UNIVERSITY OF BIRMINGHAM

A functional screen of novel haematopoietic players in haematopoietic ontogeny

Melita Tvardauskaite

A thesis submitted for the degree of MSc by Research: Cancer and Genomic Sciences

March 2023

Contents

ABSTRACT 4

1. INTRODUCTION 5

1.1 Overview of mammalian haematopoiesis 5

1.2 Specification of the HE and the endothelial to haematopoietic transition (EHT) 6

1.2.1 HSC maturation and maintenance 9

1.3 Zebrafish as a model for haematopoiesis 11

1.4 Aims and Objectives 14

2. Methods 15

Zebrafish husbandry 15

Differential gene expression analysis 15

crRNA design and RNP preparation 16

gDNA extraction 16

RNA extraction and reverse transcription 17

Whole-mount *in situ* hybridization (WISH) 17

Imaging 20

Network construction 21

3. Results 22

3.1 Candidate gene selection 22

3.1.1 Transcriptional profiling of HE and HSPCs from RNA-seq data 22

3.1.2 Selected candidate HSPC genes 27

3.1.3 Analysis of candidate gene function in haematopoiesis using CRISPR/Cas9 gene editing 29

3.2 Screen design and validation of the approach –*sart3 crispants* 30

3.3 Functional screening of candidate HSPC genes 32

3.3.1 Loss of function candidate gene crispants show impaired HE specification and HSPC emergence 32

3.3.2 Screening revealed impaired HSPC proliferation and colonisation of haematopoietic organs 36

3.4 Investigating dynamics of genes overexpressed in the HE 41

3.4.1 Differential gene expression analysis 43

3.4.2 Network construction 43

3.4.3 Genes of interest 46

3.4.4 Loss of *thy1* leads to impaired expression of haematopoietic marker genes 47

3.4.5 Loss of *ncf1* leads to impaired expression haematopoietic marker genes at 30 hpf, but not 4 dpf 49

4. Discussion 51

4.1 Loss of identified candidate genes leads to decreased HSPC emergence at 30 hpf ... 51
4.2 *spi2* is functionally co-expressed with *thy1* and *ncf1* ... 52
4.3 Limitations and future work ... 53
ACKNOWLEDGEMENTS ... 55
References ... 56

ABSTRACT

All blood cells in the adult differentiate from a common precursor, the haematopoietic stem and progenitor cell (HSPC), which arise from the haemogenic endothelium (HE) in the dorsal aorta, migrate to an intermediary haematopoietic organ (CHT in zebrafish) to proliferate late seed the definitive haematopoietic organs. Many genetic and epigenetic mechanisms regulate these processes; however many questions remain unanswered regarding the exact mechanisms of this regulation. By analysing recently published data from haematopoietic tissues including the HE and emerging HSPCs, this thesis aimed to identify novel players in HSPC ontogeny, and functionally analyse them. Seven candidate genes were selected, and functionally analysed using CRISPR-cas9. Loss of *sart3, wasla, ptpn13, tead1a, esrp1, etv5b, and spi2* all resulted in impaired haematopoietic development. Performing differential gene expression analysis, genes upregulated in the HE were identified, and their functional co-expression network constructed. This revealed a further two genes which are co-expressed together with *spi2* – *thy1* and *ncf1.* Upon knockout both genes showed haematopoietic phenotypes in terms of HSPC and lineage marker expression. Further analysis is needed to characterise the function of these candidate genes, however we conclude that our approach is effective at identifying novel players in haematopoiesis.

1. INTRODUCTION

1.1 Overview of mammalian haematopoiesis

Blood is a diverse tissue composed of various cell types that are specifically designed to perform different functions within the body. The ability to circulate blood was crucial in the evolution of complex organ systems in vertebrates, as it facilitated efficient delivery of oxygen, nutrients, and immune defence throughout the body. Throughout one's life, blood cells are continuously generated through a process called haematopoiesis (Kumar and Evans, 2010) The first blood cells in the developing embryo arise during the primitive wave of haematopoiesis and begin with the formation of extra embryonic blood islands located in the yolk sac, which give rise to primitive erythrocytes as well as smaller numbers of myeloid progenitor cells (Jagannathan-Bogdan and Zon, 2013). These cells are transitory and serve to provide oxygen and immune support to the embryo prior to the emergence of an established haematopoietic system. Despite serving the same function in the developing embryo, primitive and definitive erythrocytes are distinguishable by their size and embryonic globin expression, pointing to the differing transcriptional regulatory networks that control the development of the haematopoietic system (Palis, 2004).

The second, sometimes termed the “pro-definitive”, wave of haematopoiesis also takes place in the yolk sac and gives rise to erythroid-myeloid progenitors (EMPs)(Ciau-Uitz *et al.,* 2014). Despite sharing some similarities with the definitive wave i.e. their emergence from an endothelial precursor and the persistence of these cells beyond embryonic development as well as their emergence from an endothelial precursor located in the yolk sac, they lack the capacity to self-renew, hence the need for the final, definitive, wave which is initiated by the emergence of haematopoietic stem cells (HSCs). These cells possess the capacity to self-renew and differentiate into any haematopoietic cell type. HSCs are the source of all haematopoietic cells produced by the adult, beginning during the definitive wave of haematopoiesis in embryonic development, and continuing throughout the lifetime of the

organism, these cells have the ability to differentiate into all blood cell lineages (Costa *et al.,* 2012)*,* . Unlike the progenitors produced during the extraembryonic primitive wave, these cells have intraembryonic origins and the only time when these cells are produced *de novo* is when they arise from the aortic haemogenic endothelium (HE) in the aortic-gonads-mesonephros (AGM) region through endothelial to haematopoietic transition (EHT) (Lacaud *et al.,* 2015). Relatively few HSCs are produced this way so after their emergence from the HE, they migrate to an intermediary haematopoietic organ – the foetal liver in mammals – where they proliferate and differentiate before entering circulation again to seed the definitive haematopoietic organ, the bone marrow (Asada *et al.,* 2017). This organ remains the main source of HSCs throughout life. Additionally, a smaller number of HSCs will colonise the thymus which is the site of lymphopoiesis. This niche can only give rise to mature T-cells (Jenkinson *et al.,* 2006). The lineage specification of cells arising from the primitive and definitive waves is summarised in Figure 1.

1.2 Specification of the HE and the endothelial to haematopoietic transition (EHT)

The association between the endothelial and haematopoietic identities was first observed as early as the late 19th century, giving rise to the concept of a common precursor to both lineages. Utilising lineage tracing experiments (Zovein et al., 2008), live imaging (Boisset et al., 2010), and more recently, transcriptomic analyses (Bonkhofer et al., 2019) a model of a "haemogenic endothelium" has emerged, categorically proving that a subset of cells within the endothelial population acquire haematopoietic potential and give rise to HSCs (Kissa and Herbornel, 2010; Bertrand *et al.,* 2010). The origins of the HE are still subject to intense debate in the community, and it appears that its specification is context dependent. In vertebrates, HSCs arise from HE precursors in the main embryonic artery, and it is now widely accepted that at least in the dorsal aorta (DA) in zebrafish the HE derives from arterial endothelium (Bonkhofer *et al.,* 2019); this was also recently observed in human pluripotent

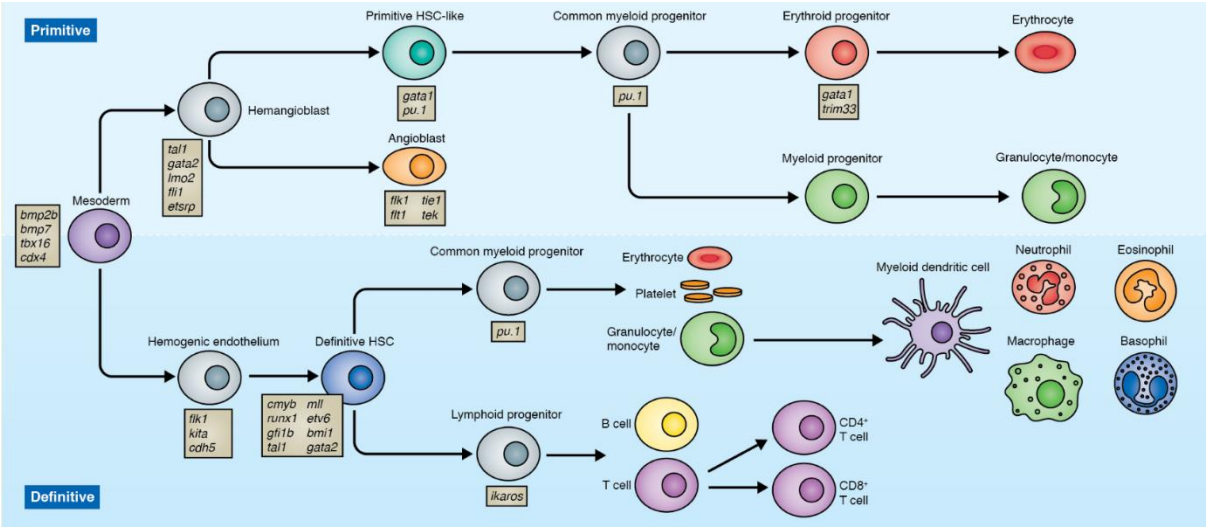


Figure 1: Haematopoietic development initiates with the primitive wave and gives rise to erythroid and myeloid progenitors which can only differentiate into mature erythrocytes and monocytes. These cells are replaced by definitive erythrocytes and monocytes produced during the definitive wave from HSCs. HSC have the ability to give rise to cells of the erythroid, myeloid and lymphoid lineage. Diagram adapted from Jagannathan-Bogdan and Zon (2013).

stem cells (Uenishi *et al.,* 2018) However, in the yolk sac where it gives rise to EMPs, it was found that HE is not specified from arterial vasculature (Frame et al., 2016). The former was recently confirmed in a transcriptome analysis of human haematopoietic tissues from the HE to birth by Calvanese *et al.,* (2022). It was found that the HE (which is characterised by the expression of *ALDH1A1* - a retinal dehydrogenase, *KCNK17* – a potassium channel protein*, and RUNX1*), derives from a population of arterial endothelial cells expressing *ILL33* and *ALDH1A1* (Calvanese *et al.,* 2022).

Even though a number of essential players in the specification of the HE have been identified, their modes of action and the exact mechanisms leading to arterial or haemogenic fate decision in endothelial cells is not clear. One such signalling pathway, which is essential for the development of embryonic vasculature, is the Notch signalling pathway (Kamano *et al.,* 2003). This type of signalling requires cells to be in direct or close contact between cells through binding of receptors and ligands between communicating cells and receptors Notch1 and Notch4 which are expressed in endothelial cells and in the dorsal aorta are able to bind

the ligands *Dll4, Jag1, and Jag2. Dll4* is a regulator of arterial fate, whereas loss of *Jag1* results in the loss of definitive haematopoiesis without affecting the endothelium (Gama-Norton *et al.,* 2015). These two ligands are in direct competition and binding of *Dll4* or *Jag1* induces high or low levels of Notch1 signal strengths respectively. In the absence of *Jag1*, *Dll4* induces high level Notch1 activity specifying the haematopoietic fate (Robert-Moreno *et al.,* 2008).

HSCs first appear in haematopoietic 'clusters' (intra-aortic clusters, or IACs) which derive from the HE in a process termed endothelial to haematopoietic transition (EHT). The involvement of Runt-related transcription factor 1 (*RUNX1)* is indispensable during this process and very few EHT events are able to initiate in *runx1* deficient zebrafish; moreover, the cells that arise from these events are abortive (Kissa *et al.,* 2010), explaining the absence of definitive haematopoietic cells in *Runx1* deficient ESC-derived HE (Lancrin et al., 2009). *Runx1* often associates with CBFβ which increases its affinity for DNA (Bresciani *et al.,* 2014). Despite forming a complex, *Runx1* and CBFβ have distinct roles in haematopoietic development as evidenced by the lack of HSCs in the absence of *runx1*, but not in *cbfb*$^{-/-}$ mutants (Bresciani *et al.,* 2014).Instead *cbfb* is believed to be essential for the release of nascent HSCs from the AGM acting downstream of the Notch pathway (Bresciani *et al.,* 2014). It is widely believed that the role of *runx1* is limited to the EHT event (Chen *et al.,* 2009). However, more recently evidence has emerged that *Runx1* has roles in the maturation of HSCs in IACs after their emergence, as is required until nascent HSCs acquire the CD43 cell surface marker (Tober *et al.,* 2013). Two direct targets of *Runx1*, Gfi1/Gfi1b, encode nuclear zinc finger proteins which act as transcriptional repressors of the endothelial programme and also upregulate the expression of haematopoiesis markers (Lancrin *et al.,* 2012).The expression and function of Runx1 is conserved across vertebrates where it's expressed in tissues involved in haematopoietic development, including the vascular endothelium, the ventral wall of the DA and definitive HSCs (Harris *et al.,* 2013; Espanola *et al.,* 2020). In the zebrafish, *runx1* is one of the earliest markers of the HE (Kalev-Zylinkska *et al.,* 2015).

Gata2 is another factor with a role in EHT. Loss of GATA2 in humans results in susceptibility to infection, autoimmunity and lymphoedema (Collin *et al.,* 2015). Similarly to *Runx1*, through loss of function experiments in human ESCs, it was found that GATA2 is required for EHT but not the specification of the HE or arterial lineages, or the proliferation of HSCs post-EHT (Kang *et al.,* 2018). Deletion of the endothelial-specific +9.5Kb enhancer of *Gata2* resulted in the loss of HSPCs together with downregulation of *runx1*, uncovering the role of *Gata2* as acting upstream of *runx1* to regulate EHT (Pater *et al.,* 2013; Gao *et al.,* 2013). An upstream regulator of both *Gata2 and Runx1* is *Sox17*, another transcriptional regulator of arterial fate (Corada *et al.,* 2013). Sox17 is a known regulator of arterial vs venous identity (Corada *et al.,* 2013). Loss of *Sox17* promotes conversion to haematopoietic cell fate and mutants exhibit a threefold increase in HE ratios. Sox17 is critical for the specification of the HE but also promotes haematopoietic fate over endothelial during EHT by regulating the commitment to the haematopoietic fate through repression of *Runx1* and *Gata2*. The role of *Sox17* is two-fold – it acts as a negative regulator of arterial gene expression, and a repressor of *Runx1* and *Gata2,* which in turn lead to the downstream repression of the endothelial programme. This allows for the progression of haematopoietic specification (Lizama *et al.,* 2015).

1.2.1 HSC maturation and maintenance

Following their release from the AGM region, HSCs migrate to a series of intermediate haematopoietic organs in the zebrafish - the caudal haematopoietic tissue (CHT), thymus, and kidney (Zhang *et al.,* 2011). An essential player in this migratory behaviour is *cmyb* – a DNA-binding transcription factor. A study by Zhang *et al.* (2011) used a Tg*(cd41*:eGFP) transgenic that labels early HSPCs and found a significant increase in GFP positive cells in the ventral wall of the DA upon loss of *cmyb*. It was found that this is due to the upregulation of stromal cell-derived factor s*df1a* that prevented HSPC migration. *cmyb* is also believed to be involved in the differentiation of HSCs, but not the specification of the HE or EHT (Lieu *et al.,* 2009). *Cmyb* is also an early marker of HSPCs in developing zebrafish embryos.

Even after their emergence, murine HSCs undergo further maturation within IACs and are defined molecularly band functionally by their embryonic stage. In mice can be categorised into three distinct steps – the pro-HSC, pre-HSC I and Pre-HSC II, characterised by their expression of cell surface marker CD41, CD43 and CD45 respectively (Rybtsov *et al.,* 2014). It has been revealed that HSCs can exist in distinct transcriptional states; from the specification of the HE, to the emergence of HSCs and their maturation to become definitive HSCs, these cells are characterised by stage-specific gene expression (McKinney-Freeman 2012; Calvanese *et al.,* 2022) which are tightly controlled by various signalling pathways.

