copied to the clipboard, or downloaded as an excel file.

This demonstrates a quick process from the initial question to the answer. The common situation is that a researcher may have access to a text file, containing the normalised read counts and possible fold change and differential expression p-values. To follow the same example of identifying highly expressed transcription factors in AoD the researcher would have to open this file in Excel or a text editor, which requires a significant amount of memory for a file of >28,000 lines. They must be familiar with using filters in excel or a text editor. Then, for genes of interest, open a web browser and suitable database to copy individual names across and search for further information. The visualisation here has been designed with common actions in mind, to reduce the time a research needs to obtain useful, potentially novel, information from the RNA-seq data.

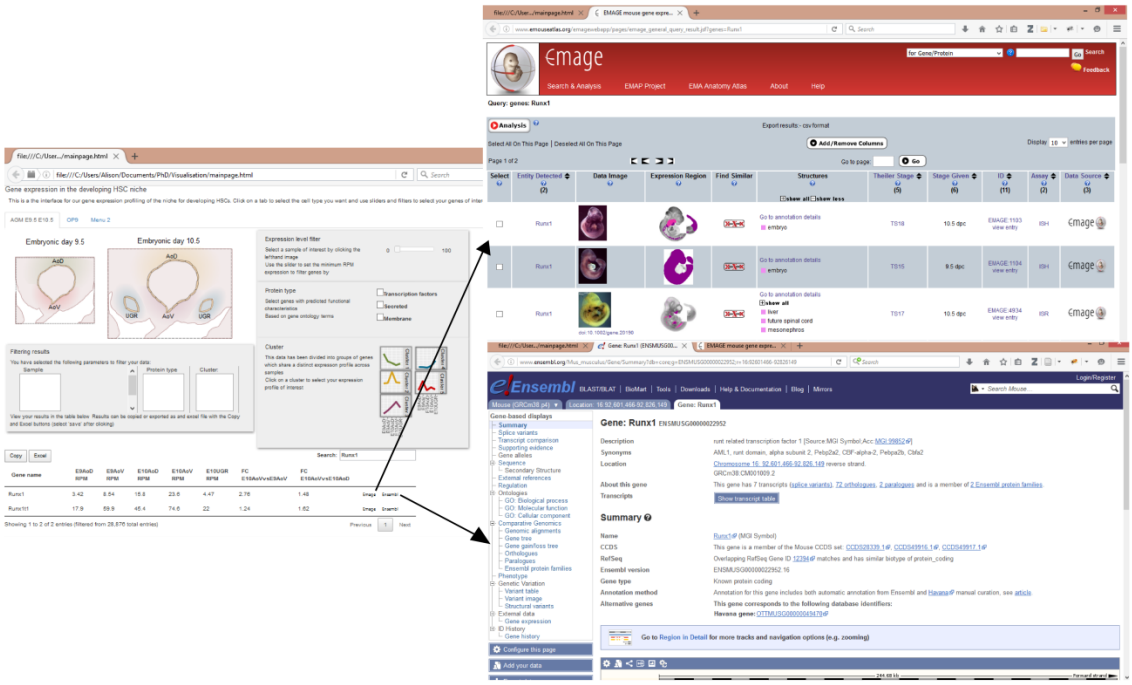


Figure 3-17 Linking transcriptional data with public resources

For a gene of interest, further information can be ascertained by clicking on the buttons titled “Emage” or “Ensembl” in the corresponding column of the table, which will link out the specific information for that gene in the Mouse Atlas Project of Ensembl web pages.

3.3 Discussion

The functional role of the AGM region in HSC maturation has previously been described in detail, demonstrating the important acquisition of an autonomous supportive environment between E9.5 and E10.5 as well as the key ventral polarity (Rybtsov et al., 2014, 2016; Souilhol et al., 2016a). Here a model of the full transcriptional landscape of these spatio-temporal transitions in the developing AGM region was produced by RNA-sequencing of precisely sub-dissected domains. Using these data it was possible to identify the gene signatures (i.e. those genes uniquely expressed) of each tissue which may underlie the functional differences in supportive activity.

These transcriptional profiles verify many previous observations about the molecular landscape of the AGM region, such as the ventral polarisation of Runx1, SCF, BMP4 as well as the dorsal polarisation of Shh (Durand et al., 2007; North et al., 1999; Souilhol et al., 2016a). However, the depth of this type of transcriptional approach provided a fuller picture of the signalling environment than has previously been possible. For example, the ventral polarisation of BMP family ligands is not limited to Bmp4, but also encompasses Bmp5, Bmp6, Inhbb, Inhba, Gdf3 and Tgfb2 as well as inhibitors of Smad6, Smad7, Smad9, Noggin, Chrdl2 and Bmper. Furthermore, a significant pro-inflammatory signature – encompassing IL6, TNF and Jak/Stat – was shown here to be unique to the E10.5 AoV environment and therefore their associated cytokines may potentially regulate HSC development. Pro-inflammatory signalling has previously been described in the environment for HSC maturation (Espín-Palazón et al., 2014; Li et al., 2014; Sawamiphak et al., 2014), but it has been shown here it is specifically enriched the E10.5 murine dorsal aorta. This signalling characterisation suggests that the HSC precursor niche is a complex environment of multiple cues, and functional validation is required to clarify which are the crucial genes important for supporting HSC maturation.

As well as the *in vivo* AGM environment, it has previously been established that OP9 cells can provide a high degree of support for HSC precursor maturation in *ex vivo* culture (Rybtsov et al., 2011). Thus the transcriptional profiling was extended to this key cell line and compared with that of the *in vivo* AGM region. The comparison of cells cultured in submersed flat conditions and those cultured as

reaggregates, demonstrates that the method of culture has a significant impact on the transcriptional output of this cell line. Moreover the reaggregate culture method used for HSC maturation assays bears a closer resemblance to E10.5 AGM (and specifically E10.5 AoV) than the conventional flat submersed culture. Indeed the transcriptional signature shared by reaggregated OP9 cells and E10.5 AoV includes the known HSC regulator, Runx1.

As these two datasets were considered to be a useful resource for the community, an interactive graphical interface to the RNA-seq data was constructed to facilitate future use of the data for a range of questions. Links to the Ensembl database as well as the Edinburgh Mouse Atlas project were included to support further exploration and validation of the genes identified in the dataset.

Chapter 4 Functional screening for candidates with a role in HSC development reveals Bmper as a novel modulator of HSC maturation

4.1 Introduction

To date the validation of transcriptome-wide expression studies in the development of haematopoietic studies has relied on mouse-knockout systems (Mascarenhas et al., 2009), zebrafish morpholinos (Kartalaei et al., 2015; McKinney-Freeman et al., 2012), or culture systems measuring expansion of HSCs rather than maturation of early precursors (Durand et al., 2007). Such methods, while yielding important insights, are limited in throughput of candidates that can be tested and, in some cases, the functional relevance to the maturation process itself.

The development of a culture-system in which early precursor cells can mature into HSCs *ex vivo* (Rybtsov et al., 2011, 2014; Sheridan et al., 2009; Taoudi et al., 2008) has facilitated analysis of regulation of HSC maturation. As with the explant culture system (Medvinsky and Dzierzak, 1996), cells are cultured in isolation from the rest of the embryo, ensuring that changes in functionality are truly due to changes occurring in the cells of the culture rather than migration from other regions. The dissociation step eliminates the variation due to slight stage differences between embryos as primary cells (measured in embryo equivalents) as control and test conditions are taken from the same pool of dissociated tissues from multiple embryos. The fact that tissues can be isolated and cultured from a stage in which no HSCs are present, ensures that the culture system is truly capturing the maturation process (not just expansion) therefore conclusions about the conditions supporting HSC maturation can be drawn from modifications of this culture. Such culture systems have previously been used to validate the specific precursor lineage from which HSCs are derived (Rybtsov et al., 2011), the role of key cytokines SCF, IL3 and FLT3 in the maturation process (Rybtsov et al., 2014), the long-range interactions between neighbouring AGM tissues (Souilhol et al., 2016a), and rigorously quantify HSC precursors during embryonic development (Rybtsov et al., 2016).

As the RNA-sequencing data here had been generated in direct relation to the functional understanding gained from cultures of E9.5 and E10.5 domains of AGM (section 3.2.1), reaggregate culture was a highly appropriate system to test the results of the transcriptional analysis. Here, a number of secreted candidates that were identified through bioinformatics analysis of AGM transcriptional data were functionally assessed using reaggregate culture followed by transplantation as an assay for maturation of HSCs from precursors.

In this chapter I will discuss the selection of secreted protein candidates; the *ex vivo* reaggregate culture assays in which 10 candidates were tested for their capacity to support HSC maturation; the output of these assays in terms of CFU-C, repopulation of irradiated mice, and multilineage engraftment of these repopulated mice; and finally demonstrate that out of these candidates BMPER is reproducibly capable of increasing the efficiency of HSC maturation from E9.5 caudal part and from E11.5 AoD.

4.2 Results

This work was conducted with the assistance of Dr Stanislav Rybtsov who assisted in the set-up of the first reaggregate and transplantation experiment, and Dr Celine Souilhol who performed the final culture experiment (section 4.2.7) of E11.5 AoV with BMPER.

4.2.1 The E9.5 reaggregate culture as a novel mode of functionally validating candidates from bioinformatic analysis

The transition between the early E9.5 environment, in which HSC precursors are detected, and the E10.5 environment, in which precursor maturation is autonomously supported, has been shown functionally (Rybtsov et al., 2014, 2016) and now transcriptionally. To identify key, novel molecular players in this process, an assay system was selected which could be tightly coupled to the bioinformatics analysis.

In this system the caudal part of E9.5 embryos, which is known not to host any mature HSCs (Rybtsov et al., 2016), is dissociated, mixed with OP9 stromal

cells, reaggregated, and cultured at the air-liquid interface on serum media containing SCF, IL3 and FLT3L for 7 days. After this culture, the system will contain transplantable HSCs, albeit in lower numbers than would arise *in vivo* after the same time period. Candidates can therefore be added to the culture media as recombinant proteins and the output of transplantable HSCs used as a measure of their effect on the efficiency of the HSC maturation process. The advantage of this system is that the number of candidates that can be screened is mainly limited only by the number of mice that can be transplanted at the end-point, as opposed to the laborious generation of transgenic mice even prior to transplantation for *in vivo* studies. This system was therefore deemed a suitable approach to functionally validate the protein product of genes which are highly expressed in E10.5 compared to E9.5.

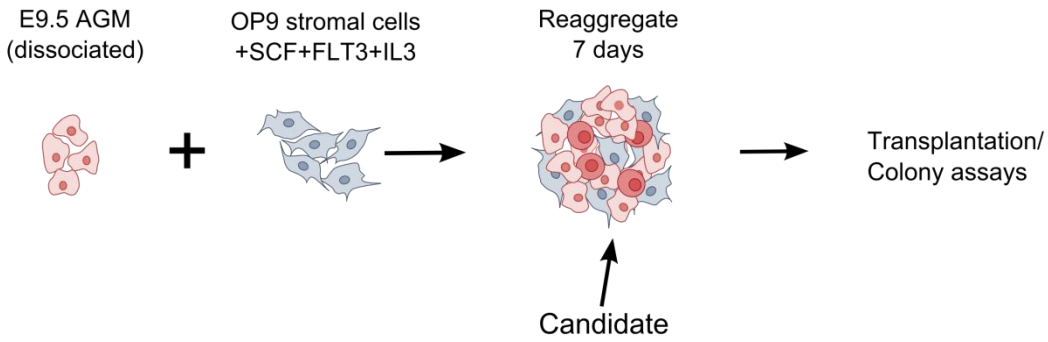


Figure 4-1 The *ex vivo* culture system to assay the effect of candidates on the maturation of precursors from E9.5 into HSCs

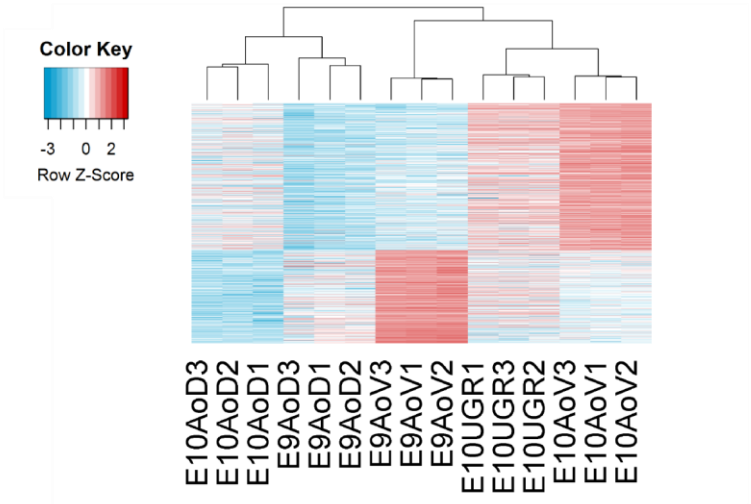
Schematic of the reaggregate culture system used to test the effect of candidates of maturation of E9.5 precursors. Caudal part from E9.5 embryos are dissected, dissociated, and reaggregated with OP9 cells. Reaggregates are cultured at the air-liquid interface of serum media to which candidate molecules can be tested. After 7 days, cultures are transplanted into irradiated mice at a dose of 0.5 or 1 ee (exact dose specified in results).

4.2.2 Selection of secreted molecules significantly differentially expressed in E10.5 AoV provides candidate regulators of HSC maturation

A refined set of genes for functional testing was selected based on the known functionality of subdomains of the AGM region and the assay in which these genes would be tested (section 4.2.1). Since E10.5 AoV is able to support a close to physiological rate of maturation of pre-HSCs whilst the E9.5 AGM region does not (Rybtsov et al., 2014), it was hypothesised that the molecules of this niche may be supplemented to the E9.5 reaggregate culture and improve the efficiency of HSC production *ex vivo*.

To identify genes which are absent or lowly expressed at E9.5 but present at E10.5, differentially expressed genes between E10.5 AoV and E9.5 AoV with > 2-fold change and FDR < 0.05 were selected and ranked by fold change (Figure 4-2A). The genes meeting this differential expression criteria show a much lower degree of differential expression between the E10.5 AoD and E9.5 AoD than E10.5 AoV and E9.5 AoV suggesting that they represent a subset specific to the supportive AoV environment at E10.5. To identify genes that promote HSC maturation, the up-regulated genes at E10.5 AoV were the particular focus. Of these 833 genes, the 119 which have a gene ontology annotation as “extracellular space” (GO:0005615) or “extracellular region” (GO:0005576) (Figure 4-2B) were selected on the basis that secreted factors are likely to be important effectors of niche signalling.

A



B

E10AoV vs E9AoV	No. of genes	No. of secreted	No. of secreted OP9 neg	No. of secreted OP9 pos	No. of secreted OP9 pos RGR up
Up	833	119	41	78	14
Down	528	89	53	36	9

Figure 4-2 Selection of genes for functional screening

(A) Heat map shows significantly differentially expressed between E9.5 AoV and E10.5 AoV with fold change ≥ 2 and FDR ≤ 0.05. (B) Table quantifies the number of differentially expressed genes in E10AoV vs E9AoV, both up and down regulated; the number of differentially expressed secreted factors (with gene ontology GO:0005576 "extracellular region" or GO:0005615 "extracellular space"); and the number not expressed in OP9 (OP9 neg), expressed in OP9 (OP9 pos) or expressed significantly higher in OP9 upon reaggregation (OP9 pos RGR up) (expression threshold > 0.5 RPM).

Since maturation of transplantable HSCs from E9.5 caudal part requires the support of OP9 co-culture (Rybtsov et al., 2014, 2016), the OP9 transcriptional data could be used to further refine the selection of candidates. The 119 secreted candidates were divided into those which are absent from OP9; those which are expressed in OP9; and those whose expression in OP9 increases upon reaggregation (Figure 4-2B). It was posited that all three categories may yield interesting candidates: those common to OP9 and AGM may be relevant as they are present in two independent supportive cell types, whilst those that are absent in OP9 may provide insight into the limitations of the culture system for early E9.5 precursors. Therefore candidates spanning all three categories were selected.

Due to the enrichment of pro-inflammatory signalling and BMP/regulation of BMP signalling in E10.5 AoV described above (sections 3.2.2 and 3.2.3), candidates belonging to these broad categories – but particularly those that are relatively little known, or have not previously been shown to affect HSC development – were selected. Selected members of the BMP/TGF-β/Nodal family comprised: BMP endothelial regulator (*Bmper*), which encodes a secreted protein capable of binding and modulating BMP ligand activity (Moser et al., 2003); Growth Differentiation Factor 3 (*Gdf3*) a regulator of cell growth in adult and embryo (Ehira et al., 2010; Levine and Brivanlou, 2006; Levine et al., 2009); Chordin-like 2 (*Chrdl2*), a paralog of the BMP inhibitor Chordin (Nakayama et al., 2004); and Inhibin Beta (*Inhbb*), which encodes an inhibin subunit capable of forming homodimers (Activin A) or heterodimers with INHBA (Activin AB) which regulate FSH secretion (Mather et al., 1997; Namwanje and Brown, 2016). Pro-inflammatory cytokines consisted of: C-X-C motif chemokine 10 (*Cxcl10*), also known as interferon gamma-induced protein 10, a chemokine secreted in response to interferon-gamma induction (Luster and Ravetch, 1987; Luster et al., 1985); and Chemokine (C-C motif) ligand 4 *(Ccl4*), also known as macrophage inflammatory protein-1β , a chemoattractant for immune cells (Bystry et al., 2001; Dorner et al., 2002). Another gene of interest was Integrin Binding Sialoprotein (*Ibsp*), which encodes a protein homologous to the immune adhesion regulator OPN (Lund et al., 2009; Tagliabracci et al., 2012) which has largely associated with bone regulation (Alford and Hankenson, 2006; Bouleftour et al., 2014) but has been implicated HSC regulation in bone marrow (Granito et al., 2015). Finally a number of secreted factors were considered interesting due to their potential roles as effectors of cell growth and differentiation: Insulin like growth factor binding protein 3 (*Igfbp3*) encodes a protein that binds and transports insulin-like growth factor (IGF-1), a hormone that potentially regulates HSC bone marrow engraftment (Baxter and Martin, 1986; Caselli et al., 2013; Martin and Baxter, 1986); NEL-like 1 (*Nell1*) produces an EGF-like repeat containing protein which has been shown to regulate cell growth and differentiation (Bokui et al., 2008; Xue et al., 2011; Zhang et al., 2002); and wingless-type MMTV integration site family, member 2B (*Wnt2b*) a member of the conserved WNT family which has been implicated in

regulating a number of developmental processes (Cho and Cepko, 2006; Kawakami et al., 2001; Ober et al., 2006) (Table 9).

For each of the selected candidates, the appropriate recombinant protein was bought and added to the culture media at two different doses. The dose was chosen based on the recommendations from the suppliers and the equivalent expression level in the dataset. The addition of two doses served as a (limited) means of titrating the protein. Each protein was further tested in the absence of SCF/IL3/FLT3 or OP9 cells in the culture environment to determine whether it could supplement these essential factors. Functional screening of these proteins was performed in two sets of experiments, and consequently these experiments are kept separately in the figures below to enable more precise comparison to the control conditions of each experiment.

Gene	Description
Cxcl10	A chemokine involved in stimulation of monocytes, natural killer and T-cell migration, and modulation of adhesion molecule expression. Homozygous knockout mice are viable, fertile, and have no overt morphological or developmental abnormalities. Homozygous KO mice have defective T cell responses, including impaired proliferation and IFN-gamma secretion following antigenic challenge (129Sv background).
Ccl4	A mitogen-inducible monokine with chemokinetic and inflammatory functions.
Gdf3	A member of the bone morphogenetic protein (BMP) family and the TGF-β superfamily. Gdf3 can act as a BMP inhibitor within its normal dose range, but at high doses may act as a nodal agonist (Levine et al., 2009). Mice homozygous for a null allele exhibit prenatal lethality and resistance to diet-induced obesity. Abnormal anterior visceral endoderm cell migration; abnormal primitive streak formation; absent mesoderm; rostral body truncation; absent anterior visceral endoderm.
Chrdl2	Chordin-like 2; may inhibit BMPs activity by blocking their interaction with their receptors.
Bmper	BMP-binding endothelial regulator Gene; Inhibitor of BMP. BMPER has been shown to be necessary for BMP-4 to exert its activating role in endothelial function and to induce Smad 1/5 activation. Vice versa, BMP-4 is necessary for BMPER activity (Heinke et al., 2008). Mice homozygous for a knock-out mutation exhibit neonatal lethality associated with abnormal lung and skeleton development. Mice heterozygous for a null allele exhibit abnormal lung development.

Inhbb	The inhibin beta B subunit joins the alpha subunit to form a pituitary FSH secretion inhibitor. Furthermore, the beta B subunit forms a homodimer, activin B, and also joins with the beta A subunit to form a heterodimer, activin AB, both of which stimulate FSH secretion.
Ibsp	The protein encoded by this gene is a major structural protein of the bone matrix. It constitutes approximately 12% of the non-collagenous proteins in human bone and is synthesized by skeletal-associated cell types, including hypertrophic chondrocytes, osteoblasts, osteocytes, and osteoclasts. The only extraskeletal site of its synthesis is the trophoblast.
Nell1	This gene encodes a cytoplasmic protein that contains epidermal growth factor (EGF)-like repeats. The encoded heterotrimeric protein may be involved in cell growth regulation and differentiation. A similar protein in rodents is involved in craniosynostosis. Alternative splicing results in multiple transcript variants.
Igfbp3	This gene is a member of the insulin-like growth factor binding protein (IGFBP) family and encodes a protein with an IGFBP domain and a thyroglobulin type-I domain. The protein forms a ternary complex with insulin-like growth factor acid-labile subunit (IGFALS) and either insulin-like growth factor (IGF) I or II. In this form, it circulates in the plasma, prolonging the half-life of IGFs and altering their interaction with cell surface receptors.
Wnt2b	Ligand for members of the frizzled family of seven transmembrane receptors.

Table 9 Selected genes for functional testing

Summary of genes selected for function testing. Gene names are based on their MGI symbols and the descriptions and knockout phenotypes compiled from information from Entrez Gene Summaries (Maglott et al., 2005), and the Mouse Genome Database[7] (Eppig et al., 2015, accessed November 2013).

4.2.3 *In vitro* haematopoietic colony formation is not affected by addition of recombinant proteins

For each of the recombinant proteins being tested, the capacity to differentiate into blood lineages – a hallmark of stem cells and progenitors – was measured by colony forming assays in methylcellulose based media. The number of haematopoietic colonies of granulocyte/monocyte (GM), granulocyte/erythrocyte/monocyte/megakaryocyte (GEMM), or proerythroblast (ER)

[7] http://www.informatics.jax.org

phenotype was not significantly altered by the culture with each of the recombinant proteins (Figure 4-3).

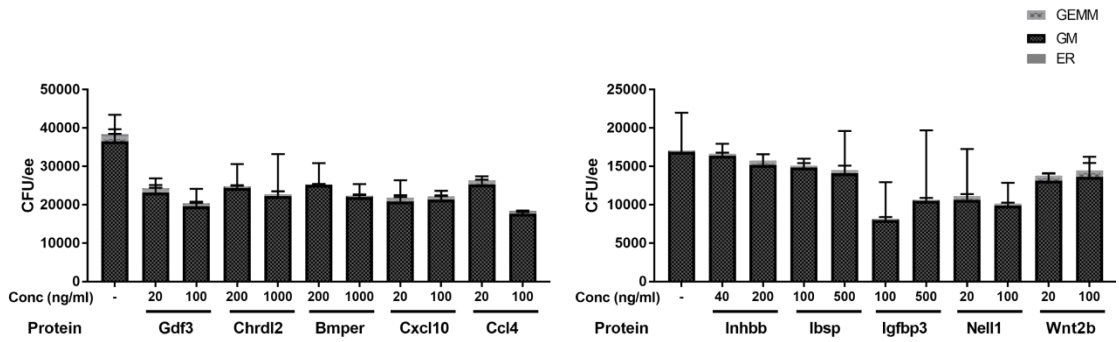


Figure 4-3 Effect of candidates on colony forming potential

Displays the number and type of colonies per embryo equivalent, produced from E9.5 caudal part cells cultured with the named recombinant protein (or in the absence of additional protein "-") at the displayed concentration for seven days with serum media, OP9 cells and SCF, IL3 and FLT3L. Number of experiments = 2. GM = granulocyte/monocyte, GEMM = granulocyte/erythrocyte/monocyte/megakaryocyte and ER = proerythroblast type colonies. Each methylcellulose culture was with 0.005 embryo equivalents, with 2 technical replicates and two biological replicates. Error bars represent the standard deviation from the mean.

