USING GENOME-WIDE DATA TO MODEL SIGNALLING-RESPONSIVE GENE REGULATORY MECHANISMS IN BLOOD DEVELOPMENT

by

ALEXANDER IAIN MAYTUM

A Thesis Submitted to the University of Birmingham for the Degree of

DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences

College of Medical and Dental Sciences

University of Birmingham

December 2022

Abstract

The control of gene expression driving developmental haematopoiesis crucially depends on distal *cis*-regulatory elements such as enhancers which directly interact with promoters in the nucleus. However, no global experiments have been conducted which identify the cell type and cell stage-specific activity of enhancers in a chromatin context. It is through these elements that lineage specific transcription factors orchestrate cell fate decisions and direct haematopoietic lineage development emerging from the mesoderm. The roles of transcriptional regulators are beginning to be understood, however, it is still unclear how the myriad of extracellular signals modulate their activity. In this work, we report a global method which enables the identification of thousands tissue-specifically active *cis*-regulatory elements able to stimulate a minimal promoter in cells representing five stages of haematopoietic specification derived from embryonic stem cells. Using serum-free differentiation culture, we demonstrate that our method can identify signalling-responsive enhancer elements and we highlight that it can be adapted to any embryonic stem cell differentiation system generating different cell types. We demonstrate that thousands of cell stage-specific sets of *cis*-elements are responsive to cytokine signals terminating at signalling-responsive transcription factors. Integrating these data with chromatin accessibility and single cell RNA-Seq data provided important new insights into the regulatory dynamics of the gene regulatory network transitions driving haematopoiesis. Our work identified the cytokine signalling-responsive transcription factors mediating responsiveness of enhancers at each developmental stage. We validated enhancers for *Sparc*, *Pxn*, *Hspg2*, *Cdh5, Dlk1* and *Mrpl15* as being signalling responsive to VEGF. We found

that the cytokine VEGF is a crucial factor that regulates the balance between endothelial and haematopoietic development and our scRNA-seq analysis revealed that in the presence of VEGF *Sox17* fails to be downregulated and *Runx1* fails to be upregulated in the haemogenic endothelium and progenitor cells. For two *Runx1* enhancers (the +23kb and +3.7kb) we studied the transcription factors motifs mediating the responsiveness of the enhancers to VEGF by mutation of these sites. Taken together, our work generated an important novel resource for future studies of haematopoietic differentiation and provides insights into how and where in the genome extrinsic signals program the cell type-specific chromatin landscape driving this process.

Acknowledgements

Firstly, I would like to thank my supervisor Professor Constanze Bonifer for her time, guidance and mentorship and importantly for giving me the opportunity to prove myself and develop as a scientist.

I thank everyone in the Bonifer-Cockerill Group for their support and guidance during the project. I especially thank Dr Ben Edginton-White for his support in the lab and with bioinformatics analyses. I also thank Dr Peter Keane for his support on bioinformatics methods and analyses.

Many people contributed to the success of this project. Dr Debbie Goode developed HM-1 cells harbouring *Galnt1* enhancer reporters. Dr Jane Gilmour and Dr Ben Edginton-White helped to set up the serum free I.V.D system used in this project. Dr Peter Keane and Dr Salam Assi performed bioinformatics analysis. I thank everyone for their time and hard work.

I thank the BBSRC and the Midlands Integrative Biosciences Training Partnership for funding my PhD.

Finally, I would like to thank my parents, my family and my friends for their continued support during my PhD and scientific training.

Contents

List of Figures

List of Tables

Abbreviations

5caC – 5-carboxylcytosine

5fC – 5-formylcytosine

5hmC – 5-hydroxymethylcytosine

5mC – 5-methylcytosine

AGM – aorta-gonad-mesonephros

BAH – bromo-adjacent homology

bHLH – basic helix-loop-helix

BMPs – bone morphogenic proteins

bp – base pair

Bry – Brachyury

bZIP – basic leucine zipper

CAP-SELECT – consecutive affinity-purification systemic evolution of ligands by exponential enrichment

CDH5- cadherin 5

CDS – coding sequence

CFI – cleavage factor I

CHID – chromodomain helicase DNA binding

CPSF – cleavage and polyadenylation factor

CRISPR-Cas9 – clustered regulatory interspaced short palindromic repeats and CRISPR-associated protein 9

CSTF – cleavage stimulation factor

CTCF – CCCTC-binding factor

CTD – carboxy-terminal domain

DBD – DNA binding domain

DHSs – DNase hypersensitive site

DIA – differentiation inhibitory activity

DNA – deoxyribonucleic acid

DNMT- DNA methyltransferase

DPE – downstream promoter element

EBs – embryoid bodies

EHT – endothelial-to-haematopoietic transition

eRNA – enhancer RNA

ES – embryonic stem

FACS – fluorescence-activated cell sorting

FCS – foetal calf serum

Fgfr1 – fibroblast growth factor receptor-1

FGFs – fibroblast growth factors

gRNAs – guide RNAs

GRNs – gene regulatory networks

GRO-seq – global run-on sequencing

GTFs – general transcription factors

H3k27 – histone H3 lysine 27

H3K3me3 – H3 lysine 4 trimethylation

HATs – histone acetyltransferases

HB – haemangioblast

HDACs – histone deacetylases

HE – haemogenic endothelium

HE1 – haemogenic endothelium type-1

HE2 – haemogenic endothelium type-2

HECs – haemogenic endothelial cells

HMG – high mobility group

HMTs – histone methyltransferases

HOX – homeotic

HP – haematopoietic progenitor

Hprt -hypoxanthine guanine phosphoribosyl transferase 1

HSC – haematopoietic stem cell

IBPs – insulator binding proteins

INR – initiator element

ISWI – imitation switch

IVD – *in vitro* differentiation

JmjC – Jumonji

LATS1/2 – large tumour suppressor ½

LIF – leukaemia inhibitory factor

MAPKs – mitogen-activated protein kinases

MEFS- mouse embryonic fibroblast

MLL – mixed lineage leukaemia

MPRAs – massively parallel reporter assays

MST1/2 – Ste20-like kinases 1/2

NHR – nuclear hormone receptor

NRLs – nucleosome-repeat lengths

OC – open complex

P-TEFb – positive transcription elongation factor b

P300/CBP – CREB binding protein

PAP – poly(A) polymerase

PAS – poly(A) signal

PcGs – polycomb group proteins

PCR1 – polycomb repressive complex 1

PIC – pre-initiation complex

PRC2 – polycomb repressive complex 2

PREs – specific polycomb response elements

PRO-seq – precision run-on sequencing

PTMs – posttranslational modifications

RTK – receptor tyrosine kinase

SCF – stem cell factor

Seq – sequencing

Ser2P – Ser2-phosphorylated

SF – serum free

Shh – sonic hedgehog

SMC – structural maintenance of chromosomes

SWI/SNF – switch/sucrose non-fermentable

TAD – transactivation domain

TADs – topologically associating domains

TAF250 – TAT-binding protein associated factor

TAFs – TBP-associated factors

TAZ – transcriptional co-activator with PDZ-binding motif

TET – Ten-Eleven Translocation

TF – transcription factor

TFIID – transcription factor IID

TFs – transcription factors

TGFβ - transforming growth factor-β

TRD – target recognition domain

TREs – Trithorax response elements

TrxG – Trithorax

TSS – transcription start site

UTR – untranslated region

UTRs – untranslated regions

VEGF – vascular endothelial growth factor

Wnt – wingless-type

YAP – YES-Associated protein

ZF – zinc finger

1. INTRODUCTION

1.1 Chromatin and nuclear organisation

1.1.1 Arrangement of DNA into chromatin

Genomic DNA (Watson and Crick, 1953, Franklin and Gosling, 1953) in eukaryotic cells in the nucleus is packaged into the nucleus with proteins which in its entirety forms chromatin. The basic unit of chromatin is the nucleosome, consisting of ~146 base pairs (bp) of DNA around a symmetrical histone octamer core. This histone octamer core comprises of two molecules of each of the histones H2A, H2B, H3 and H4 (Kornberg and Lorch, 1999, Noll and Kornberg, 1977, Bakayev et al., 1977, Richmond et al., 1984, Luger et al., 1997) (Figure 1.1).

The structure of the histone involves a H3/4 tetramer making up the inner core being initially loaded onto a central 60 bp sequence of DNA. Next, two H2A and H2B dimers are loaded above and below the H3 and H4 tetramer onto the flanking DNA regions. The majority of the genome is arranged as regularly spaced nucleosomes with a DNA repeat length of ~180-200 bp. Histone H1 and/or high mobility group (HMG) proteins are also recruited to nucleosomes and bind to the outside of the nucleosome to form the chromatosome across ~166 bp of DNA (Noll and Kornberg, 1977, Zlatanova et al., 1999, Albright et al., 1980, Simpson, 1978). Typically, nucleosomes assemble into higher order structures with core histone tails and linker histone H1 maintaining higher order chromatin condensation (Allan et al., 1982, Thoma et al., 1979, Routh et al., 2008). If histone H1 is not

present, nucleosomes arrange into a higher order fibre which is 30nm in diameter under certain physiological levels of monovalent or divalent cations (Thoma et al., 1979, Routh et al., 2008, Marsden and Laemmli, 1979). However, whether this structure exists in the interphase nucleus of a cell is still being debated. High resolution cryogenic electron microscopy (cryo-EM) studies have provided evidence for a two start zigzag helical folding of the nucleosomal filament consisting of a tetranucleosome repeat unit (Song et al., 2014). Work by Collepardo-Guevara and Schlick et al (Collepardo-Guevara and Schlick, 2014) indicates that intra-fibre nucleosome-repeat lengths (NRLs) significantly impact chromatin structure and have a wide-ranging effect on the different chromatin architectures (e.g. highly bent narrow forms, canonical and irregular zigzag fibres and polymorphic conformations). These data support previous biochemical findings of the asymmetrical bindings of linker histones to nucleosomes and the dimeric display of interactions on linker histones (Carter and van Holde, 1998). These studies challenge the concept of nucleosome acidic patch involvement in the packing of the nucleosome within the 30nm fibre (Luger et al., 1997).