Mature HSPCs are capable of differentiating into any blood lineage but also maintain their population by self-renewal and proliferation (Sandberg *et al.,* 2005). This is a tightly controlled process and many mature HSPCs are quiescent, not proliferating or contributing to haematopoiesis – these cells are termed long-term reconstituting HSCs (LT-HSCs) [Cheshier *et al.,* 1999).These two states are dynamic and the balance between the two can shift depending on the needs of an organism. In case of haemorrhage more cells will be proliferating to efficiently restore lost blood, whereas in case of infection HSCs cells will be differentiating and producing more immune cells. The decision between self-renewal and differentiation fates is controlled by complex signalling pathways from intrinsic signals as well as signals from the surrounding microenvironment. Cytokines are expressed by cells of the immune system in times of infection and a detected by HSCs. Unsurprisingly, HSCs can be affected by the presence of these cytokines such as IFN- γ, which inhibits the ability of quiescent HSCs to self-renew (de Bruin *et al.,* 2013). The role of IFN-γ is very context dependent and additionally to acting as an inhibitor of self-renewal in quiescent cells, its function can range to act as both a promoter of myelopoiesis though overexpression of Batf2 (Matatall *et al.,* 2016), as well as inhibitor of erythropoiesis by inducing the expression of *PU.1* (Libregts *et al.,* 2011). These cell extrinsic mechanisms are sometimes dispensable for haematopoietic development during which HSCs arise, but have a role in quiescence. An

example of this is *Cxcr4* and *Tie-2,* which when lost in the mouse increase cell cycle activity leading to HSC exhaustion in adult life but have no effect on mechanisms leading to the emergence of HSC (Nagasawa *et al.,* 1996; Puri *et al.,* 2003).

Two pathways that work together to maintain HSC proliferation in an undifferentiated state are Wnt and Notch. WNT proteins can promote self-renewal in murine and human HSCs *in vitro* but are not believed to aid HSC differentiation to any specific lineage (Austin *et al.,* 1997; Murdoch *et al.,* 2003). Wnt signalling can be induced by the cells themselves as well as their microenvironment i.e. WNT proteins are capable of autocrine and paracrine signalling (Nteliopoulos *et al.,* 2009) Overexpression of *Wnt-5a*, normally expressed in haematopoietic tissues in the mouse, resulted in significant HSC expansion over controls (Austin *et al.,* 1997). However, the secreted WNT proteins alone are capable of inducing HSC proliferation (Austin *et al.,* 1997). Overexpression of B-catenin, a downstream signal transducer of the Wnt signalling pathway, resulted in increased proliferation of *Hoxb4* and *Notch1* genes, both of implicated in self-renewal of HSCs (Reya *et al.,* 2003; Beslu *et al.,* 2004; Delaney *et al.,* 2005). Indeed, the activation of Notch signalling in haematopoietic cells inhibits HSPC differentiation and initiates self-renewal (Karanu *et al.,* 2000; Varnum-Finney *et al.,* 2003) and it is now clear that not only do these two pathways depend on each other, the same HSCs can utilise both pathways simultaneously, and inhibition of Notch impairs the ability of Wnt to maintain HSCs in their undifferentiated state (Duncan *et al.,* 2005). This suggests that Notch is the dominant pathway between the two at inhibiting HSC differentiation. Additionally, McKinney-Freeman *et al,* (2012) found that that HSCs derived from embryonic stem cells (ESCs) in the mouse which fail to engraft have decreased Notch target gene expression compared to *in vivo* definitive HSCs, making it an essential player in our ability to produce HSC *in vitro*.

1.3 Zebrafish as a model for haematopoiesis

The Zebrafish (*Danio rerio*) is a very useful model to study genetic mechanisms. Not only can large numbers of embryos be produced by a mating pair making them useful for large-scale genetics screen, the fertilisation and subsequent development of these embryos takes place *ex utero* enabling us to observe early embryonic development. This is especially useful as zebrafish have a short life cycle and reach sexual maturity at around 3 months post fertilisation (mpf) making it possible to observe its embryonic development in a relatively short period of time. Additionally, zebrafish embryos are small and transparent with visible internal structures.

Haematopoiesis is a very well conserved mechanism in all vertebrates; the haematopoietic niches in the three most studied models of haematopoiesis (human, mouse, and zebrafish) differ only in their location, but the processes giving rise to an established haematopoietic system are sequentially similar (Ciau-Uitz *et al.,* 2014). This is summarised in Figure … The three above mentioned haematopoietic systems all initiate with a transient primitive wave. Multi-lineage precursors are then generated in the AGM (also the placenta in mouse), which migrate to colonise the intermediary haematopoietic organ for proliferation, after which they enter circulation and seed the niche which will remain the main source of HSCs in the adult. This is summarised in Figure…)

In zebrafish, haematopoietic development begins in the lateral mesoderm with the formation of the intermediate cell mass (ICM) and gives rise to primitive erythroid progenitors at around 24 hours post fertilisation (hpf). The ICM is analogous to the extraembryonic yolk sac blood islands which give rise to primitive erythrocytes in mammals. Haematopoietic commitment in the zebrafish embryo is first observed at around 10 hpf with the expression of *gata1*, a transcription factor specific to the erythroid lineage. The remaining cells in the ICM that are not expressing *gata1* later differentiate into endothelial cells. Myeloid cells arise from the rostral blood island (RBI) between 10-11 hpf and express the myeloid-specific *pu.1.* Similarly,

B

a subset of RBI precursors which are not expressing *pu.1* differentiate into endothelial cells (Davidson and Zon, 2004). The DA in zebrafish is the site of HSC

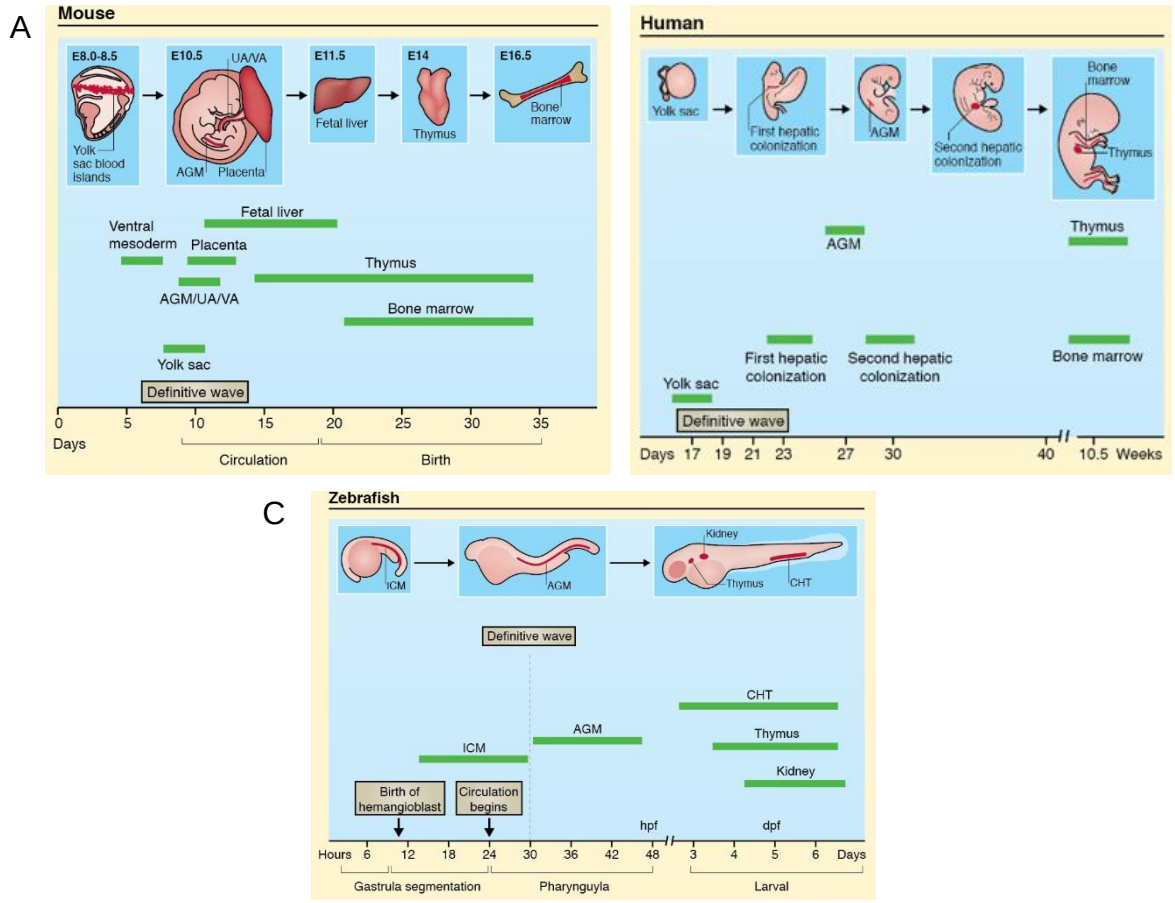


Figure 2:Haematopoiesis in vertebrates. Schematics showing the emergence the haematopoietic system in the mouse **(A)**, human **(B),** and zebrafish **(C).** All three systems initiate with the transitory extraembronic primitive wave. The definitive wave gives rise to HSCs which migrate to intermediary haematopoietic niches to proliferate and differentiate and then colonise the definitive haematopoietic niche which will remain the source of all HSCs throughout the life of the organism. Adapted from Jangannathan-Bogdan and Zon (2013).

emergence in zebrafish and is the equivalent of the AGM region identified in the mouse. These cells arise at around 30-36 hpf from the ventral wall of the DA through EHT. HSCs are multipotent and have the ability to both self-renew and differentiate into any haematopoietic cell lineage. Prior to emergence, these cells exhibit bending, contracting, and eventually budding to emerge from the ventral wall of the DA and enter the sub-aortic space between the DA and posterior cardinal vein (PCV). After their emergence, HSCs enter circulation through the PCV and migrate to the CHT where they will proliferate and later colonise the thymus for

lymphocyte development and the kidney marrow. The latter two will remain the main source of lymphocytes and HSCs throughout the adult life respectively.

1.4 Aims and Objectives

Haematopoiesis is a tightly controlled mechanism, the dysregulation of which can lead to cancerous and non-cancerous haematological disorders such as leukaemias, lymphomas, myelomas, as well as immunological disorders. Our current understanding of HSC development is insufficient for HSC reprogramming *in vitro* and subsequent transplantation for treatment. Therefore, further molecular characterization of vertebrate haematopoiesis is essential to improve understanding of the process and enable the exploration of potential new avenues for therapeutic intervention. Therefore, this project aimed to:

1. Investigate the dynamics of gene expression of HSC-specific genes during haematopoietic ontogeny and identify novel candidate HSPC genes.
2. Functionally analyse novel candidate HSPC-specific genes required for their emergence, maintenance, differentiation or migration *in vivo*.

2. Methods

Zebrafish husbandry

All work involving zebrafish followed guidelines set out under Animals Scientific Procedures Act (ASPA) 1986. Zebrafish lines used in this project were wild type, Tg (*runx1*:citrine), Tg(*lck*:GFP)…Small breeding tanks each containing a pair of one male and one female adult zebrafish separated by a divider were set up. The following morning the dividers were removed to allow the fish to mate and embryos collected.

Following collection embryos were grown in petri dishes containing E3 medium with approximately 100 embryos per dish, and kept in an incubator at 28.5°C. After approximately 8 hours, then every day until collection, any unfertilized or dead embryos and debris were removed and E3 medium replaced.

Differential gene expression analysis

Differential gene expression analysis was carried out using R (version 4.2.1). To combine the data files into a combined matrix I used edgeR (v3.40.2) which produces a matrix of raw counts with unique gene identifiers and raw read count associated with each sample. Each file was assigned sample-level information. As the raw count data doesn't into account read and library size, the data was normalised by turning the counts into log-CPM (counts per million). Genes that are lowly expressed were removed using the filterByExpr function in edgeR which keeps genes with 10 or more read counts in a minimum number of samples using CPM values. For our dataset the median library size is … so the function keeps genes that have a CPM of 0.2 or more in at least 12 samples as this is the number of replicates in the dataset. To ensure that the expression distribution is similar in all samples the samples were normalised using the calcNormFacctors in edgeR which calculates trimmed mean of M-values (TMM). The scaling factors were all close to 1 which means that the normalisation is mild. A contrast matrix

was constructed using the Limma package, after which they were compared and individually examined using an adjusted *p*-value cut off of 5%.

crRNA design and RNP preparation

The first three crRNAs suggested by the Alt-R Predesigned Cas9 crRNA selection Tool (Integrated DNA Technologies) were selected for each target gene. Upon arrival, a 200µM stock was made for each crRNA by resuspending each crRNA pellet in Duplex buffer. Each guide was then annealed at 95°C for 5 minutes to 200 µM tracrRNA in a 1:1 ratio to synthesise 57µM gRNA.

To generate RNPs for each target gene, 1µl of the generated gRNA was annealed to 1µl of 57µM Cas9 protein at 37°C for 5 minutes. Following that, RNPs generated from individual crRNAs were pooled together for each target gene, creating a 6µl pool of three RNPs.

0.6µl of Phenol Red was added to each 6µl pool, and 1 pl injected into single-cell-stage zebrafish embryos.

gDNA extraction

Hot Sodium Hydroxide and Tris Method (HotSHOT) method (Truett *et al.,* 2000) was used to extract gDNA from developing embryos. Individual embryos were transferred to PCR tubes where excess glycerol was removed and embryos washed twice with PBSTw for 5 minutes. 50µl of Lysis buffer (20 mM NaOH, 0.2 mM EDTA) was added to each tube, heated to 95°C and then cooled to room temperature. The reaction was stopped by the addition of 50µl of Neutralisation Buffer (40 mM Tris-HCl). Isolated DNA was quantified using Nanodrop and stored at -20°C.

RNA extraction and reverse transcription

To extract RNA, zebrafish embryos were collected and suspended in 300µl of TRIzol™ Reagent (Invitrogen), and macerated using a syringe and needle by aspirating up and down to form a homogenous solution. RNA was purified using the Direct-zol™ MicroPrep kit (Zymo Research), and quantified as previously stated.

To synthesise cDNA for use in qPCR and probe synthesis, 500ng of isolated RNA was reverse transcribed using SuperScript IV (Invitrogen) following the manufacturer's protocol. The synthesised cDNA was diluted 1:10 and stored at -20°C.

Whole-mount *in situ* hybridization (WISH)

Solutions and materials

Solution	**Composition**
PBSTw	1x PBS, 0.1% Tween-20
Hybe-	50% formamide, 5x SSC, 9.2mM citric acid, 0.1% Tween-20
Hybe+	50% formamide, 5x SSC, 9.2mM citric acid, 0.1% Tween-20, 0.05% tRNA, 0.005% Heparin
MABTw	0.1M Maleic Acid, 150mM NaCl, pH 7.5, 0.1% Tween-20
MAB block	2% Blocking reagent (REF 11096176001 Roche) in MAB

AP buffer	30 mL 1M Tris-HCl (pH 9.5), 10 mL 5M NaCl,50 mL 0.5 $MgCl_2$, 2.5 mL 20% Tween, 387.5 mL H_2O

Table 1 : Solutions used in WISH and their composition

Collection of Embryos

Embryos were fixed at various stages in 4% paraformaldehyde (PFA) at 4°C overnight. Embryos at earlier stages of development (24hpf, 30hpf and 48hpf) were dechorionated before being fixed in PFA.

The next day the embryos were washed in PBSTw twice for 5 minutes and then dehydrated in increasing concentrations of ethanol in PBSTw.

Antisense probe synthesis

Day 1

Embryos dehydrated in 100% ethanol were rehydrated in decreasing concentrations of ethanol in PBSTw, and washed twice for 5 minutes in PBST. Embryos fixed at 30hpf, 48hpf, 3dpf, 4dpf or 5dpf were treated with Proteinase K at room temperature. The concentration of proteinase K used, and the length of incubation was dependent on the developmental stage of the embryo, this is summarized in table…Treatment with proteinase K was not carried out for embryos at 24hpf.

Stage	Concentration	Incubation time (min)
30hpf	10cg/ml	10
48hpf	10µg/ml	15
3dpf	20µg/ml	20
4dpf	20µg/ml	28
5dpf	20µg/ml	28

Table 2: Concentration and incubation time with proteinase K at different stages of development.

To stop the treatment, embryos were washed twice for 5 minutes at room temperature with PBSTw, fixed in 4% PFA at room temperature for 20 minutes, and washed again in PBSTw for 5 minutes.

Hybridisation buffer hybe+ was preheated to 65°C, and 200µl was added to each embryo set for 1 hour at 65°C in a heat block. After 1 hour hybe+ was removed, replaced with hybe+ containing probe, and left to incubate overnight at 65°C.

Day 2

The hybe+ solution containing probe was removed, and a series of washes in decreasing concentrations of hybe- in 2x SSC were carried at 65°C. Following this a series of washes in increasing concentrations of MABtw in 0.2x SSC were carried out at room temperature.

To block, the embryos were washed for 1 hour in MAB block (2% Boehringer Blocking Reagent™ in MAB), which was replaced with MAB block containing antibody at a dilution of 1µl anti-DIG in 5ml MAB block, and incubated overnight at room temperature.

To remove the antibody the embryos were subject to eight 15 minute washes at room temperature with MABtw.

All washes were carried out with gentle agitation in a BioLane HTI 16Vx (Intavis Bioanalytica Instruments) in situ hybridization robot.