4.2.4 Long term repopulation shows functional enhancement of HSC maturation *ex vivo* by Bmper and potential effects of further molecules

For each of the recombinant proteins being tested, the production of HSCs after seven days of culture was assayed by injection of the culture into the tail-vein of sub-lethally irradiated mice. Sixteen weeks after injection, the level of peripheral blood chimerism from the donor cells was measured as the percentage of donor cells in the total peripheral blood cells (including host and transplanted carrier cells). The efficiency of the culture for HSC production was inferred from the number of mice repopulated (as an indicator of whether HSCs were generated) and the level of repopulation (which is a semi-quantitative indicator of the number of HSCs/their capacity to engraft and differentiate) in comparison to the control.

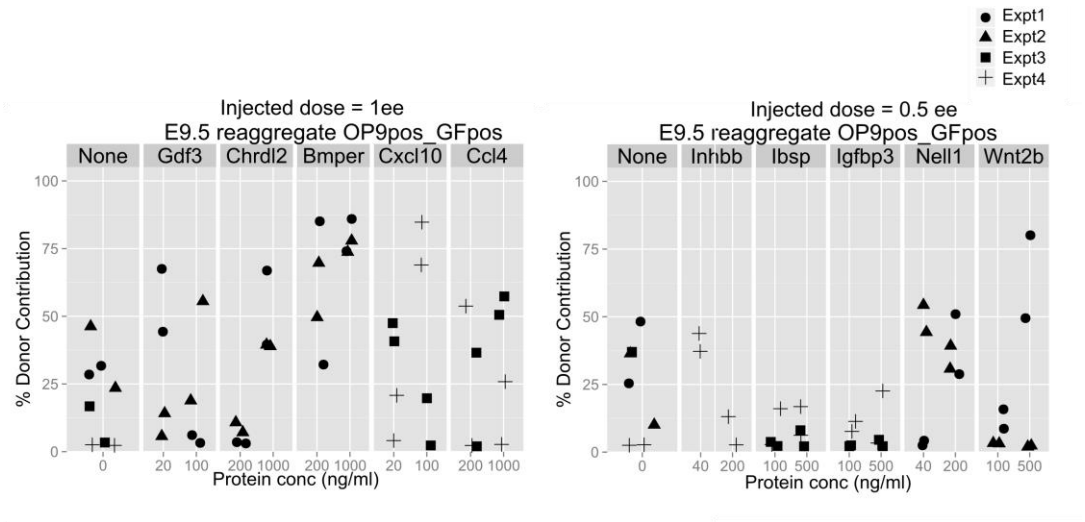


Figure 4-4 Blood repopulation results in screen for the effect of candidates on HSC maturation

Displays the percentage contribution of donor cells measured 16 weeks after injection with E9.5 caudal part cells cultured with the named recombinant protein at the displayed concentration for seven days with serum media, OP9 cells and SCF, IL3 and FLT3L (injected dose 1 ee or 0.5 ee as indicated). Each graph summarises the results of four independent experiments as indicated by the point symbols.

Two of the recombinant proteins had a noticeable effect on the donor contribution to peripheral blood, with CXCL10 giving a mild improvement by increasing the level of contribution, whilst BMPER had a more striking effect (Figure 4-4). Reaggregates that had been co-cultured with BMPER showed a significantly higher contribution to the recipient peripheral blood than untreated reaggregate cultures, with 8 out of 8 mice showing donor contribution above 25% compared to 3 out of 8 showing donor contribution above 25% from controls. Moreover, the donor contribution increased with dose such that 4 out of 4 reaggregates cultured with 1 μg/ml BMPER gave greater than 70% contribution, compared to a mean of 17% contribution from controls. Other recombinant proteins showed an increase in one experiment but reverse in another.

As BMPER had no effect on the number of haematopoietic colonies (Figure 4-3) this suggested that it had a specific role in the maturation of HSCs rather that more committed haematopoietic progenitors in reaggregate culture.

4.2.5 Multilineage reconstitution shows normal blood repopulation from HSCs matured *ex vivo*

To assess the differentiation potential of HSCs produced in this culture, the contribution of donor cells to all blood lineages was measured. The peripheral blood coming from donor cells (CD45.2+ cells) was analysed by flow cytometry to determine the proportion of myeloid (Mac1+Gr1+ and Mac1+Gr1-), B cells (B220+), T cells (CD3+) (Figure 4-5A and B). The proportion of each lineage was not significantly altered in most of the treatment conditions suggesting that HSCs produced in culture were true multipotent stem cells, capable of differentiating into all blood lineages. Multilineage analysis of mice repopulated with reaggregate cultures that had been treated with Ibsp had a larger fraction of B and T-cells, and low myeloid contribution (Figure 4-5). As these mice had been repopulated at low level, it suggests that the culture may not have produced mature HSCs, rather, reconstitution was from progenitors.

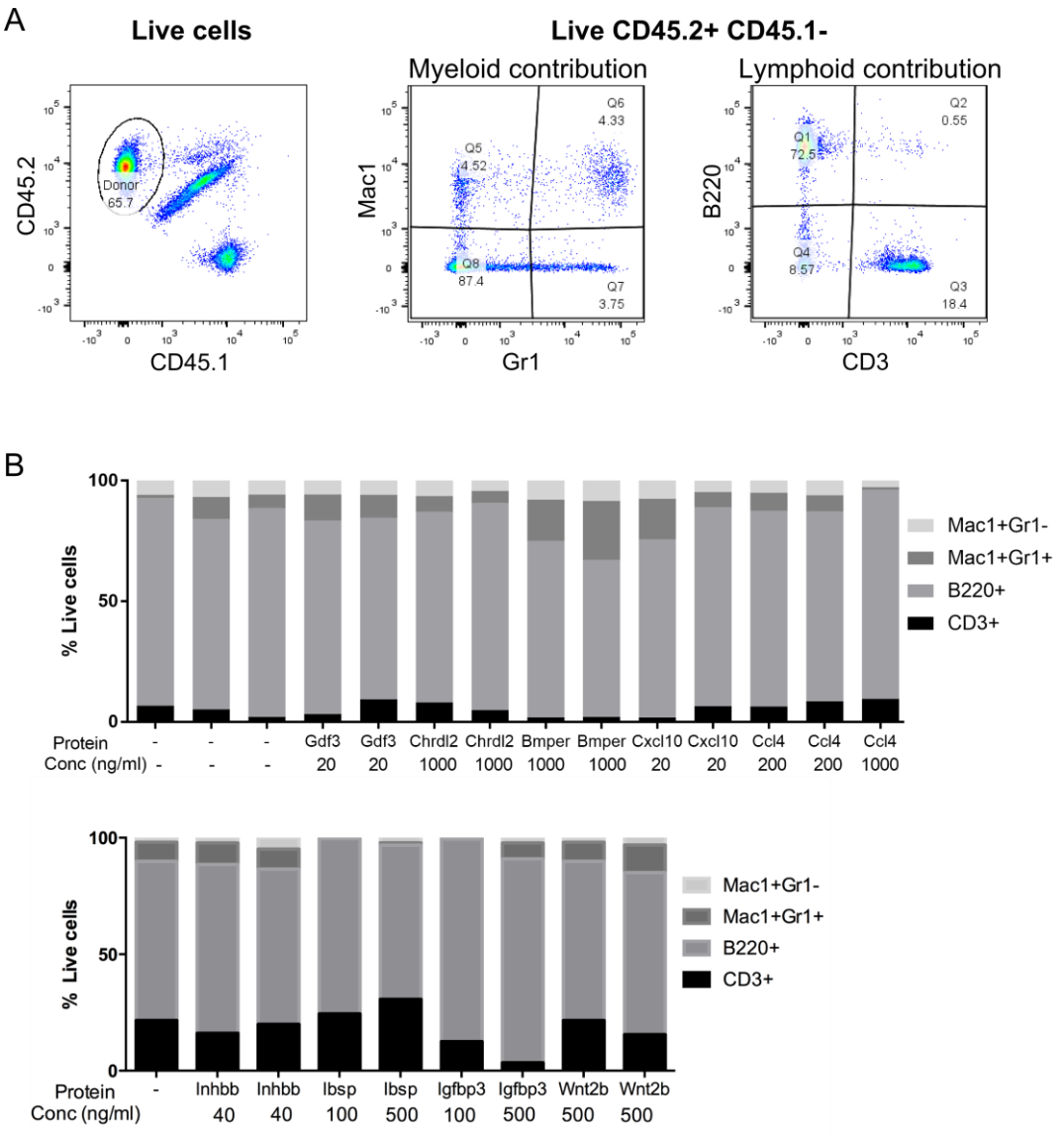


Figure 4-5 Multilineage donor contribution to repopulated mice

(A) Sorting strategy to assess percentage donor contribution from different blood lineages. CD45.2+ CD45.1- cells represent donor cells; Mac1 and Gr1 represent myeloid cell; and CD3 and B220 represent lymphoid cells. (B) Quantification of multilineage contribution from repopulated mice representing each treatment condition. Populations are a percentage of live donor cells (CD45.2+) from peripheral blood of transplanted mice: myeloid (Mac1+Gr1+ and Mac1+Gr1-), B cells (B220+), T cells (CD3+).

4.2.6 Preliminary functional studies without growth factors or cytokines suggest Bmper can rescue 3GF effect, and a potential role of IBSP in haematopoietic colony formation

The assay with E9.5 caudal part, OP9 and cytokines enables the identification of factors which affect the efficiency of HSC maturation by detecting an increase or decrease in repopulation level compared to a relatively low level of control repopulation. A more stringent mode of assay is to test an increase in repopulation using conditions in which there is no repopulation from control. In the absence of the cytokines (SCF, IL3 and FLT3) is it known that there is no (or very little) repopulation from cultured E9.5 caudal part as, particularly SCF is an essential factor in supporting HSC maturation (Rybtsov et al., 2014). Similarly absence of OP9 significantly impedes the maturation of HSCs from E9.5 precursors. Consequently, the removal of either cytokines or OP9 was used as a more stringent assay of each candidate's effect on HSC maturation.

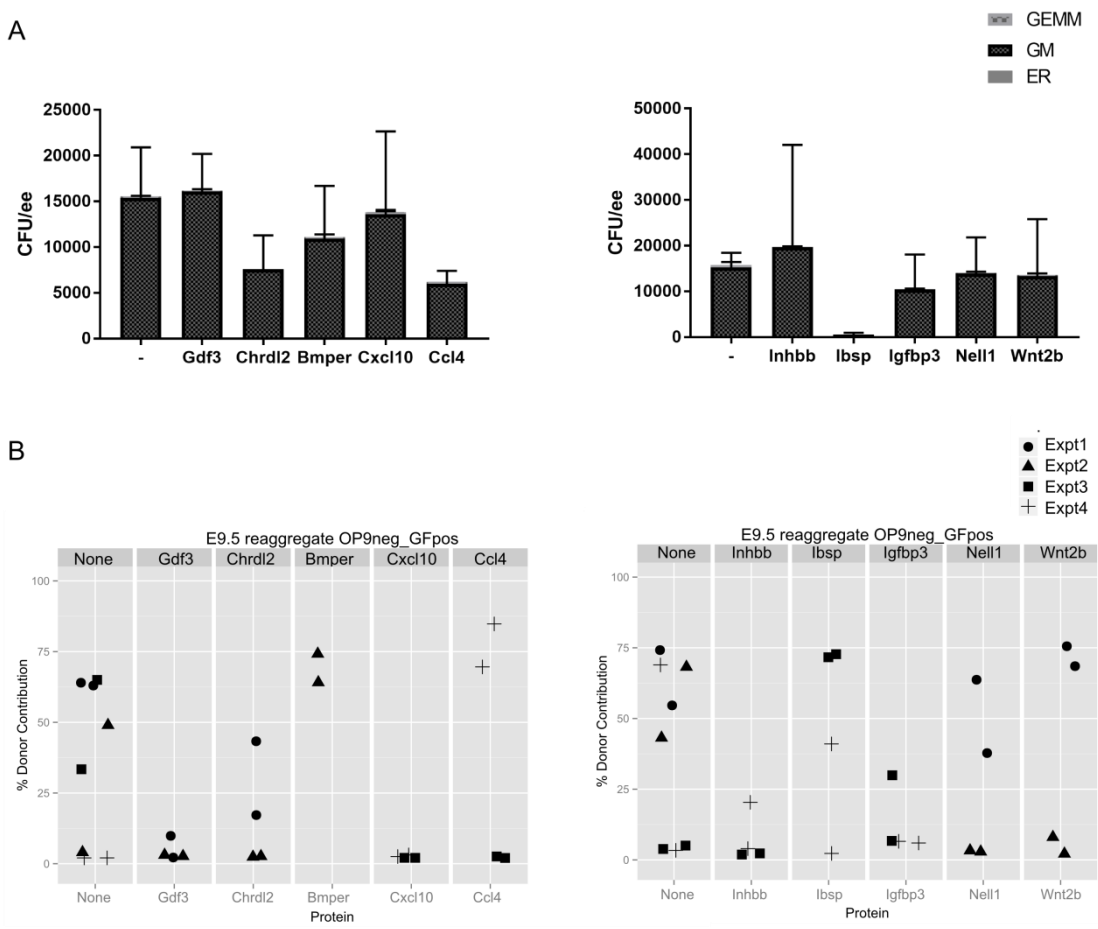


Figure 4-6 Colony forming and transplantation potential after culture in the absence of OP9

(A) Displays the number and type of colonies per embryo equivalent, produced from E9.5 caudal part cells cultured with the named recombinant protein for seven days with serum media and SCF, IL3 and FLT3L. GM = granulocyte/monocyte, GEMM = granulocyte/erythrocyte/monocyte/megakaryocyte and ER = proerythroblast type colonies. Each methylcellulose culture was with 0.005 ee equivalent, with 2 technical replicates and two biological replicates. (B) Displays the percentage contribution of donor cells measured 16 weeks after injection with E9.5 caudal part cells cultured with the named recombinant protein at the displayed concentration for seven days with serum media and SCF, IL3 and FLT3L (injected dose 1 ee).

In the absence of OP9 cells the number of CFU-Cs was similar to the number in the presence of OP9 cells. Most of the treatment conditions gave lower repopulation than when OP9 were present (Figure 4-6, Figure 4-4). More strikingly,

in the absence of cytokines the number of CFU-Cs was reduced about 10-fold to approximately 1500 CFU-C per embryo equivalent (Figure 4-7A and B) and the number of significantly repopulated mice in control reduced to 2 out of 16 (Figure 4-7C and D).

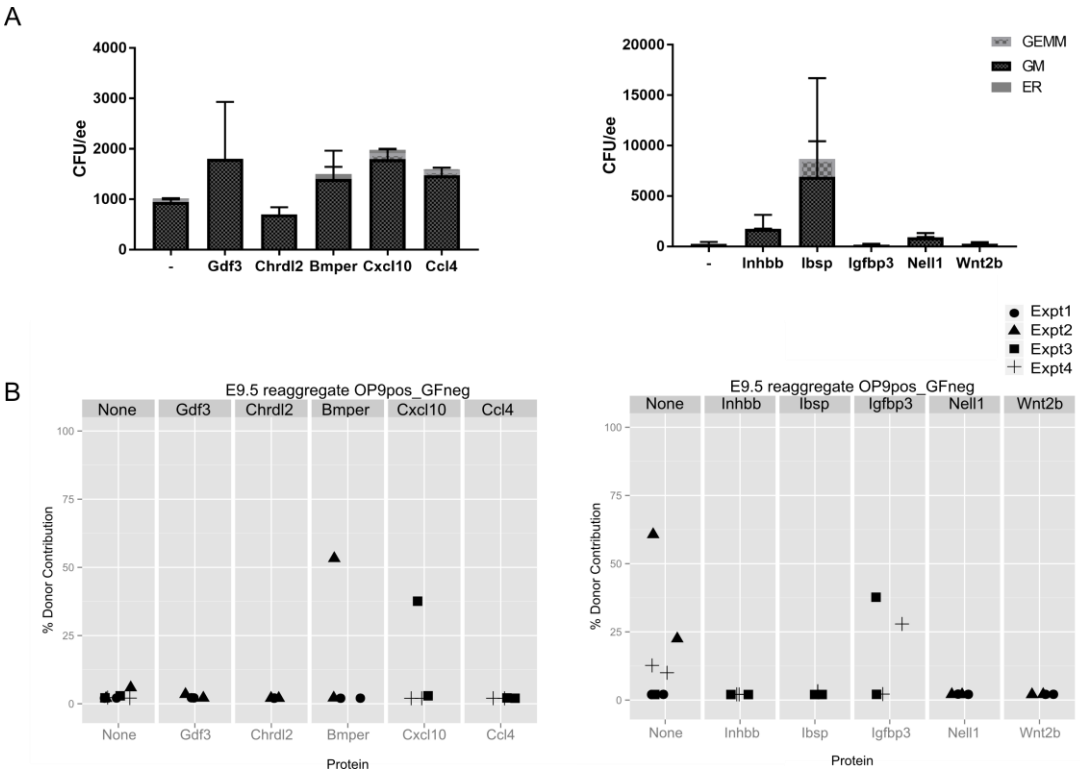


Figure 4-7 Colony forming and transplantation potential after culture in the absence of SCF/IL3/FLT3L

(A) Displays the number and type of colonies per embryo equivalent, produced from E9.5 caudal part cells cultured with the named recombinant protein for seven days with serum media and OP9 cells. GM = granulocyte/monocyte, GEMM = granulocyte/erythrocyte/monocyte/megakaryocyte and ER = proerythroblast type colonies. Each methylcellulose culture was with 0.05 ee, with 2 technical replicates and two biological replicates. (B) Displays the percentage contribution of donor cells measured 16 weeks after injection with E9.5 caudal part cells cultured with the named recombinant protein at the displayed concentration for seven days with serum media and OP9 cells (injected dose 1 ee).

In these experiments, both the absence of OP9 and the absence of cytokines respectively, BMPER showed no significant effect on colony formation compared to

control (Figure 4-6A, Figure 4-7A), similarly to in the presence of both OP9 and cytokines (Figure 4-3). The repopulation level after BMPER treatment indicated a potentially mild rescue affect compared to control (Figure 4-6B, Figure 4-7B), although this would need to be repeated to verify this significance.

Another notable result in these assays was the differing effect on CFU-Cs and repopulation after treatment with IBSP in the absence of cytokines or OP9 cells. When OP9 cells were excluded from the E9.5 culture, the addition of IBSP gave a reduced number of CFU-Cs to almost 0 colonies/ee (Figure 4-6A) whilst maintaining the repopulation level at the same level as control (Figure 4-6B). Contrastingly, when cytokines were removed, in the first experiment the number of CFU-Cs was maintained at a much higher number that in controls and in the second experiment a large number of proerythroblast colonies were produced (Figure 4-7A). In the same condition, no mice were repopulated from the culture (Figure 4-7B). This was a puzzling effect as IBSP had not induced any different CFU-Cs or repopulation compared to control when both OP9 and cytokines were present in the culture. One potential explanation is that IBSP has opposing effects on the OP9 stroma compared the stroma of SCF treated primary caudal part cells. The outcome in the presence of OP9 and cytokines may then be an average of these two affects, but when each condition is eliminated successively these opposing affects can be seen. This phenotype has not been followed up in this project, but could be of potential future interest if replicated in further experiments.

4.2.7 Bmper can enhance repopulation at E11.5

This experiment was conducted by Celine Souilhol (from Professor Alexander Medvinsky's lab) as a follow up to my findings that BMPER increases maturation of HSCs from E9.5 culture. It is included here to strengthen the support for BMPER as a regulator of HSC maturation.

As the effect of BMPER on HSC maturation was the most striking and consistent of all candidates tested, it was followed up in more detail. At E11.5 approximately 1 HSC is present in the AGM region (Kumaravelu et al., 2002; Rybtsov et al., 2016) and this is generally restricted to the ventral side (AoV) (Taoudi and Medvinsky, 2007). As the AGM region also contains a number of

precursor cells (pre-HSCs) further HSCs can be generated from both AoV and AoD after a period of culture, but the autonomous generation of HSCs from AoD is at a very low frequency (4/20 mice repopulated from AoD versus 19/28 from AoV in (Souilhol et al., 2016a)), and generally requires co-culture with AoV or OP9 and cytokines to induce further HSC production. The effect of BMPER on HSC generation in isolated E11.5 AoV and AoD was determined by addition to the 4 day culture of these tissues at two concentrations.

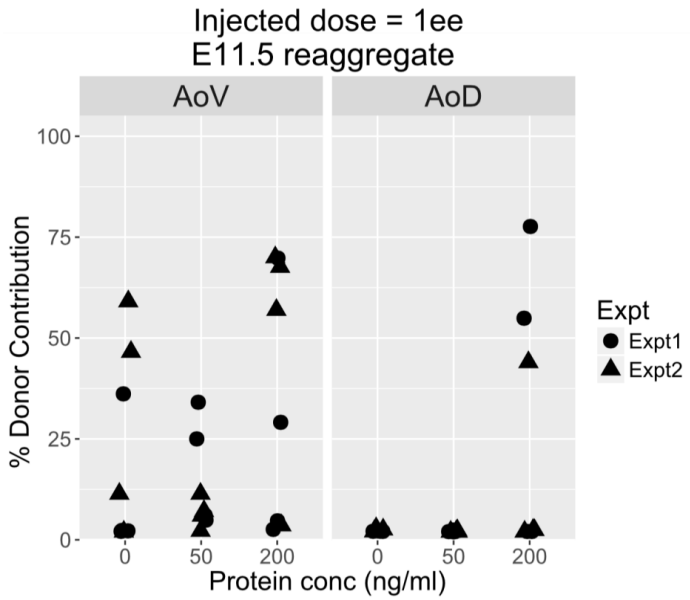


Figure 4-8 The effect of BMPER on HSC production from E11.5 AoV and AoD

Displays the percentage contribution of donor cells measured 16 weeks after injection with E11.5 AoV or AoD cells cultured with the named BMPER at the displayed concentration for five days with in serum-free, cytokine-free media (injected dose 1 ee).

The production of HSCs from these cultures was determined by measuring the peripheral blood chimerism of mice 16 weeks after they had been transplanted with the culture. The culture of AoV showed a slight increase in HSC production compared to control when the highest dose of BMPER was added with 5/8 mice repopulated, compared 3/8 in control (Figure 4-8). Furthermore the repopulation tended to be at a higher level (57 %) compared to the control (45 %). The culture of AoD with BMPER at its highest dose showed a significant difference to the culture of AoD alone. In this case 3/8 mice were repopulated at high level compared to 0/8

in control (Figure 4-8). Thus the results corroborate the effect seen at E9.5 suggesting that BMPER is capable of modulating the process of HSC maturation and when added in high levels it increases the efficiency of this process.

4.3 Discussion

The results here show that combination of transcriptional data with the reaggregate culture functional assay enabled the protein product of multiple genes to be screened for the effect on HSC maturation. Through this type of screening, a novel role for the protein BMPER was identified as modulation of HSC maturation from precursors. Further functional testing demonstrated that BMPER can maintain HSC maturation from E9.5 precursors in more stringent culture conditions (i.e. the absence of cytokines or OP9 cells) and also can increase the efficiency of maturation from E11.5 AoV and AoD tissues.

Further explored in these experiments was the role of IBSP on the production colony forming units. Contrasting effects were seen when cytokines were eliminated from the cultures, and when OP9 cells were eliminated suggesting that it was acting synergistically with other molecules, or signalling in an indirect effect through the stroma. In either case, it highlights that the functional effect measured may be the average outcome of opposing signalling which is not unmasked until separate aspects of the culture system are isolated. Indeed the significant undefined aspects of this culture system pose a limit on the inferences drawn from experiments, and the reproducibility. However as yet, this is the best known system to functionally study HSC maturation *in vitro* and achieving a more defined culture system is a long-term goal of molecular characterisation experiments.

Finally these experiments reiterated the lack of correlation between CFU-C assays and transplantation results. This demonstrates that the repopulation assay is still the main barrier to high throughput studies of the conditions/genes/proteins affecting HSC maturation.

Thus the validation, in reaggregate culture followed by transplantation, of candidates genes identified from RNA-seq of AGM domains highlighted BMPER as a novel regulator of HSC maturation, and possible roles of CXCL10 and IBSP which would require future investigation. Ultimately this suggests that the AGM region

transcriptome data may yield further candidates if functional screens were continued in the future. For the scope of this thesis, the reproducibility of the functional effect of BMPER prompted further investigation into its mechanism of action, which is discussed in detail in Chapter 5.