Short DNA segments or 'linker DNA' connect nucleosomes within nucleosomal arrays which interact with neighbouring nucleosomes to form chromatin fibres leading to a high level of compaction in the condensed chromosome. Nucleosomes can be visualised as unfolded chains of regulatory spaced 10-nm diameter particles, and is often described as a 'beads-on-a-string' organisation and can be termed the 'primary structure' of chromatin. The primary structure of chromatin can influence the subsequent secondary and tertiary higher-order chromatin structures (Woodcock and Dimitrov, 2001).

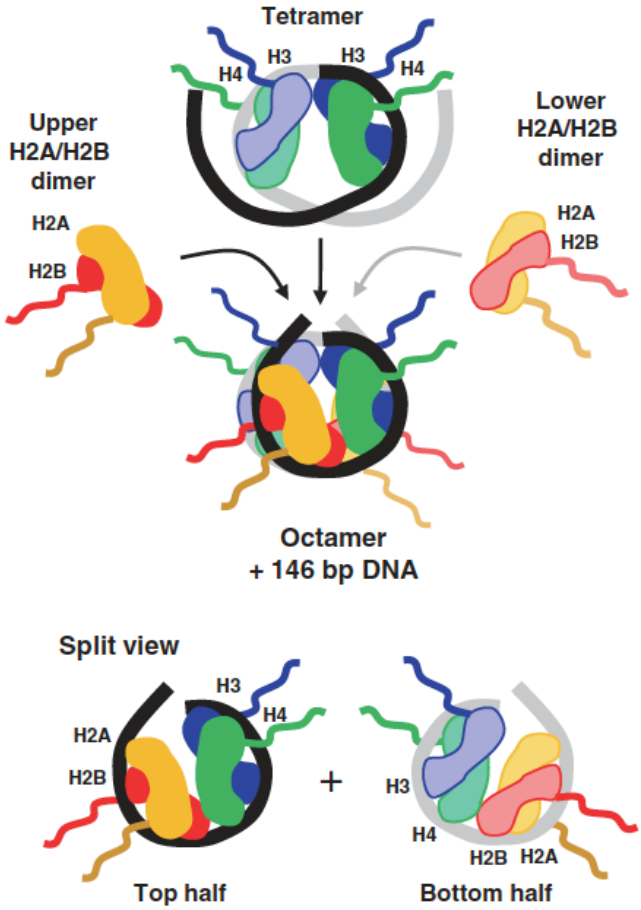


Figure 1.1- Nucleosome structure
Model of assembly of the histone octamer on DNA, showing two H3/H4 dimers with an inner core of ~60 bp of DNA. Next two H2A and H2B dimers are loaded onto the flanking DNA segments above and below the H3/H4 tetramer. Image from (Cockerill, 2011). (Permission to reproduce this figure has been granted by John Wiley and Sons.)

1.1.2 Histone modifications

DNA tightly bound to histones prevents transcription factors (TFs) from binding to the regulatory elements of genes such as enhancers and promoters and thereby prevent their transcriptional activity. The nucleosomes' ability to prevent TFs from accessing key regulatory elements represents an important regulatory mechanism for gene transcription (Lerner et al., 2020, Brahma and Henikoff, 2020).

In 1960 studies provided evidence for posttranslational modifications on histones (Allfrey et al., 1964). It is now known that a large number of posttranslational modifications (PTMs) are found on histones. Differential PTMs of nucleosomes allows for a great number of functional states and a more precise and dynamic regulation of chromatin accessibility through a number of mechanisms (Biswas and Rao, 2018). Histone modifications can result in a more accessible and dynamic chromatin environment which is optimal for transcription. The N-terminal tails of histone proteins are rich in lysine and arginine residues and are targets for various PTMs, which combined with direct modifications on DNA make up the chromatin landscape which is dynamically regulated to allow or repress transcriptional activity (Erler et al., 2014). A large number of histone modifications on histone tails affect transcriptional activation (Bannister and Kouzarides, 2011), most notably acetylation, methylation, phosphorylation, ubiquitination and SUMOylation. These histone modifications are regulated by proteins with enzymatic activities termed "readers", "writers" and "erasers".

Writers such as histone acetyltransferases (HATs), histone methyltransferases and arginine methyltransferases are enzymes that add PTMs to histones. Erasers such as histone deacetylases (HDACs) and histone demethylases are enzymes that remove PTMs from histones (Biswas and Rao, 2018). Readers are factors that recognise PTMs on histones and through binding can act as both transcriptional co-activators or co-repressors. Readers recognise histone PTMs through domains that have affinity for histone methylation (plant homeodomain (PhD), chromatin organisation modifier (chromo), Tudor, malignant Brain Tumour (MBT) or acetylation (Bromo)) (Yun et al., 2011) to regulate transcription. For example, the bromodomain family of proteins (Brd2, Brd3, Brd4 and Brdt) use their bromodomain to bind modified (acetylated) histones and regulate transcription by interacting with the transcriptional machinery (Shi and Vakoc, 2014).

Histone acetylation/deacetylation is one of the best understood forms of PTM modification on histones. Histone acetylation occurs on lysines in a dynamic fashion and is regulated by two families of enzymes, HATs and HDACs. HATs catalyse the transfer of an acetyl group to the ε-amino group of histone lysine side chains from the cofactor acetyl CoA which results in neutralisation of the lysine positive charge and a weakening of the interactions between histones and DNA. HATs are divided into two major classes; type-A and type-B. Type B HATs are localised to the cytoplasm and act on histones not associated with chromatin, they all share sequence homology with scHa1 the first identified type-B HAT. Type-A HATs have greater family diversity and can be broadly classified into three families; GNAT, MYST and CBP/p300 (Hodawadekar and Marmorstein, 2007). TFIID (transcription factor IID), subunit TAF250 (TAT-binding protein associated factor) and P300/CBP (CREB binding protein)-

associated factor are all examples of HATs (Bannister and Kouzarides, 1996, Mizzen et al., 1996, Parthun, 2007). These enzymes interact with cofactor Acetyl-CoA and act to modify sites within histone N-terminal tails to neutralise positive charges and disrupt electrostatic interactions with negatively charged phosphodiester backbone of DNA allowing for the opening up of the tightly packed nucleosome (Bannister and Kouzarides, 2011).

HDACs act to oppose the effects of HATs by removing ε-amino acetyl groups from lysine residues on histones (Gregoretti et al., 2004) resulting in the restoration of the positive charge of the histone N-terminal lysine and the ability for a stronger interaction with negatively charged DNA. Due to this action HDACs are generally considered transcriptional co-repressors. Currently there are four classes of HDAC (Yang and Seto, 2007); classes I-IV. HDACs are typically found in multiple distinct complexes with other HDAC family members (e.g. HDAC1 with HDAC2 within NuRD, Sin3a and co-REST complexes (Yang and Seto, 2008). Different HDAC enzymes are important for different processes during development, HDAC1 but not HDAC2 has been shown to control embryonic stem cell differentiation (Dovey et al., 2010).

Histone methyltransferases (HMTs) also act on the lysine residues of histones and can lay down 1, 2 or 3 methyl groups. SUB39h1 was the first HMT to be identified and belongs to the SET family (Melcher et al., 2000). Histone methylation can be associated with both transcriptional repression and activation depending on the context of methylation. Histone demethylases comprise of two families, the lysine specific demethylase LSD1 which demethylates histone H3 lysine 4 and the Jumonji (JmjC) domain containing proteins

(JHMD, JMJD3 and JMJD2D) which each target specific methyllysine groups (Zhou and Ma, 2008, Shi et al., 2004).

Modifications associated with active transcription include acetylation of the lysine tail of histone H3 including H3K9ac, H3K14ac, H3K14ac, H3K27ac and H3K122 by histone acetyltransferases and H3Kme2, H3K4me3, H3K9me1, H3K79me1 by HATs and kinases. Modifications associated with repression include H3K9me2, H3K9me3, H3K27me2 and H3K27me3 (Gillette and Hill, 2015). Figure 1.2 summarises the posttranslational modifications found of histones.

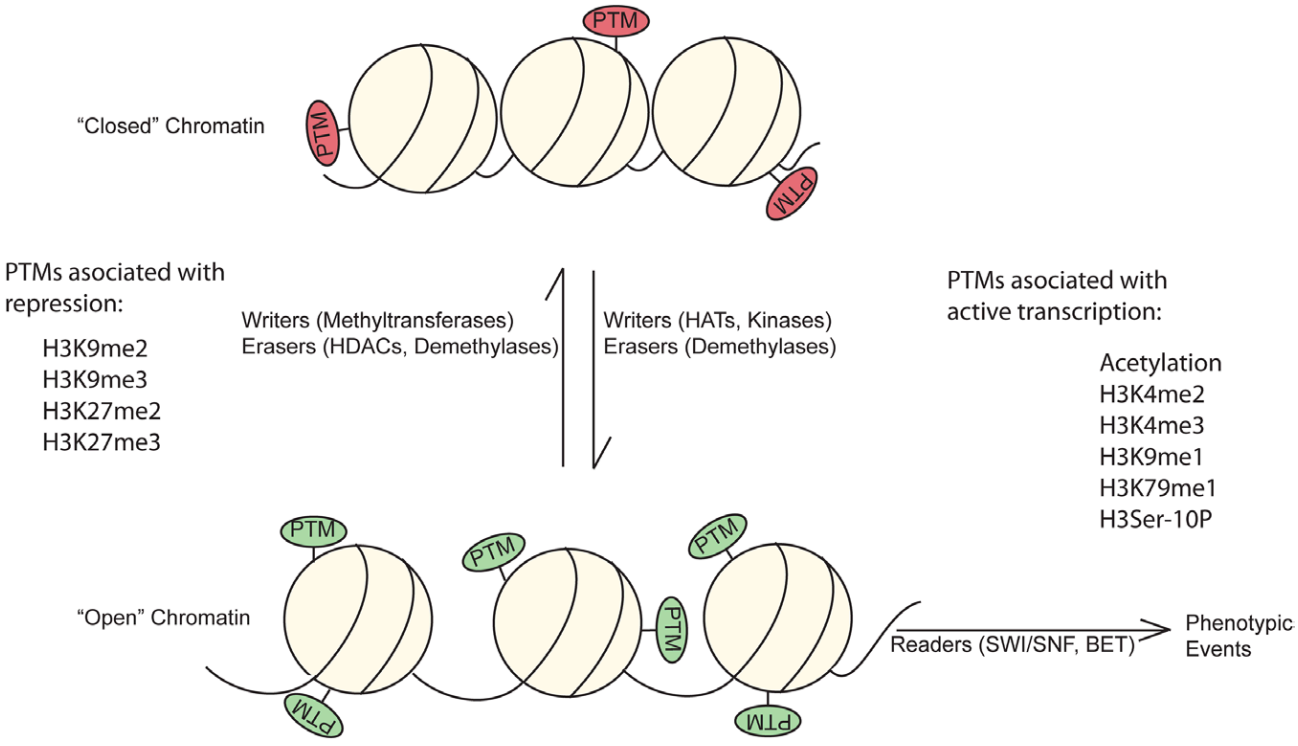


Figure 1.2- Diagram showing posttranslational modification (PTM) found at histones. PTM can induce an open chromatin (shown in green) or a repressed chromatin (shown in red) state. Histone PTM "readers" act to further increase or repress the transcription of target genes. Image from (Gillette and Hill, 2015). (Permission to reproduce this figure has been granted by Wolters Kluwer Health, Inc.)