Day 3

AP buffer was used to equilibrate by washing 3x for 5 minutes at room temperature. This was replaced with AP buffer containing a 1:1 dilution of BM Purple™ and left to stain at room temperature with gentle shaking.

Staining was stopped permanently with 4% PFA for 20 minutes at room temperature (or at 4°C overnight), the embryos were then washed 3 times for 5 minutes with PBStw and stored in 80% glycerol at 4°C.

Imaging

In situ imaging

Prior to imaging, embryos older than 24hpf were bleached on a lightbox with bleaching solution to remove all pigmentation. The bleaching solution was removed by washing the embryos three times with PBSTw for 5 minutes.

Embryos were then mounted in 100% glycerol on a glass microscope slide and imaged using a Nikon SMZ800N microscope, Nikon D5 Fi3 camera, and NIS Elements F software.

Fluorescence imaging

Embryos were anaesthetised in 0.5g/L tricaine, mounted in 3% methylcellulose, and imaged using Zeiss Lumar.12 microscope.

Image analysis

A technique previously developed by the Monteiro lab (Dobrzycki *et al.,* 2018) for the quantification of gene expression following WISH was used. For each embryo, the region of interest where signal was detected was selected by drawing around it using the freehand

selection tool (ImageJ), the image and the pixel intensity measured, to allow for a quantitative comparison of gene expression between injected embryos and their wildtype siblings.

Network construction

Genes identified as upregulated in the HE compared to the AE in the differential gene expression analysis were subject to functional co-expression interaction analysis. Normalised gene expression data for each sample was compiled in a matrix which was then subject to correlation analysis using *rcorr* function in R (v1.4.11.06). The constructed correlation matrix was filtered for correlation values of ≥0.8, and/or -0.8≤, using $p \leq 0.00001$ to select genes which are significantly positively or negatively correlated. Correlation distribution was analysed by histogram using the ggplot2 (v3.4.1) package.

3. Results

3.1 Candidate gene selection

3.1.1 Transcriptional profiling of HE and HSPCs from RNA-seq data

As mentioned previously, HSPCs can exist in a number of distinct and overlapping stages during their development. In an attempt to uncover the mechanisms behind the transition of two such states – the nascent (defined as emerging from the DA) to foetal (proliferating and differentiating mature HSPCs), Ding et al. (Ding *et al*, 2019) carried out transcriptomics analyses on zebrafish HSPCs obtained by fluorescence activated cell sorting (FACS) at 36 hpf and 3 days post fertilisation (dpf) respectively. This analysis revealed a role of epigenetic programming in the maturation of nascent HSPCs. More importantly, generating RNA-seq data allowed for differential gene expression analysis to be performed to identify potential drivers of HSPC maturation and their mechanism during this process. Similarly, in a study which aimed to uncover the origins of the HE, Bonkhofer *et al.* (Bonkhofer *et al*, 2019) carried out transcriptomics analyses on different tissue types, including the haemogenic endothelium and aortic endothelium, isolated from the dorsal aorta, at 29 hpf in developing zebrafish embryos. Combined, these data provide a snapshot of the developing haematopoietic system and allowed us to study the gene expression dynamics at different time points during zebrafish development, starting with the definitive HE at 29 hpf, followed by the emergence of HSPCs at 36 hpf as well as their maturation until 3 dpf.

RNA-seq offers powerful transcriptomics analysis. mRNA from the desired cells is isolated and reverse transcribed into cDNA which is then fragmented into smaller fragments and specific adapter sequences are annealed to the ends of each fragment generating a library. Each fragment in the library is then sequenced producing reads; these are mapped to a reference genome. These raw reads do not take into consideration the total number of reads and the transcript length for each sample, and therefore cannot be used to compare expression levels of genes between samples. To overcome this, raw read data can be

converted to transcripts per million (TPM), reads per kilobase of transcript per million reads mapped (RPKM), or fragments per kilobase of transcript per million reads mapped (FPKM). The data mentioned above were normalised to FPKM and TPM respectively and therefore could not be directly compared, however according to Wager *et al.* (2012), a conversion can be made between the two by dividing the FPKM value by the sum of all FPKM values in the sample, and dividing by one million (Figure 3A); to allow for comparison, data obtained from nascent and foetal HSPCs from the Ding et al study (Ding et al, 2019) were converted to TPM. After this conversion no skewness in the data was observed and all samples were normally distributed. It was found that a total of 2417 genes were expressed in all three tissue types (HE, nascent HSPCs and foetal HSPCs) at varying levels, therefore the expression dynamics of these genes was compared.

HSPCs arise from the ventral wall of the DA from around 30 - 36 hpf (Kissa and Herbomel, 2010). Using this as a rationale for selecting candidate genes, I was interested in genes upregulated at 36 hpf compared to 29 hpf (i.e. nascent HSPCs compared to the HE) to determine which genes are driving the transition between HE specification and HSPC emergence. From the 2417 genes that were expressed in all tissues of interest, most of them (2352) were upregulated in nascent HSPCs compared to the HE. To uncover more about the transcriptome dynamics of newly emerged HSPCs compared to their precursor the HE, I carried out a Gene Ontology term enrichment analysis using Metascape v3.5220230101 (Zhou *et al.,* 2019). The genes upregulated in nascent HSPCs compared to HE were enriched for terms involved in cell structural organisation and adhesion such as “cell adhesion”, “cell migration”, “cell-cell-junction organisation”, and “adherens junction” (Figure 3C) The smaller number of genes which were upregulated in the HE compared to nascent HSPCs were enriched for terms involving ribosomal genes as well as cell cycle progression (Figure 3B).

A

$$\mathrm{TPM}_i = \left(\frac{\mathrm{FPKM}_i}{\sum_j \mathrm{FPKM}_j} \right) \cdot 10^6$$

B

C

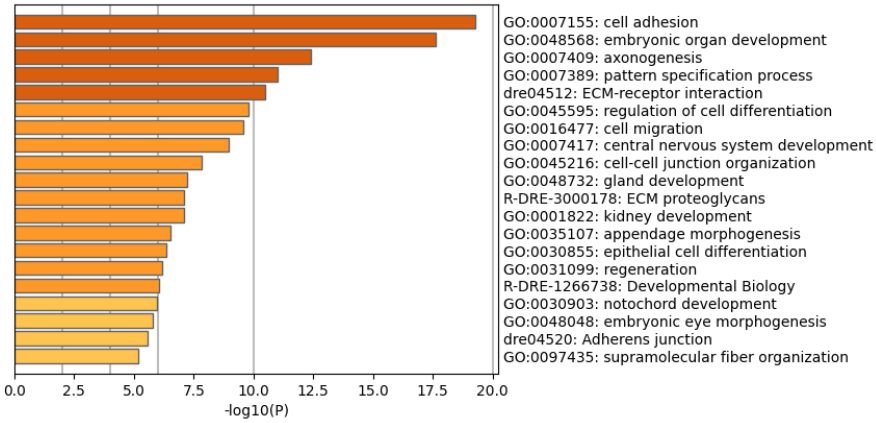


Figure 3: Gene dynamics during HSPC specification and emergence. (A) formula used for the conversion of FPKM values to TPM. Adapted from Wager *et al.* (2012) **(B)-(C)** GO term enrichment analysis for genes upregulated at **(B)** 29 hpf (haemogenic endothelium) and **(C)** 36 hpf (emerging HSPCs. Calculation of enrichment scores and p-values and visual representation of the results was performed by Metascape v3.5220230101.

Secondly, I was interested in the transcriptional changes taking place during the maturation of nascent HSCPs to foetal HSCPs. In this comparison, genes upregulated in nascent HSCs were enriched for GO terms “cell fate commitment”, “cell population proliferation”, “negative

regulation of cell differentiation" and signalling pathways "regulating pluripotency of stem cells" (Figure 4A). Interestingly, a number of these terms correspond to processes which define

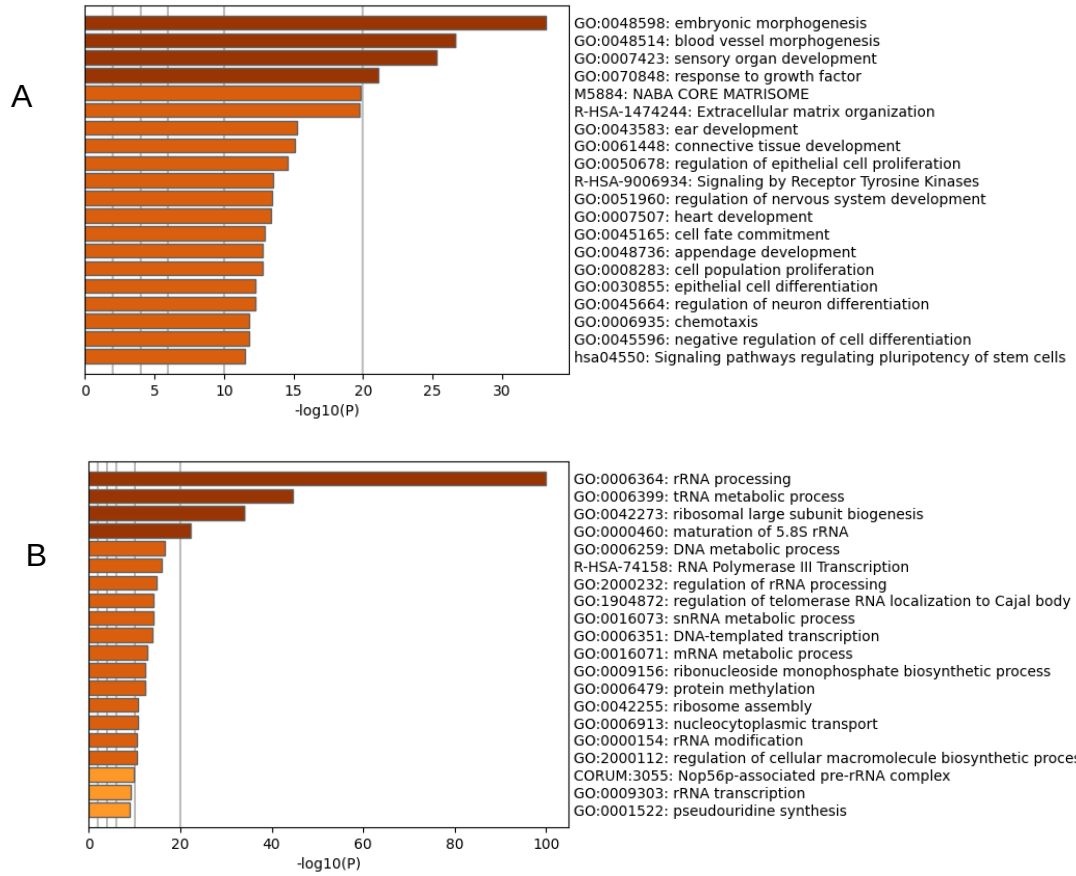


Figure 4: Gene dynamics during HSPC maturation. (A)-(B) GO term enrichment analysis for genes upregulated at **(A)** 36 hpf compared to foetal HSPCs **(B)** foetal HSPCs compared to nascent HSPCs. Calculation of enrichment scores and p-values and visual representation of the results was performed by Metascape v3.5220230101.

mature HSPCs, such as their ability to differentiate and proliferate, suggesting that these processes initiate soon after the emergence of HSCs. The term "chemotaxis" appeared in the list of enriched terms in this gene list, consistently with the need for HSPCs to migrate to intermediary haematopoietic organs soon after their emergence. Finally, I analysed the genes that are upregulated in foetal compared to nascent HSCs and found an enrichment of terms corresponding to ribosomal genes such as "rRNA processing", "ribosomal large

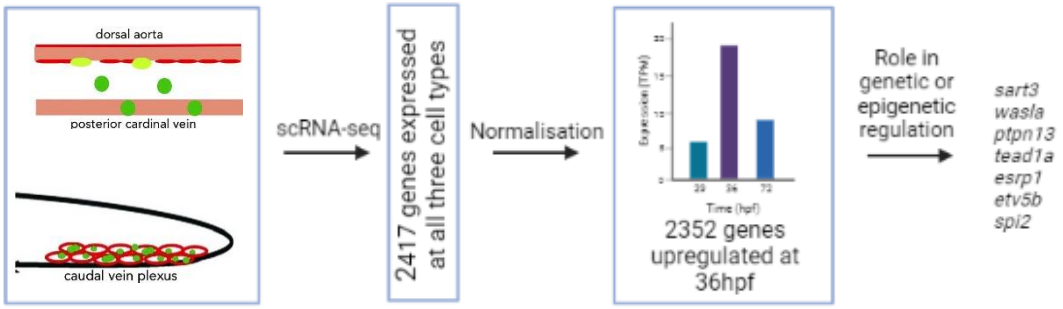


Figure 5: Selection process for identifying novel candidate genes. Transcriptional analysis was carried out on haematopoietic cells isolated at 29 hpf (haemogenic endothelium), 36 hpf (emerging, nascent HSPCs), and 3 dpf (differentiating and proliferating foetal HSPCs). 2417 genes were expressed in all three haematopoietic tissues, and after normalisation to allow comparison between samples, genes upregulated in nascent HSPCs compared to haemogenic endothelium were selected. Genes with a role in genetic or epigenetic regulation were selected for further analysis which identified seven novel players in HSPC ontogeny. Adapted from Ding *et al.* (2021).

subunit biogenesis", and "ribosome assembly" (Figure 4B). This is consistent with the increased need of proliferating cells for protein production.

Of particular interest for this project was the identification of genes driving the transition from HE specification to HSC emergence; we found a number of genes which were upregulated in nascent HSPCs compared to HE cells. Key regulators of HSPC development and early markers of HSPCs, *runx1* and *cmyb,* were not in the list of upregulated genes, however, a number of known players in haematopoietic development were found in the list. Among these were *matn4* which has been shown to have a role in HSC regulation induced by stress (Uckelmann *et al.,* 2016), *cd9b* which is overexpressed in primary myelofibrosis patients and is believed to have a role in cell migration (Bock *et al.,* 2009), and *anxa2a* which regulates the adhesion of HSPCs to osteoblasts (Jung *et al.,* 2007). To further narrow down this selection, ***only genes with a role in genetic or epigenetic regulation with no known haematopoietic phenotype were selected for further analysis***. Following the above criteria, seven potential novel players in haematopoietic development were selected (Figure 5).