Chapter 5 Characterising the *in vivo* and *in vitro* role of Bmper in HSC maturation

5.1 Introduction

Bmper (bone morphogenetic protein [BMP]-binding endothelial cell precursor-derived regulator), is a secreted glycoprotein containing five cysteine-rich domains, a von Willebrand D domain and a trypsin inhibitor domain (frequently found in extracellular proteins). Originally identified in mouse through a screen of flk1+ cells (putative endothelial bodies) vs flk- cells from embryoid bodies it has been shown to bind BMP2, BMP4, BMP6, and BMP4 ligands (Moser et al., 2003). A number of studies in mouse, zebrafish and drosophila (where it is referred to as crossveinless-2 or cv2), have shown both pro-BMP (Coles et al., 2004; Conley et al., 2000; Kamimura et al., 2004; Ralston and Blair, 2005; Rentzsch et al., 2006), and anti-BMP (Binnerts et al., 2004; Helbing et al., 2013; Moreno-Miralles et al., 2011; Moser et al., 2003) regulatory roles of Bmper. More rigorous biochemical studies of Bmper have proposed that the protein regulates a biphasic response to BMP signalling (Kelley et al., 2009; Serpe et al., 2008). In a mouse endothelial cell line, low stoichiometric ratios of Bmper promoted signalling by BMP ligands, but when Bmper exceeded a 2:1 ratio it induces endocytic uptake of Bmper and its bound ligand, thereby antagonising the BMP signalling. This likely explains the contradictory findings of different studies of Bmper's regulatory behaviour. More recently Bmper has been studied with regard to its regulation of the inflammatory response of endothelium where is has been shown to be a protective regulator of vascular inflammation (Helbing et al., 2011), constrain retinal revascularisation after injury (Moreno-Miralles et al., 2011), and retain epithelial integrity after damage (Helbing et al., 2013).

The expression distribution of *Bmper* has been described by *in situ* hybridisation in mouse (Coffinier et al., 2002). From mid-gastrulation it is expressed by the posterior primitive streak and pre-cardiac mesoderm. At E9 it is detected in the ventral tail bud, and roof of the neural tube. By E10 an expression domain appears in the mesonephric ridge as well as the forebrain, limb bud, dorsal root ganglion, otic vesicle, and nasal process. In the adult, *Bmper* expression is most

highly detected in the lung tissue. Thus previous work validates the findings here in Chapter 3 and Chapter 4 that *Bmper* is expressed in the AGM region and particularly upregulated between E9.5 and E10.5. However the functional role of Bmper within the AGM region, and particularly in regard to HSC maturation, has previously never been studied. The reproducibly positive effect on HSC maturation, shown in Chapter 4, prompted further investigation into the source of BMPER and its distribution in the AGM at both the protein and gene expression level. Furthermore, the dual agonistic and antagonistic role found in previous studies posed the question of whether Bmper's regulatory function is executed via the enhancement or inhibition of BMP signalling within the AGM region.

In this chapter I show that the protein distribution of BMPER in E10.5 AGM is ventrally polarised validating the gene level distribution detected by RNA-seq (Chapter 3); I investigate the distribution in more detail and show that the intra-aortic clusters and primordial germ cells show particularly strong staining for BMPER. I further quantify the *Bmper* transcript distribution in endothelial cells, haematopoietic cells, perivascular cells, stroma and the pre-HSC lineage by FACS and qRT-PCR observing a particularly high transcript enrichment in the non-endothelial CD146 positive cells. Finally, I compare the BMPER and pSMAD1/5/8 distribution, and the inducibility of *Bmper* by BMP4 to build a model of BMP4 inhibition by BMPER via a negative feedback interaction.

5.2 Results

5.2.1 Immunostaining reveals BMPER protein is ventrally polarised within the E10.5 AGM region

Given the findings from RNA-sequencing that *Bmper* is up-regulated between E9.5 and E10.5 and polarised in the ventral region (Chapter 3, Chapter 4) its distribution within the AGM region was validated at the protein level by immuno-staining. In transverse sections BMPER could be detected at high levels on the ventral side of the AGM region relative to the dorsal side, particularly ventro-lateral to the dorsal aorta (Figure 5-1). This is in agreement with the RNA-seq data, suggesting that both expression and protein localisation is restricted to this ventral

niche. BMPER was also present in the neural tube, and the regions lateral to the AGM region, but these tissues are removed in the dissection of the AGM region for expression studies and reaggregate culture studies, so were not considered in more detail.

At the level of individual cells, some brightly stained cells were detected in the ventral mesenchyme as well as the luminal side of the dorsal aorta, prompting further investigation of their identity.

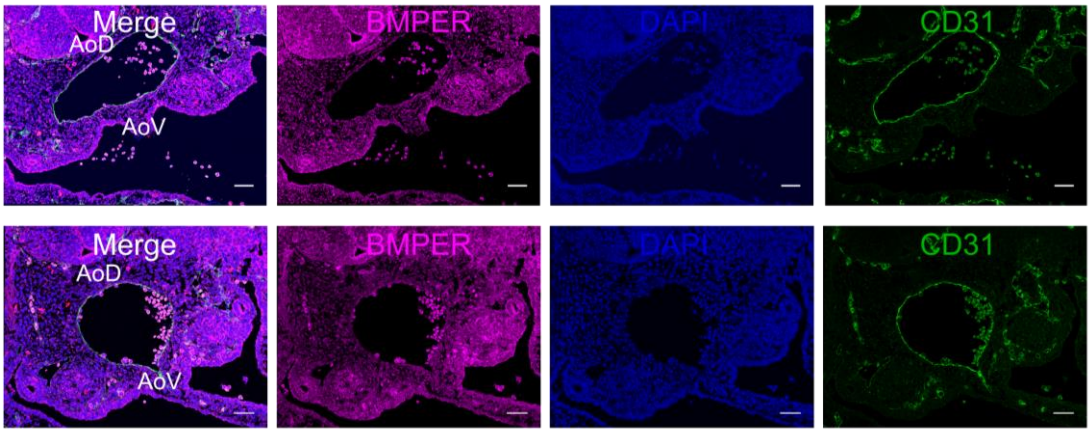


Figure 5-1 BMPER distribution in the E10.5 AGM region

Distribution of BMPER protein in the from AoD to AoV in E10.5 AGM transverse sections from two independent embryos measured by immunostaining. Green = CD31, Magenta = BMPER, Blue = DAPI. Scale bar = 50 μm.

5.2.2 The distribution of BMPER protein shows enrichment in haematopoietic clusters and primordial germ cells

The cells brightly stained with BMPER had the morphology of intra-aortic clusters – cells budding from the endothelium into the dorsal aorta which have both endothelial and haematopoietic surface markers and are believed to be the source of pre-HSC/HSCs (Boisset et al., 2010; Garcia-Porrero et al., 1995; Jaffredo et al., 1998; Tavian et al., 1996). Co-staining CD45 (a haematopoietic marker), CD31 (an endothelial marker) and BMPER, then imaging at high resolution by confocal microscopy, verified that these BMPER-bright cells express endothelial and

haematopoietic markers thus are likely to represent intra-aortic clusters (Figure 5-2A).

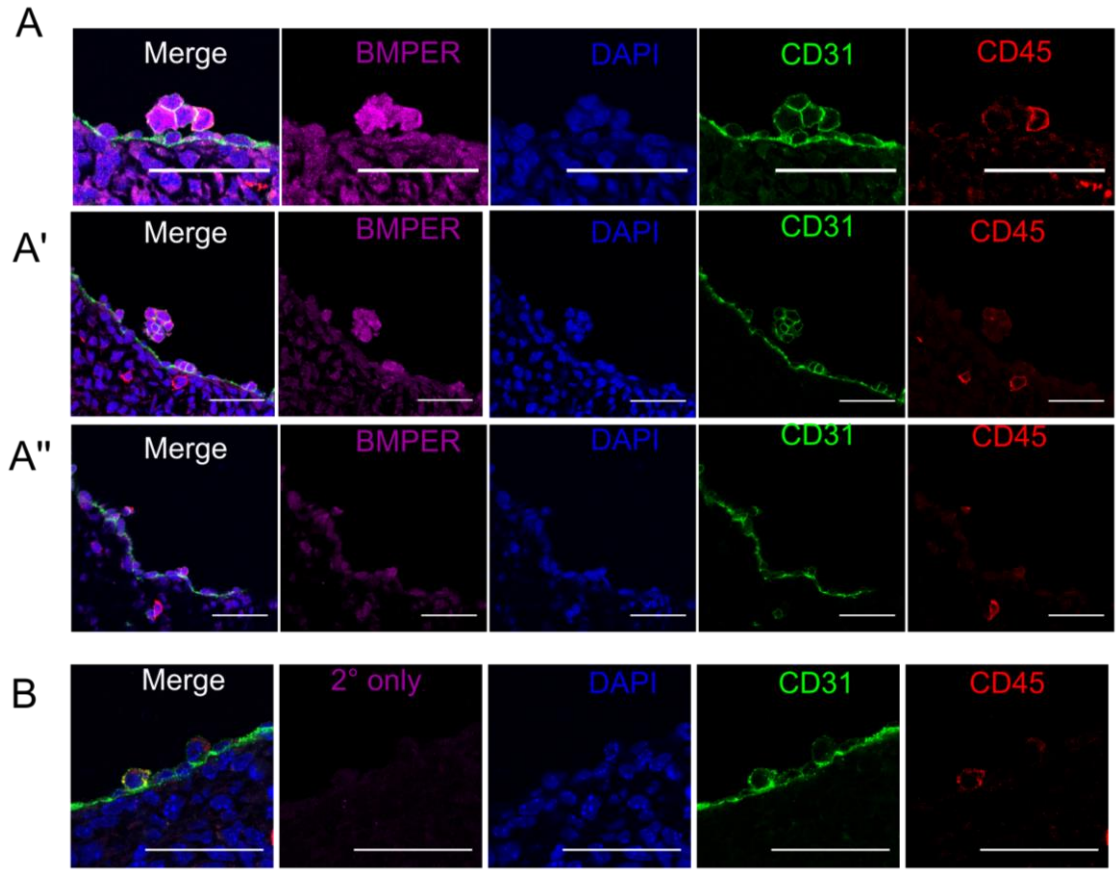


Figure 5-2 BMPER staining in intra-aorta clusters

(A) Localised BMPER protein in the hematopoietic clusters of the dorsal aorta from three different embryos in sections measured by immunostaining. (B) Secondary control immunostaining excludes anti-BMPER primary antibody. Red = CD45, Green = CD31, Magenta = BMPER, Blue = DAPI. Scale bar = 50 µm.

Staining of the protein was detected throughout the cytoplasm of these cells suggesting that either the protein is produced by these cells or is taken-up by these cells after secretion from another cell type. The endocytosis of BMPER when at high concentration has been described previously (Kelley et al., 2009) so the latter cause of cytoplasmic distribution is possible.

Another type of BMPER-bright cell was detected in the ventral mesenchyme, closer to the gut epithelium than the dorsal aorta and in CD31-positive cells (Figure

5-1). Based on a literature search of CD31 (Pecam-1)-positive cells that are not in the endothelium, and the cell types that are present in the ventral mesenchyme it was hypothesised that these cells may be primordial germ cells (Medvinsky et al., 1993; Wakayama et al., 2003; Yokomizo and Dzierzak, 2010). Co-staining with a specific surface marker of primordial germ cells, SSEA-1, suggested that these are indeed primordial germ cells (Figure 5-3), which migrate through the ventral mesenchyme at E10.5 on their journey from the gut to the urogenital ridges (Molyneaux et al., 2001).

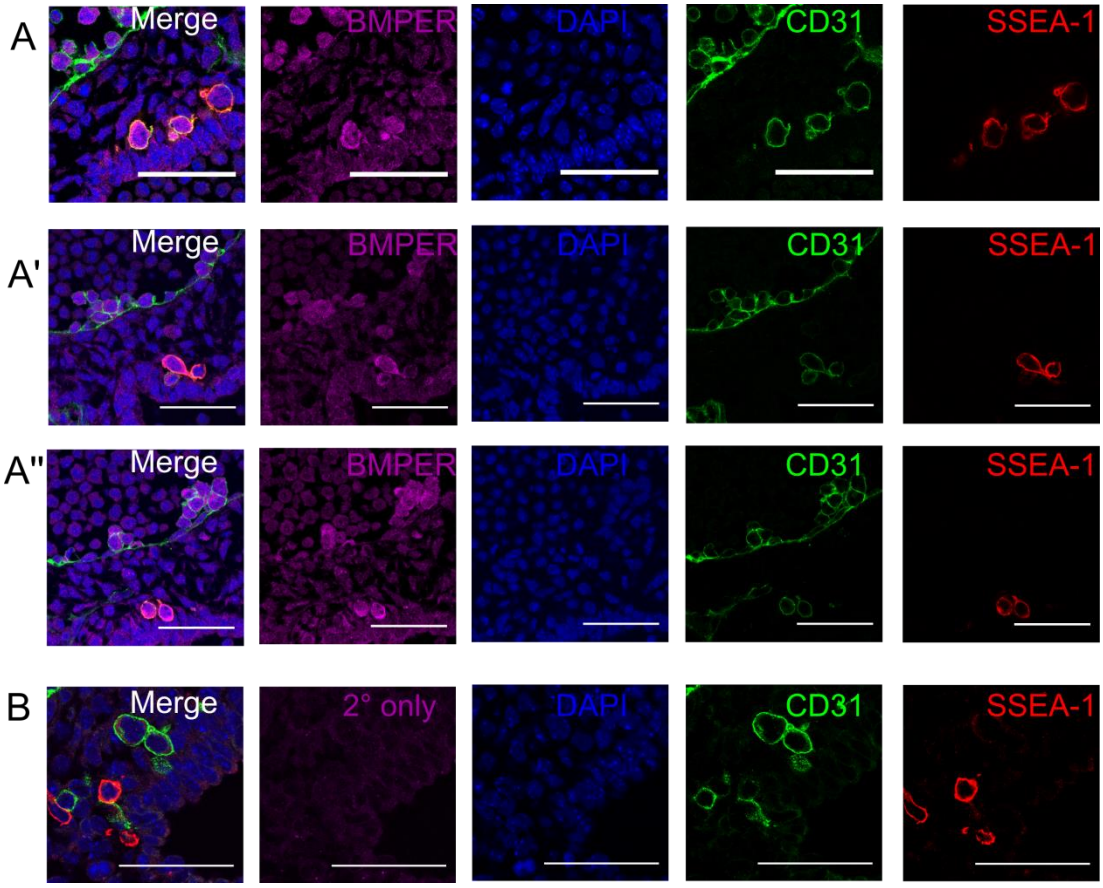


Figure 5-3 BMPER staining of primordial germs cells in the E10.5 ventral mesenchyme

(A) Localised BMPER protein in primordial germ cells in the ventral mesenchyme, in sections measured by immunostaining. (B) Secondary control immunostaining excludes anti-BMPER primary antibody. Red = SSEA-1, Green = CD31, Magenta = BMPER, Blue = DAPI. Scale bar = 50 um.

5.2.3 Bmper expression is enriched in the non-endothelial CD146+ encompassing perivascular cells, mesonephric tubules and metanephric mesenchyme

The AGM region is a heterogeneous mixture of cell types. To explore in more detail which cells are the main source of BMPER, the AGM region was sorted into broad cells types known to compose this region: Lin-CD45+VC-, representing haematopoietic cells; Lin-CD45-VC+, endothelial cells; Lin-CD45-VC-CD146+, putative perivascular cells; and Lin-CD45-VC-CD146-, remaining stroma (Figure 5-4A). The expression level of BMPER was compared between these sorted populations. The expression of *Bmper* was 5 - 25 times higher in the $Lin^{-}CD45^{-}VC^{-}CD146^{+}$ population than the other populations including the remaining stroma (Figure 5-4C), suggesting that these cells are expressing *Bmper* at a particularly high level and may be the main source of *Bmper* in the AGM region.

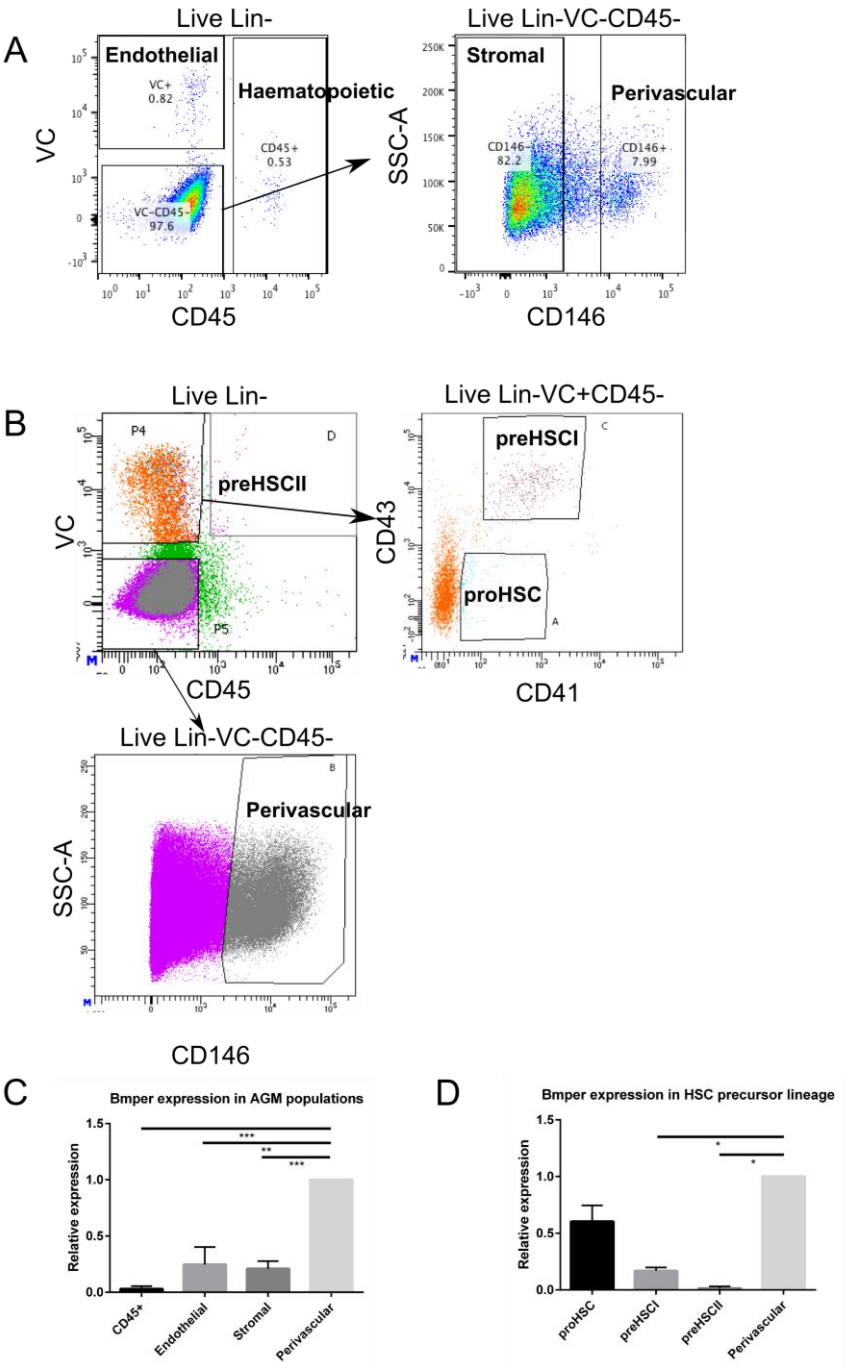


Figure 5-4 Expression of *Bmper* in sorted populations from the E10.5 AGM region

(A) Sorting strategy for isolation of AGM stromal populations: Lin$^-$CD45$^+$VC$^-$, representing haematopoietic cells; Lin$^-$CD45$^-$VC$^+$, endothelial cells; Lin$^-$CD45$^-$VC$^-$CD146$^+$, putative perivascular cells; and Lin$^-$CD45$^-$VC$^-$CD146$^-$, remaining stroma. Representative plots for three experiments. (B) Sorting strategy for isolation of AGM HSC precursor lineage populations: Lin-VC+CD45-CD43-CD41lo pro-HSC, Lin-VC+CD43+CD41+ Type I pre-HSC, and Lin-VC+CD45+ Type II pre-HSC and Lin$^-$CD45$^-$VC$^-$CD146$^+$, putative

perivascular cells. Representative plots for two experiments. Gatings were defined by fluorescence-minus-one (FMO) control samples. (C) Expression level of Bmper transcript relative to Tbp in each sorted population, normalised to expression in the $Lin^-CD45^-VC^-CD146^+$population (n=3). Significance calculated by t-test ** $p \leq 0.011$, *** $p \leq 0.001$). (D) Expression level of Bmper transcript relative to Tbp in each sorted populations, normalised to expression in perivascular population (n=2). * $p \leq 0.05$.

As BMPER protein was detected in intra-aortic clusters, which are believed to host pre-HSCs, and the markers used above would not have captured pre-HSCs in an enriched level, the expression of *Bmper* was tested in all populations of the pre-HSC lineage: $Lin^-VC^+CD45^-CD43^-CD41^{lo}$ pro-HSC, $Lin^-VC^+CD43^+CD41^+$ Type I pre-HSC, and $Lin^-VC^+CD45^+$ Type II pre-HSC and $Lin^-VC^-CD146^+$ cells to calibrate between experiments (Figure 5-4B). Again the VC^-CD146^+ population showed the highest level of *Bmper* expression compared to other populations, although there was moderate expression in the earliest precursors: pro-HSCs (Figure 5-4D). This expression enrichment suggests that the BMPER protein enrichment in the intra-aortic clusters is due to the uptake of BMPER into these cells rather than their own production. So BMPER protein is potentially acting in a non-cell autonomous fashion.

To further investigate the VC^-CD146^+ population that appear to express *Bmper* at higher levels than any other cell type, the distribution of these cells was tested in the AGM region by immunostaining. As previously described, the aorta is surrounded by a layer of cells which do not express endothelial markers but do express CD146, and are thus ascribed "perivascular cells" (Crisan et al., 2008; Sacchetti et al., 2007). However in the transverse sections, another population of $CD146^+$ cells, negative for endothelial markers was detected towards the urogenital ridges, in a more rounded structure and far from any endothelium. Sections of the embryos, taken along the length of the AGM region indicated two sources of non-endothelial and non-perivascular, $CD146^+$ cells towards the caudal end of the dorsal aorta (Figure 5-5A), and in tubule structures as the rostral end of the dorsal aorta (Figure 5-5B). Co-staining sections made at these two points with the nephric marker Pax2, suggested that these $CD146^+$ populations consist of metanephric mesenchyme

(Figure 5-5C) and mesonephric tubules (Figure 5-5D) but not the mesonephric duct. The relative contribution of mesonephric tissues and perivascular cells to the production of BMPER in the AGM region could not be easily determined from these immunostainings, and indeed they may both contribute.

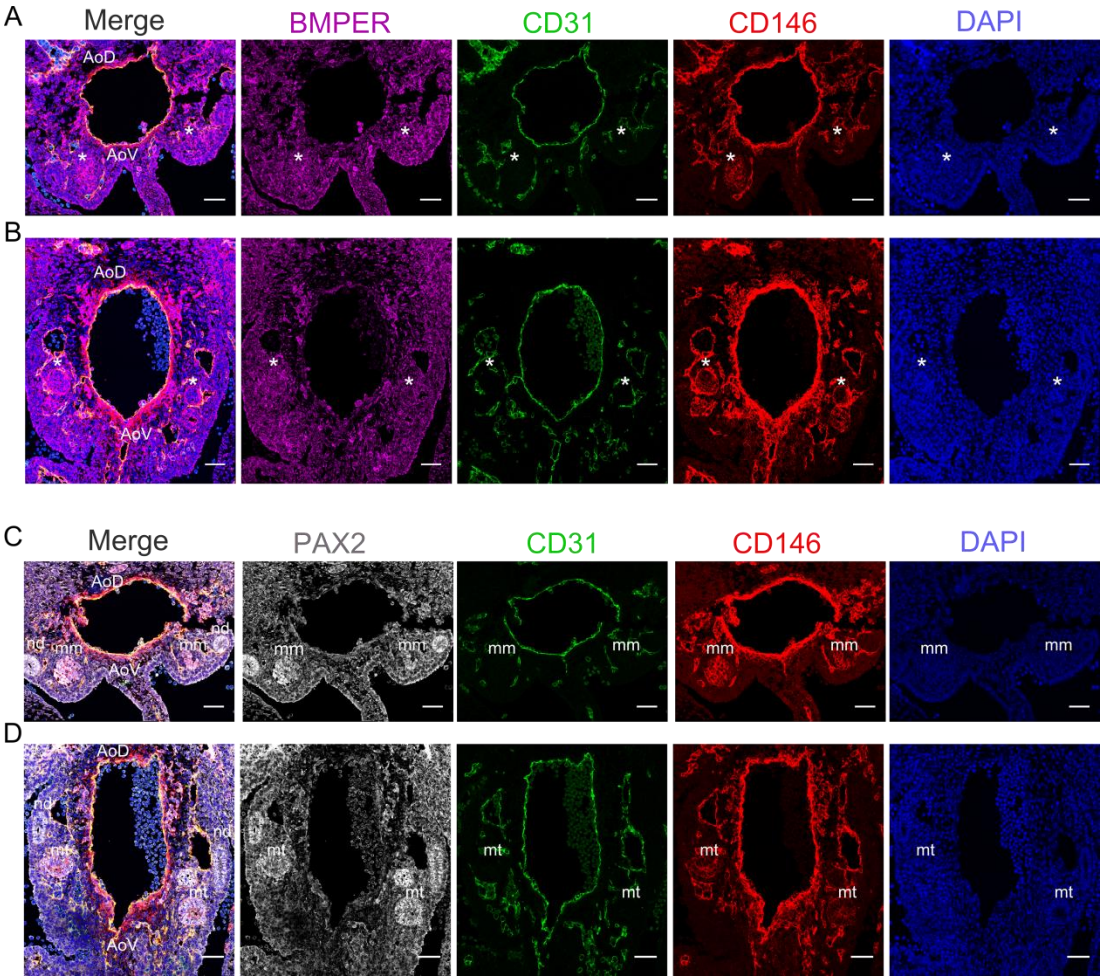


Figure 5-5 Comparison between BMPER distribution and the distribution of non-endothelial CD146$^+$ cells in the AGM region

(A) BMPER protein in CD146$^+$ cells around the dorsal aorta and urogenital ridges in the caudal part of AGM and (B) rostral part of AGM region measured by immunostaining. Red = CD146, Green = CD31, Magenta = BMPER, Blue = DAPI. Scale bar = 50 μm. * indicates clusters of non-endothelial, CD146$^+$ cells, ventrolateral to the AGM region (C) Co-staining PAX2 and CD146 validates that CD146 marks metanephric mesenchyme (mm) in the caudal AGM region and (D) mesonephric tubules (mt) in the rostral AGM region but not nephric duct (nd). Red = CD146, Green = CD31, Grey = PAX2, Blue = DAPI. Scale bar = 50 μm.