1.1.3 DNA modifications

Along with histone modifications chemical alteration of the DNA sequence itself can also impact on transcriptional activity. The first early evidence for modified cytosine in DNA was provided by Hotchkiss in 1948 by paper chromatography after observing that this fraction was 5-methylcytosine because it separated from cytosine similarly to how thymine (or methyluracil) separated from uracil (Hotchkiss, 1948). Later studies showed that DNA methylation was important for gene regulation and cell differentiation (Holliday and Pugh, 1975, Compere and Palmiter, 1981). In total four different types of cytosine residue modifications have been identified in DNA; methylation, hydroxy-methylation, formylation and carboxylation.

DNA methylation can occur at the carbon-5 position of cytosines within CpG dinucleotides producing 5-methylcytosine (5mC) by DNA methyltransferase enzymes (DNMT) which transfer methyl groups from S-adenosyl methionine (Li et al., 1992). DNMT3a and DMNT3b establish an initial methylation pattern at unmodified DNA and are known as *de novo* Dnmt. DNMT1 is considered the major maintenance DNMT functioning during DNA replication in humans and contains a C-terminal methyltransferase domain and an N-terminal regulatory domain, linked by a conserved (Gly-Lys)n repeat. The N-terminal domain of DNMT1 contains sequences that regulate interactions with other proteins (Rountree et al., 2000, Chuang et al., 1997), a target recognition sequence which associates DNMT1 to the DNA replication fork (Leonhardt et al., 1992), a zinc finger CXXC domain which can identify unmethylated CpG DNA (Lee et al., 2001, Pradhan et al., 2008) and a pair of bromo-adjacent homology (BAH) domains (Callebaut et al., 1999). DNMT1s' methyltransferase domain is

folding into two subdomains termed the target recognition domain (TRD). While tethered to the DNA replication fork, DNMT1s' TRD domain recognises hemimethylated CpG sites emerging from the replication complex and adds a methyl group to the daughter strand (Song et al., 2012). DNA methylation is generally assumed to be associated with transcriptional repression and it was shown that DNA methylation prevents transcription factors and co-factors from interacting with cis-regulatory elements of genes (Smith and Meissner, 2013, Domcke et al., 2015, Yin et al., 2017). In addition, DNA-methylation is read by methyl-binding proteins which recruit co-repressors such as HDACs and this maintains the transcriptionally silent state (Jones et al., 1998).

5mC is considered a stable DNA state, however, it can be reversed to its unmodified state in a number of ways leading to changes in DNA accessibility and increased gene expression. Dilution of 5mC can occur as a result of a reduction or loss of functional DNA methylation maintenance machinery, this process is known as passive DNA demethylation. Alternatively, Ten-Eleven Translocation (TET) enzymes are involved in the removal of methyl groups from DNA and induce the iterative oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in a process called active DNA demethylation (Wu and Zhang, 2017). Pioneer transcription factors (discussed further in Introduction 1.2.2) such as C/EBP alpha, KLF4, FOXA1 and others, can interact with and recruit TET enzymes (such as TET2) to cis-regulatory elements (enhancers and promoters) of target genes resulting in the loss of 5mC at these sites (Sardina et al., 2018).

1.1.4 Polycomb complexes maintain heritable patterns of gene expression

Polycomb group proteins (PcGs) are important for the control of the stepwise activation of developmentally associated genes. PcG proteins also act to establish heritable chromatin states within cells, leading to the transmission of cell type-specific gene expression patterns from one cell generation to the next which is important in the formation of stable tissues, organs and morphologies. Epigenetic states refer to the distinct gene expression and phenotypic states of genetically identical cells and develop without alteration to the DNA sequence and are maintained in the absence of extra-cellular and intra-cellular signalling (Gottschling, 2004, Ringrose and Paro, 2004).

PcG proteins were first identified as lineage specification regulating factors in *Drosophila melanogaster*. Members of the PcG proteins were first identified by mutations that resulted in de-repression in *D. melanogaster* homeotic (Hox) genes (Schwartz and Pirrotta, 2013). High-throughput studies have demonstrated that PcGs not only act on homeotic genes to provide dynamic repression but also other genes that are involved in the control of development, such as transcription factors or morphogens (Schwartz and Pirrotta, 2013).

PcGs repressive mechanisms involve two complexes, polycomb repressive complex 2 (PRC2), which methylates histone H3 lysine 27 (H3k27), and polycomb repressive complex 1 (PCR1) which can ubiquitylate histone H2A lysine 119 (Schwartz and Pirrotta, 2013). Specific polycomb response elements (PREs), which are recruitment and binding sites for PRC1 and PRC2 have been identified at PcGs target genes in *Drosophila*. Neither PCR1 nor PCR2 bind directly to DNA, what determines whether they bind to specific target genes is

the prior chromatin state of that gene (Schwartz and Pirrotta, 2013). PREs contain binding sites for the YY1 transcription factors and GAGA-related factors (Sing et al., 2009, Woo et al., 2010).

The repressive action of PcGs is not well understood in detail, however, any combination of chromatin condensation, interference with nucleosome remodelling or transcription initiation, inhibition of transcription elongation by H2Aub or blocking of H3K27 acetylation may play a role (Schwartz and Pirrotta, 2013). Within many organisms including insects and mammals, PcGs act to induce a repressed or poised promoter state during development. Promoters bound by PcGs bind to the non-elongating form of RNA-polymerase II and exist in a bivalent state containing the active histone H3 lysine 4 trimethylation (H3K4me3) mark alongside the PcG-associated repressive H3K27me3 mark (Voigt et al., 2013). Through the action of preventing transcriptional elongation, PcGs complexes effectively poise promoters for rapid reactivation of transcription at a defined and appropriate time (Stock et al., 2007).

Gene targeting methods which result in the removal of PcG complexes from promoters result in premature gene activation (Mazzarella et al., 2011) and global changes to gene expression patterns (Lee et al., 2015). Genome wide studies have shown that PcGs complexes display a highly dynamic pattern of binding during development, with promoters being bound and thereby repressed by PcGs at one stage, then derepressed at the next stage resulting in transcriptional activation and then being rebound by PcGs to induce gene silencing and stop transcription at the final stage (Goode et al., 2016, Kloet et

al., 2016).

Both *Drosophila* and mammalian PcGs complexes have been shown to associate with noncoding RNAs, which may act as important scaffolding RNAs to establish and maintain gene silencing (Guenther and Young, 2010, Hekimoglu and Ringrose, 2009). After recruitment of PRC2 and H3K27 methylation, the maintenance of H3K27 methylation may require interactions between PCR1, PCR2, methylated H3k27 and PRE-binding proteins. During embryogenesis, transcription within PRE regions inactivates PRE induced gene silencing by the establishment of transcription-associated histone modifications (Schmitt et al., 2005, Hogga and Karch, 2002). This mechanism along with the absence of H3K27n methylation blocks cooperative interaction between nucleosomal histones, PRE-bound proteins and RNAs prevents the recruitment of polycomb complexes during later chromatin replication cycles.

While polycomb complexes give stable silent or poised gene expression states, Trithorax (TrxG) group proteins oppose the action of polycomb complexes and act to induce active gene expression states. TrxG proteins bind to Trithorax response elements (TREs) (Sneppen and Ringrose, 2019). In vertebrates, the most prominent TrxG proteins are the H3K4me3 histone methyl-transferases of the MLL (mixed lineage leukaemia) class which are required for transcriptional elongation (Sneppen and Ringrose, 2019, Kim and Kingston, 2022).

1.1.6 TADs and nuclear organisation

The three-dimensional folding of the eukaryotic genome in the nucleus is a highly organised and tightly regulated process. The position of genes within nuclei can correlate with transcriptional activity with genes located in the nuclear interior often being active while repressed or heterochromatic nuclear regions are found closer to the periphery (Bickmore, 2013). Interestingly, chromosomes occupy distinct subnuclear territories in an evolutionarily conserved and cell type specific manner (Cremer and Cremer, 2010). Transcriptionally active loci within chromosomes are found at the chromosome surface (Shah et al., 2018). Chromosomes separate into regions forming two types of chromatin, termed type "A" and "B" compartments. A compartments comprise of gene-rich and transcriptionally active chromatin. In contrast B compartments are found to be enriched in repressive chromatin (Rao et al., 2014, Lieberman-Aiden et al., 2009). Chromosomes fold into domains forming intradomain interactions over interdomain interactions, these contact domains are termed topologically associating domains (TADs) (Figure 1.3a) (Nora et al., 2012).