3.1.2 Selected candidate HSPC genes

wasla

WASP-like actin nucleation-promoting factor a (WASL in humans) encodes a protein in the Wiskott-Aldrich syndrome protein family, and plays an essential role in actin assembly (Millard and Machesky, 2001) - the coordination of which is necessary for cell motility and its ability to maintain cell shape (Dogterom & Koenderink, 2019). In humans, WASL has been implicated in a number of cancer types and their clinical outcomes; lower expression or loss of WASL has resulted in improved pancreatic (Hidalgo-Satre *et al.,* 2020), cervical (Chen *et al.,* 2021) and breast cancer (Schwickert *et al.,* 2015) outcomes by inhibiting the migration and subsequent invasion of cancerous cells (Wang *et al.,* 2018; Schwickert *et al.,* 2015). Actin cytoskeleton organisation is also required for T-cell effector function (Zhang *et al.,* 2009) and Wiskott-Aldrich syndrome patients often suffer from immunodeficiency due to reduced T- and B-lymphocyte function (Ochs *et al.,* 1980; Westerberg et al., 2005). Additionally, a loss of function screen of potential haematopoietic players in zebrafish found that a knockdown of wasla produced “mild” or “severe” phenotypes in terms of thrombocyte and macrophage numbers respectively (Bielczyk-Maczynska *et al.,* 2014).

ptpn13

Protein tyrosine phosphatase non-receptor type 13 (ptpn13) is a member of the protein tyrosine phosphatase (PTP) family, a group of enzymes responsible for catalysing the dephosphorylation of tyrosine residues on proteins – a common posttranslational modification. Specifically, PTPN13 is believed to have a role in Fas mediated programmed cell death by interacting with the Fas receptor C-terminus via it’s PDZ domain (Yanagisawa et al., 1997). This interaction has also been studied in the haematopoietic system, where PTPN13 has been found to have an inhibitory role of FAS-induced apoptosis in myeloid progenitor cells expressing the Bcr-abl Oncogene – a characteristic of Chronic myeloid leukaemia (CML) (Huang and Eklund, 2013). Interestingly, inhibiting PTPN13 together with BCR/ARL in murine

CML cells results in better treatment outcomes compared to BCR/ARL inhibition alone (Huang et al., 2016). Additionally, PTPN13 has been found to have a regulatory role in megakaryocyte differentiation, through phosphorylation of β-catenin (Sardina et al., 2014).

tead1a

TEA domain family member 1a (tead1a) is a transcriptional enhancer factor containing two binding domains shared by all members of the TEA domain family. The TEA domain contains three α-helices/β-sheets, of which the first and third are required for sequence-specific binding of DNA (Hwang et al., 1993). However, this interaction alone is not enough to regulate transcription and TEAD proteins require coactivator binding; two such transcriptional co-activators identified are YAP, and its paralogue TAZ. It is believed that the binding site required for the formation of this complex is located at the carboxy-terminal 75% portion of TEAD (Vassilev et al., 2001; Li et al., 2010). The interaction between these three proteins is most notable downstream of the Hippo pathway, where its function ranges from the induction of epithelial-to-mesenchymal transition (EMT), to more recently the re-programming of differentiated cells into their progenitors (Wang et al., 2015; Panciera et al., 2016).

esrp1

Epithelial splicing regulatory protein 1 (esrp1) in mammals is an RNA-binding protein responsible for the splicing of FGFR2 to favour the formation of IIIb, the variant found exclusively in epithelial cells, during epithelial-to-mesenchymal transition (EMT) (Warzecha et al., 2009). Unsurprisingly, as switching from epithelial to mesenchymal specification in cells increases their motility, esrp1 has been implicated in many types of cancers and their ability to form metastases. A high expression of ESRP1 has been associated with a favourable outcome following diagnosis of pancreatic cancer (Ueda et al., 2014), colorectal cancer (Deloria et al., 2016), breast cancer, and clear-cell renal cell carcinoma (Lu et al., 2015).

etv5b

ETS Variant Transcription Factor 5 belongs to a family of transcription factors containing the ETS domain which, through its winged helix-turn-helix structure, allows for the binding to DNA sites containing a GGA (A/T) sequence to take place. Even though this transcription factor is best characterised in spermatogenesis, its role in angiogenesis has also been described mammal and zebrafish models (Wythe *et al.,* 2013). More specifically, ETV5 has a role in EMT through the activation of Zeb1, a negative regulator of E-Cadherin; endometrial cells undergoing EMT regulated by ETV5 are more migratory and invasive (Colas *et al.,* 2012). Additionally, other ETV factors in the arterial endothelium are involved in Vegf/MAPK-dependent pathway to specify arterial identity (Wythe *et al.,* 2013). Similarly to TEA domain factors, it is unlikely that ETS transcription factors are able to regulate the transcription of their target genes through binding alone, and require the association of transcriptional partners.

spi2

Spi-2 proto-oncogene is another factor in the ETS transcription factor family which is exclusively found in ray-finned fish. *spi2* is a newly discovered gene and its function is largely uncharacterised., however its role in haematopoietic development has been recently uncovered. It has been found to be differentially expressed in the HE compared to endothelial cells from the roof of the DA as well as being expressed in the AGM at 24hpf and CHT at 2dpf. Loss of *spi2* has little effect on the formation of the HE or the emergence of HSPCs, however HSPCs in the CHT showed compromised proliferation and myeloid lineage defects; it is believed to regulate this process by acting downstream of *runx1* (Zhao *et al.,* 2022). Spi2 has been found to share 34-40% protein sequence alignment with other paralogs, and 55-61% with other members of the ETS domain family (Zhao *et al.,* 2022).

3.1.3 Analysis of candidate gene function in haematopoiesis using CRISPR/Cas9 gene editing

To investigate how candidate HSPC-enriched genes affect haematopoiesis, we used the AltR CRISPR-Cas9 system. In particular, we used a two-part chemically modified guide RNA

(crRNA:tracrRNA) combined with *S. Pyogenes* Cas9, injected as a ribonucleoprotein complex (RNP). This method of gene knockout was chosen due to its high efficiency in zebrafish (Hoshijima et al., 2019; Kroll *et al.,* 2021), which allows for rapid, small-scale screening in F0 embryos without the need to establish mutant lines. These F0 mutants are commonly referred to as *crispants*. To target the loss of function of candidate genes in haematopoietic cells specifically, we aimed to generate a zebrafish transgenic line using a *cmyb* enhancer to drive Cas9 expression in haematopoietic cells in the embryo prior to HSPC emergence, as well as a dll4 enhancer which will drive Cas9 expression in all arterial populations, including the HE. A GFP reporter was integrated in both constructs to identify transgenic animals. The constructs were generated using the AcDs transposition system, previously used in zebrafish (Burgess *et al.,* 2013). The construct was injected into single-stage embryos and observed until 5 dpf for GFP expression in haematopoietic tissues. Unfortunately, no GFP expression corresponding to *cmyb* expression was observed at any time point (data not shown) and this approach was deemed unsuitable. The CRISPR-cas9 was therefore carried out in wildtype embryos.

3.2 Screen design and validation of the approach –*sart3 crispants*

Two time points were selected to test the efficiency of our strategy to capture changes in haematopoietic marker expression upon candidate gene knockout – 30 hpf and 4 dpf. These two time points allowed for two essential stages of haematopoietic development to be studied - the initiation of the definitive wave commencing with the specification of the HE, as well as the ability of mature HSPCs to proliferate and seed haematopoietic organs.

To validate our approach, we required a gene with known function in haematopoiesis to function as a positive control. Spliceosome associated factor 3 (also known as squamous cell carcinoma antigen recognised by T cells 3), or *sart3*, is a component of the spliceosome which is involved in eukaryotic pre-mRNA splicing through the reassembly of U4/U6 small nuclear ribonucleoproteins (snRNPs) (Bell, 2002; Liu et al., 2015). The role of *sart3* in post

translational modification of histones has also been noted. Sart3 has histone chaperone properties, and is able to bind H2A/H2B dimers, H3/H4 tetramers, histone octamers, but not mono-nucleosomes, acting as a substrate recruiter for Usp15 - a histone deubiquitinase (Long *et al.,* 2014). A study by Zhao *et al.,* (Zhao et al, 2021) uncovered the role of spliceosome factor *sart3* as a regulator of HSPC development through the p53 pathway. They found that *cmyb* expression was affected in heterozygous (*sart3*^{smu471}*)* mutants compared to their wildtype siblings: while at 2dpf the expression of *cmyb* was comparable to wildtype siblings, it steadily decreased from 2.5 dpf onwards and was almost absent at 4 dpf. This phenotype was not exacerbated by the deletion of both alleles and biallelic (*sart3*$^{smu471/\ \Delta 14+4}$*)* mutants experienced the same level of haematopoietic impairment.

cmyb is an early HSPCs marker and as they arise from the ventral wall of the DA at around 36 hpf, *cmyb* expression is located in the AGM and marks the HE and emerging HSPCs. Single cell stage zebrafish embryos were injected with an RNP complex containing three sgRNAs and cas9 protein to target *sart3.* The injected embryos, together with their wildtype siblings, were raised and collected at 30hpf, fixed in 4% PFA, stained using a newly synthesised *cmyb* antisense probe, and imaged. The expression of *cmyb* was quantified using the ImageJ freehand selection tool to measure pixel intensity in the AGM region compared to the background (Dobrzycki et al, 2018); this correlates to relative *cmyb* expression in the AGM and allowed for the average levels of *cmyb* expression between wildtype and injected (*sart3* crispant) embryos to be compared. One-way (unpaired) T-test was chosen as an appropriate statistical test for this analysis, with the assumption that the collected data is normally distributed. Accepted significance value for this analysis was $p \leq 0.05$. Consistently with previous findings by Zhao *et al.* (2021), the one-way t-test revealed that at 30 hpf there was no significant difference in *cmyb* expression between *sart3* crispants and their wildtype siblings. The second developmental stage to be characterised in this study was 4 dpf. As previously, single cell stage embryos were injected with an RNP complex targeting *sart3*, raised until 4 dpf, fixed in 4% PFA, stained using *cmyb* anti-sense probe and imaged. At this stage we detected an almost complete loss of *cmyb* expression in the CHT. Because the

phenotype was so severe, no quantification analysis was carried out. These results were consistent with previous finding by Zhao *et al.,* (Zhao *et al*, 2021) as described above. We therefore concluded that the CRISPR-cas9 system we selected could induce an efficient loss of function and is an appropriate approach for a large-scale loss of function genetic screen, capable of mimicking a phenotype observed in stable mutants.

3.3 Functional screening of candidate HSPC genes

3.3.1 Loss of function candidate gene crispants show impaired HE specification and HSPC emergence

As previously mentioned, *cmyb* and *runx1* are essential drivers of HSPC biology and are early markers of the HE and HSPCs. At 30 hpf both markers are expressed in the AGM region, including the ventral wall of the DA; *runx1* is exclusively expressed in the DA and marks the HE Therefore, to characterize the effect of loss of candidate gene function, we started by characterizing expression of the haematopoietic markers *cmyb* and *runx1* at 30 hpf, which marks the beginning of the definitive wave of haematopoiesis and gives rise to HSPCs. All selected candidate genes showed some impairment in *cmyb* and/or *runx1* expression at this time point.

Wasla – *wasla* crispants showed a significant increase in *cmyb* expression at 30hpf (Figure 6A). This increase was specific to the AGM region of the embryo where HSPCs arise, and the expression of *cmyb* in the head, and the rest of the trunk were comparable to that seen in wildtype siblings when adjusted for the background. This is indicative of increased *cmyb* expression specifically in HSPCs either as a result of increased HSPC numbers or increased expression of *cmyb* in individual cells. The same analysis revealed that *runx1* expression was highly variable between (even sibling) embryos, however the spread of values corresponding to *runx1* expression was much more variable in *wasla* crispants compared to the control group, with a small number of embryos expressing barely detectable levels of *runx1*¸ and others

expressing higher levels than wildtype controls. Overall, we detected no significant changes in *runx1* expression between *wasla* crispants and their wildtype siblings (Figure 6B).

Tead1a - *Tead1a* is a transcriptional enhancer in the TEA domain family which is required for DNA-specific binding (Hwang et al., 1993). At 30 hpf no impaired expression of *cmyb* was observed (Figure 6E), however there was a significant increase in *runx1* expression in the AGM region (Figure 6F). *Tead1a* is a known regulator of EMT, which shares many similarities with EHT; during EHT, HE cells lose their endothelial identity and become less adhesive and more motile. It was recently found that that YAP is a co-activator of *tead1a* and together they form a complex that acts downstream of the Hippo pathway. Indeed, it was found that *yap1* and *mst1* (the zebrafish orthologue of hippo), are both upregulated in nascent HSPCs at 36 hpf together with *tead1a* providing further evidence of the role of Hippo signalling in HSPC ontogeny*.*

Esrp1 - Loss of *esrp1,* an epithelial cell-specific splicing regulatory protein(Warzecha *et al.,* 2009), resulted in a significant increase in *cmyb* and *runx1* expression at 30 hpf in the AGM region (Figures 6G and H). *Esrp1* and its paralogue *esrp2* have been shown to alternatively splice FGFR2, a fibroblast growth factor receptor, in epithelial cells to favour the formation of FGFR2 IIIb, the splicing variant found exclusively in epithelial cells, and therefore suppress the specification of the mesenchymal programme during EMT. Considering the results, it can be hypothesised that mechanistically, *esrp1* acts similarly on FGFR2 expressed in endothelial cells to reduce the expression of adhesion molecule related genes, shifting the balance away from the endothelial specification and in turn towards the haematopoietic specification during EHT. Contrary to what as observed in *wasla* crispants, the increase in *cmyb* expression wasn't specific to the AGM region and was also seen in the trunk, head

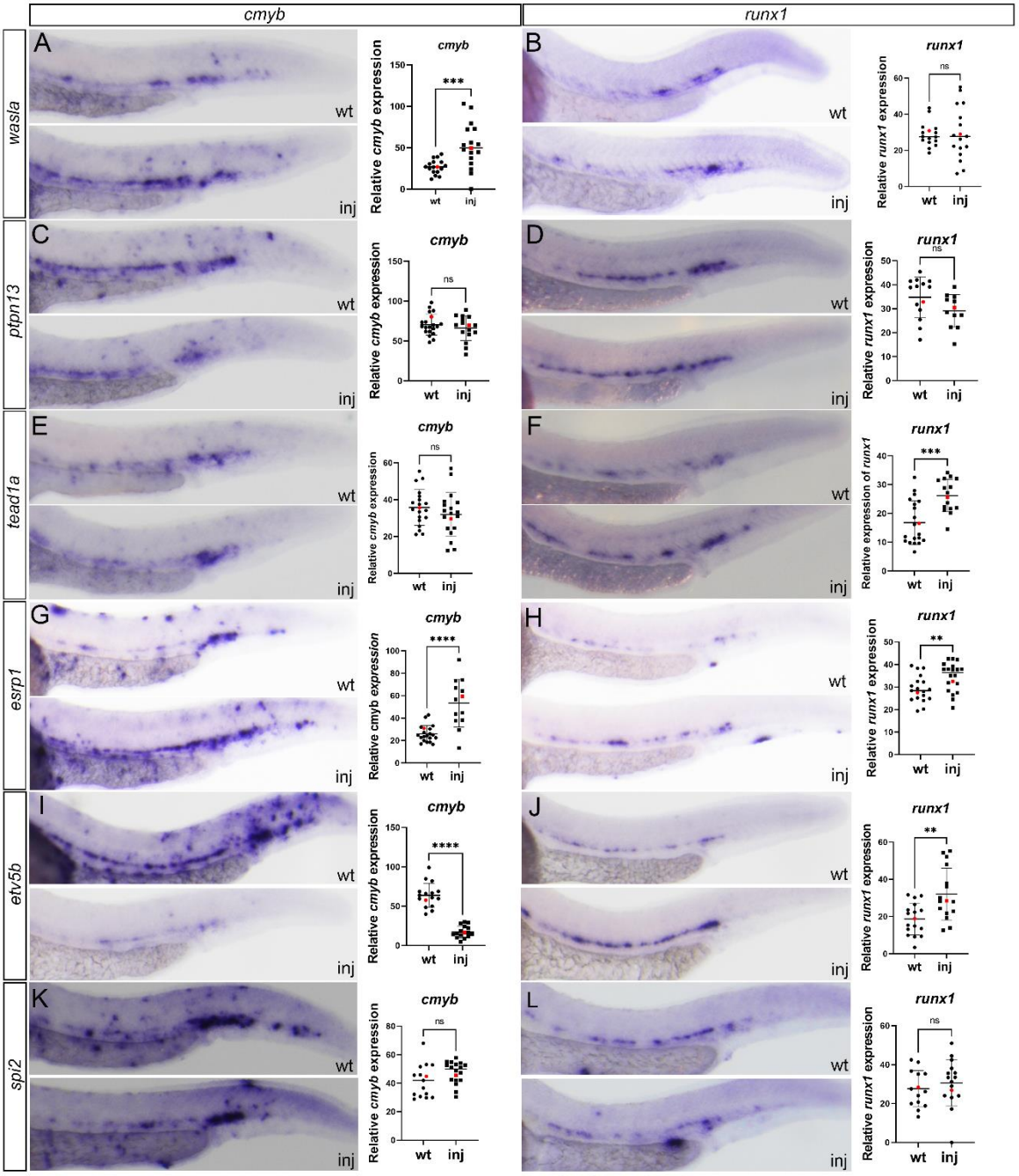


Figure 6: Early haematopoietic marker expression upon loss of candidate genes. Expression of early haematopoietic markers *cmyb* **(A), (C), (E), (G), (I)** and *runx1* **(B), (D), (F), (H), (J)** analysed by WISH at 30 hpf in wildtype embryos and candidate gene crispants. Graphs next to each set of images represent the expression level of *cmyb* and *runx1*, each red data point corresponding to the image shown. Data represents mean ±SD; parametric t-test was used to assess differences between treatment groups; accepted significance value for this analysis was $p \leq 0.05$.

and tail of the embryos. This could be indicative of vasculature defects in the embryo and led us to believe that this increase in *cmyb* expression could have been the result of impaired

circulation in the embryo and the accumulation of cells expressing *cmyb* in the DA that are unable to enter circulation, instead of an increase of HSPCs. *cmyb* is also a known regulator of and is expressed in primitive erythrocytes (Bartunek *et al.,* 2003). However, *esrp1* crispants were observed until 5 dpf and showed no impaired circulation or developmental delay, therefore the defects in vasculature, if present, are very minor and the increase in *cmyb* and *runx1* likely reflects an increase in specification of HE.