5.2.4 qRT-PCR, immunostaining and flow cytometry demonstrate that Bmper is expressed in OP9 cells and increases in level upon reaggregation

As *Bmper* expression had been detected from RNA-seq data of OP9 cells and was particularly up-regulated upon reaggregate culture, this expression was validated

by qRT-PCR. Adding in the time points 12h, 24h and 48h after reaggregation, and 0h (no reaggregation), *Bmper* was upregulated in OP9 from as early as 12h after reaggregation (Figure 5-6A), verifying that it is truly responsive to the reaggregation and/or culture method.

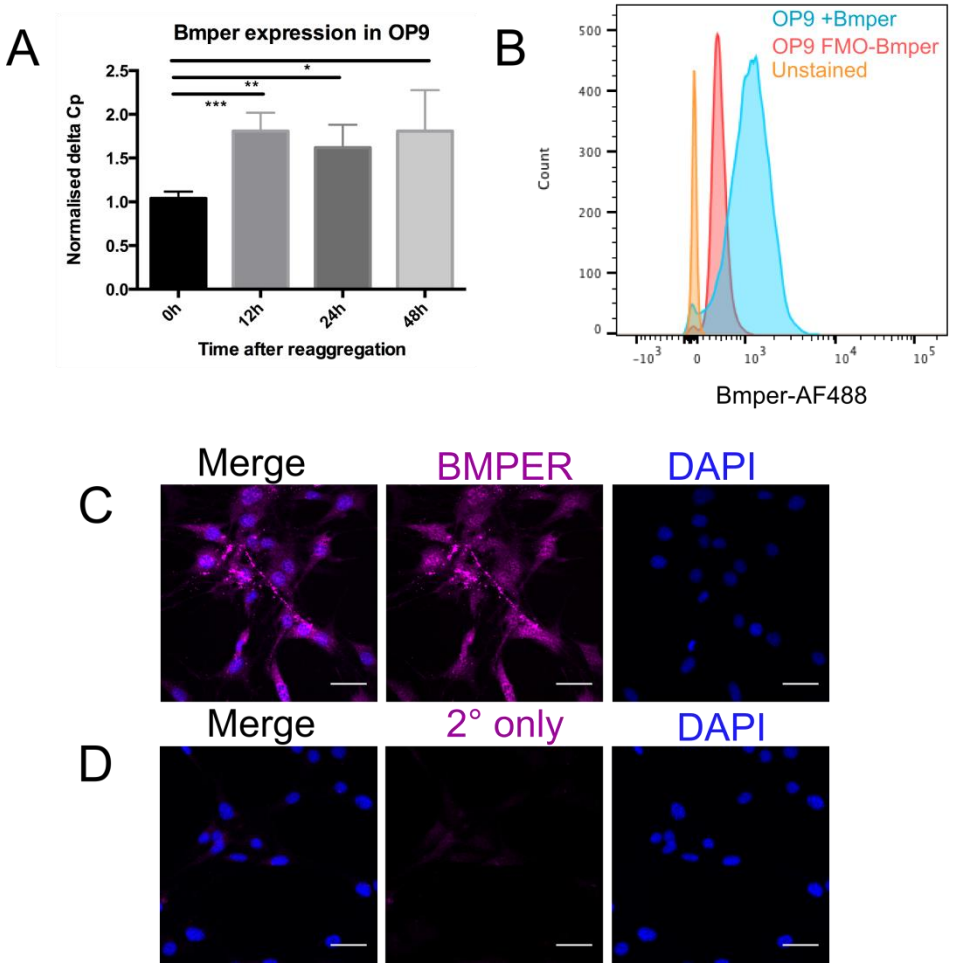


Figure 5-6 Bmper gene and protein expression OP9 cells

(A) Bmper transcript levels measured in OP9 cells after reaggregation for 0h, 12h, 24h, and 48h. Measured by qPCR relative to Tbp (n=4). *** p = 0.0005, ** p = 0.0055, * p = 0.0177. (B) BMPER positive OP9 cells measured by flow cytometry after intracellular staining with anti-BMPER antibody. (C) Intracellular and secreted BMPER from OP9 cells grown in submersed culture. (D) Secondary control immunostaining excludes anti-BMPER primary antibody. Magenta = BMPER, Blue = DAPI. Scale bar = 50 µm.

Expression of BMPER was then tested at the protein level by flow cytometry and by immunostaining (Figure 5-6B, C). All OP9 cells show a shift in fluorescence and all images showed positive BMPER staining (relative to secondary antibody-only control), suggesting that it is homogeneously expressed throughout this cell population. Further inspection of immuno-stained OP9 cells shows conglomerates of BMPER protein outside the cells, demonstrating its secretion (Figure 5-6C).

5.2.5 An anti-correlative distribution of pSMAD1/5/8 and BMPER staining suggests BMPER is acting as an inhibitor of BMP signalling in the intra-aortic clusters

Previous studies have suggested that BMPER can act as an antagonist or an agonist of BMP signalling depending on the context and level (Kelley et al., 2009; Serpe et al., 2008). Given the supportive role of BMPER in HSC maturation, the expression of BMP ligands in the AGM region, and the role of BMP signalling in HSC development it was deemed important to establish the role of BMPER in the AGM region.

A downstream target of the activated BMP pathway is the SMAD family of proteins. Following engagement of BMP ligands with the receptors BMPR1 and BMPR2, SMAD proteins 1, 5 and 8 become phosphorylated by the kinase domain of the receptor. After phosphorylation, these SMADs translocate into the nucleus, where they co-activate expression of target genes. Thus SMAD1/5/8 phosphorylation is a good indicator of BMP signalling activation, as it represents an early response which is independent of time-delays due, for example, to transcription or translation.

To establish the mechanism of BMPER on BMP signalling modulation, the distribution of BMPER protein was compared with the distribution of regions activated for BMP signalling (i.e. positive for pSMAD1/5/8). BMPER showed high staining in intra-aortic clusters. As has been shown previously (Souilhol et al., 2016a) and was validated here (Figure 5-7), these intra-aortic clusters have low pSMAD1/5/8 staining relative the underlying endothelium. Importantly the nuclei of these cells are completely lacking pSMAD1/5/8 staining.

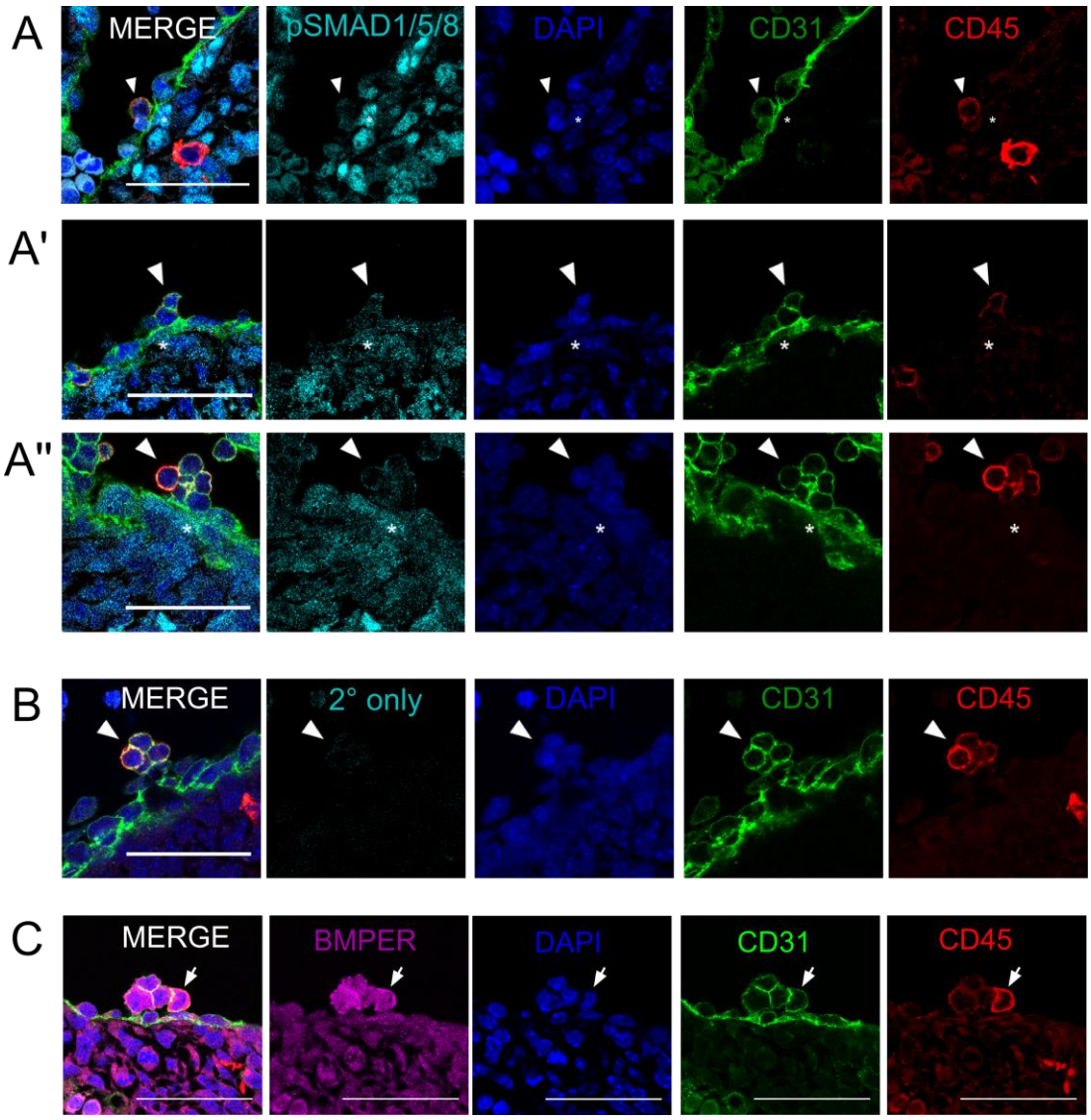


Figure 5-7 Co-staining intra-aortic clusters for BMPER and pSMAD1/5/8

(A), (A') and (A'') show three examples of the localisation of pSMAD1/5/8 around CD45+CD31+ haematopoietic clusters. (B) Secondary control immunostaining excludes anti-pSMAD1/5/8 primary antibody. Red = CD45, Green = CD31, cyan = pSMAD1/5/8, Blue = DAPI. (C) Localisation of BMPER in CD45+CD31+ haematopoietic clusters. Red = CD45, Green = CD31, Magenta = BMPER, Blue = DAPI. Scale bar = 50 μm.

The other cells with high BMPER signal, SSEA1 positive primordial germ cells, show low pSMAD1/5/8 signal compared to the surrounding cells and, importantly, no nuclear staining (Figure 5-8). In combination this suggests that high BMPER acts as an inhibitor of BMP signalling in the AGM region, which is in

keeping with previous functional demonstration that high BMP4 signalling is inhibitory to HSC maturation (Souilhol et al., 2016a).

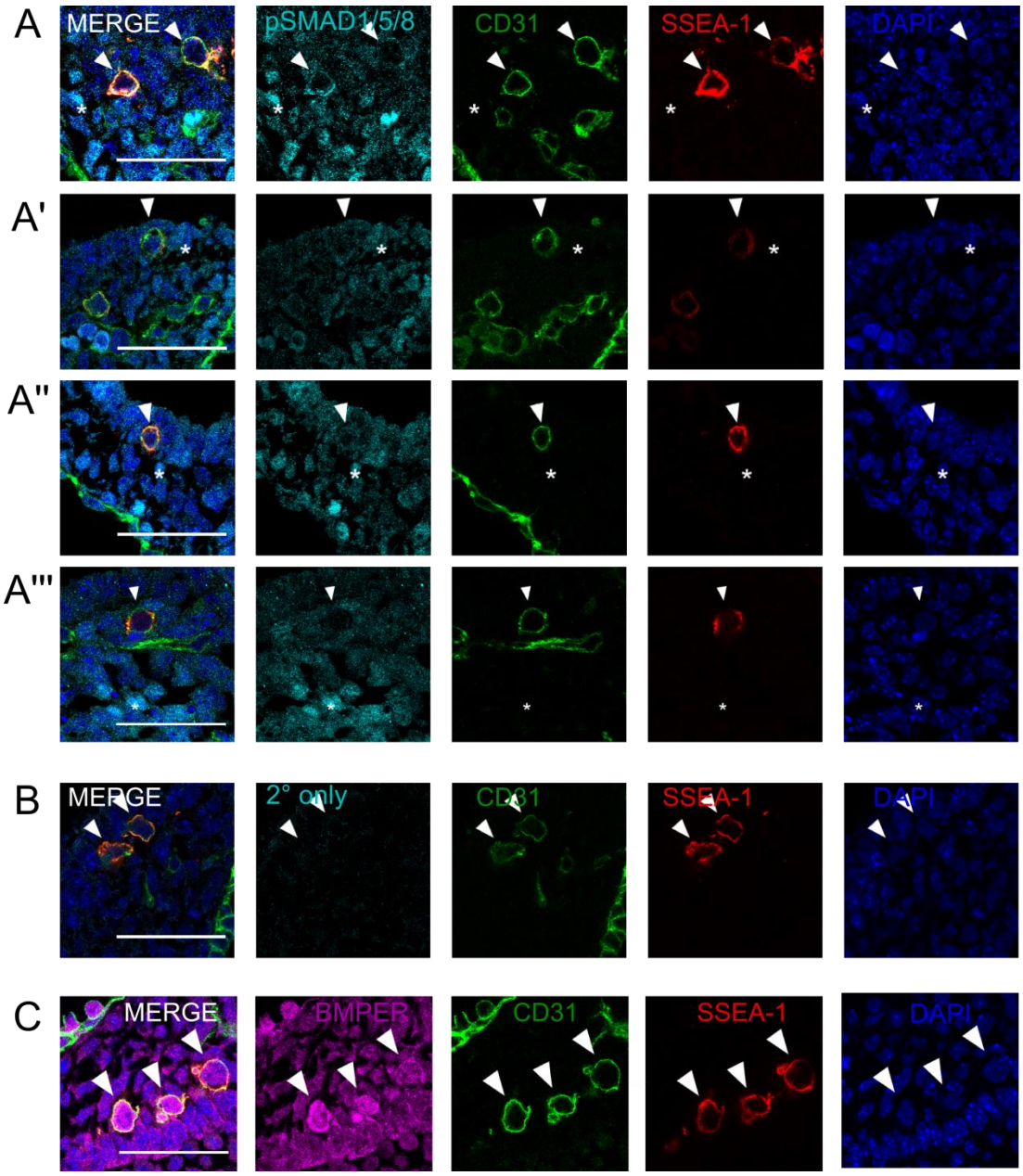


Figure 5-8 Co-staining primordial germ cells for BMPER and pSMAD1/5/8

(A), (A'), (A'') and (A''') Show four examples of the localisation of pSMAD1/5/8 around SSEA-1+CD31+ primordial germ cells. (B) Secondary control immunostaining excludes anti-pSMAD1/5/8 primary antibody. Red = SSEA-1, Green = CD31, cyan = pSMAD1/5/8, Blue = DAPI. (C) Localisation of BMPER in SSEA-1+CD31+ primordial germ cells. Red = SSEA-1, Green = CD31, Magenta = BMPER, Blue = DAPI. Scale bar = 50 um.

5.2.6 Expression analysis after AGM explant culture or OP9 culture with BMP4 suggests Bmper expression is induced by BMP4 activation

The potential model of BMPER as an inhibitor of BMP activity in the environment of developing HSCs prompted further investigation of whether there is a regulatory relationship between BMP4, which is expressed in the ventral side of the AGM (Durand et al., 2007; Marshall et al., 2000; Souilhol et al., 2016a) and BMPER. A comparison of the expression of *Bmp4* and *Bmper* across a range of cell types (Figure 5-9B) revealed a positive correlation between these two genes which tended to be expressed in the same cell types. As *Bmp4* is expressed at high levels in E9.5, E10.5 and E11.5 AGM regions, while *Bmper* is expressed only after E10.5 it was hypothesised that BMP4 might induce *Bmper* expression (Figure 5-9A).

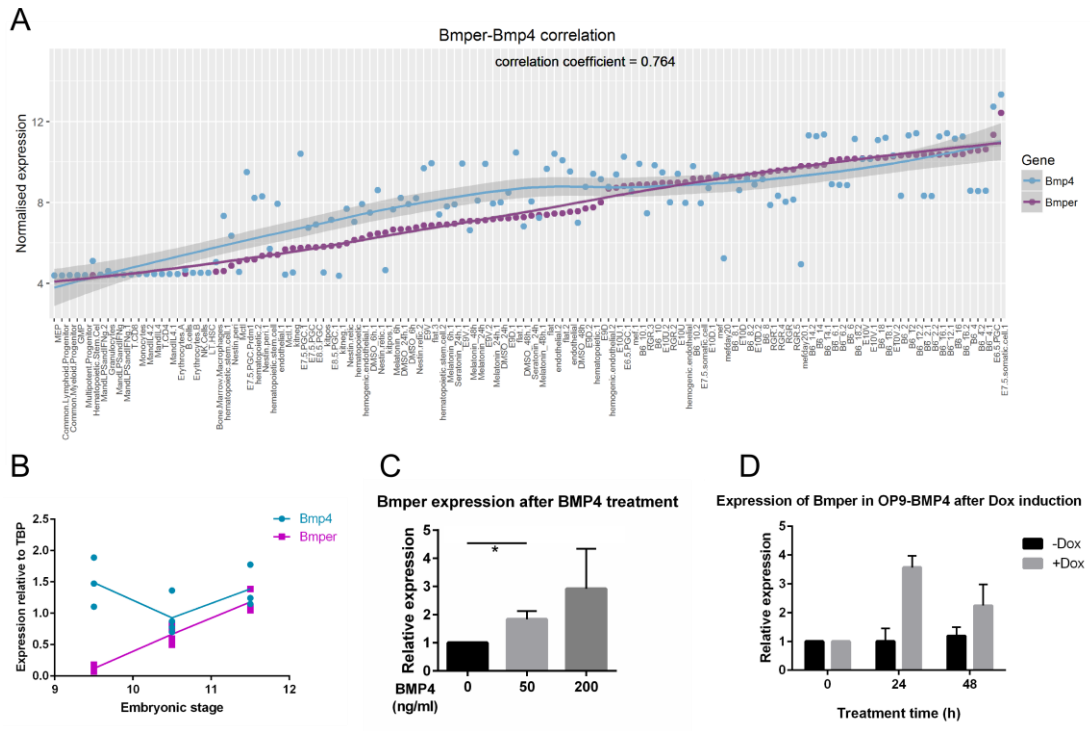


Figure 5-9 A regulatory relationship between Bmper and Bmp4

(A) Normalised expression of Bmper (purple) and Bmp4 (blue) across all 124 AGM and related niche cell type gene expression samples. Correlation coefficient is Spearman's correlation coefficient between *Bmper* and *Bmp4* across all datasets. Each point represents one embryo. (B) Expression of *Bmper* and *Bmp4* in dissected AGM at E9.5, E10.5 and E11.5. Expression is normalised to *Tbp*. (C) Expression of *Bmper* in E11.5 AGM explants after 24h culture with BMP4 at displayed dose, IMDM without cytokines or serum (n=2). * p = 0.0354. (D) Expression of *Bmper* in OP9-BMP4 after culture with doxycycline to induce *Bmp4* overexpression (n=2) Error bars represent the standard deviation from the mean.

To test a potentially regulatory role between BMP4 and Bmper, BMP4 was added to an explant culture of E11.5 AGM (in the absence of serum or cytokines) and after 24h the expression of *Bmper* was upregulated approximately 3-fold (Figure 5-9C). In another cell context, OP9 transfected with a doxycycline inducible *Bmp4* expression construct were treated with dox for 24h or 48h and the expression of *Bmper* measured (Figure 5-9D). Again *Bmper* expression was induced more than three-fold in the presence of BMP4 suggesting that there is a regulatory relationship between these molecules, and that *Bmper* expression may be induced as a consequence of high BMP4 levels.

5.3 Discussion

The results here validate the identification of a functional molecule, BMPER, from RNA-sequencing but additionally explore its distribution at higher resolution. The BMPER protein distribution is indeed ventrally polarised, but also enriched in the intra-aortic clusters, placing it directly in the niche in which HSCs are believed to mature. An additional interesting observation was the high enrichment of BMPER in the PGCs of the ventral mesenchyme, highlighting further potential parallels between these two stem cell systems, which are already known to share a dependence on SCF signalling,(Broudy, 1997; Dolci et al., 1991; Medvinsky et al., 1993; Rybtsov et al., 2014). The primary source of *Bmper* in the AGM region still remains unclear, but is apparently from a VC-CD146+ population which includes cells perivascular to the dorsal aorta, or from the ventro-lateral mesonephric tubules and metanephric mesenchyme, or both.

The distribution of BMPER at the transcript and protein level shows some discrepancy, as pre-HSC populations express very low levels of the transcript while the protein is relatively enriched in the intra-aortic clusters (putative pre-HSCs). This could be interpreted as an issue of antibody specificity, but the broad enrichment of anti-BMPER staining in the AoV relative to the AoD as well as the detection of protein secreted from OP9 cells is in agreement with transcriptional data. Given that BMPER is a secreted protein, the distribution of transcript and protein may be expected to differ and indeed BMPER has previously been shown *in vitro* to be selectively accumulated by certain cell types (Kelley et al., 2009).

The anti-correlative distribution of BMPER and pSMAD1/5/8 suggests that in the context of haematopoietic clusters and PGCs, BMPER is exerting an antagonistic affect towards BMP signalling, in agreement with the known inhibitory effect of high BMP4 signalling on HSC maturation (Souilhol et al., 2016a). Moreover the increase in expression following BMP4 treatment in OP9 and AGM reaggregates demonstrates that in mouse, like in zebrafish and drosophila (Rentzsch et al., 2006; Serpe et al., 2008), that Bmper expression is induced by BMP4.

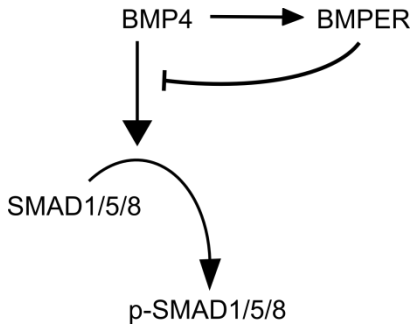


Figure 5-10 Proposed model of interaction between BMP4 and BMPER in the AGM

Represents the action of BMP4 to drive phosphorylation of pSMAD1/5/8 which then translocate the nucleus and promotes transcription. BMP4 also (through an unknown mechanism) promotes the transcription of Bmper which, in the intra-aortic clusters, inhibits the BMP4-driven phosphorylation of pSMAD1/5/8.