TADs were originally identified by Hi-C analysis of mouse and human embryonic stem cells and human fibroblast cells. These studies indicated that large regions of the genome are characterised by high levels of chromatin interactions only occurring in the large regions or domain. In between these large domains, smaller regions are found with fewer numbers of chromatin interactions (Figure 1.3b) (Dixon et al., 2012). Further Hi-C analysis of the mouse genome indicated that around 2,200 TADs exist accounting for around 95% of the mouse genome and with an average size of 880kb. The assembly of the genome into TADs is in

part conserved between the human and mouse genomes (Dixon et al., 2012). The identification and characterisation of TAD boundaries is challenging and is dependent on the resolution of Hi-C data and the method of TAD annotation used.

CCCTC-binding factor (CTCF) has been found at TAD boundaries and has been deemed to be a mediator of TAD boundary insulation blocking interactions across adjacent TADs (Dixon et al., 2012). The altering of TAD boundaries resulting in ectopic contacts between cis-regulating elements and gene promoters can result in inappropriate gene expression leading to developmental defects and cancer (Groschel et al., 2014, Taberlay et al., 2016). Within TADs gene promoters are associated with their appropriate enhancer elements, with proper enhancer-promoter interactions within the TADs being facilitated by the presence of cohesin and CTCF (Szabo et al., 2019).

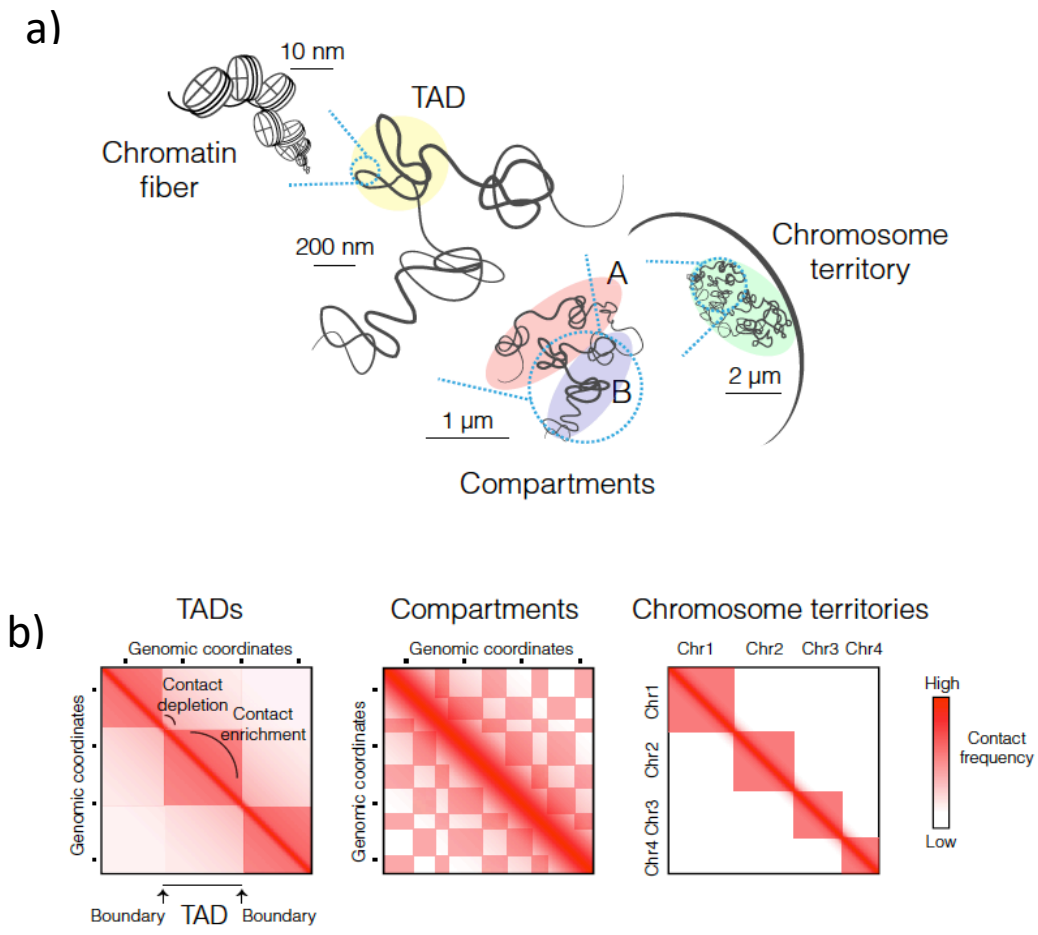


Figure 1.3- Diagram showing the hierarchical folding of the eukaryotic genome inside the nucleus.
a) Schematic overview of folding of the chromosome within the nucleus into chromatin fibers, TADs, compartments and then showing chromosome territories. b) Schematic of Hi-C data maps at different genomic scales and indicating the various levels of higher-order chromosome folding in the nucleus. Image from (Szabo et al., 2019). (Permission to reproduce this figure has been provided under CC BY-NC 4.0 licence (http://creativecommons.org/licenses/by-nc/4.0/).)

It has been hypothesised that TAD structures are formed through a loop extrusion process.

The loop extrusion hypothesis states that long-range *cis*-interaction within a DNA molecule

are generated by loop extrusion factors which bind to DNA and fold flanking regions of the

same DNA molecule into a loop. The loop initially would be small, however, as the loop extrusion factor moves more chromatin into the loop, the loop would get larger (Davidson and Peters, 2021). Studies of type 1 bacterial restriction enzymes, which can fold short stretches of DNA into loops in order to cut DNA in variable positions, provided evidence for DNA folding (Rosamond et al., 1979). Structural maintenance of chromosomes (SMC) protein complexes have been identified which mediate the action of loop formation. The first gene described as an SMC protein was *mukB*, analysis of the protein sequence predicted amino- and carboxy- terminal globular ATPase domains flanking a long stretch of coiled-coil, which was later confirmed by electron micrographs of the MukB protein (Niki et al., 1992). These findings were followed by the identification of similar proteins in yeast; SMC1 (Strunnikov et al., 1993), *cut3+*, *cut14+* (Saka et al., 1994), Smc4, and Smc2 (Hirano and Mitchison, 1994). It was thought that these SMC proteins were key components of chromosome structure.

In eukaryotes SMC proteins form the core of three multi-subunit protein complexes. Notably two SMC proteins, Smc1 and Smc3, are subunits of the cohesin complex, which acts to induce chromosome condensation and Polycomb complex formation (described previously) (Uhlmann, 2016). The cohesion complex forms a ring that associates with DNA and starts to extrude DNA into a loop symmetrically. Cohesin continues to incorporate more DNA into the loop until it encounters convergently orientated CTCF molecules which stops loop extrusion leaving cohesion complex and convergently oriented CTCF at the loops base (Davidson and Peters, 2021).

Evidence of loop extrusion in mammals was provided through studies of the diversity and joining of different segments of mouse immunoglobulin heavy chain genes (Wood and Tonegawa, 1983). The loop extrusion model has also been discussed in the context of how enhancers act over large genomic distances to act on gene promoters regulating their activity and *cis*-spreading of X-chromosome inactivation (Riggs, 1990). Other studies have revealed how condensing complexes, as part of the loop extrusion process, may fold DNA in mitotic chromosomes (Nasmyth, 2001).

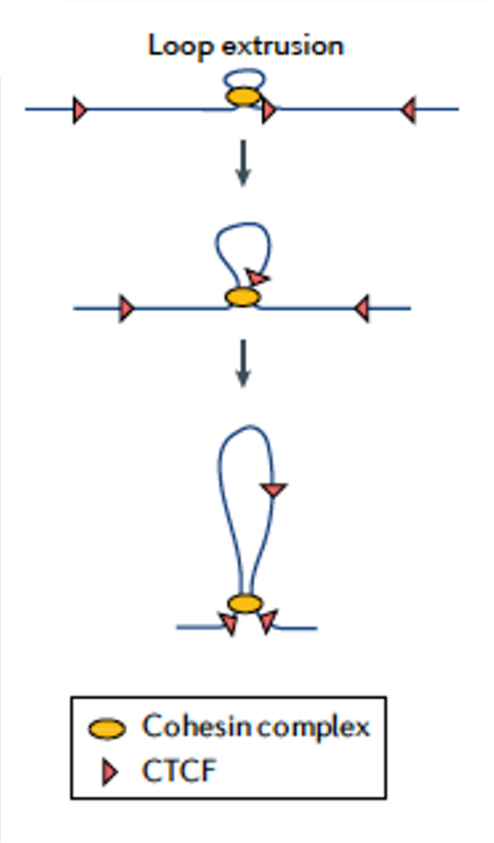


Figure 1.4- Model of DNA loop extrusion conducted by cohesin.
Cohesin binds to DNA initiating symmetrical loop formation. The loop expansion stops when cohesin encounters two convergently orientated CTCF molecules. Image from (Davidson and Peters, 2021). (Permission to reproduce this figure has been granted by Springer Nature.)

Several studies have tested the roles of cohesin and CTCF in TAD organisation and function. Knockout of cohesin had no effect on TAD shape with architectural compartments being maintained, however, reduced interactions within TADs were evident between cohesin bound sites (Seitan et al., 2013, Zuin et al., 2014). CTCF depletion resulted in reduced interactions within TADs but also resulted in an increase in the number of interdomain interactions (Zuin et al., 2014). It is possible that different cell-types recruit cohesins or CTCF binding in different dynamic fashions in order to maintain specific TADs structures.

1.2 Eukaryotic transcriptional regulation

1.2.1 Transcription factors

Transcription factors (TFs) with their associated cofactors and chromatin regulators control the chromatin structure and the transcriptional activity of genes in a complex system that regulates the overall activity of the genome. The key concepts of transcriptional control were established in bacteria over half a century ago (Jacob and Monod, 1961), however, to this day it still remains a challenge to determine the exact mechanisms by which genomic binding sites of TFs are specified, how TFs recruit cofactors and then in turn regulate gene expression levels. Ultimately, the activity of TFs underpins most aspects of mammalian physiology and disease. Currently, the term transcription factor described proteins which can (1) bind DNA in a sequence-specific manner and (2) regulate transcription (Fulton et al., 2009, Vaquerizas et al., 2009). TFs may act as “master regulators” of significant developmental processes (such as RUNX1 during the endothelial-to-haematopoietic transition discussed in detail in 1.4.3) or as “selector genes” (such as homeobox genes

(Mann and Carroll, 2002)).