Etv5b - The most severe phenotype in terms of *cmyb* expression at 30 hpf was observed upon knockout of *etv5b*, a transcription factor containing an ETS DNA-binding domain (Figure 6I). This increase was
observed in the floor of the DA but also overall the whole embryo (Figure 6I). Interestingly, the opposite – a significant increase – was observed in terms of *runx1* expression (Figure 6J). This is indicative that *etv5b* is not a direct regulator of HSPC emergence and is instead acting on mechanisms prior to the definitive haematopoietic wave, more specifically a mechanism dependent on *cmyb* expression. As primitive erythrocytes also express *cmyb, etv5b* could be acting on primitive erythropoiesis which would explain the decrease in *cmyb* but not *runx1* expression. u

At the time of the initial data analysis and selection of candidate HSPC genes, *spi2* was found to be enriched in the 36 hpf nascent HSPC population together with the other candidate genes selected for this screen and was largely uncharacterised in terms of its function. Since then, evidence has emerged of its involvement in HSPC development (Zhao *et al.,* 2022). It was found that despite being enriched in the HE, it played no role on the specification of the HE or the emergence of HSPCs. This was consistent with our results, where we found no impairment in *cmyb* or *runx1* expression between *spi2* crispants and their wildtype siblings at 30hpf (Figure 6K and L).

3.3.2 Screening revealed impaired HSPC proliferation and colonisation of haematopoietic organs

Following the emergence of HSPCs from the ventral wall of the DA, HSPCs travel to the intermediate haematopoietic organ – the CHT, where they proliferate before entering circulation and seeding the kidney marrow, where they will remain for the duration of the adult life and give rise to all blood cell lineages. At the same time, however, a smaller number of HSPCs (T-cell progenitors) will travel from the CHT to the thymus where they will differentiate into mature T-cells. These cells begin seeding the CHT and the thymus at 48 hpf and 54 hpf dpf respectively (Chen and Zon, 2009; Kissa *et al.,* 2008). We aimed to uncover the roles of candidate genes and their role in multiple stages of haematopoietic stem cell ontogeny, including their ability to proliferate and seed the haematopoietic organs. As previously, wildtype embryos were injected with an RNP mix targeting three regions of the gene of interest, but this time collected and fixed at 4 dpf, and stained for *cmyb* (HSPC marker) and *rag1* (T-cell progenitor marker) expression. At this stage *cmyb* is still expressed in HSPCs which have colonised the CHT and the thymus, and *rag1* exclusively marks T-cells and their progenitors.

Ptpn13 is a protein tyrosine phosphatase, which dephosphorylates tyrosine residues on proteins, most notably FAS receptor. At 30 hpf *ptpn13* crispants showed slightly impaired *cmyb,* but not *runx1* expression. At 4 dpf, no differences between *ptpn13* knockouts and their wildtype siblings were observed in terms of *cmyb* expression (Figure 7C). However, a complete absence of *rag1* expression was observed in the thymus at 4 dpf (Figure 7D) Lymphopoiesis begins in the zebrafish embryo at around 3 dpf, and this complete absence of *rag1* expression at 4 dpf could indicate delayed seeding of the thymus by progenitors or impaired T-cell gene expression. To distinguish between loss of T cell progenitors and loss of *rag1* expression, we repeated the experiment in the Tg(*lck*:GFP) zebrafish transgenic line which labels T-cell progenitors in the thymus (Shimizu *et al.,* 2001). As previously, Tg(*lck*:GFP) embryos were injected with an RNP mix containing three sgRNAs targeting

ptpn13 and imaged live at 5 dpf. For consistency it was intended for these embryos to be imaged at 4 dpf as previously collected for WISH, however no *lck* expression was detected at 4 dpf in either the injected embryos or their siblings. Performing a screen in this transgenic line revealed a significant reduction in the size of the area where *lck* is being expressed in, but not a complete absence of GFP expression (Figure 8F). We concluded that seeding of the thymus was likely delayed. In addition, given that

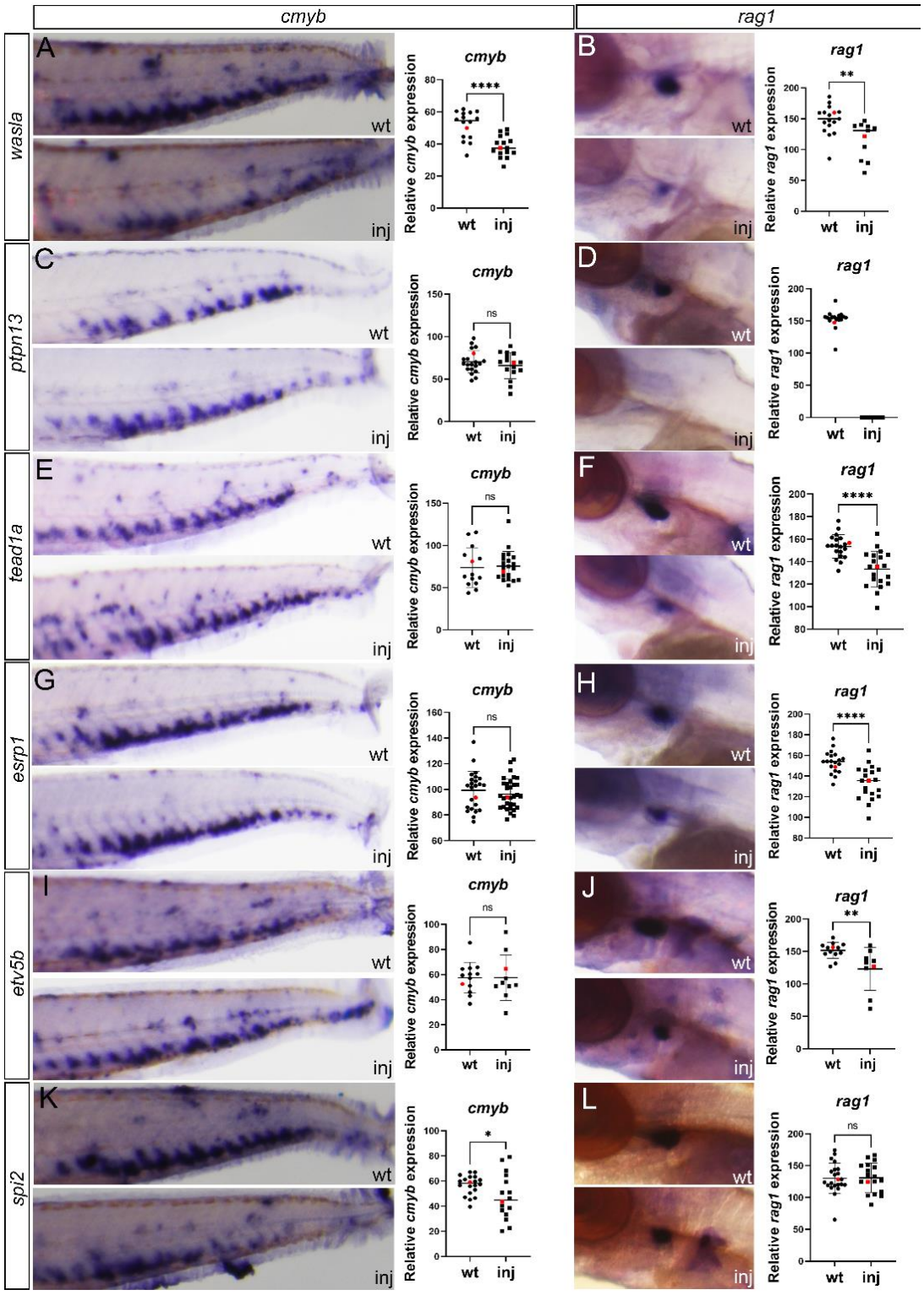


Figure 7: Colonisation of haematopoietic organs upon loss of candidate genes. Expression of the haematopoietic marker *cmyb* **(A), (C), (E), (G), (I)** and T-cell marker *rag1* **(B), (D), (F), (H), (J)** analysed by WISH at 4 dpf in wildtype embryos and candidate gene crispants. Graphs next to each set of images represent the expression level of *cmyb* and *rag1*, each red data point corresponding to the image shown. Data represents mean ±SD; parametric t-test was used to assess differences between treatment groups; accepted significance value for this analysis was $p \leq 0.05$.

lck-GFP positive cells were present in the thymus, rag1 expression was severely impaired, suggesting that the T cell lineage was not properly specified.

Similarly, no impairment of *cmyb* expression were observed in *tead1a* crispants at 4 dpf in the CHT (Figure 7E), however in a number of embryos there is an increase in *cmyb* expression in the trunk of the embryo. By scoring the level of *cmyb* expression in the trunk of the embryo as "low" i.e. comparable to wildtype expression (Figure 8B), "medium" (Figure 8C) or "high" (Figure 8D) we found that a higher number of *tead1a* crispants were expressing "high" levels of *cmyb* compared to wildtype levels. At 4 dpf *cmyb* is also expressed in definitive erythrocytes which might be the source of *cmyb* expression as previously hypothesised for *wasla* crispants at 30 hpf. There was also a significant reduction in *rag1* expression in the thymus (Figure 7F). This was observed in terms of intensity of expression and the size of the area where it's being expressed. *Rag1* expression in the thymus was measured using ImageJ by selecting the area of a representative thymus, which was used as a reference for all other images that were analysed, therefore taking into account both the intensity of expression of *rag1* and the size of the area in which its being expressed i.e. the thymus. There is evidence for a role in the differentiation of leukocyte development of non-canonical hippo signalling, an upstream of regulator of *tead1a* expression (Kurz *et al.,* 2018), therefore the expression of *lymphocyte cytosolic protein 1* (*lcp1*/*L-plastin*), a pan-leukocyte marker in *tead1a* crispants was explored. There was a significant increase in *l-plastin* expression at 4 dpf in *tead1a*-injected embryos compared to their wildtype siblings (Figure 8E).

Contrary to the significant increase in *cmyb* and *runx1* expression observed at 30 hpf in the AGM, the level of *cmyb* expression in the CHT appeared to be comparable to wildtype levels at 4 dpf in *esrp1* crispants (Figure 7G). This return to normal expression levels could be indicative of a number of things. Firstly, an increase in *cmyb* expression does not automatically suggest an increase in the number of HSPCs arising from the DA, but could be indicative of higher transcript levels of *cmyb* in the same number of expressing cells. Additionally, there

was a significant decrease in *rag1* expression (Figure 7H) in *esrp1* crispants, suggesting a defect in specification or T-cell progenitors or expression of T-cell specific gene programmes.

At 4 dpf, loss of *etv5b* lead to reduced *rag1* expression in the thymus, while *cmyb* expression in the CHT was comparable to wildtype (Figure7I), in contrast to the effect observed at 30hpf (Figure 6I). Overall, loss of *etv5b* lead to defective *rag1* expression in the thymus, and a potential role in the programming of this lineage. However, the recovery of *cmyb* expression by 4dpf suggested that *etv5* plays a more important role prior to HSPC emergence in the initial regulation of gene expression in the HE. It is possible that the increased *runx1* expression at 30hpf overrides the initial loss of *cmyb*. Indeed, previous research indicates that overexpression of one of the runx1 targets in the HE, *dnmt3bb.1*, upregulates *cmyb* expression in the CHT through methylation of its regulatory sequences (Gore *et al.,* 2016).

Confirming previous findings by Zhao et al. (2021), loss of *spi2* appears to impair HSPCs proliferation as seen by the reduction of *cmyb* expression in the CHT region although not as severely as seen in stable *spi2* mutants. The previous study only investigated the effect of *spi2* loss on the ability of HSPCs to seed the CHT and proliferate, not their ability to colonise the thymus or give rise to mature T-cells or their progenitors. In our hands, the expression of *rag1* in *spi2* crispants at 4 dpf showed no impairment in *rag1* expression, suggesting no differentiation defects for this lineage.

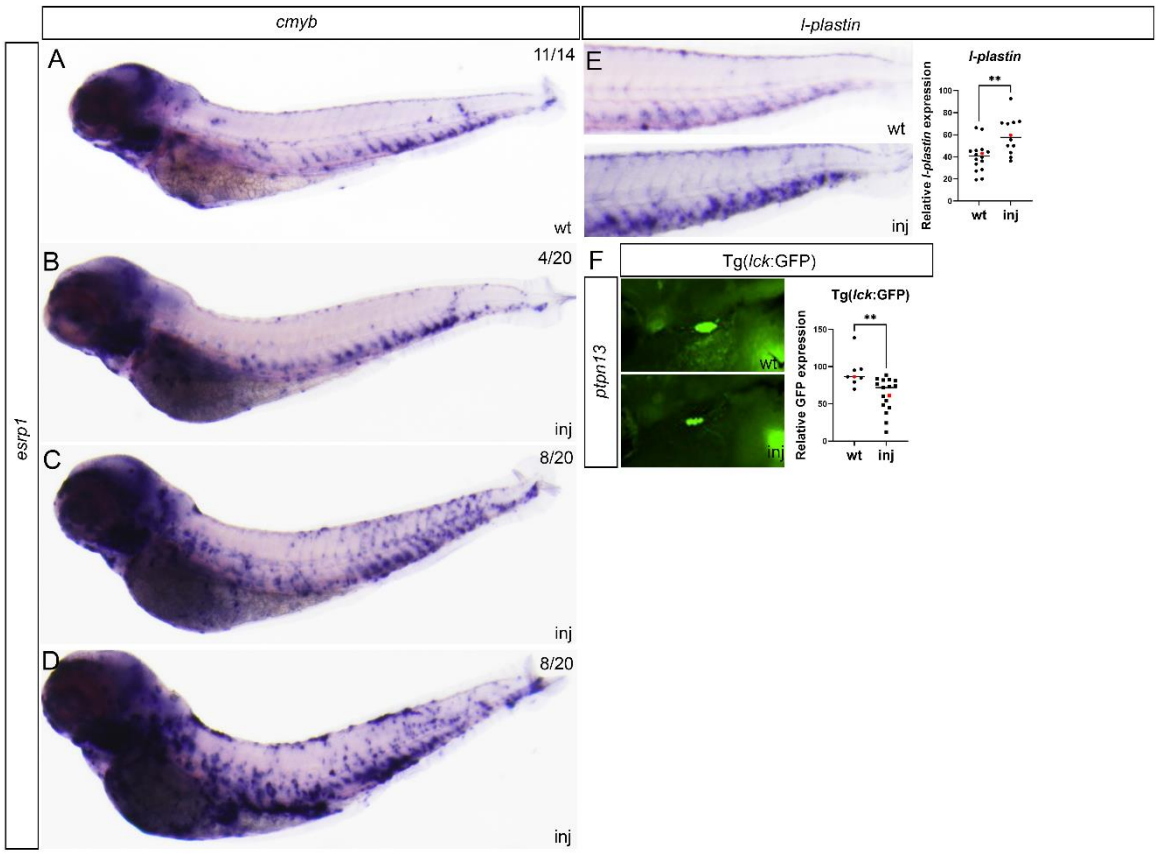


Figure 8: Loss of *t* and ptpn13 leads to impaired haematopoietic gene expression. (A)-(D) Expression of the haematopoietic marker *cmyb* as analysed by WISH at 4 dpf in wildtype embryos and *esrp1* crispants. Injected embryos showed a higher variation of *cmyb* expression in the trunk and head but not the CHT. The level of *cmyb* expression can be categorised as either **(B)** low i.e. comparable to wildtype **(A), (C)** "medium", or **(D)** "high". The numbers at the bottom of each image indicate the number of embryos that exhibit that phenotype in each treatment group. **(E)** *l-plastin* expression in wildtype embryos and *esrp1* crispants. **(F)** Expression of *lck* in wildtype and *ptpn13* injected 5 dpf embryos as visualised by GFP expression. Graphs next to each set of images represent the expression level of *cmyb* and *rag1*, each red data point corresponding to the image shown. Data represents mean ±SD; parametric t-test was used to assess differences between treatment groups; accepted significance value for this analysis was $p \leq 0.05$.