In summary, from the transcriptional profiling of the AGM region (Chapter 3), and functional validation of the capacity to support HSC maturation (Chapter 4), it has been possible in this chapter to investigate the potential mechanism of antagonism of BMP signalling by BMPER in facilitating HSC maturation. Furthermore, due to the induction of *Bmper* by BMP4, the model can be proposed that high BMP4 signalling just outside of the dorsal aorta endothelium promotes the expression of Bmper, which restricts the level of BMP signalling in the local pre-HSC environment, facilitating their maturation (Figure 5-10).

Chapter 6 Discussion

6.1 Summary

The main goal of this project was to improve our molecular understanding of the niche environment in which HSCs develop. I have described how a reference for this type of analysis was established through the production of a transcriptome-wide dataset of sub-domains of the AGM region by RNA-sequencing 3.2.1. By processing this data and identifying sets of genes specific to each subdomain of the AGM region I have tried to rationalise the large amount of data gathered into an interpretable format (sections 3.2.2 and 3.2.7). The enrichment of gene ontologies and pathways in agreement with existing understanding of the AGM regions (sections 3.2.2 and 3.2.3), suggests that this information represents a reliable reflection of the tissues.

Building on this approach a second RNA-seq dataset was produced from the supportive stromal cells – OP9 – cultured in a variety of conditions (section 3.2.5). By comparison of these conditions I identified the major transcriptional change when these cells are cultured in reaggregate conditions, and found that there was a significant commonality between this profile, and the *in vivo* profile of the E10.5 AoV (section 3.2.6).

Drawing together this data, I selected a number of potential candidates which, due to their temporal and spatial positioning, may have an active role in regulation of HSC specification (section 4.2.2). By adding these candidates to a culture system which captures the transition between HSC precursors and transplantable HSCs, I was able to identify a novel positive regulator of HSC maturation called Bmper (section 4.2.4). In a slightly different culture system, but one which still captures the maturation of HSC precursors (in this case from later HSC precursors, predominantly Type II pre-HSCs), Dr Celine Souilhol independently verified this positive regulatory role of Bmper (section 4.2.7).

Finally through characterisation of the gene-level and protein-level distribution of BMPER in the AGM region, I identified that the population expressing *Bmper* at most abundant levels is a VC-CD146+ population constituted by perivascular cells, metanephric mesenchyme and mesonephric tubules (section 5.2.3). Moreover, BMPER protein appears to accumulate in the intra-aortic clusters

and primordial germ cells, potentially indicating the main targets of its function (section 5.2.2). The distribution of pSMAD1/5/8 appears to indicate that BMPER is exerting an antagonistic effect on BMP signalling within the developing HSC niche (section 5.2.5), and indeed the upregulation of *Bmper* in the presence of BMP4 (section 5.2.6) suggests that this is part of an autoregulatory feedback system to maintain precise control of BMP signalling in the developing HSC environment (Figure 5-10).

6.2 Elucidating the signalling environment of the HSC niche

6.2.1 Expression profiling of the AGM region captures key signalling and cellular compositions of different functional domains

Transcriptional profiling was able to capture many dynamic changes within the AGM region during development. The signatures of spatially polarised tissues such as the neural tube (in AoD), nephric system (in UGR), and transient cell contributions such as PGCs (in E10.5 AoV) were evident from the tissue specific gene clusters (section 3.2.2). This provided a validation of the relevance of this dataset, and the potential to be used for discovery of previously unknown molecular changes in the AGM region.

Perhaps more importantly for the focus of this thesis was the identification of a number of signalling pathways that increase between E9.5 and E10.5 and have a degree of ventral polarisation: SCF signalling, the TGF-β superfamily family and a pro-inflammatory signalling family. These are in keeping with previous studies proposing their functional roles (Li et al., 2014; Rybtsov et al., 2014; Souilhol et al., 2016a), but also provide potential candidate ligands for functional analysis. In the case of the TGF-β superfamily, it was possible to verify the utility of this dataset in identifying of novel regulators of HSC development through the functional increase in HSCs driven by BMPER. Importantly, this highlighted a potential inhibitory branch of this pathway in the AGM region (discussed further in 6.5). Moreover some pro-inflammatory pathway members such as CXCL10 and CCL4 showed promising potential regulatory roles, but require further replication. As such, other members of the TGF-β superfamily, expressed in E10.5 AoV, such as BMP5, INHBA, and TGFB2 may be of interest for follow up studies.

Notably some signalling pathways, which are known to regulate HSC development, did not appear in this analysis. Although being able to promote HSC expansion in reaggregate culture (Taoudi et al., 2008), Il3 expression was not detected in the AGM region and its receptor Il3ra was not differentially expressed in any condition. Similarly Flt3l was not detected, while its receptor, Flt3 is expressed at a very low level and is ventrally polarised (Figure 8-2B and C). These could be expressed in very small numbers of cells which couldn't be captured amongst the large numbers of cells of the AGM region. Alternatively these signals may be found in the blood/serum and therefore wouldn't have been captured by the tissue dissection.

The role of Notch signalling, which is required for early stages of HSC specification but must be reduced to facilitate progression from Type I pre-HSCs to Type II pre-HSCs (Souilhol et al., 2016b), was also not apparent from this transcriptional data. Although the pathway is likely to be blocked at E10.5 to facilitate pre-HSC progression, none of the pathway molecules (except Dll1) were differentially expressed between E9.5 and E10.5 (Figure 8-2A). This may relate to the nature of Notch signalling as ligands are presented through direct cell-cell interactions so regulation may be very localised and therefore would not be apparent in analysis of the bulk population. Moreover any protein level regulation would not be detectable here. Interestingly, the Notch pathway was apparently up-regulated in OP9 cells after reaggregation, particularly due to upregulation of the receptors Notch 1-4 (Figure 3 13). This may indicate a role in non-cell autonomous regulation via the stroma although further investigation is required (discussed further in 6.3).

Similarly, the retinoic acid pathway did not show significant differential expression despite having been shown previously to regulate HSC maturation (Chanda et al., 2013). Such results highlight the limitation of this reverse genetic approach to identify regulators that change in correlation with functional changes. To identify regulators whose dynamics don't necessary match the functional changes, but are still essential, a forward screen would be necessary, whereby phenotypes of impaired haematopoiesis at E10.5 were traced to their genetic aberrations. This has, to some extent, been performed at the level of individual genes (discussed in section 1.5.1), but a transcriptome-wide functional screen will be challenging *in vivo* in

mouse. The development of CRISPR/Cas9 technology for efficient genetic modification may hold the potential to do this in the future (Chen et al., 2015).

Therefore the transcriptional data produced here (Chapter 3) was able to capture important signalling changes in the developing HSC environment, and novel pathway members which may be executing this function. Indeed, the ability to identify novel pathway members gave insight into the controversy of BMP's role in the AGM region suggesting that a number of BMP inhibitors are expressed in the ventral region at E10.5. While limitations in this approach mean that some signalling pathways were missed, this data is likely to yield further insightful signalling mechanisms upon additional functional screening.

6.2.2 Technology frontiers and the potential for future insight into HSC development

Matching the transcriptional profiling to the exact tissues which have been functionally assayed was an important way to define meaningful comparisons of expression and ultimately find functionally relevant molecules. However the AGM region is a complex mixture of cells types, and some information is lost in sequencing a bulk population (for example whether a signalling pathway is restricted to particular cell types, or diffuse across a whole region). An interesting, and potentially informative next step would be to deconvolute these bulk expression signatures i.e. partition the expression signature into subgroups based on cell type of origin. Deconvolution may facilitate insight into molecules that act together as a signalling pathway, and how transcriptional networks can be regulated by external signals. This is theoretically possible computationally, and to some extent the differential enrichment of gene sets or gene ontologies helps partition genes into groups which potentially act together (section 3.2.2). A more quantitative deconvolution is possible if marker profiles of the constituent cell types are known, or the relative proportions of cell mixtures are known (Gaujoux and Seoighe, 2012, 2013; Kuhn et al., 2011; Shen-Orr et al., 2010). In the case of the AGM region this is likely obscured by its poorly characterised composition by a mixture of cell types in unknown proportions. An alternative approach would be to derive full expression profiles of the constituent cell types of the AGM region and use these in

deconvolution of the average expression domains. Again this may be restricted by a limited understanding of the cellular constituents of the niche.

A more unbiased approach would be single-cell expression profiling of the entire niche, in which cell types could be defined by clustering their expression profiles. The feasibility of this is currently limited by throughput of single cell technology. Libraries still require a degree of manual preparation, so most current single-cell RNA-seq studies have measured cell numbers in the order of 100-1000 cells (Paul et al., 2015; Proserpio et al., 2016; Scialdone et al., 2016; Wilson et al., 2015). Since the AGM region is composed of approximately 50,000 cells, but within that has approximately 1 HSC and 50 pre-HSCs, a representative single-cell transcriptional profile may not be met by current technology. A new frontier in large scale library preparation on single cells is the use of nanoliter droplets in combination with DNA-barcoded beads which are able to prepare tens of thousands of libraries (Klein et al., 2015; Macosko et al., 2015). Hence single cell sequencing may be a promising future methodology in characterising the developing HSC niche.

An alternative means of deconvoluting the bulk expression profiles is to characterise the spatial compartmentalisation of gene expression. The AGM region, like most developmental structures, is a precisely spatially defined tissue. If expression information can be measured on a large scale, while retaining positional information, the local environment(s) in which HSC precursors reside may be observed as well as the communication between the many constituent cells that orchestrate this environment. Some approaches which may enable spatially resolved transcriptional profiling include laser-capture microdissection (Espina et al., 2006; Morton et al., 2014) and *in situ* sequencing (Crosetto et al., 2015; Ke et al., 2013; Lee et al., 2014).

A final limitation on the high resolution profiling of the HSC niche is the biological limitation of defining HSCs by phenotype. The use of FACS index sorting combined with single cell sequencing and functional study has recently provided an effective way to combine phenotypic information and functional assay (Wilson et al., 2015). If this could be combined with retention of spatial information, it might be key to resolving the precise location of the HSC niche.

6.3 The OP9 expression profile

The identification of a significant transcriptional shift in OP9 cells after reaggregation was notable, particularly as a number of regulators of HSC development were included in this, such as *Runx1*, *Bmper* and Notch receptors. From this study it was not possible to ascertain which aspect of the culture induces these transcriptional changes. Two possibilities are: the culture at the air-liquid interface, a common procedure for explants and reaggregates, compared to culture submersed in media; or the process of reaggregation in which cells undergo a degree of compressive forces. As Notch signalling is known to be induced by direct cell-cell interactions (Artavanis-Tsakonas et al., 1999), the latter could provide an explanation, as compression of cells into reaggregate may increase the number of cell-cell interactions.

A second striking aspect of the expression signature of OP9 cells was that the genes upregulated after reaggregation showed a significant overlap with the transcriptional signature of E10.5 AoV. This was not expected, as the two tissues/cell types have very different origins. A possible explanation, given the observation that *Runx1* was upregulated, and that *Runx1* can be induced by Notch signalling (Burns et al., 2005; Nakagawa et al., 2006; Richard et al., 2013), is that the activation of Notch pathway caused increased expression of *Runx1* which could elicit a haematopoietic-supportive transcriptional program similar to that of the AGM region, where *Runx1* is expressed in the ventral stroma (North et al., 2002). To what extent the common transcriptional program is necessary for the supportive function of OP9 would need to be interrogated further through knockouts/knockdowns of these genes. Moreover, whether the common transcriptional program is predominantly initiated by *Runx1* would be interesting to ascertain, as well as the contribution of *Bmper*, which was also upregulated in after reaggregation.

6.4 Screening secreted factors in reaggregate and assay limitations

In this study the use of reaggregate cultures followed by transplantation allowed the testing of a number of candidates for their role in HSC maturation. This culture system has been used before to demonstrate the functional role of secreted

regulators in the AGM region such as SCF, BMP4, Noggin and Shh (Rybtsov et al., 2014; Souilhol et al., 2016a), however the approach in this thesis differed in that it used a more data-driven selection of candidates for functional screening. Although this enabled the identification of Bmper as a regulator of HSC development, a number of other candidates such as Cxcl10 showed potential effects on HSC development, but would need further replication to assess significance as there was a high degree of variation in the system. Hence, this reaggregate culture system is a useful, tractable assay for study of molecular regulator, but has a number of undefined aspects which introduce unaccountable variations that limit the conclusions that can be drawn. The requirement for foetal calf serum to support maturation from early precursors represents one of these undefined elements. A further investigation of the characteristics of serums that support this culture may yield significant insight, as well as clarify the currently unknown interactions between the recombinant proteins being tested and serum proteins.

A further limitation of this screening approach is that the application of high levels of a recombinant protein may not be representative of the physiological situation. To robustly define the essential physiological regulators of HSC maturation, a knockout/knockdown approach could be used with OP9 cells (although the heterogeneity of the OP9 population would have to be captured). On the other hand, if the goal of the study is to find the minimal requirements to support the maturation of HSCs *in vitro*, the physiological relevance of the conditions may be considered a secondary issue.

Ultimately the transplantation assay poses the most significant limitation to the throughput of screens of HSC maturation. In the absence of phenotypic characteristics that can truly correlate with transplantation output, the study of molecular regulators remains, time-consuming, costly, and inherently variable.

6.5 The functional role of Bmper in HSC maturation

This study showed a positive role on HSC maturation by Bmper in two culture systems: 7-day culture of E9.5 caudal part with serum, OP9 and SCF/IL3/FLT3L; and 5-day culture of E11.5 AoD in the absence of serum, OP9 or cytokines. However as both of these cultures likely support a multi-step process of

maturation (Rybtsov et al., 2011, 2014, 2016; Taoudi et al., 2008), it is not possible to infer from these experiments, which aspect of HSC maturation is modulated by Bmper. This would be interesting to follow up with timed addition of the protein, to establish whether regulation is in an early or late stage of HSC maturation.

The production of HSCs from E11.5 AoD in the presence of BMPER (Figure 4-8), is a notably similar phenotype to the increase in production of HSCs from E11.5 AoD when cultured with E11.5 AoV (Souilhol et al., 2016a Figure 2C) or when E11.5 AoD is cultured with Noggin (Souilhol et al., 2016a Figure 7D). This suggests that Bmper may be one of the mediators of support of HSC maturation from AoV at E11.5, although it may have some functional redundancy with other proteins such as Noggin. Whether this is acting through proliferation of early precursors, induction of maturation of precursors, or relief of a repressive signalling environment is not clear from these experiments. This could be addressed to some extent by a newly established limiting dilution assay, in which the number of pre-HSCs can be compared with the number of HSCs (Rybtsov et al., 2016). If the number of pre-HSCs is increasing as well as the number of HSCs, then the mechanism may be increasing the *de novo* specification of pre-HSCs or increasing their proliferation. If the number of pre-HSCs remains the same, but HSC numbers are increased, then the mechanism is either inducing the final maturation step, de-repression or proliferation of HSCs. This may be an important step in understanding the regulation of the multistep maturation process.

6.6 Bmper as a modulator of BMP signalling

In this study, the regulatory effect of BMPER during HSC maturation appears to be via antagonism of BMP signalling, particularly in the intra-aortic clusters, as part of a feedback mechanism to BMP4 signalling (Figure 5-10). These findings support the argument that BMP4 antagonises HSC development, and the similarity to the action of Noggin in supporting HSC maturation from E11.5 AoD (Souilhol et al., 2016a) suggests that BMP antagonism is an important process regulated by multiple molecules. The antagonistic role of BMPER could in future be supported with functional experiments, to test whether BMPER is able to rescue the inhibitory effect of addition of BMP4 to cultures of E10.5 and E11.5 AGM region.

The likely antagonistic effect of BMPER in BMP signalling in the intra-aortic clusters does not exclude the possibility that it could be eliciting a biphasic role in regulating BMP signalling within the AGM region. This biphasic control of BMP signalling has been modelled as an important way of producing a sharp boundary between cells that receive a very high and very low levels of a signal (Kelley et al., 2009; Serpe et al., 2008; Umulis et al., 2006). In the AGM region, this may be both a temporal and a spatial boundary (Figure 6-1). For example, between E9.5 and E10.5 there may be a rapid switch from pro-BMP to anti-BMP signalling (particularly in the pre-HSC environment) (Figure 6-1A). As well as this, in the E10.5 AGM region, there may be a spatial boundary between the intra-aortic clusters and the underlying mesenchyme. BMPER could act to enhance the BMP signalling response in the ventral sub-aortic mesenchyme, where BMP4 is high, but towards the luminal budding intra-aortic clusters the differing BMP4-BMPER stoichiometry may induce endocytosis of BMP4, abrogating its effect (Figure 6-1B). However a lack of clarity of the spatial distribution of pre-HSCs precludes a full understanding of the importance of the spatial boundary.

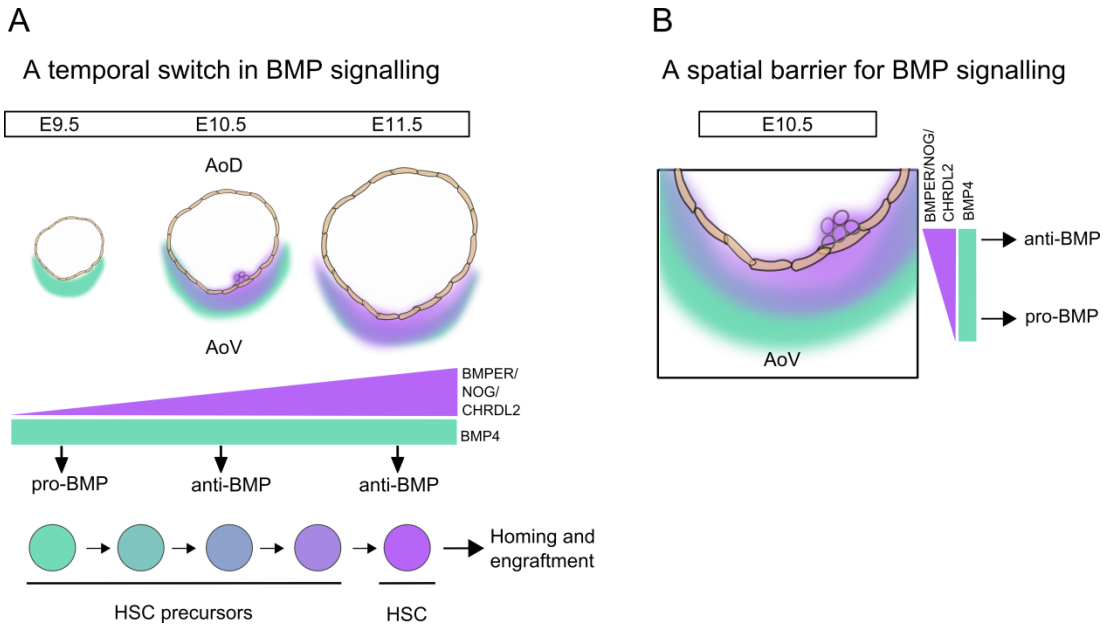


Figure 6-1 The interaction between BMP4 and BMPER creates temporal and spatial boundaries between BMP signalling environments

(A) Schematic representation of the increasing ventral expression of BMP antagonists Bmper, Noggin, Chrdl2 in the AoV between E9.5 and E11.5, while the expression of Bmp4 remains relatively constant. The balance of BMP agonists and antagonists therefore shifts to

create an anti-BMP environment (particularly in the intra-aortic clusters) between E9.5 and E10.5. Therefore the HSC precursor lineage experiences a transition to anti-BMP signalling environment as they progress developmentally. (B) The distribution of BMPER and BMP4 in the E10.5 AoV also creates spatially restricted zones of pro-BMP signalling (in the sub-aortic mesenchyme) and anti-BMP signalling (in the intra-aortic clusters).

If pre-HSCs rapidly transition between an area of high BMP signalling to low BMP signalling, this could reconcile the apparently contradictory findings that HSCs in the AGM region have a history of BMP activation (Crisan et al., 2015), but high BMP4 inhibits HSC maturation (Souilhol et al., 2016a). A possible early requirement for BMP signalling in the AGM region may be explained by the capacity of BMP4 to regulate *Scl* and *Runx1* expression (Mead et al., 1998; Pimanda et al., 2007), which could initiate the haematopoietic program. BMP4 has also been shown to regulate the essential HSC growth factor SCF (Dudley et al., 2007). Functional experiments, for example chimeric embryos with a mixture of Smad or Bmp receptor knockout cells could help distinguish between a necessary or incidental role for BMP signalling in the AGM region.

The effect, seen in adult, that antagonism of BMP signalling leads to increased HSC homing via induction of Cxcl12 (Khurana et al., 2014), could also explain the importance of BMP inhibition in the AGM region, particularly as a mechanism in the latter stages of HSC maturation. Indeed the antagonism of BMP signalling may be the mechanism by which pre-HSCs embedded in the endothelium acquire the capacity to home to the adult niche. In agreement with this, expression data from this thesis shows that Cxcl12 expression dramatically increases between E9.5 and E10.5 concomitant with the increase in Bmper expression (Figure 8-2).

At a broader level, the relative contribution of BMPER to haematopoiesis *in vivo* remains unclear. Previous studies of homozygous knockout mice show lethality immediately after birth (Ikeya et al., 2006), suggesting that BMPER is not essential to haematopoietic development as embryos survive beyond the stage of initiation of definitive haematopoiesis. Likely, the modulation of BMP signalling by BMPER is co-regulated by multiple molecules (e.g. Noggin, Chordin or other BMP modulators) which can provide some functional redundancy. Therefore the *in vivo* understanding

of the requirement for BMPER, or indeed BMP antagonism, to permit HSC maturation may require combinatorial knockouts of several genes.

6.7 Prospects for directed differentiation

Given the capacity of BMPER to increase the efficiency of maturation of transplantable HSCs, it may be a suitable molecule to test in protocols for the directed differentiation of ES cells towards HSCs. BMP4 has previously been applied to embryoid body culture to produce haematopoietic cells (Chadwick et al., 2003). In the light of the study here and recent work (Souilhol et al., 2016a), the role of BMP signalling in HSC maturation appears to be complex. The application of BMP may be transiently required but must be followed by inhibition at the final stages, thus it may be appropriate to apply BMPER at the latter stages of the differentiation protocol.

Other molecules from the expression data produced in this study may also be effective molecules in directing differentiation towards the HSC fate. However due to the complex nature of signalling within the AGM region stroma and pre-HSC lineage, it will be important to first establish signalling mechanisms that have a direct functional effect on HSC precursor maturation.

6.8 Summary and perspectives

The study of HSC ontogeny faces a number of challenges, namely the rarity of HSCs and the requirement for their assay by transplantation; the complexity of the AGM environment with regard to both the cell constituents and signalling landscape; the multistep process of lineage progression; and a lack of a fully defined culture system to capture this process. Yet the constant developments in methods for molecular, biochemical, cellular and bioinformatic approaches to the study of developmental haematopoiesis are facilitating continuous progress in our understanding.

This thesis provides a more comprehensive characterisation of the functional aspects of the AGM region and OP9 cells than has been previously performed, which should provide a useful reference for future research into the signalling that drives

HSC maturation. Although the technology used here was not able to capture the full heterogeneity of the niche, the application of more recent technological developments (discussed in section 6.2.2) which facilitate more precise, high-resolution, profiling will be an important advance of this work to refine the transcriptional signatures generated from the AGM region subdomains.

In this thesis, I have demonstrated the capacity to identify novel regulators of HSC development, such as BMPER, by combining transcriptional profiling and a targeted functional assay. Although limited in its throughput and subject to inherent variation, continued screening of candidates from transcriptional profiles is likely to identify further novel regulators that can improve HSC maturation in *ex vivo* reaggregate culture and may be applicable to protocols for differentiation of ES cells towards an HSC fate. In this manner, the goals of deriving HSCs from alternative sources *in vitro* to improve clinical studies and potential therapeutic interventions into haematopoietic diseases may hopefully be attainable in the not too distant future.

Chapter 7 References

Agarwal, A., Koppstein, D., Rozowsky, J., Sboner, A., Habegger, L., Hillier, L.W., Sasidharan, R., Reinke, V., Waterston, R.H., and Gerstein, M. (2010). Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. BMC Genomics *11*, 383.

Aggarwal, B.B., Gupta, S.C., and Kim, J.H. (2012). Historical perspectives on tumor necrosis factor and its superfamily: 25 years later, a golden journey. Blood *119*, 651–665.

Akagi, K., Suzuki, T., Stephens, R.M., Jenkins, N.A., and Copeland, N.G. (2004). RTCGD: retroviral tagged cancer gene database. Nucleic Acids Res. *32*, D523–D527.

Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics *22*, 1600–1607.

Alexa A and Rahnenfuhrer J (2016). topGO: Enrichment Analysis for Gene Ontology. R Package Version 2240.

Alford, A.I., and Hankenson, K.D. (2006). Matricellular proteins: Extracellular modulators of bone development, remodeling, and regeneration. Bone *38*, 749–757.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Aplan, P.D., Lombardi, D.P., Ginsberg, A.M., Cossman, J., Bertness, V.L., and Kirsch, I.R. (1990). Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. Science *250*, 1426–1429.