TFs protein sequences and their binding sites are highly conserved among metazoans (Carroll, 2008, Bejerano et al., 2004). Variation and mutation in their coding sequences or in their transcription factor (TF) binding sites or direct mis-regulation through inappropriate signalling can cause a broad range of diseases indicating the high importance of understanding their precise function (Lambert et al., 2018).

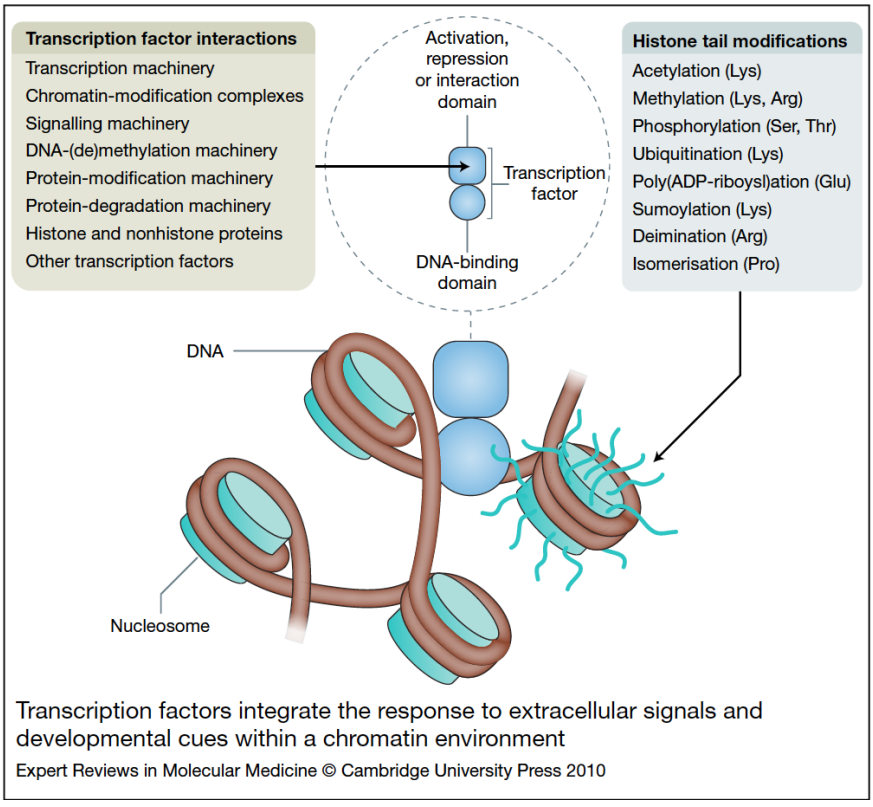


Figure 1.5- Transcription factors bind to DNA within a chromatin environment to integrate signals and developmental cues.
The figure indicates the interaction of molecules and complexes which bind with transcription factors at DNA. Image from (Bonifer and Bowen, 2010). (Permission to reproduce this figure has been provided under CC BY 4.0 licence (http://creativecommons.org/licenses/by/4.0/).)

TFs have a modular structure containing a DNA binding domain (DBD), a transactivation domain (TAD) and a dimerisation/interaction domain (Brent and Ptashne, 1985, Frankel and Kim, 1991). Substantial efforts have identified ~1600 TFs in the human genome which have been structurally classified into TF families (Lambert et al., 2018), often distinguished by the type of DBD, such as C2H2-zinc finger (ZF), homeodomain, basic helix-loop-helix (bHLH), basic leucine zipper (bZIP), and nuclear hormone receptor (NHR) (Johnson and McKnight, 1989). TFs recognise and bind to short degenerate DNA sequences 6-12 bp long and can have over a 100-fold preference for binding to these sequences relative to other sequences (Geertz et al., 2012). These binding sites are clustered in control elements (*cis*-regulatory elements) that recruit the transcription apparatus or modulate the rate of transcription of their target genes (enhancers and other distal elements) (Lambert et al., 2018, Lee and Young, 2013). Significant effort has resulted in the development of binding motif collections such as TRANSFAC (Matys et al., 2006), JASPAR (Mathelier et al., 2016), HT-SELEX (Jolma et al., 2013), UniPROBE (Hume et al., 2015) and CisBP (Weirauch et al., 2014). TF motifs have low sequence specificity in the genome, as a result additional levels of regulation are required to ensure specific TF binding such as open chromatin and accessibility to the site, cofactor binding and combinatorial TF binding (Biggin, 2011).

TFs have the capacity to interact and thus impact on each other's binding and recruit activating or repressive co-factors (Figure 1.5). TFs can cooperatively bind to increase the affinity of such a complex for a DNA sequence or they can bind synergistically to activate the transcription of target genes to a greater extent than any single TF alone (Spitz and Furlong, 2012). Very few TFs occupy all of their TF motifs present in the genome (an idea

termed the ‘futility theorem’ (Wasserman and Sandelin, 2004)), and it is believed that in most cases other conditions must be met for TFs to bind their specific motifs such as cooperativity with other TFs. Some TFs can directly recruit RNA polymerase (e.g., TBP) or accessory factors which promote transcription. They may also recruit co-activators or co-repressors to alter chromatin (Lambert et al., 2018).

Functional and structural studies on TF-TF-DNA complexes have investigated different mechanisms by which TFs can cooperate. A study of 9400 TF-TF-DNA interactions using consecutive affinity-purification systematic evolution of ligands by exponential enrichment (CAP-SELEX) indicated that DNA dependent TF-TF interactions are very common and that interactions can occur between different TF families. The study revealed 315 TF-TF interactions binding to 618 heterodimeric motifs. The researchers estimated that ~25000 distinct TF pair specificities can contribute to TF-TF-DNA interactions (Jolma et al., 2015). TF-TF cooperativity can occur in the presence of or in the absence of DNA (Morgunova and Taipale, 2017).

The activity of TFs can be regulated through intra-cellular and extra-cellular signalling mechanisms. A common mode for the control of TFs in both prokaryotes and eukaryotes is protein phosphorylation. For example, signals transmitted from growth-factor receptors at the cell surface upon receptor ligand binding induces a phosphorylation cascade which terminate at transcription factors altering their activity. NFkB was the first TF whose activity was shown to be regulated this way (Sen and Baltimore, 1986). The signalling cascades and TFs they act on which pertain the most to the work reported in this thesis are described in

detail in Introduction 1.4.4-1.4.8. TFs can interact with serine/threonine kinases, lysine and arginine methyltransferases, lysine acetyltransferases and HDACs resulting in PTMs of the TF. For example the TF RUNX1 has been shown to be phosphorylated within its transactivation domain by ERK1/2 serine/threonine kinases in response to IL-3 (Tanaka et al., 1996), thrombopoietin (Hamelin et al., 2006) and EGF induced signalling (Imai et al., 2004).

1.2.2 Chromatin remodellers and pioneer factors

Nucleosome positioning or depletion can occur through ATP-dependent nucleosome remodelling complexes (Ho and Crabtree, 2010). These complexes cooperate with site sequence-specific transcription factors and histone modifiers to displace or eject histones to allow for further TF binding to DNA and regulate nearly all chromosomal processes. Their deregulation results in a variety of diseases including cancer (Wilson and Roberts, 2011). Remodellers are specialised to perform one of three functions: chromatin access, nucleosome eviction or nucleosome assembly and organisation. All ATP-ase chromatin remodellers have been classified as the RNA/DNA helicase superfamily 2 which is further split into four sub-families of chromatin-remodelling enzymes: chromodomain helicase DNA binding (CHID), imitation switch (ISWI), switch/sucrose non-fermentable (SWI/SNF) and INO80. All remodellers have an ATPase domain which breaks histone-DNA contacts to move DNA along the histone surface (Clapier et al., 2017).

The SWI/SNF remodeller complexes facilitate chromatin access by sliding or ejecting nucleosomes leading to either gene activation or repression. SWI/SNF remodellers contain

two RecA-like lobes which flank an N-terminal helicase/SANT-associated domain which can bind to actin or actin-related proteins, an adjacent post-HAS domain, AT-hooks and a C-terminal bromodomain (Schubert et al., 2013). The use of different core paralogues within SWI/SNF complexes results in tissue/developmental specific SWI/SNF subtypes (Ho and Crabtree, 2010). TFs termed “pioneer factors” recruit SWI/SNF complexes to chromatin which increases chromatin accessibility allowing for additional TFs to bind to DNA.

Pioneer factors are a special class of TFs which are capable of accessing their DNA target sites in compacted chromatin prior to transcriptional activation when other TFs cannot (Figure 1.6). Pioneer factors not only take part in the activation of transcription, but they have been shown to be physically bound to the genome for a period before transcriptional activation of the gene and before other factors bind (Zaret and Carroll, 2011). ChIP-based approaches and in particular *in vivo* footprinting have been used to identify which transcription factors initially bind silent chromatin before any target gene activation and therefore can be classed as a pioneer transcription factor (Zaret and Carroll, 2011).

The first detected pioneer factors were FOXA and GATA factors which bind to a specific enhancer of the *Alb1* gene. In nascent liver buds six TF binding sites were occupied (Liu et al., 1991) as measured by *in vivo* footprinting. In undifferentiated gut endoderm cells however, where *Alb1* gene is not expressed, only two of the TF binding sites were occupied (Gualdi et al., 1996). These binding sites were FOXA and GATA factors which are both expressed in endoderm and the nascent liver and whose binding to the enhancer element are required for *Abl1* transcription. FoxA was the first binding factor to fulfil the criteria of

a pioneer factor (Ang et al., 1993, Liu et al., 1991).

Another example found at an intronic enhancer for the *Pax5* gene in multipotent haematopoietic progenitor cells before *Pax5* transcription (Decker et al., 2009). PU.1 can expand the linker region between nucleosomes and induces local modifications of histones allowing additional factor binding (Ghisletti et al., 2010). Many developmental genes in ES cells may be marked for potential activity by pioneer factors. These data indicate that pioneer factors act more to initiate developmental lineage commitment than to maintain it (Zaret and Carroll, 2011). Pioneer factors can act through active or passive means. Through the passive model the binding of a pioneer factor may have no effect on the chromatin structure but may allow for the recruitment of further factors associating to it. In the active model pioneer factor binding helps to reorganise the chromatin environment allowing for other TFs, chromatin modifiers and nucleosome remodellers to bind (Zaret and Carroll, 2011).