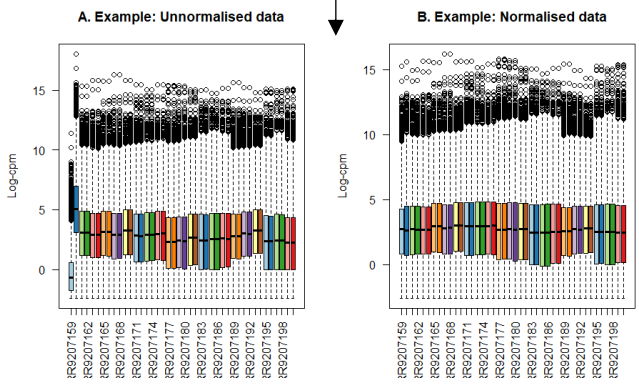
3.4 Investigating dynamics of genes overexpressed in the HE

HE AE NAE NET

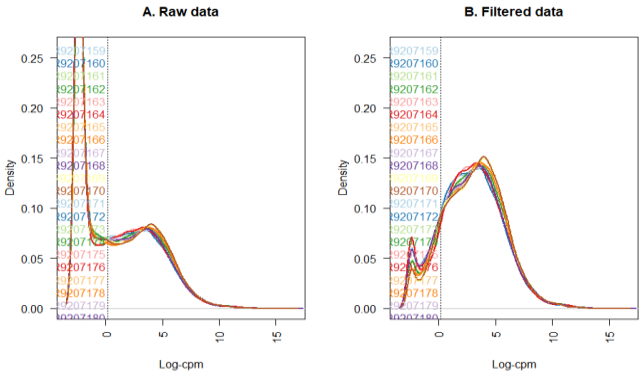
edgeR

	HE	AE	NAE	NET
gene1				
gene2				

Normalization



Removing lowly expressed genes



Differential gene expression analysis

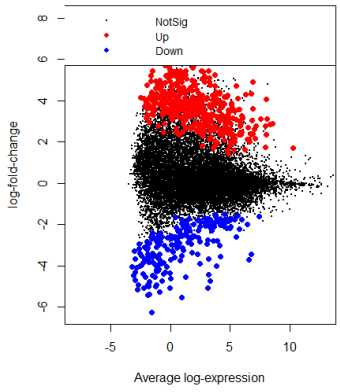


Figure 9: Workflow of differential gene expression analysis. Raw read counts from four tissue types were compiled into a matrix using edgeR. The data was then normalised and lowly expressed genes removed. 1641 upregulated genes in the HE were identified.

3.4.1 Differential gene expression analysis

As previously mentioned, Bonkhofer *et al.* (2021) carried out scRNA-seq analysis on tissues isolated from the DA. These included the haemogenic endothelium (HE), aortic endothelium (AE), non-aortic endothelium, as well as non-endothelial tissue. Differential gene expression analysis was carried out to uncover enriched genes in each tissue type when compared to each other tissue. Because the HE derives directly from the AE, of particular interest were the transcriptional changes that are taking place during this transition from aortic to haematopoietic states.

This analysis was carried out using R. The workflow of the analysis is summarized in Figure… A total of 1641 genes were enriched in the HE compared to AE, among these were the essential haematopoietic players *gata2b, runx1, cmyb*. These genes were subject to functional enrichment analysis by constructing protein interaction network in String (v11.5) using the "Proteins with Values/Ranks" tool to input the gene list together with their level of enrichment. It was found that these genes were enriched for terms involved in ribosomal synthesis and modification. Similarly, genes upregulated in aortic, compared to haemogenic, endothelial cells were also subject to functional enrichment analysis and revealed an enrichment in terms for structure and adhesion.

3.4.2 Network construction

Genes that were enriched in the HE compared to the AE were subject to further analysis to explore their correlations and functional interactions. A correlation matrix was constructed using R to discover networks of interacting genes; a total of 17539 nodes were constructed in the network. Upon looking more closely, the distribution of the correlation revealed a subgroup

(cluster 1) (Figure 10A) which contained the largest number of genes with only 1 other node, and the number of nodes per gene followed an inverse relationship i.e. the largest number of genes had fewest connections. No genes were found to have between 1205 and 3537 connections, and a second, larger subgroup (cluster 2) emerged, which contained 4609 genes which all contained around 4000 nodes each. Upon further analysis using String, a significant enrichment in ribosomal genes was identified in subset 2, however the ribosomal genes were the only ones interacting as visualised by the network. The remaining genes showed no enrichment and produced very few interactions (Figure10C).

Cluster 2, however produced a more connected network of interacting genes which were enriched for terms involved in cell cycle and replication https://version-11-5.string-db.org/cgi/network?networkId=bP0h01DPD2Bz. It was decided that this cluster gives more meaningful results and this is what we concentrated on for this analysis. From this cluster we aimed to discover a network of genes which are interacting to each other on a functional level, more specifically a network for our genes of interest being screened. No network was identified for *sart3, wasla, ptpn13, tead1a,* or *esrp1,* most likely due to the strict significance requirements and high correlation used in this analysis. Analysing interactions with *etv5b* and *spi2,* revealed that *etv5b* is communicating to another 98 genes which are enriched for terms involved in cell cycle and replication such as “negative regulation of sister chromatic cohesion”, “kinetochore assembly”, and

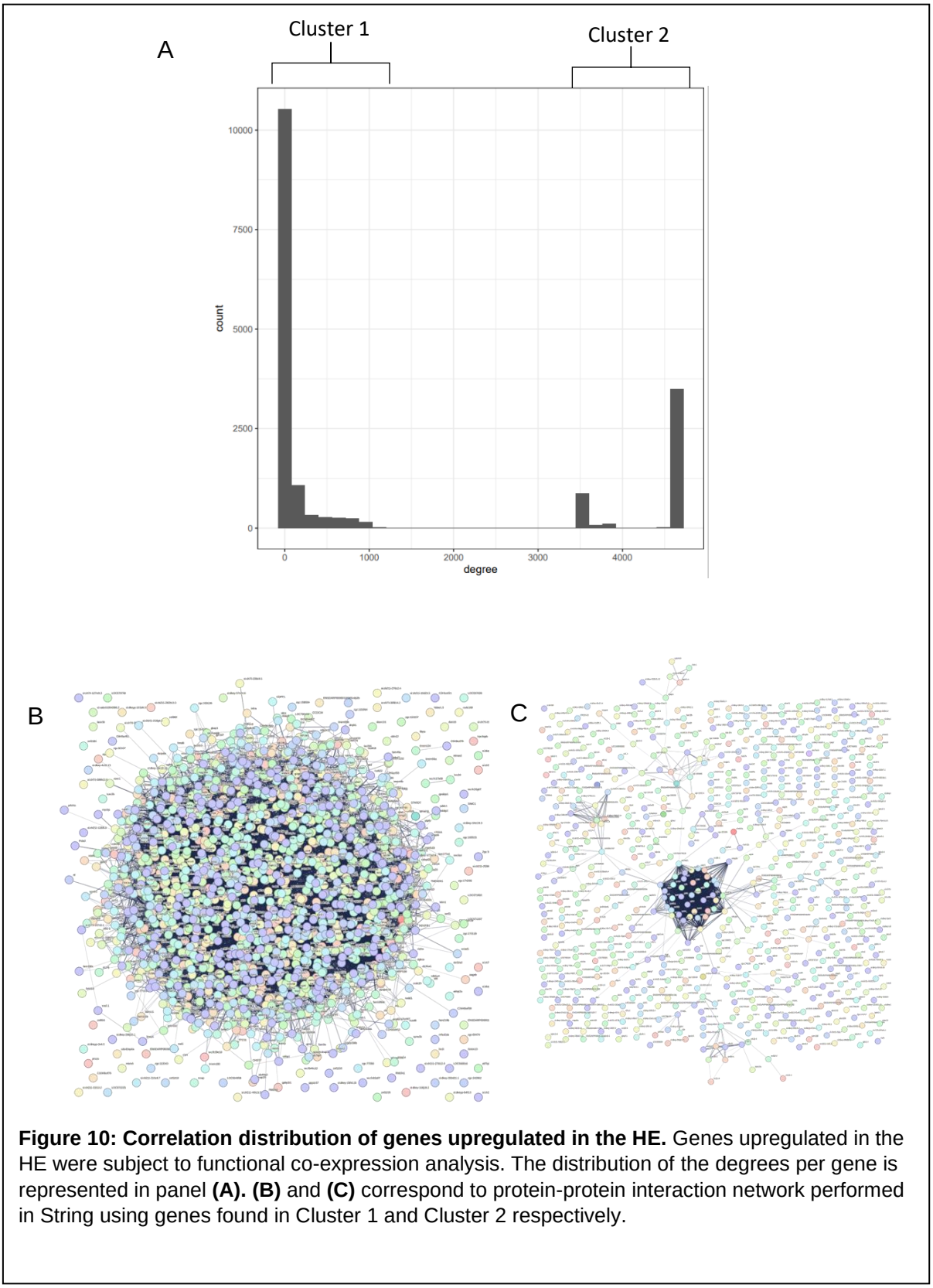


Figure 10: Correlation distribution of genes upregulated in the HE. Genes upregulated in the HE were subject to functional co-expression analysis. The distribution of the degrees per gene is represented in panel **(A). (B)** and **(C)** correspond to protein-protein interaction network performed in String using genes found in Cluster 1 and Cluster 2 respectively.

“mitotic metaphase plate congression”. Upon analysis of *spi2* interactions, it was found that it is only interacting with three other genes – *spi1* (*pu.1)* the paralogue of *spi2* with a knock function in haematopoietic stem cell emergence and a regulator of myeloid specification (Nerlov and Graf, 1998), *thy1,* and *ncf1*. These are overviewed below.

3.4.3 Genes of interest

thy1

Thy-1 cell surface antigen is a small cell surface glycoprotein most notably involved in cell-to-cell adhesion and communication. Thy-1 is unique in its ability to act as a receptor and a ligand, and its cellular effects and signalling pathways vary not only between cell types but also within (Herrera-Molina *et al,* 2013; Herrera-Molina *et al.,* 2012). Its versatility is also reflected in its ability to regulate processes in many cell types including neurons, mesenchymal stem cells (Choi *et al.,* 2008), endothelial cells (Wandel *et al.,* 2012), fibroblasts (Koumas *et al.,* 2003), and haematopoietic progenitor cells (Craig *et al.,* 1993). In humans Thy1 has been used as a marker for HSPCs (Calvanese *et al.,* 2022) however its specific mechanism of action in haematopoiesis has been proven difficult to establish, possibly due to its varying expression distribution between (even closely related) species (Haeryfar and Hoskin, 2004).

ncf1

Neutrophilic cytosolic factor 1 (*ncf1*), encodes for p47-phox, a cytosolic subunit of the neutrophilic NADPH oxidase (Zhao *et al.,* 2017). It is a regulator of the NOX2 complex which is responsible for the production of reactive oxygen species (ROS) (Sareila *et al.,* 2017). Because of this *ncf1* has been implicated in a number of autoimmune diseases including rheumatoid arthritis (Zhao *et al.,* 2017). It was also recently found that NCF1 (together with NCF2 and 4) is a marker of poor prognostic outcomes in kidney clear cell carcinoma cases (Chen *et al.,* 2021).

3.4.4 Loss of *thy1* leads to impaired expression of haematopoietic marker genes

The newly selected candidates *thy1* and *ncf1* were screened using the same CRISPR-cas9 approach to investigate whether, similarly to spi2, it has an effect on haematopoietic

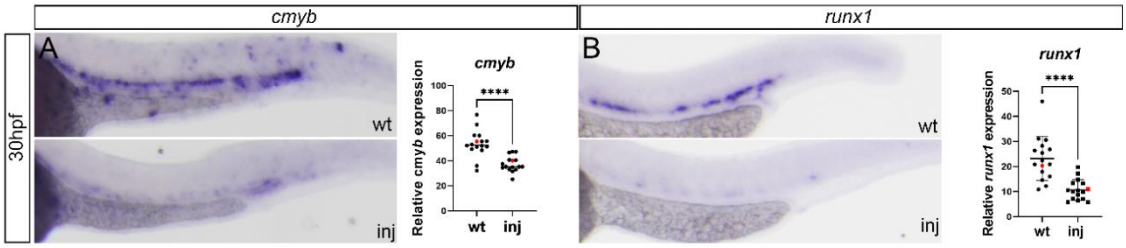


Figure 11: Early haematopoietic marker expression upon loss of *thy1*. Expression of early haematopoietic markers *cmyb* **(A)** and *runx1* **(B)** analysed by WISH at 30 hpf in wildtype embryos and *thy1* crispants. Graphs next to each set of images represent the expression level of *cmyb* and *runx1*, each red data point corresponding to the image shown. Data represents mean ±SD; parametric t-test was used to assess differences between treatment groups; accepted significance value for this analysis was $p \leq 0.05$.

development. Upon loss of *thy1* the expression of early haematopoietic marker genes *cmyb* and *runx1* was severely decreased in the AGM region at 30 hpf (Figures 11A and B). This phenotype persisted until 4 dpf where a significant reduction in the average expression of *cmyb* and *rag1* in the CHT and the thymus respectively (Figures 12D and H). Interestingly, in both cases, a very varied phenotype was observed at 4 dpf. *Cmyb* expression was significantly reduced in 7 out of 16 *thy1* injected embryos compared to wildtype siblings (Figure12 C), however the remaining 9 embryos exhibited *cmyb* expression levels comparable to wildtype.

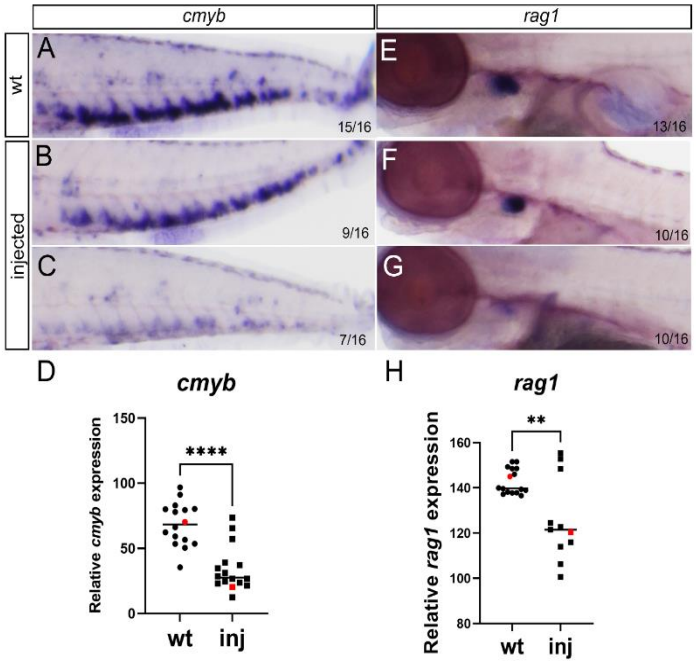


Figure 12: Colonisation of haematopoietic organs upon loss of candidate genes. Expression of the haematopoietic marker *cmyb* **(A), (B), (C), (D)** and T-cell marker *rag1* **(E), (F), (G), (H)** analysed by WISH at 4 dpf in wildtype embryos and candidate gene crispants. Panel **(A)** shows *cmyb* expression in the CHT region of a 4 dpf wildtype embryo, **(B)** *thy1* crispants expressing "high" levels of *cmyb* in the CHT **(C)** *thy1* crispants expressing "low" levels of *cmyb* in the CHT **(E)** *rag1* expression in the thymus of wildtype embryos **(F)** *thy1* crispants expressing "high" levels of *rag1* in the thymus **(C)** *thy1* crispants expressing "low" levels of *rag1* in the thymus.

embryos (Figure 12B). A similar phenomenon was observed in terms of *rag1* expression in the thymus. 10 out of 16 embryos showed significantly reduced *rag1* staining at 4 dpf (Figure…G), whereas the remaining 6 embryos shows no impaired *rag1* expression (Figure…F). This severe reduction at the time of HSPC emergence, but a return to normal

expression levels in some of the embryos at 4 dpf suggests the ability of those embryos to "rescue" the mechanism impaired by loss of *thy1*.