Ara, T., Tokoyoda, K., Sugiyama, T., Egawa, T., Kawabata, K., and Nagasawa, T. (2003). Long-term hematopoietic stem cells require stromal cell-derived factor-1 for colonizing bone marrow during ontogeny. Immunity *19*, 257–267.

Armstrong, L., Stojkovic, M., Dimmick, I., Ahmad, S., Stojkovic, P., Hole, N., and Lako, M. (2004). Phenotypic Characterization of Murine Primitive Hematopoietic Progenitor Cells Isolated on Basis of Aldehyde Dehydrogenase Activity. STEM CELLS *22*, 1142–1151.

Armulik, A., Genové, G., and Betsholtz, C. (2011). Pericytes: Developmental, Physiological, and Pathological Perspectives, Problems, and Promises. Dev. Cell *21*, 193–215.

Artavanis-Tsakonas, S., Rand, M.D., and Lake, R.J. (1999). Notch Signaling: Cell Fate Control and Signal Integration in Development. Science *284*, 770–776.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Atkins, H.L., Bowman, M., Allan, D., Anstee, G., Arnold, D.L., Bar-Or, A., Bence-Bruckler, I., Birch, P., Bredeson, C., Chen, J., et al. (2016). Immunoablation and autologous haemopoietic stem-cell transplantation for aggressive multiple sclerosis: a multicentre single-group phase 2 trial. Lancet Lond. Engl. *388*, 576–585.

Axelrad, A.A. (1974). Properties of cells that produce erythrocytic colonies in vitro. Hemopoiesis Cult.

Baldridge, M.T., King, K.Y., Boles, N.C., Weksberg, D.C., and Goodell, M.A. (2010). Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. Nature *465*, 793–797.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. *41*, D991–D995.

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

Baxter, R.C., and Martin, J.L. (1986). Radioimmunoassay of growth hormone-dependent insulinlike growth factor binding protein in human plasma. J. Clin. Invest. *78*, 1504–1512.

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.

Becker, A.J., Mcculloch, E.A., Siminovitch, L., and Till, J.E. (1965). The Effect of Differing Demands for Blood Cell Production on DNA Synthesis by Hemopoietic Colony-Forming Cells of Mice. Blood *26*, 296–308.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.

Bentley, D.R. (2006). Whole-genome re-sequencing. Curr. Opin. Genet. Dev. *16*, 545–552.

Bertrand, J.Y., Kim, A.D., Teng, S., and Traver, D. (2008). CD41+ cmyb+ precursors colonize the zebrafish pronephros by a novel migration route to initiate adult hematopoiesis. Development *135*, 1853–1862.

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

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.

Bianco, P., Cao, X., Frenette, P.S., Mao, J.J., Robey, P.G., Simmons, P.J., and Wang, C.-Y. (2013). The meaning, the sense and the significance: translating the science of mesenchymal stem cells into medicine. Nat. Med. *19*, 35–42.

Binnerts, M.E., Wen, X., Canté-Barrett, K., Bright, J., Chen, H.-T., Asundi, V., Sattari, P., Tang, T., Boyle, B., Funk, W., et al. (2004). Human Crossveinless-2 is a novel inhibitor of bone morphogenetic proteins. Biochem. Biophys. Res. Commun. *315*, 272–280.

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.

Bokui, N., Otani, T., Igarashi, K., Kaku, J., Oda, M., Nagaoka, T., Seno, M., Tatematsu, K., Okajima, T., Matsuzaki, T., et al. (2008). Involvement of MAPK signaling molecules and Runx2 in the NELL1-induced osteoblastic differentiation. FEBS Lett. *582*, 365–371.

Bouleftour, W., Boudiffa, M., Wade-Gueye, N.M., Bouët, G., Cardelli, M., Laroche, N., Vanden-Bossche, A., Thomas, M., Bonnelye, E., Aubin, J.E., et al. (2014). Skeletal Development of Mice Lacking Bone Sialoprotein (BSP) - Impairment of Long Bone Growth and Progressive Establishment of High Trabecular Bone Mass. PLoS ONE *9*, e95144.

Bowie, M.B., McKnight, K.D., Kent, D.G., McCaffrey, L., Hoodless, P.A., and Eaves, C.J. (2006). Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. J. Clin. Invest. *116*, 2808–2816.

Bowie, M.B., Kent, D.G., Dykstra, B., McKnight, K.D., McCaffrey, L., Hoodless, P.A., and Eaves, C.J. (2007a). Identification of a new intrinsically timed developmental checkpoint that reprograms key hematopoietic stem cell properties. Proc. Natl. Acad. Sci. U. S. A. *104*, 5878–5882.

Bowie, M.B., Kent, D.G., Copley, M.R., and Eaves, C.J. (2007b). Steel factor responsiveness regulates the high self-renewal phenotype of fetal hematopoietic stem cells. Blood *109*, 5043–5048.

Bradley, T.R., and Metcalf, D. (1966). The growth of mouse bone marrow cells in vitro. Immunol. Cell Biol. *44*, 287–300.

Brasel, K., McKenna, H.J., Morrissey, P.J., Charrier, K., Morris, A.E., Lee, C.C., Williams, D.E., and Lyman, S.D. (1996). Hematologic effects of flt3 ligand in vivo in mice. Blood *88*, 2004–2012.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Bresnick, E.H., Katsumura, K.R., Lee, H.-Y., Johnson, K.D., and Perkins, A.S. (2012). Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. Nucleic Acids Res. gks281.

Broudy, V.C. (1997). Stem Cell Factor and Hematopoiesis. Blood *90*, 1345–1364.

Brown, L., Cheng, J.T., Chen, Q., Siciliano, M.J., Crist, W., Buchanan, G., and Baer, R. (1990). Site-specific recombination of the tal-1 gene is a common occurrence in human T cell leukemia. EMBO J. *9*, 3343–3351.

Bruce, W.R., Meeker, B.E., and Valeriote, F.A. (1966). Comparison of the Sensitivity of Normal Hematopoietic and Transplanted Lymphoma Colony-Forming Cells to Chemotherapeutic Agents Administered In Vivo. J. Natl. Cancer Inst. *37*, 233–245.

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

de Bruijn, M.F.T.R., 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.

Burda, P., Laslo, P., and Stopka, T. (2010). The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. Leukemia *24*, 1249–1257.

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.

Buza-Vidas, N., Cheng, M., Duarte, S., Charoudeh, H.N., Jacobsen, S.E.W., and Sitnicka, E. (2009). FLT3 receptor and ligand are dispensable for maintenance and posttransplantation expansion of mouse hematopoietic stem cells. Blood *113*, 3453–3460.

Bystry, R.S., Aluvihare, V., Welch, K.A., Kallikourdis, M., and Betz, A.G. (2001). B cells and professional APCs recruit regulatory T cells via CCL4. Nat. Immunol. *2*, 1126–1132.

Calvi, L.M., Adams, G.B., Weibrecht, K.W., Weber, J.M., Olson, D.P., Knight, M.C., Martin, R.P., Schipani, E., Divieti, P., Bringhurst, F.R., et al. (2003). Osteoblastic cells regulate the haematopoietic stem cell niche. Nature *425*, 841–846.

Capron, C., Lécluse, Y., Kaushik, A.L., Foudi, A., Lacout, C., Sekkai, D., Godin, I., Albagli, O., Poullion, I., Svinartchouk, F., et al. (2006). The SCL relative LYL-1 is required for fetal and adult hematopoietic stem cell function and B-cell differentiation. Blood *107*, 4678–4686.

Caselli, A., Olson, T.S., Otsuru, S., Chen, X., Hofmann, T.J., Nah, H.-D., Grisendi, G., Paolucci, P., Dominici, M., and Horwitz, E.M. (2013). IGF-1-mediated osteoblastic niche expansion enhances long-term hematopoietic stem cell engraftment after murine bone marrow transplantation. Stem Cells Dayt. Ohio *31*, 2193–2204.

Chadwick, K., Wang, L., Li, L., Menendez, P., Murdoch, B., Rouleau, A., and Bhatia, M. (2003). Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. Blood *102*, 906–915.

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, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544–547.

Charbord, P., Pouget, C., Binder, H., Dumont, F., Stik, G., Levy, P., Allain, F., Marchal, C., Richter, J., Uzan, B., et al. (2014). A Systems Biology Approach for Defining the Molecular Framework of the Hematopoietic Stem Cell Niche. Cell Stem Cell *15*, 376–391.

Charron, F., and Nemer, M. (1999). GATA transcription factors and cardiac development. Semin. Cell Dev. Biol. *10*, 85–91.

Chen, W.V., and Maniatis, T. (2013). Clustered protocadherins. Dev. Camb. Engl. *140*, 3297–3302.

Chen, X.D., and Turpen, J.B. (1995). Intraembryonic origin of hepatic hematopoiesis in Xenopus laevis. J. Immunol. *154*, 2557–2567.

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, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R., et al. (2015). Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. Cell *160*, 1246–1260.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell *133*, 1106–1117.

Chinwalla, A.T., Cook, L.L., Delehaunty, K.D., Fewell, G.A., Fulton, L.A., Fulton, R.S., Graves, T.A., Hillier, L.W., Mardis, E.R., McPherson, J.D., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Cho, S.-H., and Cepko, C.L. (2006). Wnt2b/beta-catenin-mediated canonical Wnt signaling determines the peripheral fates of the chick eye. Dev. Camb. Engl. *133*, 3167–3177.

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.

Chow, A., Lucas, D., Hidalgo, A., Méndez-Ferrer, S., Hashimoto, D., Scheiermann, C., Battista, M., Leboeuf, M., Prophete, C., van Rooijen, N., et al. (2011). Bone marrow CD169+ macrophages promote the retention of hematopoietic stem and progenitor cells in the mesenchymal stem cell niche. J. Exp. Med. *208*, 261–271.

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.

Ciau-Uitz, A., Walmsley, M., and Patient, R. (2000). Distinct Origins of Adult and Embryonic Blood in Xenopus. Cell *102*, 787–796.

Coffinier, C., Ketpura, N., Tran, U., Geissert, D., and De Robertis, E.M. (2002). Mouse Crossveinless-2 is the vertebrate homolog of a Drosophila extracellular regulator of BMP signaling. Mech. Dev. *119*, *Supplement*, S179–S184.

Coles, E., Christiansen, J., Economou, A., Bronner-Fraser, M., and Wilkinson, D.G. (2004). A vertebrate crossveinless 2 homologue modulates BMP activity and neural crest cell migration. Development *131*, 5309–5317.

Compeau, P.E.C., Pevzner, P.A., and Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. Nat. Biotechnol. *29*, 987–991.

Conley, C.A., Silburn, R., Singer, M.A., Ralston, A., Rohwer-Nutter, D., Olson, D.J., Gelbart, W., and Blair, S.S. (2000). Crossveinless 2 contains cysteine-rich domains and is required for high levels of BMP-like activity during the formation of the cross veins in Drosophila. Development *127*, 3947–3959.

Consortium, I.H.G.S. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931–945.

Consortium, T.G.O. (2015). Gene Ontology Consortium: going forward. Nucleic Acids Res. *43*, D1049–D1056.

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

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561–563.

Crisan, M., Yap, S., Casteilla, L., Chen, C.-W., Corselli, M., Park, T.S., Andriolo, G., Sun, B., Zheng, B., Zhang, L., et al. (2008). A Perivascular Origin for Mesenchymal Stem Cells in Multiple Human Organs. Cell Stem Cell *3*, 301–313.

Crisan, M., Kartalaei, P.S., Vink, C.S., 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.

Crosetto, N., Bienko, M., and van Oudenaarden, A. (2015). Spatially resolved transcriptomics and beyond. Nat. Rev. Genet. *16*, 57–66.

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.

Davidson, A.J. (2008). Mouse kidney development. In StemBook, (Cambridge (MA): Harvard Stem Cell Institute), p.

Díaz-Flores, L., Gutiérrez, R., Madrid, J.F., Varela, H., Valladares, F., Acosta, E., Martín-Vasallo, P., and Díaz-Flores, L. (2009). Pericytes. Morphofunction, interactions and pathology in a quiescent and activated mesenchymal cell niche. Histol. Histopathol. *24*, 909–969.

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

Dieterlen-Lièvre, F., and Martin, C. (1981). Diffuse intraembryonic hemopoiesis in normal and chimeric avian development. Dev. Biol. *88*, 180–191.

Dieterlen-Lièvre, F., Pouget, C., Bollérot, K., and Jaffredo, T. (2006). Are intra-aortic hemopoietic cells derived from endothelial cells during ontogeny? Trends Cardiovasc. Med. *16*, 128–139.

Ding, L., Saunders, T.L., Enikolopov, G., and Morrison, S.J. (2012). Endothelial and perivascular cells maintain haematopoietic stem cells. Nature *481*, 457–462.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dolci, S., Williams, D.E., Ernst, M.K., Resnick, J.L., Brannan, C.I., Lock, L.F., Lyman, S.D., Boswell, H.S., and Donovan, P.J. (1991). Requirement for mast cell growth factor for primordial germ cell survival in culture. Nature *352*, 809–811.

Dorner, B.G., Scheffold, A., Rolph, M.S., Hüser, M.B., Kaufmann, S.H.E., Radbruch, A., Flesch, I.E.A., and Kroczek, R.A. (2002). MIP-1α, MIP-1β, RANTES, and ATAC/lymphotactin function together with IFN-γ as type 1 cytokines. Proc. Natl. Acad. Sci. *99*, 6181–6186.

Dressler, G.R. (2006). The Cellular Basis of Kidney Development. Annu. Rev. Cell Dev. Biol. *22*, 509–529.

Dudley, B.M., Runyan, C., Takeuchi, Y., Schaible, K., and Molyneaux, K. (2007). BMP signaling regulates PGC numbers and motility in organ culture. Mech. Dev. *124*, 68–77.

Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science *344*, 1156–1160.

Durand, C., Robin, C., Bollerot, K., Baron, M.H., Ottersbach, K., and Dzierzak, E. (2007). Embryonic stromal clones reveal developmental regulators of definitive hematopoietic stem cells. Proc. Natl. Acad. Sci. U. S. A. *104*, 20838–20843.

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

Eaves, C.J. (2015). Hematopoietic stem cells: concepts, definitions, and the new reality. Blood *125*, 2605–2613.

Ehira, N., Oshiumi, H., Matsumoto, M., Kondo, T., Asaka, M., and Seya, T. (2010). An embryo-specific expressing TGF-β family protein, growth-differentiation factor 3 (GDF3), augments progression of B16 melanoma. J. Exp. Clin. Cancer Res. *29*, 135.

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

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

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., et al. (2012). Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. PLOS ONE *7*, e47768.

Eppig, J.T., Blake, J.A., Bult, C.J., Kadin, J.A., Richardson, J.E., and Group, T.M.G.D. (2015). The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic Acids Res. *43*, D726–D736.

Espina, V., Wulfkuhle, J.D., Calvert, V.S., VanMeter, A., Zhou, W., Coukos, G., Geho, D.H., Petricoin, E.F., and Liotta, L.A. (2006). Laser-capture microdissection. Nat. Protoc. *1*, 586–603.

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.

Essers, M.A.G., Offner, S., Blanco-Bose, W.E., Waibler, Z., Kalinke, U., Duchosal, M.A., and Trumpp, A. (2009). IFNα activates dormant haematopoietic stem cells in vivo. Nature *458*, 904–908.

Ewing, B., and Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. Genome Res. *8*, 186–194.

Ferkowicz, M.J., and Yoder, M.C. (2005). Blood island formation: longstanding observations and modern interpretations. Exp. Hematol. *33*, 1041–1047.

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. Dev. Camb. Engl. *130*, 4393–4403.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell *11*, 477–490.

Fitch, S.R., Kimber, G.M., Wilson, N.K., Parker, A., Mirshekar-Syahkal, B., Göttgens, B., Medvinsky, A., Dzierzak, E., and Ottersbach, K. (2012). Signaling from the sympathetic nervous system regulates hematopoietic stem cell emergence during embryogenesis. Cell Stem Cell *11*, 554–566.

Fleischman, R.A., Custer, R.P., and Mintz, B. (1982). Totipotent hematopoietic stem cells: Normal self-renewal and differentiation after transplantation between mouse fetuses. Cell *30*, 351–359.

Ford, C.E., Hamerton, J.L., Barnes, D.W.H., and Loutit, J.F. (1956). Cytological Identification of Radiation-Chimæras. Nature *177*, 452–454.

Gao, J., Graves, S., Koch, U., Liu, S., Jankovic, V., Buonamici, S., El Andaloussi, A., Nimer, S.D., Kee, B.L., Taichman, R., et al. (2009). Hedgehog signaling is dispensable for adult hematopoietic stem cell function. Cell Stem Cell *4*, 548–558.

Garcia-Porrero, J.A., Godin, I.E., and Dieterlen-Lièvre, F. (1995). Potential intraembryonic hemogenic sites at pre-liver stages in the mouse. Anat. Embryol. (Berl.) *192*, 425–435.

Gaujoux, R., and Seoighe, C. (2012). Semi-supervised Nonnegative Matrix Factorization for gene expression deconvolution: a case study. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. *12*, 913–921.

Gaujoux, R., and Seoighe, C. (2013). CellMix: a comprehensive toolbox for gene expression deconvolution. Bioinformatics *29*, 2211–2212.

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.

Gentleman R, Carey V, Huber W and Hahne F (2016). genefilter: genefilter: methods for filtering genes from high-throughput experiments. R Package Version 1542.

Gering, M., and Patient, R. (2005). Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos. Dev. Cell *8*, 389–400.

Gilks, C.B., Bear, S.E., Grimes, H.L., and Tsichlis, P.N. (1993). Progression of interleukin-2 (IL-2)-dependent rat T cell lymphoma lines to IL-2-independent growth following activation of a gene (Gfi-1) encoding a novel zinc finger protein. Mol. Cell. Biol. *13*, 1759–1768.

Gilliland, D.G., and Griffin, J.D. (2002). The roles of FLT3 in hematopoiesis and leukemia. Blood *100*, 1532–1542.

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.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Granito, R.N., Bouleftour, W., Sabido, O., Lescale, C., Thomas, M., Aubin, J.E., Goodhardt, M., Vico, L., and Malaval, L. (2015). Absence of bone sialoprotein (BSP) alters profoundly hematopoiesis and upregulates osteopontin. J. Cell. Physiol. *230*, 1342–1351.

Grün, D., and van Oudenaarden, A. (2015). Design and Analysis of Single-Cell Sequencing Experiments. Cell *163*, 799–810.

Hadland, B.K., Huppert, S.S., Kanungo, J., Xue, Y., Jiang, R., Gridley, T., Conlon, R.A., Cheng, A.M., Kopan, R., and Longmore, G.D. (2004). A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. Blood *104*, 3097–3105.

Halbritter, F., Vaidya, H.J., and Tomlinson, S.R. (2012). GeneProf: analysis of high-throughput sequencing experiments. Nat. Methods *9*, 7–8.

Halbritter, F., Kousa, A.I., and Tomlinson, S.R. (2014). GeneProf data: a resource of curated, integrated and reusable high-throughput genomics experiments. Nucleic Acids Res. *42*, D851-858.

Halbritter, F., (2012). Genome-Scale Transcriptomic and Epigenomic Analysis of Stem Cells.

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R., et al. (2009). Oct4 and LIF/Stat3

Additively Induce Krüppel Factors to Sustain Embryonic Stem Cell Self-Renewal. Cell Stem Cell *5*, 597–609.

Hebenstreit, D., Fang, M., Gu, M., Charoensawan, V., Oudenaarden, A. van, and Teichmann, S.A. (2011). RNA sequencing reveals two major classes of gene expression levels in metazoan cells. Mol. Syst. Biol. *7*, 497.

Helbing, T., Rothweiler, R., Ketterer, E., Goetz, L., Heinke, J., Grundmann, S., Duerschmied, D., Patterson, C., Bode, C., and Moser, M. (2011). BMP activity controlled by BMPER regulates the proinflammatory phenotype of endothelium. Blood *118*, 5040–5049.

Helbing, T., Herold, E.-M., Hornstein, A., Wintrich, S., Heinke, J., Grundmann, S., Patterson, C., Bode, C., and Moser, M. (2013). Inhibition of BMP activity protects epithelial barrier function in lung injury. J. Pathol. *231*, 105–116.

Hess, D.A., Meyerrose, T.E., Wirthlin, L., Craft, T.P., Herrbrich, P.E., Creer, M.H., and Nolta, J.A. (2004). Functional characterization of highly purified human hematopoietic repopulating cells isolated according to aldehyde dehydrogenase activity. Blood *104*, 1648–1655.

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

Hock, H., Meade, E., Medeiros, S., Schindler, J.W., Valk, P.J.M., Fujiwara, Y., and Orkin, S.H. (2004b). Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. Genes Dev. *18*, 2336–2341.

Hodgson, G.S., and Bradley, T.R. (1979). Properties of haematopoietic stem cells surviving 5-fluorouracil treatment: evidence for a pre-CFU-S cell? Nature *281*, 381–382.

Hoggatt, J., Kfoury, Y., and Scadden, D.T. (2016). Hematopoietic Stem Cell Niche in Health and Disease. Annu. Rev. Pathol. Mech. Dis. *11*, 555–581.

Hu, Y., and Smyth, G.K. (2009). ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J. Immunol. Methods *347*, 70–78.

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

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol. *19*, 342–347.

Ikeya, M., Kawada, M., Kiyonari, H., Sasai, N., Nakao, K., Furuta, Y., and Sasai, Y. (2006). Essential pro-Bmp roles of crossveinless 2 in mouse organogenesis. Development *133*, 4463–4473.

Ikuta, K., and Weissman, I.L. (1992). Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation. Proc. Natl. Acad. Sci. *89*, 1502–1506.

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.

Jacobson, L.O., Simmons, E.L., Marks, E.K., and Eldredge, J.H. (1951). Recovery from Radiation Injury. Science *113*, 510–511.

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

Jain, R.K. (2003). Molecular regulation of vessel maturation. Nat. Med. *9*, 685–693.

Jaroscak, J., Goltry, K., Smith, A., Waters-Pick, B., Martin, P.L., Driscoll, T.A., Howrey, R., Chao, N., Douville, J., Burhop, S., et al. (2003). Augmentation of umbilical cord blood (UCB) transplantation with ex vivo–expanded UCB cells: results of a phase 1 trial using the AastromReplicell System. Blood *101*, 5061–5067.

Jeannet, G., Scheller, M., Scarpellino, L., Duboux, S., Gardiol, N., Back, J., Kuttler, F., Malanchi, I., Birchmeier, W., Leutz, A., et al. (2008). Long-term, multilineage hematopoiesis occurs in the combined absence of β-catenin and γ-catenin. Blood *111*, 142–149.

Jones, R.J., Wagner, J.E., Celano, P., Zicha, M.S., and Sharkis, S.J. (1990). Separation of pluripotent haematopoietic stem cells from spleen colony-forming cells. Nature *347*, 188–189.

Kamimura, M., Matsumoto, K., Koshiba-Takeuchi, K., and Ogura, T. (2004). Vertebrate crossveinless 2 is secreted and acts as an extracellular modulator of the BMP signaling cascade. Dev. Dyn. *230*, 434–445.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. *44*, D457-462.

Kartalaei, P.S., Yamada-Inagawa, T., Vink, C.S., Pater, E. de, Linden, R. van der, Marks-Bluth, J., Sloot, A. van der, Hout, M. van den, Yokomizo, T., Schaick-Solernó, M.L. van, 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.

Kaufman, D.S., Hanson, E.T., Lewis, R.L., Auerbach, R., and Thomson, J.A. (2001). Hematopoietic colony-forming cells derived from human embryonic stem cells. Proc. Natl. Acad. Sci. *98*, 10716–10721.

Kawakami, Y., Capdevila, J., Büscher, D., Itoh, T., Rodríguez Esteban, C., and Izpisúa Belmonte, J.C. (2001). WNT signals control FGF-dependent limb initiation and AER induction in the chick embryo. Cell *104*, 891–900.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods *10*, 857–860.

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.

Kelley, R., Ren, R., Pi, X., Wu, Y., Moreno, I., Willis, M., Moser, M., Ross, M., Podkowa, M., Attisano, L., et al. (2009). A concentration-dependent endocytic trap and sink mechanism converts Bmper from an activator to an inhibitor of Bmp signaling. J. Cell Biol. *184*, 597–609.

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.

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

Khurana, S., Melacarne, A., Yadak, R., Schouteden, S., Notelaers, T., Pistoni, M., Maes, C., and Verfaillie, C.M. (2014). SMAD Signaling Regulates CXCL12 Expression in the Bone Marrow Niche, Affecting Homing and Mobilization of Hematopoietic Progenitors. STEM CELLS *32*, 3012–3022.