Pioneer factors may also act to establish stably silence domains of the genome. FoxA can act with Groucho/TLE corepressor proteins (Fu et al., 2000) allowing FoxA to recruit TLE3/Grg3 to target sites resulting in the closing of chromatin and the repression of transcription (Sekiya and Zaret, 2007).

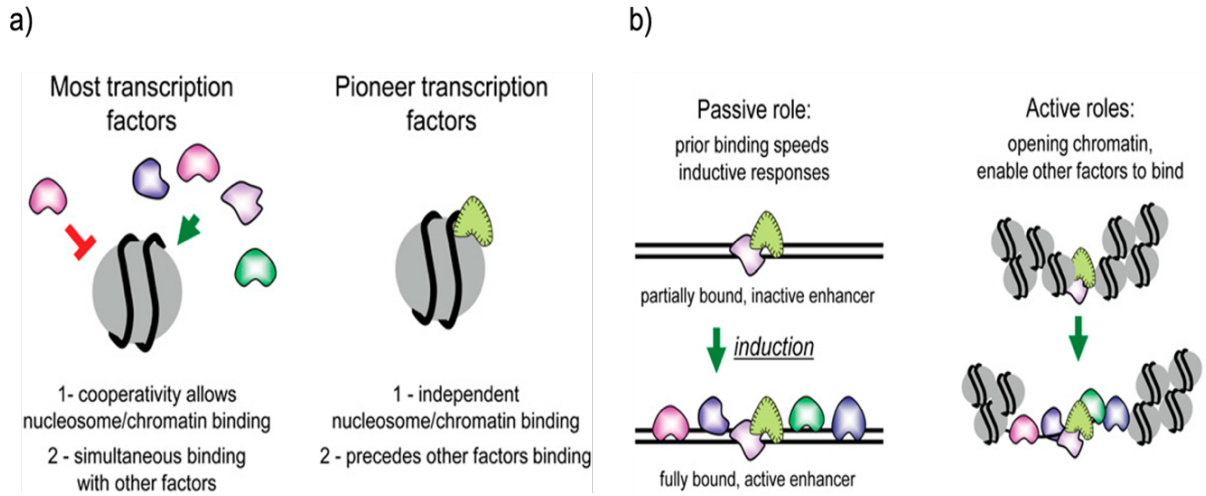


Figure 1.6- The action of pioneer factors and their passive and active roles.
a) Non-pioneer transcription factors can not bind to their target sequences in compacted chromatin or on nucleosomes unless binding in a highly cooperative fashion with additional transcription factors. Pioneer factors however can bind their target sequences on some forms of compacted chromatin and on nucleosomes allowing for other transcription factors to bind. b) Through the passive role binding of a pioneer factor to a regulatory sequence reduces the number of other transcription factors needed to bind to activate the regulatory element. Through the active role, pioneer factors directly facilitate additional transcription factor binding to nucleosomal DNA or open local chromatin allow other factors to bind. Image from (Zaret and Carroll, 2011). (Permission to reproduce this figure has been provided under CC BY 4.0 licence (http://creativecommons.org/licenses/by/4.0/).)

1.2.3 Promoters

Transcription usually initiates at the transcription start site (TSS) located at the 5' end of genes. The TSS is located within a 'core promoter', which is a sequence about 50 bp upstream and downstream of the TSS. The transcriptional machinery is recruited to the TSS, initially the pre-initiation complex (PIC) followed by RNA polymerase II, which recognises and binds to the TSS. The core promoter is able to direct low level basal transcription activity. This basal activity can be reduced by chromatin remodelling or

activated by *cis*-regulatory elements called enhancers.

A proximal promoter site sits a few hundred bases upstream of the core promoter and contains DNA binding motifs for transcription factors which act to influence transcription initiated at the core promoter. Combined, the core and proximal promoters typically span less than 1 kb (Maston et al., 2006). The core promoter contains the TATA box motif (TATAA) about 30 bp upstream of the TSS, which is essential for initiation of transcription of a subset of genes within the genome. Deletion of the TATA box abolishes transcription of associated genes.

TFIID is considered the primary factor thought to initiate the assembly of the PIC (Patel et al., 2020). The TFIID complex includes the TATA-binding protein TBP, which recognises the TATA-box sequence upstream of the TSS, as well as 13 TBP-associated factors (TAFs) which bind downstream promoter elements including the initiator element (INR) and the downstream promoter element (DPE) motifs. Not all promoters contain a TATA-box or CCAAT element (Wiley et al., 1992). Promoters enriched in unmethylated CG dinucleotide repeats called CpG islands which lack the TATA-box sequence have the capacity to initiate transcription (Blake et al., 1990).

1.2.4 CpG islands

CpG islands are short DNA sequences that are significantly different from the typical genomic sequence by being GC-rich or CpG-rich. It is believed that most CG dinucleotides

have been depleted through evolution of the human genome due to the high rate of mutation of methyl-cytosine to thymine by deamination as they can undergo DNA methylation at the 5' carbon of the pyrimidine ring of cytosine (Bird, 1980). However, CG dinucleotide repeats are still over-represented in the human genome and are termed CpG islands as they are usually unmethylated. CpG islands are commonly located in promoter regions where they are unmethylated in transcriptionally active genes allowing for the access of transcriptional machinery to the TSS (Deaton and Bird, 2011). The presence of CpG islands is considered to be a strong indicator of a promoter element, with promoters containing CpG islands being associated to broadly expressed housekeeping genes (Deaton and Bird, 2011). Promoters that are enriched for CpG islands are unlikely to contain core promoter motifs such as a TATA box (Reynolds et al., 1984) and require TFs to bind to recruit the transcriptional complex (Deaton and Bird, 2011).

CpG islands in promoter regions are often marked by the H3K4me3 histone modification which is enriched for at the TSS and indicates actively transcribed genes (Bernstein et al., 2005, Schneider et al., 2004). H3K4me3 interacts with the transcriptional machinery directly as the core TF TFIID (TAF3) can recognise the H3K4me3 mark (van Ingen et al., 2008, Vermeulen et al., 2007). The core transcriptional machinery also has the capacity to recruit H3K4 methyltransferases to chromatin meaning that the act of transcription may in fact contribute to H3K4me3 maintenance at CpG islands (Ruthenburg et al., 2007).

1.2.5 Enhancers and enhancer-promoter interactions

Enhancers are regulatory elements that recruit transcription factors and cofactors to regulate expression of their target genes (Banerji et al., 1981). They act within dynamic gene regulatory networks to regulate tissue and developmental stage-specific gene expression giving rise to specific patterns of temporal and spatial activity (Spitz and Furlong, 2012). Enhancers can be located upstream or downstream of their target genes promoter elements or within the introns of a gene (Bonifer and Cockerill, 2017). Enhancer sequences house TF binding motifs for TFs that recruit co-activators and co-repressors. Enhancer activity has also been correlated with certain states of chromatin; active enhancers typically lack nucleosomes so that DNA is accessible and can therefore be bound by TFs and cofactors. Nucleosomes close to active enhancer elements tend to contain histones with specific post-translational modifications, such as histone H3 lysine 4 monomethylation and H3K27 acetylation at their amino termini (Bonifer and Cockerill, 2017).

A notable hallmark of enhancers which has been repeatedly demonstrated is that they act independently of the distance from and orientation to their associated genes (Field and Adelman, 2020). Gribnau et al (Gribnau et al., 2000) demonstrated that the human β-globin locus is divided into three differentially activated chromatin subdomains (Gribnau et al., 2000). It was then shown at the β-globin locus that long-range gene regulation *in vivo* requires interactions between transcriptional elements with intermittent chromatin looping out (Tolhuis et al., 2002). Due to chromatin looping, enhancer elements can be located at large distances and up to several hundred kilobases or even megabases away

from the promoter they regulate. One promoter can interact with multiple enhancer elements in a looping fashion. Additionally, enhancers maintain their ability to stimulate a promoter independently of sequence context, such as when an enhancer is placed into heterologous reporter constructs (Shlyueva et al., 2014).

Interestingly, in a number of gene clusters such as the β-globin and *HOX* genes which are developmentally regulated, the order of genes within the clusters on the chromosome matches the order of expression of genes, with the enhancer elements that specify these patterns being located within these loci (Strouboulis et al., 1992, Izpisua-Belmonte et al., 1991). Altering the specific order of genes on the chromosome changes the temporal activation of the individual genes on the chromosomes (Soshnikova and Duboule, 2009, Hanscombe et al., 1991) demonstrating that the linear order of genes on the chromosome is essential for their correct developmental control.

A number of studies have shown that cell-stage-specific activation of a gene locus in development is a stepwise process which requires the functional and physical interaction of multiple cis-regulatory elements. The interacting cis-regulatory elements may form active chromatin hubs or domains (Bonifer, 2000, Patrinos et al., 2004) incorporating several DNA elements within active gene loci. These hubs or clusters of regulatory elements have recently been termed “super enhancers” (Whyte et al., 2013) and have been demonstrated to play an important role in development and in disease states such as cancer (Hnisz et al., 2013).

Some enhancers at tissue specifically active gene loci show signs of activation before expression of the gene they are associated to. Chromatin priming was first demonstrated by single gene analysis (Gualdi et al., 1996, Kontaraki et al., 2000, Tagoh et al., 2004, Xu et al., 2007, Walter et al., 2008) and then demonstrated in numerous system-wide studies (Pimkin et al., 2014, Goode et al., 2016, Mercer et al., 2011, Wang et al., 2015, Wamstad et al., 2012, Gifford et al., 2013, Lara-Astiaso et al., 2014). Chromatin priming establishes a ready state for genes in precursor cells before signalling responsive commitment decisions occur (Figure 1.7). Genes with cis-regulatory elements in this primed state may be associated with low levels of gene expression or no gene expression at all. Specific transcription factors maintain the primed or subsequently committed states of cells during differentiation (Goode et al., 2016, Tsankov et al., 2015).