3.4.5 Loss of *ncf1* leads to impaired expression haematopoietic marker genes at 30 hpf, but not 4 dpf

The opposite is seen in *ncf1*-injected embryos. At 30hpf, a slight decrease in *runx1* was observed in

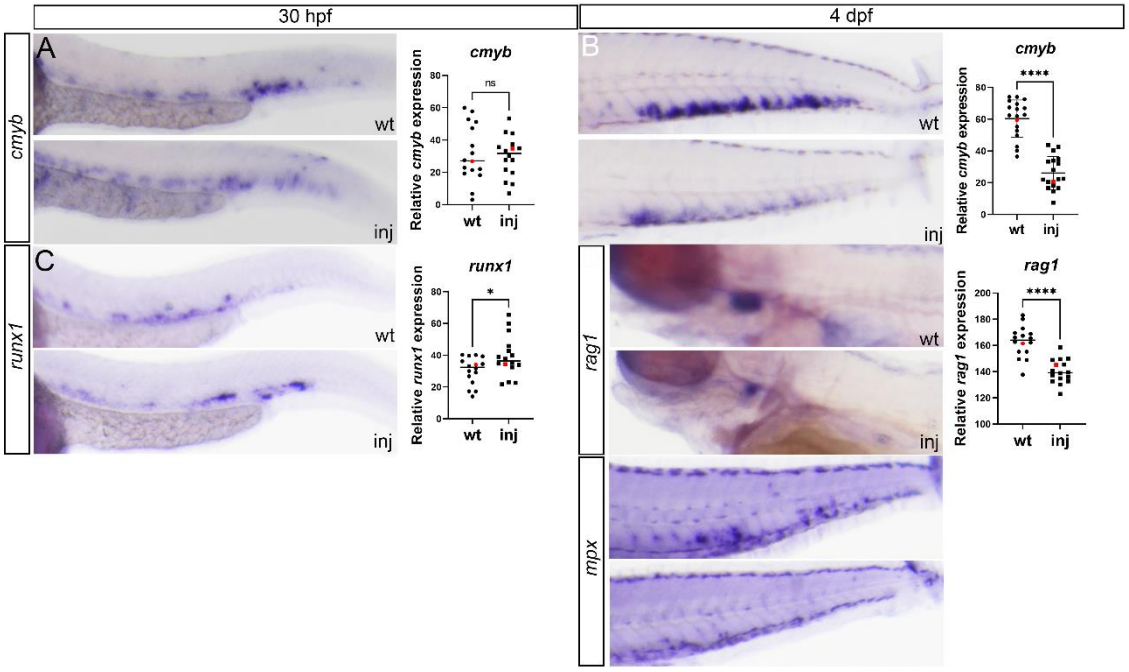


Figure 13: Loss of *ncf1* results in impairment of haematopoietic marker expression. Expression of haematopoietic marker gene expression in wildtype and *ncf1* crispant embryos at 30 hpf and 4 dpf. Expression of *cmyb* at 30 hpf **(A),** *cmyb* at 4 dpf **(B),** *runx1* at 30 hpf **(C),** *rag1* at 4 dpf **(D),** *mpx* at 4dpf **(E).**

ncf1 crispants compared to their wildtype siblings (Figure 13C), but *cmyb* expression remained

unaffected at this stage (Figure 13A). However, at 4 dpf when HSPCs are colonising the CHT and thymus, both *cmyb* and *rag1* were significantly reduced in both intermediary haematopoietic tissues respectively. This implies that loss of *ncf1* does not affect the emergence of HSPCs and EHT can initiate and complete normally, but these cells are mis-programmed at some point during or after their emergence, affecting their ability to migrate and seed the haematopoietic tissues, or their ability to proliferate. Alternatively, because *ncf1* encodes one of the neutrophilic cytosolic subunits of NADPH, we hypothesised that it might have a role in neutrophil differentiation which would offer an alternative explanation to the loss of multipotent HSPCs in the CHT, and in turn, the thymus. To test this hypothesis, 4 dpf *ncf1*-injected embryos were stained for *mpx*, a neutrophil marker, and imaged. This revealed slightly decreased expression of *mpx* in *ncf1* crispants (Figure 13E). This suggests that the decrease in haematopoietic marker gene expression is (at least not in the case of neutrophils) not a result of increased differentiation vs proliferation fate, but most likely represents a decrease in HSPC number.

4. Discussion

4.1 Loss of identified candidate genes leads to decreased HSPC emergence at 30 hpf

Firstly, it was important to validate my approach of using CRISPR-cas9 to knock out the target genes of interest. For this I used a recently published player in haematopoiesis *sart3* and successfully replicated previous findings. Upon knockdown of my genes of interest using this approach, all seven exhibited a haematopoietic phenotype of varying degrees.

Loss of *wasla*, a gene encoding a member of the Wiskott-Aldrich syndrome protein family resulted in a significant increase in HSPC emergence from the HE, a phenotype which persisted into later stages of haematopoietic development and at 4 dpf a decrease in HSPCs colonising the CHT and thymus was observed. *Ptpn13* was determined to be dispensable in for the emergence of HSPCs from the DA and their colonisation of the CHT, but without which these HSPCs are unable to produce lymphocytes in the thymus by 4 dpf as seen by the complete absence of *rag1* expression in the thymus. *Tead1a* knockout resulted in increased expression of *cmyb* in the trunk and head of the embryo, but not the CHT, potentially suggesting vasculature defects and the accumulation of primitive erythrocytes expressing *cmyb* in the embryo. *Tead1a* is also believed to play a role in leukocyte development and differentiation; this is consistent with the significant increase seen in *l-plastin* expression in *tead1a* knockouts. *Esrp1* is a known positive regulator of endothelial cell fate during EMT. Our results suggest a similar role of *esrp1* in EHT; a significant increase in HSC was observed upon *esrp1* knockout. At 4 dpf this returned to wildtype levels possibly pointing to failed EHT events at earlier stages of development, or mis-programming of emerging HSPCs. At 30 hpf *etv5b* knockout resulted in slightly increased *runx1* levels and severely decreased *cmyb* levels, which returned to normal by 4 dpf. Because *cmyb* is expressed in primitive erythrocytes this might suggest a possible role of *etv5b* in primitive, but not definitive, haematopoiesis.

In conclusion, I was able to confirm the efficiency of my approach to perform a large-scale functional screen by confirming findings from stable mutants. All seven selected genes of interest are believed to have a role in haematopoietic development. It was found that all candidate genes with a known function in EMT (*wasla, tead1a, esrp1, and etv5b),* a process closely resembling EHT which gives rise to HSPCs, showed impaired early haematopoietic cell marker expression at the initiation of the definitive wave of haematopoiesis and are therefore believed to have a role in the emergence of HSPCs. This is unsurprising as EMT leads to the loss of the epithelial identity by downregulation of genes responsible for cell adhesion.

4.2 *spi2* is functionally co-expressed with *thy1* and *ncf1*

To further analyse the gene dynamics during HSPC ontogeny, this project carried out differential gene expression analysis in samples isolated from tissues in the DA. Of particular interest was the identification of differentially expressed genes in the haemogenic endothelium (HE) compared to the aortic endothelium (AE) as potential drivers of HSPC emergence. A total of … genes were overexpressed in the HE compared to the AE, and these genes were enriched for terms involved in… Genes overexpressed in the AE were enriched for terms … The identified genes and their expression profiles were used to construct a correlation matrix to identify functionally co-expressed genes. Two such genes, *thy1* and *ncf1* were identified as co-expressed together with *spi2,* a known player in haematopoietic development. Upon knockdown *thy1* produced a strong haematopoietic phenotype as seen by the reduction of *cmyb* and *runx1* expression in the AGM region at 30 hpf. This reduction was also seen at 4 dpf in terms of *cmyb* and *rag1* expression, however this was only observed in about half of embryos imaged. *Ncf1* is believed to be redundant for the initiation of the definitive wave but is required after the emergence of HSPCs as seen by the significant reduction in *cmyb* and *rag1* expression at 4 dpf.

4.3 Limitations and future work

The aims intended for this project and its nature are broad to allow for a larger screen to be carried out and increase the likelihood of identifying novel players in haematopoietic development; these approaches, often lack depth and require follow-up experiments to explain the exact mechanisms of action that are taking place. This project, although identified seven potential novel players in EHT or expansion/colonisation of HSCs, was unable to pin-point the exact developmental stages and mechanism where the impairment caused by loss of candidate genes originates. It would be of interest to investigate developmental stages between the broad window of 30 hpf and 4 dpf. A useful tool for such studies is the use of transgenic zebrafish lines which allow for the events giving rise to HSPCs to be visualised in real time. In the lab we have a number of transgenic lines labelling specific haematopoietic and vascular lineages, including Tg(*runx1:*citrine*; kdrl:*mCherry), Tg(*cmyb:*GFP), Tg(*gata1:*dsRed), and Tg(*lck*:GFP) specific to *runx1* positive endothelial cells (useful for identifying the HE), HSPCs, erythrocytes, and thymocytes respectively. This project successfully identified genes which may play a role in the emergence of HSPCs and the colonisation of the haematopoietic organs CHT and thymus, which in turn takes into account their ability to proliferate. Originally in this project it was aimed to also use specific lineage markers to investigate the loss of candidate genes on the erythroid lineage (*alas2*), myeloid lineage (*mpx*), and leukocytes (*l-plastin*). The latter two were only possible in genes with a suspected function in myeloid or leukocyte differentiation. Additionally, some phenotypes resulted in a partial phenotype where a varying proportion of injected embryos showed an impairment in haematopoietic development. To test whether this is due to varying efficiency of crRNAs injected together with cas9, I attempted to genotype single embryos and match them to the phenotype, however this proved difficult to do after staining. In the future older embryos can be genotyped by fin clip and fixed individually before staining.

As well was broadening the scope of the initial screen, it would also be of interest to perform further experiments tailored to each candidate gene. *Ptpn13* was shown to induce developmental defects in the thymus, and is believed to dephosphorylate the Fas receptor which has a role in apoptosis. It would be of interest to investigate this further and perform apoptosis assays to determine whether this phenotype is a result of cell death upon HSPC emergence. Interestingly, a preliminary experiment using acridine orange to stain apoptotic cells suggests that there might be an increase in cell death in the CHT at this stage, but not the thymus. *Spi2* is a known regulator of HSPC emergence, and two genes were identified which are co-expressed together with *spi2* on a functional level in the data analysed in this project. However, we don't know if these genes act upstream or downstream of *spi2* expression. To uncover more about their interaction, WISH and qPCR experiments could be carried out on *spi2*, *thy1*, and *ncf1* knockouts to investigate the resulting phenotype in terms of the other co-expressed genes.

ACKNOWLEDGEMENTS

I cannot thank everyone who has supported me through this exciting journey enough. I will start by thanking my supervisor Dr Rui Monteiro for his unwavering support and always making time to help me with whatever problem I was facing. I would also like to thank Dr Roland Arnold, my co-supervisor, for his help with analysing and interpreting bioinformatics data used in this project. You were always here to answer questions, no matter how minor they might have seemed to you. Thank you also to everyone around me who supported me in a non-academic capacity; my family and boyfriend, Alexander, for listening to my stories even when they didn't understand, and all the friends I made at the University of Birmingham along the way, you made my time here so fun and I will never forget it.

References

Austin, T.W., Solar, G.P., Ziegler, F.C., Liem, L. and Matthews, W., 1997. A role for the Wnt gene family in hematopoiesis: expansion of multilineage progenitor cells. *Blood, The Journal of the American Society of Hematology*, *89*(10), pp.3624-3635.

Bell, M., Schreiner, S., Damianov, A., Reddy, R. and Bindereif, A., 2002. p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *The EMBO journal*, *21*(11), pp.2724-2735.

Beslu, N., Krosl, J., Laurin, M., Mayotte, N., Humphries, K.R. and Sauvageau, G., 2004. Molecular interactions involved in HOXB4-induced activation of HSC self-renewal. *Blood*, *104*(8), pp.2307-2314.

Bielczyk-Maczyńska, E., Serbanovic-Canic, J., Ferreira, L., Soranzo, N., Stemple, D.L., Ouwehand, W.H. and Cvejic, A., 2014. A loss of function screen of identified genome-wide association study Loci reveals new genes controlling hematopoiesis. *PLoS genetics*, *10*(7), p.e1004450.

Bock, O., Muth, M., Theophile, K., Winter, M., Hussein, K., Büsche, G., Kröger, N. and Kreipe, H., 2009. Identification of new target molecules PTK2, TGFBR2 and CD9 overexpressed during advanced bone marrow remodelling in primary myelofibrosis. *British journal of haematology*, *146*(5), pp.510-520.

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

Chen, Y., He, F., Wang, R., Yao, M., Li, Y., Guo, D. and He, S., 2021. NCF1/2/4 are prognostic biomarkers related to the immune infiltration of kidney renal clear cell carcinoma. *BioMed Research International*, *2021*.

Choi, Y.S., Noh, S.E., Lim, S.M., Lee, C.W., Kim, C.S., Im, M.W., Lee, M.H. and Kim, D.I., 2008. Multipotency and growth characteristic of periosteum-derived progenitor cells for chondrogenic, osteogenic, and adipogenic differentiation. *Biotechnology letters*, *30*, pp.593-601.

Colas, E., Muinelo-Romay, L., Alonso-Alconada, L., Llaurado, M., Monge, M., Barbazan, J., Gonzalez, M., Schoumacher, M., Pedrola, N., Ertekin, T. and Devis, L., 2012. ETV5 cooperates with LPP as a sensor of extracellular signals and promotes EMT in endometrial carcinomas. *Oncogene*, *31*(45), pp.4778-4788.

Costa, G., Kouskoff, V. and Lacaud, G., 2012. Origin of blood cells and HSC production in the embryo. *Trends in immunology*, *33*(5), pp.215-223.

Craig, W., Kay, R., Cutler, R.L. and Lansdorp, P.M., 1993. Expression of Thy-1 on human hematopoietic progenitor cells. *The Journal of experimental medicine*, *177*(5), pp.1331-1342.

de Bruin, A.M., Demirel, Ö., Hooibrink, B., Brandts, C.H. and Nolte, M.A., 2013. Interferon-γ impairs proliferation of hematopoietic stem cells in mice. *Blood, The Journal of the American Society of Hematology*, *121*(18), pp.3578-3585.

Delaney, C., Varnum-Finney, B., Aoyama, K., Brashem-Stein, C. and Bernstein, I.D., 2005. Dose-dependent effects of the Notch ligand Delta1 on ex vivo differentiation and in vivo marrow repopulating ability of cord blood cells. *Blood*, *106*(8), pp.2693-2699.

Deloria, A.J., Höflmayer, D., Kienzl, P., Łopatecka, J., Sampl, S., Klimpfinger, M., Braunschmid, T., Bastian, F., Lu, L., Marian, B. and Stättner, S., 2016. Epithelial splicing regulatory protein 1 and 2 paralogues correlate with splice signatures and favorable outcome in human colorectal cancer. *Oncotarget*, *7*(45), p.73800.

Ding, Y., Wang, W., Ma, D., Liang, G., Kang, Z., Xue, Y., Zhang, Y., Wang, L., Heng, J., Zhang, Y. and Liu, F., 2021. Smarca5-mediated epigenetic programming facilitates fetal HSPC development in vertebrates. *Blood*, *137*(2), pp.190-202.

Duncan, A.W., Rattis, F.M., DiMascio, L.N., Congdon, K.L., Pazianos, G., Zhao, C., Yoon, K., Cook, J.M., Willert, K., Gaiano, N. and Reya, T., 2005. Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. *Nature immunology*, *6*(3), pp.314-322.

Fathman, C.G., Small, M., Herzenberg, L.A. and Weissman, I.L., 1975. Thymus cell maturation: II. Differentiation of three “mature” subclasses in Vivo. *Cellular immunology*, *15*(1), pp.109-128.

Gama-Norton, L., Ferrando, E., Ruiz-Herguido, C., Liu, Z., Guiu, J., Islam, A.B., Lee, S.U., Yan, M., Guidos, C.J., López-Bigas, N. and Maeda, T., 2015. Notch signal strength controls cell fate in the haemogenic endothelium. *Nature communications*, *6*(1), p.8510.

Herrera-Molina, R., Frischknecht, R., Maldonado, H., Seidenbecher, C.I., Gundelfinger, E.D., Hetz, C., Aylwin, M.D.L.L., Schneider, P., Quest, A.F. and Leyton, L., 2012. Astrocytic αVβ3 integrin inhibits neurite outgrowth and promotes retraction of neuronal processes by clustering Thy-1. *PLoS One*, *7*(3), p.e34295.

Herrera-Molina, R., Valdivia, A., Kong, M., Alvarez, A., Cárdenas, A., Quest, A.F. and Leyton, L., 2013. Thy-1-interacting molecules and cellular signaling in cis and trans. *International review of cell and molecular biology*, *305*, pp.163-216.

Hidalgo-Sastre, A., Desztics, J., Dantes, Z., Schulte, K., Ensarioglu, H.K., Bassey-Archibong, B., Öllinger, R., Engleiter, T., Rayner, L., Einwächter, H. and Daniel, J.M., 2020. Loss of Wasl improves pancreatic cancer outcome. *JCI insight*, *5*(10).

Hoshijima, K., Jurynec, M.J., Shaw, D.K., Jacobi, A.M., Behlke, M.A. and Grunwald, D.J., 2019. Highly efficient CRISPR-Cas9-based methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish. *Developmental cell*, *51*(5), pp.645-657.

Hou, J., Chen, C., Hu, Y., Gong, Q., Gan, L. and Xu, Y., 2021. Identify Function of WASL in Prognosis of Cervical Cancer Based on Omics Data. *Frontiers in Cell and Developmental Biology*, *9*, p.670890.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L. and McLaren, S., 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, *496*(7446), pp.498-503.

Huang, W., Bei, L. and Eklund, E.A., 2013. Fas-associated phosphatase 1 mediates Fas resistance in myeloid progenitor cells expressing the Bcr–abl oncogene. *Leukemia & lymphoma*, *54*(3), pp.619-630.

Huang, W., Luan, C.H., Hjort, E.E., Bei, L., Mishra, R., Sakamoto, K.M., Platanias, L.C. and Eklund, E.A., 2016. The role of Fas-associated phosphatase 1 in leukemia stem cell persistence during tyrosine kinase inhibitor treatment of chronic myeloid leukemia. *Leukemia*, *30*(7), pp.1502-1509.