Kiel, M.J., Yilmaz, O.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.

Kiel, M.J., Acar, M., Radice, G.L., and Morrison, S.J. (2009). Hematopoietic stem cells do not depend on N-cadherin to regulate their maintenance. Cell Stem Cell *4*, 170–179.

Kieusseian, A., Brunet de la Grange, P., Burlen-Defranoux, O., Godin, I., and Cumano, A. (2012). Immature hematopoietic stem cells undergo maturation in the fetal liver. Dev. Camb. Engl. *139*, 3521–3530.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kim, I., Saunders, T.L., and Morrison, S.J. (2007). Sox17 Dependence Distinguishes the Transcriptional Regulation of Fetal from Adult Hematopoietic Stem Cells. Cell *130*, 470–483.

Kim, P.G., Canver, M.C., Rhee, C., Ross, S.J., Harriss, J.V., Tu, H.-C., Orkin, S.H., Tucker, H.O., and Daley, G.Q. (2016). Interferon-α signaling promotes embryonic HSC maturation. Blood *128*, 204–216.

Kirstetter, P., Anderson, K., Porse, B.T., Jacobsen, S.E.W., and Nerlov, C. (2006). Activation of the canonical Wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. Nat. Immunol. *7*, 1048–1056.

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

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell *161*, 1187–1201.

Koch, U., Wilson, A., Cobas, M., Kemler, R., MacDonald, H.R., and Radtke, F. (2008). Simultaneous loss of β- and γ-catenin does not perturb hematopoiesis or lymphopoiesis. Blood *111*, 160–164.

Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., et al. (2015). ArrayExpress update--simplifying data submissions. Nucleic Acids Res. *43*, D1113-1116.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. Mol. Cell *58*, 610–620.

Korchynskyi, O., and Dijke, P. ten (2002). Identification and Functional Characterization of Distinct Critically Important Bone Morphogenetic Protein-specific Response Elements in the Id1 Promoter. J. Biol. Chem. *277*, 4883–4891.

Kuhn, A., Thu, D., Waldvogel, H.J., Faull, R.L.M., and Luthi-Carter, R. (2011). Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. Nat. Methods *8*, 945–947.

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

Lakshmanan, G., Lieuw, K.H., Lim, K.-C., Gu, Y., Grosveld, F., Engel, J.D., and Karis, A. (1999). Localization of Distant Urogenital System-, Central Nervous System-, and Endocardium-Specific Transcriptional Regulatory Elements in the GATA-3 Locus. Mol. Cell. Biol. *19*, 1558–1568.

Lampugnani, M.G., Resnati, M., Raiteri, M., Pigott, R., Pisacane, A., Houen, G., Ruco, L.P., and Dejana, E. (1992). A novel endothelial-specific membrane protein is a marker of cell-cell contacts. J. Cell Biol. *118*, 1511–1522.

Lancrin, C., Mazan, M., Stefanska, M., Patel, R., Lichtinger, M., Costa, G., Vargel, Ö., Wilson, N.K., Möröy, 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.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

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.

Lantz, C.S., Boesiger, J., Song, C.H., Mach, N., Kobayashi, T., Mulligan, R.C., Nawa, Y., Dranoff, G., and Galli, S.J. (1998). Role for interleukin-3 in mast-cell and basophil development and in immunity to parasites. Nature *392*, 90–93.

Lassila, O., Eskola, J., Toivanen, P., Martin, C., and Dieterlen-Lievre, F. (1978). The origin of lymphoid stem cells studied in chick yolk sac–embryo chimaeras. Nature *272*, 353–354.

Lassila, O., Martin, C., Toivanen, P., and Dieterlen-Lievre, F. (1982). Erythropoiesis and lymphopoiesis in the chick yolk-sac-embryo chimeras: contribution of yolk sac and intraembryonic stem cells. Blood *59*, 377–381.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly Multiplexed Subcellular RNA Sequencing in Situ. Science *343*, 1360–1363.

Leinonen, R., Sugawara, H., and Shumway, M. (2011). The Sequence Read Archive. Nucleic Acids Res. *39*, D19–D21.

Lessard, J., and Sauvageau, G. (2003). Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature *423*, 255–260.

Levine, A.J., and Brivanlou, A.H. (2006). GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. Dev. Camb. Engl. *133*, 209–216.

Levine, A.J., Levine, Z.J., and Brivanlou, A.H. (2009). GDF3 is a BMP inhibitor that can activate Nodal signaling only at very high doses. Dev. Biol. *325*, 43–48.

Levine, J.E., Uberti, J.P., Ayash, L., Reynolds, C., Ferrara, J.L.M., Silver, S.M., Braun, T., Yanik, G., Hutchinson, R., and Ratanatharathorn, V. (2003). Lowered-

intensity preparative regimen for allogeneic stem cell transplantation delays acute graft-versus-host disease but does not improve outcome for advanced hematologic malignancy. Biol. Blood Marrow Transplant. *9*, 189–197.

Li, S., Miao, T., Sebastian, M., Bhullar, P., Ghaffari, E., Liu, M., Symonds, A.L.J., and Wang, P. (2012). The transcription factors Egr2 and Egr3 are essential for the control of inflammation and antigen-induced proliferation of B and T cells. Immunity *37*, 685–696.

Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C.-C., and Kain, S.R. (1998). Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter. J. Biol. Chem. *273*, 34970–34975.

Li, Y., McClintick, J., Zhong, L., Edenberg, H.J., Yoder, M.C., and Chan, R.J. (2005). Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. Blood *105*, 635–637.

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.

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

Ling, K.-W., Ottersbach, K., Hamburg, J.P. van, 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.

Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. Nat. Genet. *21*, 20–24.

Liu, Y., Zhou, J., and White, K.P. (2013). RNA-seq differential expression studies: more sequence or more replication? Bioinforma. Oxf. Engl.

Lo Celso, C., Fleming, H.E., Wu, J.W., Zhao, C.X., Miake-Lye, S., Fujisaki, J., Côté, D., Rowe, D.W., Lin, C.P., and Scadden, D.T. (2009). Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. Nature *457*, 92–96.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. *14*, 1675–1680.

Lorenz, E., Uphoff, D., Reid, T.R., and Shelton, E. (1951). Modification of irradiation injury in mice and guinea pigs by bone marrow injections. J. Natl. Cancer Inst. *12*, 197–201.

Luis, T.C., Weerkamp, F., Naber, B.A.E., Baert, M.R.M., Haas, E.F.E. de, Nikolic, T., Heuvelmans, S., Krijger, R.R.D., Dongen, J.J.M. van, and Staal, F.J.T. (2009). Wnt3a deficiency irreversibly impairs hematopoietic stem cell self-renewal and leads to defects in progenitor cell differentiation. Blood *113*, 546–554.

Lund, S.A., Giachelli, C.M., and Scatena, M. (2009). The role of osteopontin in inflammatory processes. J. Cell Commun. Signal. *3*, 311–322.

Luster, A.D., and Ravetch, J.V. (1987). Biochemical characterization of a gamma interferon-inducible cytokine (IP-10). J. Exp. Med. *166*, 1084–1097.

Luster, A.D., Unkeless, J.C., and Ravetch, J.V. (1985). Gamma-interferon transcriptionally regulates an early-response gene containing homology to platelet proteins. Nature *315*, 672–676.

M, O. (1999). Stochastic model revisited. Int. J. Hematol. *69*, 2–5.

Ma, X., Robin, C., Ottersbach, K., and Dzierzak, E. (2002). The Ly-6A (Sca-1) GFP Transgene is Expressed in all Adult Mouse Hematopoietic Stem Cells. STEM CELLS *20*, 514–521.

Mach, N., Lantz, C.S., Galli, S.J., Reznikoff, G., Mihm, M., Small, C., Granstein, R., Beissert, S., Sadelain, M., Mulligan, R.C., et al. (1998). Involvement of Interleukin-3 in Delayed-Type Hypersensitivity. Blood *91*, 778–783.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Maéno, M., Tochinai, S., and Katagiri, C. (1985). Differential participation of ventral and dorsolateral mesoderms in the hemopoiesis of Xenopus, as revealed in diploid-triploid or interspecific chimeras. Dev. Biol. *110*, 503–508.

Magli, M.C., Iscove, N.N., and Odartchenko, N. (1982). Transient nature of early haematopoietic spleen colonies. Nature *295*, 527–529.

Maglott, D., Ostell, J., Pruitt, K.D., and Tatusova, T. (2005). Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. *33*, D54–D58.

Maillard, I., Koch, U., Dumortier, A., Shestova, O., Xu, L., Sai, H., Pross, S.E., Aster, J.C., Bhandoola, A., Radtke, F., et al. (2008). Canonical notch signaling is dispensable for the maintenance of adult hematopoietic stem cells. Cell Stem Cell *2*, 356–366.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376–380.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. *18*, 1509–1517.

Marshall, C.J., Kinnon, C., and Thrasher, A.J. (2000). Polarized expression of bone morphogenetic protein-4 in the human aorta-gonad-mesonephros region. Blood *96*, 1591–1593.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell *11*, 491–504.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J. *32*, 2561–2574.

Martin, J.L., and Baxter, R.C. (1986). Insulin-like growth factor-binding protein from human plasma. Purification and characterization. J. Biol. Chem. *261*, 8754–8760.

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.

Mather, J.P., Moore, A., and Li, R.H. (1997). Activins, inhibins, and follistatins: further thoughts on a growing family of regulators. Proc. Soc. Exp. Biol. Med. Soc. Exp. Biol. Med. N. Y. N *215*, 209–222.

Matthews, D.C., Appelbaum, F.R., Eary, J.F., Fisher, D.R., Durack, L.D., Hui, T.E., Martin, P.J., Mitchell, D., Press, O.W., Storb, R., et al. (1999). Phase I study of I-131-anti-CD45 antibody plus cyclophosphamide and total body irradiation for advanced acute leukemia and myelodysplastic syndrome. Blood *94*, 1237–1247.

Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. *74*, 560–564.

Maximow, A.A. (1924). RELATION OF BLOOD CELLS TO CONNECTIVE TISSUES AND ENDOTHELIUM. Physiol. Rev. *4*, 533–563.

McKenna, H.J., Stocking, K.L., Miller, R.E., Brasel, K., Smedt, T.D., Maraskovsky, E., Maliszewski, C.R., Lynch, D.H., Smith, J., Pulendran, B., et al. (2000). Mice lacking flt3 ligand have deficient hematopoiesis affecting hematopoietic progenitor cells, dendritic cells, and natural killer cells. Blood *95*, 3489–3497.

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.

McSweeney, P.A., Niederwieser, D., Shizuru, J.A., Sandmaier, B.M., Molina, A.J., Maloney, D.G., Chauncey, T.R., Gooley, T.A., Hegenbart, U., Nash, R.A., et al. (2001). Hematopoietic cell transplantation in older patients with hematologic malignancies: replacing high-dose cytotoxic therapy with graft-versus-tumor effects. Blood *97*, 3390–3400.

Mead, P.E., Kelley, C.M., Hahn, P.S., Piedad, O., and Zon, L.I. (1998). SCL specifies hematopoietic mesoderm in Xenopus embryos. Development *125*, 2611–2620.

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

Medvinsky, A., Taoudi, S., Mendes, S., and Dzierzak, E. (2007). Analysis and Manipulation of Hematopoietic Progenitor and Stem Cells from Murine Embryonic Tissues. In Current Protocols in Stem Cell Biology, (John Wiley & Sons, Inc.), p.

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

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

Medvinsky, A.L., Gan, O.I., Semenova, M.L., and Samoylina, N.L. (1996). Development of day-8 colony-forming unit-spleen hematopoietic progenitors during early murine embryogenesis: spatial and temporal mapping. Blood *87*, 557–566.

Meirelles, L. da S., Chagastelles, P.C., and Nardi, N.B. (2006). Mesenchymal stem cells reside in virtually all post-natal organs and tissues. J. Cell Sci. *119*, 2204–2213.

Mendes, S.C., Robin, C., and Dzierzak, E. (2005). Mesenchymal progenitor cells localize within hematopoietic sites throughout ontogeny. Development *132*, 1127–1136.

Méndez-Ferrer, S., Michurina, T.V., Ferraro, F., Mazloom, A.R., MacArthur, B.D., Lira, S.A., Scadden, D.T., Ma'ayan, A., Enikolopov, G.N., and Frenette, P.S. (2010). Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. Nature *466*, 829–834.

Metcalf, D., MacDonald, H.R., Odartchenko, N., and Sordat, B. (1975). Growth of mouse megakaryocyte colonies in vitro. Proc. Natl. Acad. Sci. U. S. A. *72*, 1744–1748.

Mikheyev, A.S., and Tin, M.M.Y. (2014). A first look at the Oxford Nanopore MinION sequencer. Mol. Ecol. Resour. *14*, 1097–1102.

Mikkola, H.K.A., Klintman, J., Yang, H., Hock, H., Schlaeger, T.M., Fujiwara, Y., and Orkin, S.H. (2003). Haematopoietic stem cells retain long-term repopulating

activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. Nature *421*, 547–551.

Miles, C., Sanchez, M.J., Sinclair, A., and Dzierzak, E. (1997). Expression of the Ly-6E.1 (Sca-1) transgene in adult hematopoietic stem cells and the developing mouse embryo. Dev. Camb. Engl. *124*, 537–547.

Milner, L.A., and Bigas, A. (1999). Notch as a Mediator of Cell Fate Determination in Hematopoiesis: Evidence and Speculation. Blood *93*, 2431–2448.

Minehata, K., Mukouyama, Y., Sekiguchi, T., Hara, T., and Miyajima, A. (2002). Macrophage colony stimulating factor modulates the development of hematopoiesis by stimulating the differentiation of endothelial cells in the AGM region. Blood *99*, 2360–2368.

Mirshekar-Syahkal, B., Fitch, S.R., and Ottersbach, K. (2014). Concise Review: From Greenhouse to Garden: The Changing Soil of the Hematopoietic Stem Cell Microenvironment During Development. STEM CELLS *32*, 1691–1700.

Mitchell, R., Szabo, E., Shapovalova, Z., Aslostovar, L., Makondo, K., and Bhatia, M. (2014). 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.

Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., and Ohki, M. (1991). t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. Proc. Natl. Acad. Sci. *88*, 10431–10434.

Molkentin, J.D. (2000). The Zinc Finger-containing Transcription Factors GATA-4, -5, and -6 UBIQUITOUSLY EXPRESSED REGULATORS OF TISSUE-SPECIFIC GENE EXPRESSION. J. Biol. Chem. *275*, 38949–38952.

Molyneaux, K.A., Stallock, J., Schaible, K., and Wylie, C. (2001). Time-lapse analysis of living mouse germ cell migration. Dev. Biol. *240*, 488–498.

Molyneaux, K.A., Zinszner, H., Kunwar, P.S., Schaible, K., Stebler, J., Sunshine, M.J., O'Brien, W., Raz, E., Littman, D., Wylie, C., et al. (2003). The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. Dev. Camb. Engl. *130*, 4279–4286.

Monteiro, R.M., de Sousa Lopes, S.M.C., Bialecka, M., de Boer, S., Zwijsen, A., and Mummery, C.L. (2008). Real time monitoring of BMP Smads transcriptional activity during mouse development. Genesis *46*, 335–346.

Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Mach. Learn. *52*, 91–118.

Moore, M. a. S., and Owen, J.J.T. (1967). Chromosome Marker Studies in the Irradiated Chick Embryo. Nature *215*, 1081–1082.

Moore, M.A., 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.

Moreno-Miralles, I., Ren, R., Moser, M., Hartnett, M.E., and Patterson, C. (2011). Bone Morphogenetic Protein Endothelial Cell Precursor–Derived Regulator Regulates Retinal Angiogenesis In Vivo in a Mouse Model of Oxygen-Induced Retinopathy. Arterioscler. Thromb. Vasc. Biol. *31*, 2216–2222.

Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pagès, H., and Gentleman, R. (2009). ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinforma. Oxf. Engl. *25*, 2607–2608.

Morozova, O., Hirst, M., and Marra, M.A. (2009). Applications of New Sequencing Technologies for Transcriptome Analysis. Annu. Rev. Genomics Hum. Genet. *10*, 135–151.

Morrison, S.J., and Spradling, A.C. (2008). Stem Cells and Niches: Mechanisms That Promote Stem Cell Maintenance throughout Life. Cell *132*, 598–611.

Morrison, S.J., and Weissman, I.L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity *1*, 661–673.

Morrison, S.J., Hemmati, H.D., Wandycz, A.M., and Weissman, I.L. (1995). The purification and characterization of fetal liver hematopoietic stem cells. Proc. Natl. Acad. Sci. *92*, 10302–10306.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Morton, M.L., Bai, X., Merry, C.R., Linden, P.A., Khalil, A.M., Leidner, R.S., and Thompson, C.L. (2014). Identification of mRNAs and lincRNAs associated with lung cancer progression using next-generation RNA sequencing from laser micro-dissected archival FFPE tissue specimens. Lung Cancer *85*, 31–39.

Moser, M., Binder, O., Wu, Y., Aitsebaomo, J., Ren, R., Bode, C., Bautch, V.L., Conlon, F.L., and Patterson, C. (2003). BMPER, a Novel Endothelial Cell Precursor-Derived Protein, Antagonizes Bone Morphogenetic Protein Signaling and Endothelial Cell Differentiation. Mol. Cell. Biol. *23*, 5664–5679.

Mukouyama, Y., Hara, T., Xu, M., Tamura, K., Donovan, P.J., Kim, H., Kogo, H., Tsuji, K., Nakahata, T., and Miyajima, A. (1998). In Vitro Expansion of Murine Multipotential Hematopoietic Progenitors from the Embryonic Aorta–Gonad–Mesonephros Region. Immunity *8*, 105–114.

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

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. Science *320*, 1344–1349.

Nakagawa, M., Ichikawa, M., Kumano, K., Goyama, S., Kawazu, M., Asai, T., Ogawa, S., Kurokawa, M., and Chiba, S. (2006). AML1/Runx1 rescues Notch1-null mutation-induced deficiency of para-aortic splanchnopleural hematopoiesis. Blood *108*, 3329–3334.

Nakano, T., Kodama, H., and Honjo, T. (1994). Generation of lymphohematopoietic cells from embryonic stem cells in culture. Science *265*, 1098–1101.

Nakayama, N., Han, C.E., Cam, L., Lee, J.I., Pretorius, J., Fisher, S., Rosenfeld, R., Scully, S., Nishinakamura, R., Duryea, D., et al. (2004). A novel chordin-like BMP inhibitor, CHL2, expressed preferentially in chondrocytes of developing cartilage and osteoarthritic joint cartilage. Dev. Camb. Engl. *131*, 229–240.

Namwanje, M., and Brown, C.W. (2016). Activins and Inhibins: Roles in Development, Physiology, and Disease. Cold Spring Harb. Perspect. Biol. *8*, a021881.

Nirenberg, M.W., and Matthaei, J.H. (1961). THE DEPENDENCE OF CELL- FREE PROTEIN SYNTHESIS IN E. COLI UPON NATURALLY OCCURRING OR SYNTHETIC POLYRIBONUCLEOTIDES. Proc. Natl. Acad. Sci. U. S. A. *47*, 1588–1602.

Nishikawa, S.I., Nishikawa, S., Hirashima, M., Matsuyoshi, N., and Kodama, H. (1998a). Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. Development *125*, 1747–1757.

Nishikawa, S.-I., Nishikawa, S., Kawamoto, H., Yoshida, H., Kizumoto, M., Kataoka, H., and Katsura, Y. (1998b). In Vitro Generation of Lymphohematopoietic Cells from Endothelial Cells Purified from Murine Embryos. Immunity *8*, 761–769.

Nishimura, D. (2001). BioCarta. Biotech Softw. Internet Rep. *2*, 117–120.

Nombela-Arrieta, C., Pivarnik, G., Winkel, B., Canty, K.J., Harley, B., Mahoney, J.E., Park, S.-Y., Lu, J., Protopopov, A., and Silberstein, L.E. (2013). Quantitative imaging of haematopoietic stem and progenitor cell localization and hypoxic status in the bone marrow microenvironment. Nat. Cell Biol. *15*, 533–543.

North, T., Gu, T.L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marin-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.T.R., 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., Zandi, S., Takayama, N., Dobson, S., Gan, O.I., Wilson, G., Kaufmann, K.B., McLeod, J., Laurenti, E., Dunant, C.F., et al. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science *351*, aab2116.

Ober, E.A., Verkade, H., Field, H.A., and Stainier, D.Y.R. (2006). Mesodermal Wnt2b signalling positively regulates liver specification. Nature *442*, 688–691.

Ogawa, M., Porter, P.N., and Nakahata, T. (1983). Renewal and commitment to differentiation of hemopoietic stem cells (an interpretive review). Blood *61*, 823–829.

Orkin, S.H. (1995). Transcription Factors and Hematopoietic Development. J. Biol. Chem. *270*, 4955–4958.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell *132*, 631–644.

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

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.

Pardanaud, L., Luton, D., Prigent, M., Bourcheix, L.M., Catala, M., and Dieterlen-Lievre, F. (1996). Two distinct endothelial lineages in ontogeny, one of them related to hemopoiesis. Dev. Camb. Engl. *122*, 1363–1371.

Park, I., Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., and Clarke, M.F. (2003). Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. Nature *423*, 302–305.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell *163*, 1663–1677.

Peeters, M., Ottersbach, K., Bollerot, K., Orelio, C., de Bruijn, M., Wijgerde, M., and Dzierzak, E. (2009). Ventral embryonic tissues and Hedgehog proteins induce early AGM hematopoietic stem cell development. Dev. Camb. Engl. *136*, 2613–2621.

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.

Pesce, M., Farrace, M.G., Piacentini, M., Dolci, S., and Felici, M.D. (1993). Stem cell factor and leukemia inhibitory factor promote primordial germ cell survival by suppressing programmed cell death (apoptosis). Development *118*, 1089–1094.

Phillips, R.L., Ernst, R.E., Brunk, B., Ivanova, N., Mahan, M.A., Deanehan, J.K., Moore, K.A., Overton, G.C., and Lemischka, I.R. (2000). The Genetic Program of Hematopoietic Stem Cells. Science *288*, 1635–1640.

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

Pluznik, D.H., and Sachs, L. (1965). The cloning of normal "Mast" cells in tissue culture. J. Cell. Comp. Physiol. *66*, 319–324.

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.

Privratsky, J.R., and Newman, P.J. (2014). PECAM-1: regulator of endothelial junctional integrity. Cell Tissue Res. *355*, 607–619.

Proserpio, V., Piccolo, A., Haim-Vilmovsky, L., Kar, G., Lönnberg, T., Svensson, V., Pramanik, J., Natarajan, K.N., Zhai, W., Zhang, X., et al. (2016). Single-cell analysis of CD4+ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation. Genome Biol. *17*, 103.

Purton, L.E., Dworkin, S., Olsen, G.H., Walkley, C.R., Fabb, S.A., Collins, S.J., and Chambon, P. (2006). RARγ is critical for maintaining a balance between hematopoietic stem cell self-renewal and differentiation. J. Exp. Med. *203*, 1283–1293.

R Core Team (2015). R: A Language and Environment for Statistical Computing.

Ralston, A., and Blair, S.S. (2005). Long-range Dpp signaling is regulated to restrict BMP signaling to a crossvein competent zone. Dev. Biol. *280*, 187–200.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., and Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol. *14*, R95.

Rentzsch, F., Zhang, J., Kramer, C., Sebald, W., and Hammerschmidt, M. (2006). Crossveinless 2 is an essential positive feedback regulator of Bmp signaling during zebrafish gastrulation. Development *133*, 801–811.

Reya, T., Duncan, A.W., Ailles, L., Domen, J., Scherer, D.C., Willert, K., Hintz, L., Nusse, R., and Weissman, I.L. (2003). A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature *423*, 409–414.

Rich, I.N. (1995). Primordial germ cells are capable of producing cells of the hematopoietic system in vitro. Blood *86*, 463–472.

Richard, C., Drevon, C., Canto, P.-Y., Villain, G., Bollérot, K., Lempereur, A., Teillet, M.-A., Vincent, C., Rosselló Castillo, C., Torres, M., et al. (2013). Endothelio-mesenchymal interaction controls runx1 expression and modulates the notch pathway to initiate aortic hematopoiesis. Dev. Cell *24*, 600–611.

Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y., Moss, J., Graham, L., Burton, N., Hill, B., Rao, J., Baldock, R.A., et al. (2013). EMAGE mouse embryo spatial gene expression database: 2014 update. Nucleic Acids Res. gkt1155.

Rieger, M.A., Hoppe, P.S., Smejkal, B.M., Eitelhuber, A.C., and Schroeder, T. (2009). Hematopoietic Cytokines Can Instruct Lineage Choice. Science *325*, 217–218.