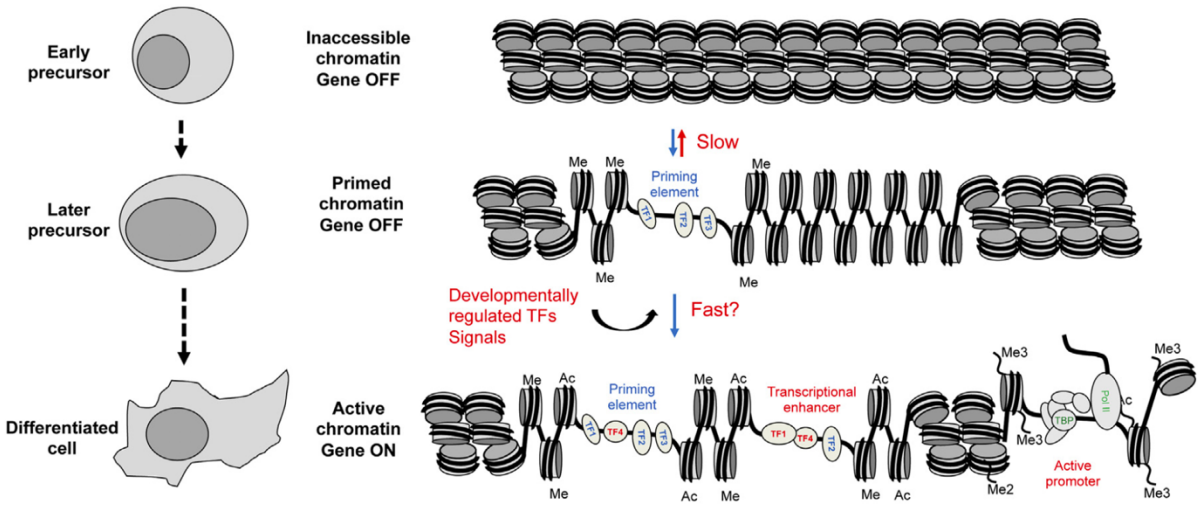


Figure 1.7- Model of chromatin priming and transcription factors recruitment to enhancer and promoter regulatory elements in differentially regulated genes.
In response to signalling, more TFs recognise and bind to priming and enhancer *cis*-regulatory elements. These bound factors influence promoter activity and result in an active chromatin site defined by histones carrying acetylation (Ac) or lysine 4 trimethyl (me3) marks. Image from (Bonifer and Cockerill, 2017). (Permission to reproduce this figure has been provided under CC BY 4.0 licence (http://creativecommons.org/licenses/by/4.0/).)

Gualdi et al (Gualdi et al., 1996) demonstrated that TF-binding motifs were occupied at sequences of an albumin enhancer in precursor gut endoderm where the gene is not expressed. The albumin gene is only active when additional factors bind to nearby sites after hepatic cell commitment (Gualdi et al., 1996). Studies of the PU.1 (*Spi1*) locus showed that prior to transcriptional activation of this gene selective demethylation of specific TF-binding sites at one of the genes enhancer elements occurs. Simultaneously DNase I accessibility increased as a result of the activity of the TF RUNX1 leading to the assembly of a stable factor complex and the activation of gene transcription (Hoogenkamp et al., 2009) alongside the activation of other transcriptional enhancers. Moreover, mutation of the demethylated TF binding site led to a delay in the transcriptional activation of the locus (Lichtinger et al., 2012). The importance of chromatin priming for the developmental control of gene expression at a global level is still largely unknown.

Pol II ChIP-seq studies have shown substantial Pol II occupancy of intergenic loci, indicating that transcription can occur at enhancers (Carroll et al., 2006). More recently studies have confirmed that transcription is a general feature of active enhancers (Kim et al., 2010) as a result of TF occupancy of enhancer regions. Enhancer transcription produces enhancer RNA (eRNA) which are typically short (less than 200 nucleotides long) eRNAs are unstable and prone to rapid degradation by the exome (Andersson et al., 2014) (Figure 1.8). The instability of eRNAs means they are challenging to detect by steady-state RNA-seq methods, therefore, specific methods such as global run-on sequencing (GRO-seq) (Core et al., 2008), precision run-on sequencing (PRO-seq) (Kwak et al., 2013) and Start-seq (Nechaev et al., 2010) are employed to detect eRNAs. These sequencing methods have

been used to identify putative enhancers. For example, a study in *Drosophila* found that ~94% of enhancers previously defined through massively parallel reporter assays (MPRAs) underwent transcription (Henriques et al., 2018). However, whether enhancer transcription is necessary or a by-product of enhancer activity is still unclear (Field and Adelman, 2020).

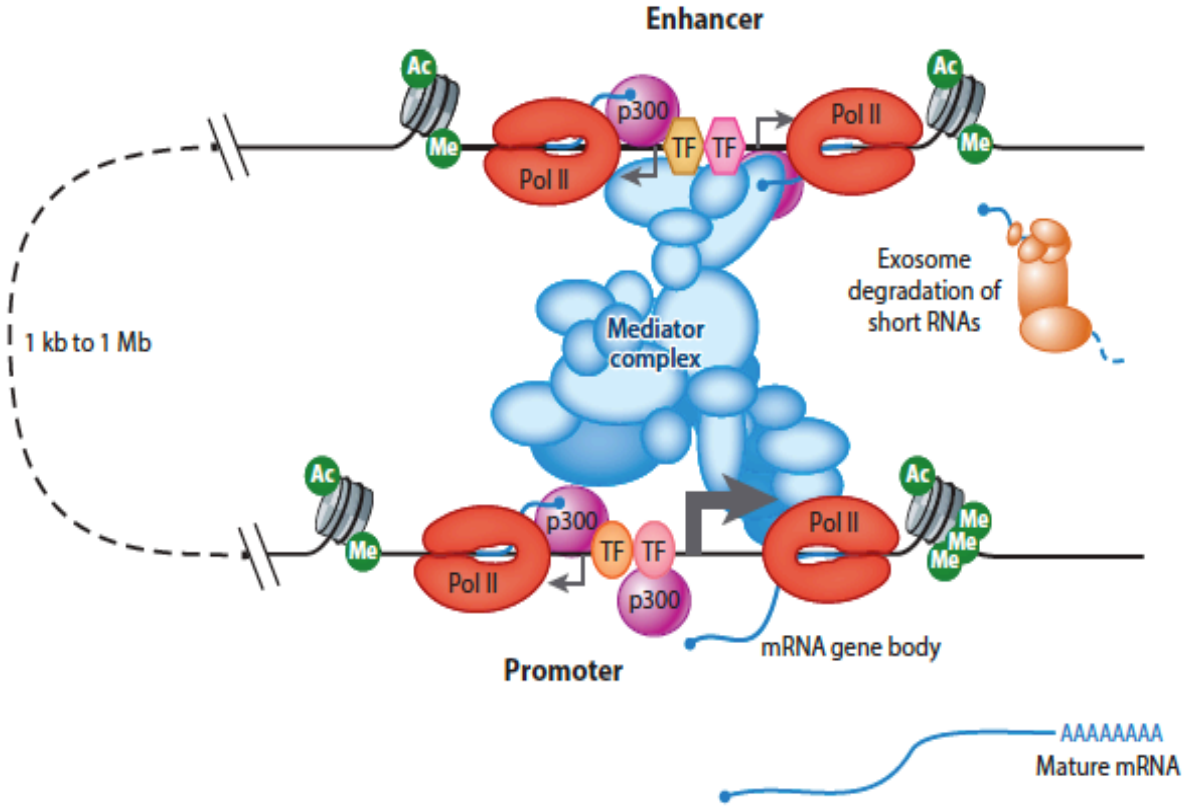


Figure 1.8- Diagram showing the features of an active enhancer.
The molecular features of both the enhancer element (top) and the target promoter (bottom) shown. The transcription start sites are shown by grey arrows. The enhancer and the promoter are associated with Pol II (red), the mediator complex (blue) and p300 (purple). Histone modifications are labelled Ac (histone H3 lysine 27 acetylation (H3K27ac) or Me (histone 3 lysine 4 monomethylation/trimethylation (H3K4me1/me3)) (green). RNA extends from Pol II and enhancers (blue). Image from (Field and Adelman, 2020). (Permission to reproduce this figure has been granted by Annual Reviews, Inc.)

1.2.6 Insulators

Insulators effectively prevent promoter elements being inappropriately stimulated by enhancers with which they are not associated. This regulatory safety mechanism exists in flies and mammals and involves the recruitment of protein factors called insulator binding proteins (IBPs) to these dedicated DNA elements. Currently only one IBP has been identified in mammals called the CCCTC-binding factor (CTCF), although in *Drosophila* more IBPs have been identified. The regulatory mechanism by which insulators prevent inappropriate promoter activation by enhancer elements is unknown, however, it is thought that IBPs induce folding of chromosomes into conformations which affect enhancer-promoter interactions (Bell et al., 1999).

Evidence for insulators came from the study of the *Hox* gene *Abdominal-B* (*Abd-B*) in *Drosophila* where a chromosomal deletion removed the boundary between two enhancer domains which both direct *Abd-B* expression in separate body segments. Deletion of the boundary between these two enhancers resulted in a phenotype whereby the gene activation occurred by the incorrect enhancer in the wrong body segment (Gyurkovics et al., 1990, Ozdemir and Gambetta, 2019). Study of the *gypsy* transposable element in *Drosophila* revealed that transposition of *gypsy* in between a gene and its distal tissue-specific enhancers stopped gene expression in this tissue (Peifer and Bender, 1986, Jack et al., 1991, Geyer and Corces, 1992). Reporter assays developed in 1991 allowed for numerous studies of insulator activity in flies (Kellum and Schedl, 1991, Kellum and Schedl, 1992) and vertebrates (Chung et al., 1993, Ozdemir and Gambetta, 2019). Candidate insulators flanking a transgene which was then inserted into chromosomal locations

randomly induced similar levels of expression by protecting them from inappropriate activation/repression from local regulatory elements (Kellum and Schedl, 1991). These studies were conducted using reporter gene assays and currently very few insulators have been functionally tested in their endogenous loci (Ozdemir and Gambetta, 2019).