Hwang, J.J., Chambon, P. and Davidson, I., 1993. Characterization of the transcription activation function and the DNA binding domain of transcriptional enhancer factor-1. *The EMBO journal*, *12*(6), pp.2337-2348.

Ishii, K., Kanatsu-Shinohara, M., Toyokuni, S. and Shinohara, T., 2012. FGF2 mediates mouse spermatogonial stem cell self-renewal via upregulation of Etv5 and Bcl6b through MAP2K1 activation. *Development*, *139*(10), pp.1734-1743.

Jung, Y., Wang, J., Song, J., Shiozawa, Y., Wang, J., Havens, A., Wang, Z., Sun, Y.X., Emerson, S.G., Krebsbach, P.H. and Taichman, R.S., 2007. Annexin II expressed by osteoblasts and endothelial cells regulates stem cell adhesion, homing, and engraftment following transplantation. *Blood, The Journal of the American Society of Hematology*, *110*(1), pp.82-90.

Koumas, L., Smith, T.J., Feldon, S., Blumberg, N. and Phipps, R.P., 2003. Thy-1 expression in human fibroblast subsets defines myofibroblastic or lipofibroblastic phenotypes. *The American journal of pathology*, *163*(4), pp.1291-1300.

Kumano, K., Chiba, S., Kunisato, A., Sata, M., Saito, T., Nakagami-Yamaguchi, E., Yamaguchi, T., Masuda, S., Shimizu, K., Takahashi, T. and Ogawa, S., 2003. Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells. *Immunity*, *18*(5), pp.699-711.

Kurz, A.R., Catz, S.D. and Sperandio, M., 2018. Noncanonical Hippo signalling in the regulation of leukocyte function. *Trends in immunology*, *39*(8), pp.656-669.

Li, H., Wang, H. and Ren, Z., 2018. MicroRNA-214-5p inhibits the invasion and migration of hepatocellular carcinoma cells by targeting Wiskott-Aldrich syndrome like. *Cellular Physiology and Biochemistry*, *46*(2), pp.757-764.

Li, Z., Zhao, B., Wang, P., Chen, F., Dong, Z., Yang, H., Guan, K.L. and Xu, Y., 2010. Structural insights into the YAP and TEAD complex. *Genes & development*, *24*(3), pp.235-240.

Libregts, S.F., Gutiérrez, L., de Bruin, A.M., Wensveen, F.M., Papadopoulos, P., van Ijcken, W., Özgür, Z., Philipsen, S. and Nolte, M.A., 2011. Chronic IFN-γ production in mice induces anemia by reducing erythrocyte life span and inhibiting erythropoiesis through an IRF-1/PU. 1 axis. *Blood, The Journal of the American Society of Hematology*, *118*(9), pp.2578-2588.

Liu, Y., Liu, J., Wang, Z. and He, J.J., 2015. Tip110 binding to U6 small nuclear RNA and its participation in pre-mRNA splicing. *Cell & Bioscience*, *5*, pp.1-12.

Long, L., Thelen, J.P., Furgason, M., Haj-Yahya, M., Brik, A., Cheng, D., Peng, J. and Yao, T., 2014. The U4/U6 recycling factor SART3 has histone chaperone activity and associates with USP15 to regulate H2B deubiquitination. *Journal of Biological Chemistry*, *289*(13), pp.8916-8930.

Lu, Z.X., Huang, Q., Park, J.W., Shen, S., Lin, L., Tokheim, C.J., Henry, M.D. and Xing, Y., 2015. Transcriptome-wide landscape of pre-mRNA alternative splicing associated with metastatic colonization. *Molecular Cancer Research*, *13*(2), pp.305-318.

Matatall, K.A., Jeong, M., Chen, S., Sun, D., Chen, F., Mo, Q., Kimmel, M. and King, K.Y., 2016. Chronic infection depletes hematopoietic stem cells through stress-induced terminal differentiation. *Cell reports*, *17*(10), pp.2584-2595.

Murdoch, B., Chadwick, K., Martin, M., Shojaei, F., Shah, K.V., Gallacher, L., Moon, R.T. and Bhatia, M., 2003. Wnt-5A augments repopulating capacity and primitive hematopoietic

development of human blood stem cells in vivo. *Proceedings of the National Academy of Sciences*, *100*(6), pp.3422-3427.
Panciera, T., Azzolin, L., Fujimura, A., Di Biagio, D., Frasson, C., Bresolin, S., Soligo, S., Basso, G., Bicciato, S., Rosato, A. and Cordenonsi, M., 2016. Induction of expandable tissue-specific stem/progenitor cells through transient expression of YAP/TAZ. *Cell stem cell*, *19*(6), pp.725-737.

Nagasawa, T., Hirota, S., Tachibana, K., Takakura, N., Nishikawa, S.I., Kitamura, Y., Yoshida, N., Kikutani, H. and Kishimoto, T., 1996. Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature*, *382*, pp.635-638.

Puri, M.C. and Bernstein, A., 2003. Requirement for the TIE family of receptor tyrosine kinases in adult but not fetal hematopoiesis. *Proceedings of the National Academy of Sciences*, *100*(22), pp.12753-12758.

Robert-Moreno, À., Guiu, J., Ruiz-Herguido, C., López, M.E., Inglés-Esteve, J., Riera, L., Tipping, A., Enver, T., Dzierzak, E., Gridley, T. and Espinosa, L., 2008. Impaired embryonic haematopoiesis yet normal arterial development in the absence of the Notch ligand Jagged1. *The EMBO journal*, *27*(13), pp.1886-1895.

Sardina, J.L., López-Ruano, G., Prieto-Bermejo, R., Sánchez-Sánchez, B., Pérez-Fernández, A., Sánchez-Abarca, L.I., Pérez-Simón, J.A., Quintales, L., Sánchez-Yagüe, J., Llanillo, M. and Antequera, F., 2014. PTPN13 regulates cellular signalling and β-catenin function during megakaryocytic differentiation. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1843*(12), pp.2886-2899.

Schrader, J.W., Battye, F. and Scollay, R., 1982. Expression of Thy-1 antigen is not limited to T cells in cultures of mouse hemopoietic cells. *Proceedings of the National Academy of Sciences*, *79*(13), pp.4161-4165.

Schwickert, A., Weghake, E., Brüggemann, K., Engbers, A., Brinkmann, B.F., Kemper, B., Seggewiß, J., Stock, C., Ebnet, K., Kiesel, L. and Riethmüller, C., 2015. microRNA miR-142-3p inhibits breast cancer cell invasiveness by synchronous targeting of WASL, integrin alpha V, and additional cytoskeletal elements. *PloS one*, *10*(12), p.e0143993.

Tober, J., McGrath, K.E. and Palis, J., 2008. Primitive erythropoiesis and megakaryopoiesis in the yolk sac are independent of c-myb. *Blood, The Journal of the American Society of Hematology*, *111*(5), pp.2636-2639.
Uckelmann, H., Blaszkiewicz, S., Nicolae, C., Haas, S., Schnell, A., Wurzer, S., Wagener, R., Aszodi, A. and Essers, M.A.G., 2016. Extracellular matrix protein Matrilin-4 regulates stress-induced HSC proliferation via CXCR4. *Journal of experimental medicine*, *213*(10), pp.1961-1971.
Ueda, J., Matsuda, Y., Yamahatsu, K., Uchida, E., Naito, Z., Korc, M. and Ishiwata, T., 2014. Epithelial splicing regulatory protein 1 is a favorable prognostic factor in pancreatic cancer that attenuates pancreatic metastases. *Oncogene*, *33*(36), pp.4485-4495.

Uenishi, G.I., Jung, H.S., Kumar, A., Park, M.A., Hadland, B.K., McLeod, E., Raymond, M., Moskvin, O., Zimmerman, C.E., Theisen, D.J. and Swanson, S., 2018. NOTCH signaling specifies arterial-type definitive hemogenic endothelium from human pluripotent stem cells. *Nature communications*, *9*(1), p.1828.

Vassilev, A., Kaneko, K.J., Shu, H., Zhao, Y. and DePamphilis, M.L., 2001. TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. *Genes & development*, *15*(10), pp.1229-1241.

Wagner, G.P., Kin, K. and Lynch, V.J., 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in biosciences*, *131*, pp.281-285.

Wandel, E., Saalbach, A., Sittig, D., Gebhardt, C. and Aust, G., 2012. Thy-1 (CD90) is an interacting partner for CD97 on activated endothelial cells. *The Journal of Immunology*, *188*(3), pp.1442-1450.

Wang, Q., Xu, Z., An, Q., Jiang, D., Wang, L., Liang, B. and Li, Z., 2015. TAZ promotes epithelial to mesenchymal transition via the upregulation of connective tissue growth factor expression in neuroblastoma cells. *Molecular Medicine Reports*, *11*(2), pp.982-988.

Warzecha, C.C., Sato, T.K., Nabet, B., Hogenesch, J.B. and Carstens, R.P., 2009. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Molecular cell*, *33*(5), pp.591-601.

Wu, X., Goodyear, S.M., Tobias, J.W., Avarbock, M.R. and Brinster, R.L., 2011. Spermatogonial stem cell self-renewal requires ETV5-mediated downstream activation of Brachyury in mice. *Biology of reproduction*, *85*(6), pp.1114-1123.

Yanagisawa, J., Takahashi, M., Kanki, H., Yano-Yanagisawa, H., Tazunoki, T., Sawa, E., Nishitoba, T., Kamishohara, M., Kobayashi, E., Kataoka, S. and Sato, T., 1997. The molecular interaction of Fas and FAP-1: a tripeptide blocker of human Fas interaction with FAP-1 promotes Fas-induced apoptosis. *Journal of Biological Chemistry*, *272*(13), pp.8539-8545.

Zhang, J., Dong, B. and Siminovitch, K.A., 2009. Contributions of Wiskott–Aldrich syndrome family cytoskeletal regulatory adapters to immune regulation. *Immunological reviews*, *232*(1), pp.175-194.

Zhang, Y., Jin, H., Li, L., Qin, F.X.F. and Wen, Z., 2011. cMyb regulates hematopoietic stem/progenitor cell mobilization during zebrafish hematopoiesis. *Blood, The Journal of the American Society of Hematology*, *118*(15), pp.4093-4101.
Zhao, S., Zhang, A., Zhu, H. and Wen, Z., 2022. The ETS transcription factor Spi2 regulates hematopoietic cell development in zebrafish. *Development*, *149*(19), p.dev200881.

Zhao, S., Zhang, A., Zhu, H. and Wen, Z., 2022. The ETS transcription factor Spi2 regulates hematopoietic cell development in zebrafish. *Development*, *149*(19), p.dev200881.

Zhao, Y., Wu, M., Li, J., Meng, P., Chen, J., Huang, Z., Xu, J., Wen, Z., Zhang, W. and Zhang, Y., 2021. The spliceosome factor sart3 regulates hematopoietic stem/progenitor cell development in zebrafish through the p53 pathway. *Cell death & disease*, *12*(10), p.906.

Kalev-Zylinska, M.L., Horsfield, J.A., Flores, M.V.C., Postlethwait, J.H., Vitas, M.R., Baas, A.M., Crosier, P.S. and Crosier, K.E., 2002. Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis.

De Pater, E., Kaimakis, P., Vink, C.S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P.S., Camper, S.A., Speck, N. and Dzierzak, E., 2013. Gata2 is required for HSC generation and survival. *Journal of Experimental Medicine*, *210*(13), pp.2843-2850.

Zhang, Y., Jin, H., Li, L., Qin, F.X.F. and Wen, Z., 2011. cMyb regulates hematopoietic stem/progenitor cell mobilization during zebrafish hematopoiesis. *Blood, The Journal of the American Society of Hematology*, *118*(15), pp.4093-4101.

Sandberg, M.L., Sutton, S.E., Pletcher, M.T., Wiltshire, T., Tarantino, L.M., Hogenesch, J.B. and Cooke, M.P., 2005. c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Developmental cell*, *8*(2), pp.153-166.

Cheshier, S.H., Morrison, S.J., Liao, X. and Weissman, I.L., 1999. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proceedings of the National Academy of Sciences*, *96*(6), pp.3120-3125.

Nteliopoulos, G., Marley, S.B. and Gordon, M.Y., 2009. Influence of PI-3K/Akt pathway on Wnt signalling in regulating myeloid progenitor cell proliferation. Evidence for a role of autocrine/paracrine Wnt regulation. *British journal of haematology*, *146*(6), pp.637-651.

Ciau-Uitz, A., Monteiro, R., Kirmizitas, A. and Patient, R., 2014. Developmental hematopoiesis: ontogeny, genetic programming and conservation. *Experimental hematology*, *42*(8), pp.669-683.

Burgess‡, C.W., Emelyanov, A., Parinov, S., Sampath, K., Renn, J., Jun, Y., Brocher, J., Willems, B., Sheng, Y., Rajaei, F. and Roy, R.V., 2013. A Multifunctional Mutagenesis System for Analysis of Gene Function in Zebrafish.

Dobrzycki, T., Krecsmarik, M., Bonkhofer, F., Patient, R. and Monteiro, R., 2018. An optimised pipeline for parallel image-based quantification of gene expression and genotyping after in situ hybridisation. *Biology open*, *7*(4), p.bio031096.

Bartůněk, P., Kralova, J., Blendinger, G., Dvořák, M. and Zenke, M., 2003. GATA-1 and c-myb crosstalk during red blood cell differentiation through GATA-1 binding sites in the c-myb promoter. *Oncogene*, *22*(13), pp.1927-1935.

Kissa, K., Murayama, E., Zapata, A., Cortés, A., Perret, E., Machu, C. and Herbomel, P., 2008. Live imaging of emerging hematopoietic stem cells and early thymus colonization. *Blood, The Journal of the American Society of Hematology*, *111*(3), pp.1147-1156

Chen, A.T. and Zon, L.I., 2009. Zebrafish blood stem cells. *Journal of cellular biochemistry*, *108*(1), pp.35-42..

Shimizu, C., Kawamoto, H., Yamashita, M., Kimura, M., Kondou, E., Kaneko, Y., Okada, S., Tokuhisa, T., Yokoyama, M., Taniguchi, M. and Katsura, Y., 2001. Progression of T cell lineage restriction in the earliest subpopulation of murine adult thymus visualized by the expression of lck proximal promoter activity. *International immunology*, *13*(1), pp.105-117.

Kroll, F., Powell, G.T., Ghosh, M., Gestri, G., Antinucci, P., Hearn, T.J., Tunbak, H., Lim, S., Dennis, H.W., Fernandez, J.M. and Whitmore, D., 2021. A simple and effective F0 knockout method for rapid screening of behaviour and other complex phenotypes. *elife*, *10*, p.e59683.

Kumar, R. and Evans, T., 2015. Haematopoiesis. *eLS*.

Ciau-Uitz, A., Monteiro, R., Kirmizitas, A. and Patient, R., 2014. Developmental hematopoiesis: ontogeny, genetic programming and conservation. *Experimental hematology*, *42*(8), pp.669-683.

Lacaud, G., Thambyrajah, R., Mazan, M., Patel, R., Moignard, V., Stefanska, M., Marinopoulou, E., Li, Y., Lancrin, C., Clapes, T. and Möröy, T., 2015. New insights into the endothelial-to-haematopoietic transition leading to HSC emergence. *Experimental Hematology*, *43*(9), p.S41.

Asada, N., Takeishi, S. and Frenette, P.S., 2017. Complexity of bone marrow hematopoietic stem cell niche. *International journal of hematology*, *106*, pp.45-54.

Jenkinson, E.J., Jenkinson, W.E., Rossi, S.W. and Anderson, G., 2006. The thymus and T-cell commitment: the right niche for Notch?. *Nature Reviews Immunology*, *6*(7), pp.551-555.

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

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

Gore, A.V., Athans, B., Iben, J.R., Johnson, K., Russanova, V., Castranova, D., Pham, V.N., Butler, M.G., Williams-Simons, L., Nichols, J.T. and Bresciani, E., 2016. Epigenetic regulation of hematopoiesis by DNA methylation. *elife*, *5*, p.e11813.

Sareila, O., Hagert, C., Kelkka, T., Linja, M., Xu, B., Kihlberg, J. and Holmdahl, R., 2017. Reactive oxygen species regulate both priming and established arthritis, but with different mechanisms. *Antioxidants & Redox Signaling*, *27*(18), pp.1473-1490.

Zhao, J., Ma, J., Deng, Y., Kelly, J.A., Kim, K., Bang, S.Y., Lee, H.S., Li, Q.Z., Wakeland, E.K., Qiu, R. and Liu, M., 2017. A missense variant in NCF1 is associated with susceptibility to multiple autoimmune diseases. *Nature genetics*, *49*(3), pp.433-437.