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.

Robb, L., Lyons, I., Li, R., Hartley, L., Köntgen, F., Harvey, R.P., Metcalf, D., and Begley, C.G. (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. Proc. Natl. Acad. Sci. U. S. A. *92*, 7075–7079.

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

Robert-Moreno, A., 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.

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.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl. *26*, 139–140.

Roeder, I., and Glauche, I. (2006). Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors GATA-1 and PU.1. J. Theor. Biol. *241*, 852–865.

Ruau, D., Ng, F.S.L., Wilson, N.K., Hannah, R., Diamanti, E., Lombard, P., Woodhouse, S., and Göttgens, B. (2013). Building an ENCODE-style data compendium on a shoestring. Nat. Methods *10*, 926–926.

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.

Rybtsov, S., Ivanovs, A., Zhao, S., and Medvinsky, A. (2016). Concealed expansion of immature precursors underpins acute burst of adult HSC activity in foetal liver. Dev. Camb. Engl. *143*, 1284–1289.

Sabin, F. (1920). Studies on the origin of blood vessels and of red corpuscles as seen in the living blastoderm of the chick during the second day of incubation. Contrib. Embryol. *9*, 213–262.

Sacchetti, B., Funari, A., Michienzi, S., Di Cesare, S., Piersanti, S., Saggio, I., Tagliafico, E., Ferrari, S., Robey, P.G., Riminucci, M., et al. (2007). Self-Renewing Osteoprogenitors in Bone Marrow Sinusoids Can Organize a Hematopoietic Microenvironment. Cell *131*, 324–336.

Saliba, A.-E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. *42*, 8845–8860.

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.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. *74*, 5463–5467.

Santos, G.W., Tutschka, P.J., Brookmeyer, R., Saral, R., Beschorner, W.E., Bias, W.B., Braine, H.G., Burns, W.H., Elfenbein, G.J., and Kaizer, H. (1983). Marrow transplantation for acute nonlymphocytic leukemia after treatment with busulfan and cyclophosphamide. N. Engl. J. Med. *309*, 1347–1353.

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.

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

Schaefer, C.F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., and Buetow, K.H. (2009). PID: the Pathway Interaction Database. Nucleic Acids Res. *37*, D674–D679.

Scheller, M., Huelsken, J., Rosenbauer, F., Taketo, M.M., Birchmeier, W., Tenen, D.G., and Leutz, A. (2006). Hematopoietic stem cell and multilineage defects generated by constitutive beta-catenin activation. Nat. Immunol. *7*, 1037–1047.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc. Natl. Acad. Sci. *93*, 10614–10619.

Schindler, C., Levy, D.E., and Decker, T. (2007). JAK-STAT Signaling: From Interferons to Cytokines. J. Biol. Chem. *282*, 20059–20063.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol. Biol. *7*, 3.

Schütte, J., Wang, H., Antoniou, S., Jarratt, A., Wilson, N.K., Riepsaame, J., Calero-Nieto, F.J., Moignard, V., Basilico, S., Kinston, S.J., et al. (2016). An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. eLife *5*, e11469.

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N.K., Macaulay, I.C., Marioni, J.C., and Göttgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. Nature *535*, 289–293.

Serpe, M., Umulis, D., Ralston, A., Chen, J., Olson, D.J., Avanesov, A., Othmer, H., O'Connor, M.B., and Blair, S.S. (2008). The BMP-Binding Protein Crossveinless 2

Is a Short-Range, Concentration-Dependent, Biphasic Modulator of BMP Signaling in Drosophila. Dev. Cell *14*, 940–953.

Seyednasrollah, F., Laiho, A., and Elo, L.L. (2015). Comparison of software packages for detecting differential expression in RNA-seq studies. Brief. Bioinform. *16*, 59–70.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. Science *309*, 1728–1732.

Shen-Orr, S.S., Tibshirani, R., Khatri, P., Bodian, D.L., Staedtler, F., Perry, N.M., Hastie, T., Sarwal, M.M., Davis, M.M., and Butte, A.J. (2010). Cell type–specific gene expression differences in complex tissues. Nat. Methods *7*, 287–289.

Sheridan, J.M., Taoudi, S., Medvinsky, A., and Blackburn, C.C. (2009). A novel method for the generation of reaggregated organotypic cultures that permits juxtaposition of defined cell populations. Genes. N. Y. N 2000 *47*, 346–351.

Shivdasani, R.A., Mayer, E.L., and Orkin, S.H. (1995). Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature *373*, 432–434.

Shuai, K., and Liu, B. (2003). Regulation of JAK–STAT signalling in the immune system. Nat. Rev. Immunol. *3*, 900–911.

Siminovitch, L., McCulloch, E.A., and Till, J.E. (1963). The distribution of colony-forming cells among spleen colonies. J. Cell. Comp. Physiol. *62*, 327–336.

Singbrant, S., Karlsson, G., Ehinger, M., Olsson, K., Jaako, P., Miharada, K., Stadtfeld, M., Graf, T., and Karlsson, S. (2010). Canonical BMP signaling is dispensable for hematopoietic stem cell function in both adult and fetal liver hematopoiesis, but essential to preserve colon architecture. Blood *115*, 4689–4698.

Sitnicka, E., Bryder, D., Theilgaard-Mönch, K., Buza-Vidas, N., Adolfsson, J., and Jacobsen, S.E.W. (2002). Key Role of flt3 Ligand in Regulation of the Common Lymphoid Progenitor but Not in Maintenance of the Hematopoietic Stem Cell Pool. Immunity *17*, 463–472.

Smith, R.A., and Glomski, C.A. (1982). "Hemogenic endothelium" of the embryonic aorta: Does it exist? Dev. Comp. Immunol. *6*, 359–368.

Snick, J.V. (1990). Interleukin-6: An Overview. Annu. Rev. Immunol. *8*, 253–278.

Snoeck, H.W., Bockstaele, D.R.V., Nys, G., Lenjou, M., Lardon, F., Haenen, L., Rodrigus, I., Peetermans, M.E., and Berneman, Z.N. (1994). Interferon gamma selectively inhibits very primitive CD342+CD38- and not more mature CD34+CD38+ human hematopoietic progenitor cells. J. Exp. Med. *180*, 1177–1182.

Souilhol, C., Gonneau, C., Lendinez, J.G., Batsivari, A., Rybtsov, S., Wilson, H., Morgado-Palacin, L., Hills, D., Taoudi, S., Antonchuk, J., et al. (2016a). Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells. Nat. Commun. *7*, 10784.

Souilhol, C., Lendinez, J.G., Rybtsov, S., Murphy, F., Wilson, H., Hills, D., Batsivari, A., Binagui-Casas, A., McGarvey, A.C., MacDonald, H.R., et al. (2016b). Developing HSCs become Notch independent by the end of maturation in the AGM region. Blood.

Souroullas, G.P., Salmon, J.M., Sablitzky, F., Curtis, D.J., and Goodell, M.A. (2009). Adult Hematopoietic Stem and Progenitor Cells Require Either Lyl1 or Scl for Survival. Cell Stem Cell *4*, 180–186.

Southern, E.M., Maskos, U., and Elder, J.K. (1992). Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: Evaluation using experimental models. Genomics *13*, 1008–1017.

Spangrude, G.J., Heimfeld, S., and Weissman, I.L. (1988). Purification and characterization of mouse hematopoietic stem cells. Science *241*, 58–62.

Speck, N.A., and Gilliland, D.G. (2002). Core-binding factors in haematopoiesis and leukaemia. Nat. Rev. Cancer *2*, 502–513.

Stark, G.R., and Darnell Jr., J.E. (2012). The JAK-STAT Pathway at Twenty. Immunity *36*, 503–514.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. *16*, 133–145.

Stephenson, J.R., Axelrad, A.A., McLeod, D.L., and Shreeve, M.M. (1971). Induction of colonies of hemoglobin-synthesizing cells by erythropoietin in vitro. Proc. Natl. Acad. Sci. U. S. A. *68*, 1542–1546.

Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G., and Bayley, H. (2009). Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc. Natl. Acad. Sci. *106*, 7702–7707.

Stoughton, R.B. (2005). Applications of Dna Microarrays in Biology. Annu. Rev. Biochem. *74*, 53–82.

Subbiah, K., Hamlin, D.K., Pagel, J.M., Wilbur, D.S., Meyer, D.L., Axworthy, D.B., Mallett, R.W., Theodore, L.J., Stayton, P.S., and Press, O.W. (2003). Comparison of immunoscintigraphy, efficacy, and toxicity of conventional and pretargeted radioimmunotherapy in CD20-expressing human lymphoma xenografts. J. Nucl. Med. *44*, 437–445.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene

set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Sun, J., Ramos, A., Chapman, B., Johnnidis, J.B., Le, L., Ho, Y.-J., Klein, A., Hofmann, O., and Camargo, F.D. (2014). Clonal dynamics of native haematopoiesis. Nature *514*, 322–327.

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.

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.

Szilvassy, S.J., Humphries, R.K., Lansdorp, P.M., Eaves, A.C., and Eaves, C.J. (1990). Quantitative assay for totipotent reconstituting hematopoietic stem cells by a competitive repopulation strategy. Proc. Natl. Acad. Sci. *87*, 8736–8740.

Tagliabracci, V.S., Engel, J.L., Wen, J., Wiley, S.E., Worby, C.A., Kinch, L.N., Xiao, J., Grishin, N.V., and Dixon, J.E. (2012). Secreted Kinase Phosphorylates Extracellular Proteins That Regulate Biomineralization. Science *336*, 1150–1153.

Tai, C.-I., and Ying, Q.-L. (2013). Gbx2, a LIF/Stat3 target, promotes reprogramming to and retention of the pluripotent ground state. J Cell Sci *126*, 1093–1098.

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.

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. U. S. A. *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.

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

The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Thomas, E.D., Lochte, H.L.J., Lu, W.C., and Ferrebee, J.W. (1957). Intravenous Infusion of Bone Marrow in Patients Receiving Radiation and Chemotherapy. N. Engl. J. Med. *257*, 491–496.

Thomas, E.D., Lochte, H.L., Cannon, J.H., Sahler, O.D., and Ferrebee, J.W. (1959). SUPRALETHAL WHOLE BODY IRRADIATION AND ISOLOGOUS MARROW TRANSPLANTATION IN MAN*. J. Clin. Invest. *38*, 1709–1716.

Thomas, E.D., Buckner, C.D., Banaji, M., Clift, R.A., Fefer, A., Flournoy, N., Goodell, B.W., Hickman, R.O., Lerner, K.G., Neiman, P.E., et al. (1977). One hundred patients with acute leukemia treated by chemotherapy, total body irradiation, and allogeneic marrow transplantation. Blood *49*, 511–533.

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.

Till, J.E., McCulloch, E.A., and Siminovitch, L. (1964). A STOCHASTIC MODEL OF STEM CELL PROLIFERATION, BASED ON THE GROWTH OF SPLEEN COLONY-FORMING CELLS*. Proc. Natl. Acad. Sci. U. S. A. *51*, 29–36.

Toles, J.F., Chui, D.H., Belbeck, L.W., Starr, E., and Barker, J.E. (1989). Hemopoietic stem cells in murine embryonic yolk sac and peripheral blood. Proc. Natl. Acad. Sci. *86*, 7456–7459.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinforma. Oxf. Engl. *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.

Travnickova, J., Tran Chau, V., Julien, E., Mateos-Langerak, J., Gonzalez, C., Lelièvre, E., Lutfalla, G., Tavian, M., and Kissa, K. (2015). Primitive macrophages control HSPC mobilization and definitive haematopoiesis. Nat. Commun. *6*.

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.

Umulis, D.M., Serpe, M., O'Connor, M.B., and Othmer, H.G. (2006). Robust, bistable patterning of the dorsal surface of the Drosophila embryo. Proc. Natl. Acad. Sci. *103*, 11613–11618.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K., et al. (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res. *18*, 1051–1063.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.

Vize, P.D., Woolf, A.S., and Bard, J.B.L. (2003). The Kidney: From Normal Development to Congenital Disease (Academic Press).

Vo, L.T., and Daley, G.Q. (2015). De novo generation of HSCs from somatic and pluripotent stem cell sources. Blood *125*, 2641–2648.

Vodyanik, M.A., Bork, J.A., Thomson, J.A., and Slukvin, I.I. (2005). Human embryonic stem cell–derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. Blood *105*, 617–626.

Wagner, J.E., Barker, J.N., DeFor, T.E., Baker, K.S., Blazar, B.R., Eide, C., Goldman, A., Kersey, J., Krivit, W., MacMillan, M.L., et al. (2002). Transplantation of unrelated donor umbilical cord blood in 102 patients with malignant and nonmalignant diseases: influence of CD34 cell dose and HLA disparity on treatment-related mortality and survival. Blood *100*, 1611–1618.

Wakayama, T., Hamada, K., Yamamoto, M., Suda, T., and Iseki, S. (2003). The expression of platelet endothelial cell adhesion molecule-1 in mouse primordial germ cells during their migration and early gonadal formation. Histochem. Cell Biol. *119*, 355–362.

Walkley, C.R., Olsen, G.H., Dworkin, S., Fabb, S.A., Swann, J., McArthur, G., Westmoreland, S.V., Chambon, P., Scadden, D.T., and Purton, L.E. (2007). A Microenvironment-Induced Myeloproliferative Syndrome Caused by Retinoic Acid Receptor γ Deficiency. Cell *129*, 1097–1110.

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

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. (1996b). The CBFβ Subunit Is Essential for CBFα2 (AML1) Function In Vivo. Cell *87*, 697–708.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Warren, A.J., Colledge, W.H., Carlton, M.B.L., Evans, M.J., Smith, A.J.H., and Rabbitts, T.H. (1994). The Oncogenic Cysteine-rich LIM domain protein Rbtn2 is essential for erythroid development. Cell *78*, 45–57.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737–738.

Weissman, I. (1978). Fetal hematopoietic origins of the adult hematolymphoid system. Differ. Norm. Neoplast. Hematop. Cells *5*, 33–47.

Whichard, Z.L., Sarkar, C.A., Kimmel, M., and Corey, S.J. (2010). Hematopoiesis and its disorders: a systems biology approach. Blood *115*, 2339–2347.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

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.

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.

Wilson, N.K., Kent, D.G., Buettner, F., Shehata, M., Macaulay, I.C., Calero-Nieto, F.J., Sánchez Castillo, M., Oedekoven, C.A., Diamanti, E., Schulte, R., et al. (2015). Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell *16*, 712–724.

Wu, A.M., Till, J.E., Siminovitch, L., and McCulloch, E.A. (1967). A cytological study of the capacity for differentiation of normal hemopoietic colony-forming cells. J. Cell. Physiol. *69*, 177–184.

Wu, A.M., Till, J.E., Siminovitch, L., and McCulloch, E.A. (1968). Cytological evidence for a relationship between normal hemotopoietic colony-forming cells and cells of the lymphoid system. J. Exp. Med. *127*, 455–464.

Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M.-L., Visvader, J.E., and Smyth, G.K. (2010). ROAST: rotation gene set tests for complex microarray experiments. Bioinformatics *26*, 2176–2182.

Xue, J., Peng, J., Yuan, M., Wang, A., Zhang, L., Liu, S., Fan, M., Wang, Y., Xu, W., Ting, K., et al. (2011). NELL1 promotes high-quality bone regeneration in rat femoral distraction osteogenesis model. Bone *48*, 485–495.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. *44*, D710–D716.

Yoder, M.C., and Hiatt, K. (1997). Engraftment of Embryonic Hematopoietic Cells in Conditioned Newborn Recipients. Blood *89*, 2176–2183.

Yokomizo, T., and Dzierzak, E. (2010). Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos. Dev. Camb. Engl. *137*, 3651–3661.

Yokomizo, T., Ogawa, M., Osato, M., Kanno, T., Yoshida, H., Fujimoto, T., Fraser, S., Nishikawa, S., Okada, H., Satake, M., et al. (2001). Requirement of Runx1/AML1/PEBP2αB for the generation of haematopoietic cells from endothelial cells. Genes Cells *6*, 13–23.

Zhang, C.C., Kaba, M., Ge, G., Xie, K., Tong, W., Hug, C., and Lodish, H.F. (2006). Angiopoietin-like proteins stimulate ex vivo expansion of hematopoietic stem cells. Nat. Med. *12*, 240–245.

Zhang, J., Niu, C., Ye, L., Huang, H., He, X., Tong, W.-G., Ross, J., Haug, J., Johnson, T., Feng, J.Q., et al. (2003). Identification of the haematopoietic stem cell niche and control of the niche size. Nature *425*, 836–841.

Zhang, X., Kuroda, S. 'ichi, Carpenter, D., Nishimura, I., Soo, C., Moats, R., Iida, K., Wisner, E., Hu, F.-Y., Miao, S., et al. (2002). Craniosynostosis in transgenic mice overexpressing Nell-1. J. Clin. Invest. *110*, 861–870.

Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., Li, Z., et al. (2016). Tracing haematopoietic stem cell formation at single-cell resolution. Nature *533*, 487–492.

Zhu, J., Guo, L., Min, B., Watson, C.J., Hu-Li, J., Young, H.A., Tsichlis, P.N., and Paul, W.E. (2002). Growth Factor Independent-1 Induced by IL-4 Regulates Th2 Cell Proliferation. Immunity *16*, 733–744.

European Nucleotide Archive, EMBL-EBI.

Chapter 8 Appendix

8.1 Interactive visualisation code

The code for the interactive data visualisation is included as an electronic supplementary file in the attached CD. This includes the input SVG and TXT files; the JAVA files and classes; the JavaScript file encoding the interactive components (main.js); the CSS attributes of the visualisation (main.css); all additional plug-ins; and a build (build.xml) and jar (VisualisationDemo.jar) to enable deployment.

8.2 Induction of Bmp4 after doxycycline treatment

Control demonstrates the induced high level of expression of *Bmp4* in OP9-BMP4 cell line in response to doxycycline treatment.

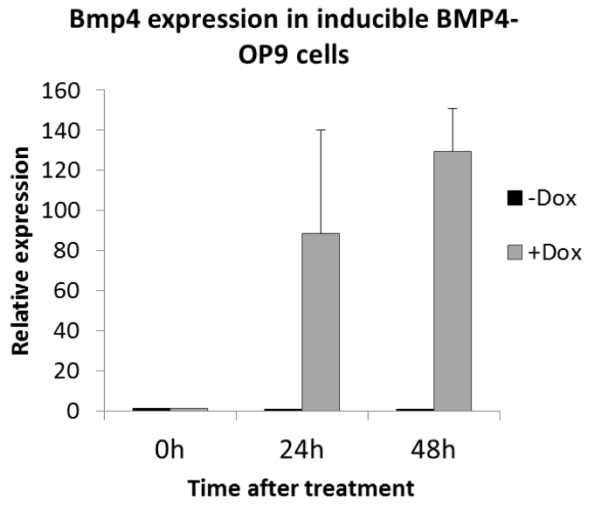


Figure 8-1 Control expression of Bmp4 after induction of OP9-BMP4 cells with doxycycline

Expression of *Bmp4* detected by qRT-PCR in OP9 cells which had been transfected with an expression construct containing inducible expression of *Bmp4* (described in section 2.4.4), and treated with 1µg/ml doxycycline in reaggregate conditions. Expression normalised to Tbp relative to expression at 0 h. n=2. Error bar represents standard deviation of the mean.

8.3 Expression of genes of interest in the AGM RNA-seq dataset

The expression of a small number of genes of interest mentioned in Chapter 6 are displayed below.

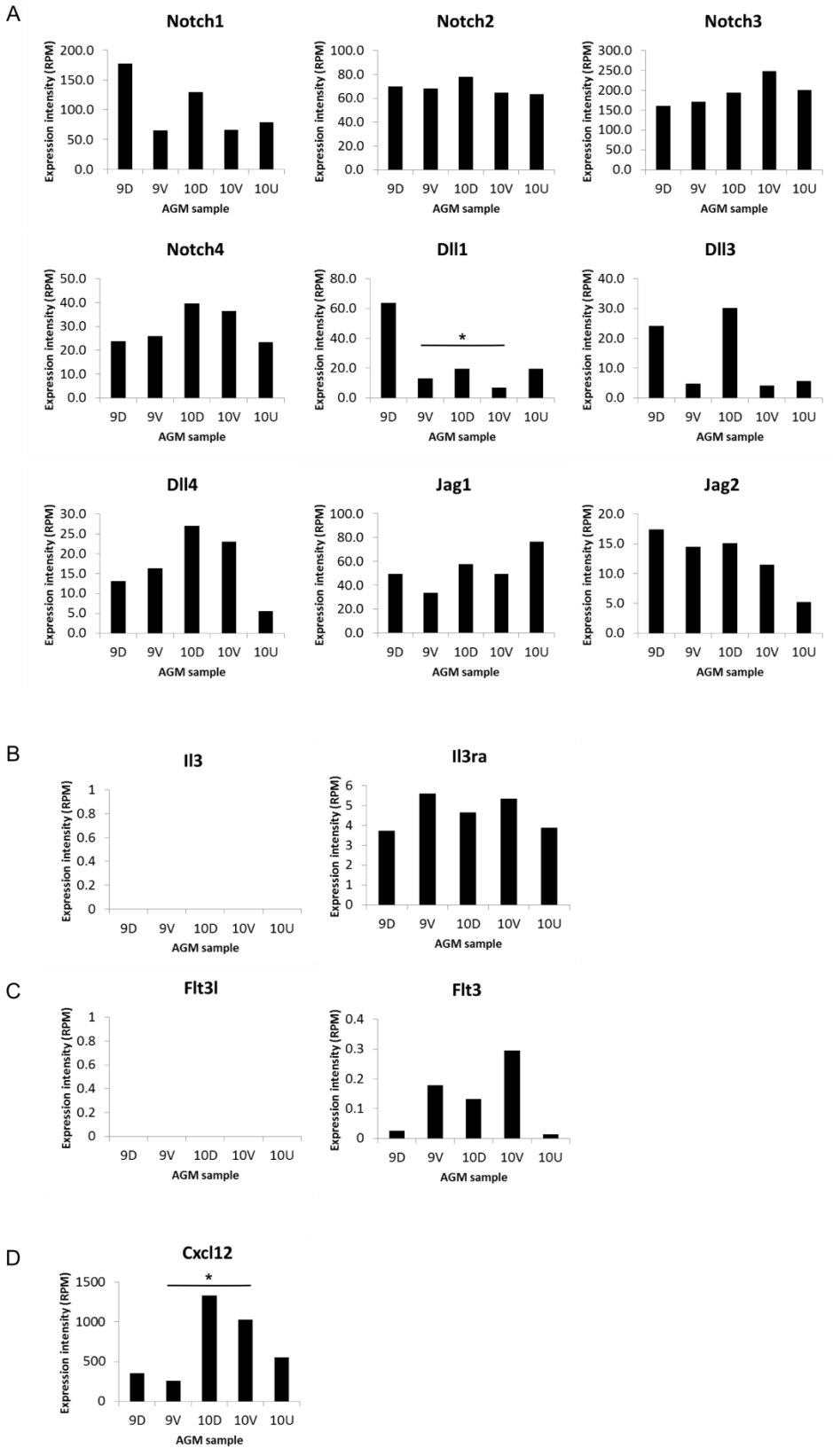


Figure 8-2 Expression of genes of interest in AGM RNA-seq dataset

Expression intensity in RPM of genes of interest in the AGM RNA-seq dataset (A) Notch receptors and ligands (B) Il3 ligand and its receptor Il3ra (C) Flt3l ligand and its receptor Flt3 (D) and Cxcl12. * indicates where there is a significant difference between E9.5 AoV and E10.5 AoV (absolute fold change > 2, FDR ≤ 0.05 from DESeq). Expression of Il3 ligand and Flt3l was not detected in the AGM RNA-seq dataset.

8.4 Additional outputs: publication in preparation and external meetings

Manuscript in preparation:

- McGarvey, A., Rybtsov, S., Souilhol, C., Godwin, D., Tomlinson, S.R., Medvinsky, A., A molecular roadmap of the spatio-temporal transitions in the developing HSC niche reveals regulation of HSC maturation by BMPER. *Manuscript under review*.

External meetings:

- Talk: "A genome-wide approach to characterising the developing HSC niche" (2016). Copenhagen Bioscience Conference: Stem Cell Niche - Development & Disease. Reviewed in:
  Kirkeby, A., Perlmann, T., and Pereira, C.-F. (2016). The stem cell niche finds its true north. Development 143, 2877–2881.
- Poster: "A spatio-temporal molecular framework of the developing haematopoietic stem cell niche" (2016). ISSCR Annual Meeting, San Francisco.
- Poster: "Global gene expression analysis in elucidating the molecular characteristics of emerging hematopoietic stem cells" (2013). Hydra IX – The European Summer School on Stem Cells & Regenerative Medicine, Hydra, Greece.