1.2.7 Reporter constructs and reporters in chromatin

Studying the mechanisms by *cis*-regulatory elements function is highly important for understanding how dynamic gene regulatory networks control embryonic development processes such as early haematopoiesis. There is no single assay to definitively discover enhancers and their target genes. To identify enhancers a number of assays can be used in combination.

The gold standard for experimentally studying *cis*-regulatory elements is considered to be transgenic animal models such as mouse models. These are powerful tools as they reveal the precise spatial and temporal activity of enhancers. However, these models can be expensive and low throughput as there are challenges with integrating variable copy numbers of the transgene and with positional effects, requiring multiple transgenic lines to be produced to achieve a reproducible expression pattern (Wilkinson et al., 2013, Bonifer, 2000).

The most common assay used to assess if a sequence has enhancer capabilities are reporter

assays. The assay uses a construct containing a reporter gene (luciferase or GFP for example) driven by a minimal promoter element which cannot drive the activity of the reporter gene alone. A sequence to test can then be inserted into the construct and the effect on the reporter activity measured. However, reporter-based assays are not able to reliably link enhancers to their target genes or reveal information about their activity in their endogenous locus (Field and Adelman, 2020, Arnold et al., 2013).

A recent method for studying *cis*-regulatory elements involves developing embryonic stem cell lines which carry a defective *hypoxanthine guanine phosphoribosyl transferase 1 (Hprt)* gene (Wilkinson et al., 2013) (Figure 1.9). Targeting transgenes with flanking HPRT sequences restores *Hprt* function and allows for accurate single copy integrates. It has been demonstrated that insertion of tissue-specific *cis*-regulatory elements into *Hprt* targeting constructs results in transgene expression under control of the exogenous regulatory elements. Each of *Flt-1*, *vWF* and *Tie-2* regulatory elements inserted into the *Hprt* gene locus resulted in appropriate expression patterns in transgenic mice (Evans et al., 2000, Minami et al., 2002).

Wilkinson and colleagues (Wilkinson et al., 2013) developed an ES cell line carrying a more advanced enhancer reporter system for the interrogation of developmentally associated *cis*-regulatory elements, consisting of a *LacZ* reporter gene with the *LacZ* reporter replaced by a fluorescent reporter gene (*Venus-YFP*) (Figure.10). The *Hprt* locus also contains a Hsp68 minimal promotor upstream of the fluorescent reporter gene. This arrangement allows for the cloning of *cis*-regulatory elements upstream this promotor by Gateway ®

cloning (Hartley et al 2000) and, via differentiation of ES cells into various tissues, for the assessment of tissue-specific enhancer activity.

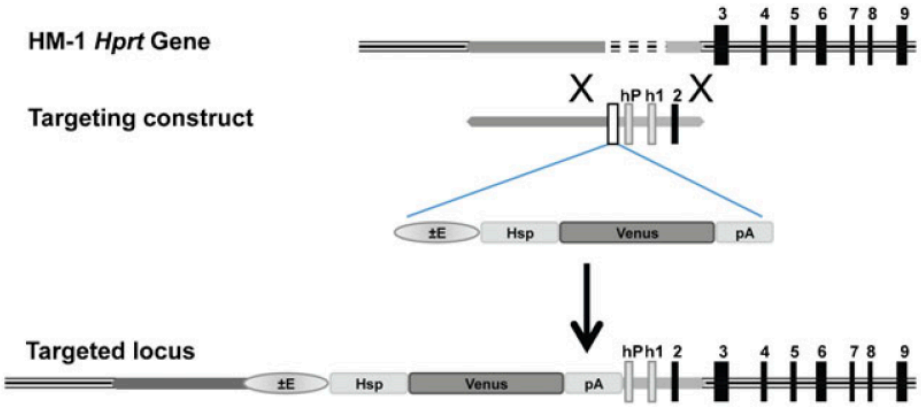


Figure 1.9- *Hprt* locus in mouse HM-1 cells.
Hprt targeting vector containing the enhancer of interest, *Hsp68* minimal promoter and *Venus* fluorescent reporter gene upstream of a *HPRT* promoter. Homologous recombination of this targeting construct results in reconstitution of the *Hprt* locus and insertion of the enhancer*/Hsp68/Venus* reporter. Functional enhancers then drive the expression of an expression cassette coding for a Venus fluorochrome. Image from (Wilkinson et al., 2013). (Permissions to reproduce this figure has been provided under CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).)

To study how the endogenous chromatin environment of an enhancer influences its biological role assays that use enhancer traps or sensors have been developed. A transposase inserts randomly a reporter gene construct at sites in the genome. If an insert occurs close to an enhancer element it will be activated in a way that reports the spatial and temporal activity of the enhancer. Reporter activity is then measured by staining embryos for reporter gene at specific developmental stages. The insertion sites are then mapped to reveal developmentally regulated genomic regions (O'Kane and Gehring, 1987).

Enhancer traps identify genomic regions which are potentially influenced by regulatory elements in the host genome while enhancer reporters identify linked regions that have enhancer activity. However, reporter assays and enhancer traps cannot assess the requirement of an element for appropriate gene expression within a whole gene locus. To study this issue, the putative enhancers must be perturbed in their genomic environment. This can be conducted through the use of clustered regulatory interspaced short palindromic repeats and CRISPR-associated protein 9 (CRISPR-Cas9) technology directed by guide RNAs (gRNAs) (Pickar-Oliver and Gersbach, 2019). Such a technique was recently developed to examine the role of CTCF in the activity of enhancers in a haematopoietic cell line (Qi et al., 2021).

1.3 mRNA transcription

1.3.1 Transcription- initiation

During transcription initiation RNA Pol II assembles with the general transcription factors (GTFs) TFIIB, TFIID, TFIIE, TFIIF and TFIIH at the core promoter in a stepwise fashion to form the pre-initiation complex (PIC). In TATA-box containing promoters, PIC formation initiates with TFIID binding to the TATA box, initiator and/or a downstream promoter element (DPE) (Thomas and Chiang, 2006). TFIID contains TBP and 12-13 TBP associated factors (TAFs). TFIIA binds first, which in turn recruits TFIIB which then recruits the pre-formed TFIIF-RNAPII complex. After PIC formation a number of DNA-GTF interactions hold Pol II to double stranded promoter DNA sitting outside of the RNA Poll II active site cleft. TFIIH then

induced ATP-dependent unwinding of ~10 base pairs of promoter DNA resulting in an open complex (OC) state. TFIIB and TFIIF then orientate the DNA template leading to Pol II being able to locate the TSS and progress into an elongation phase (Grunberg and Hahn, 2013, Tirode et al., 1999).

The TFIID complex which contains TBP can act to drive transcription at a promoter even if it doesn't contain a TATA-box (Pugh and Tjian, 1991). Some TFIID subunits are capable of recognising short promoter elements such as INR, MTE and DPE and it is believed that sequence recognition of these motifs by TFIID subunits allow for PIC formation at TATA-less promoters (Donczew and Hahn, 2018). Genome wide analysis of PICs in yeast identified that PICs are often found at TATA-like sequences, deviating from typical TATA consensus sequence by 1 or 2 mismatches (Rhee and Pugh, 2012).

1.3.2 Transcription- elongation

Once the initiation complex has been established the initiation complex moves into the gene body. TFIIA, TFIID, TFIIH and TFIIE remain bound to the gene promoter effectively priming the promoter for reinitiating (Yudkovsky et al., 2000). RNA Poll II transcribes around 20-120 bp downstream of the TSS. At this point TFIIH induces Poll II to pause. Pausing is an important regulatory step to ensure that the mRNA is capped at the 5'end. For elongation to continue Poll II requires the activity of the highly regulated positive transcription elongation factor b (P-TEFb). Once P-TEFb is recruited by cofactors it phosphorylates NELF, DSIF, and RNAPII on Ser2 leading to recruitment of elongation factors and the release of

Poll II from the PIC and a paused state. The rate of elongation varies between genes and within genes (Jonkers et al., 2014) and appears to play a role in co-transcriptional processes such as splicing and maintenance of genome stability. Features such as histone modifications and modifications of the gene body itself can impact on the rate of elongation (Alexander et al., 2010, Saponaro et al., 2014).

1.3.3 Transcription- termination and polyadenylation

Transcription termination and polyadenylation are required for the production of stable transcripts which are then transported to the cytoplasm for translation. Termination occurs when the Poll II and the newly synthesised nascent RNA are released from the DNA template. The termination of transcription depends on three complexes conserved in eukaryotes; CPF, CFIA and CFIB (Kuehner et al., 2011). Components of these complexes recognise the 3' untranslated region (UTR) of nascent RNA. The CTD-interacting domain of the Pcf1 1 subunit recognises the Ser2-phosphorylated (Ser2P) carboxy-terminal domain (CTD) of the largest Pol II subunit (Kim et al., 2006). Pcf1 1 and Clp1 then bind to the body of the polymerase and subsequently the RNA is cleaved by the CPF endonuclease Ysh1 at the poly(A) site. Adenosine nucleotides are then added to the free hydroxyl group at the 3' end creating a poly(A) tail, which is then bound by Poly(A)-binding protein. It has been shown that Pcf1 1 can terminate elongation complex *in vitro* by destabilising the elongation complex (Zhang et al., 2005). Another model suggests that after transcription of the poly(A) site, termination complex binding induces conformational changes to the elongation complex which decreases processivity leading to termination (Richard and Manley, 2009).

Many human genes express distinct mRNA transcript isoforms in a cell type dependent manner through differential RNA processing. Both alternative promoters or transcript termination sites can control the boundaries of 5' and 3' untranslated regions (UTRs) and give rise to a large proportion of the variation between isoforms (Reyes and Huber, 2018). Distinct UTRs can however drive isoforms with the same coding region to be differentially processed. The 3' end of RNA polymerase II (RNAP II)-derived transcripts result from endonucleolytic cleavage at a polyadenylation site defined by sequence motifs. Subsequently a poly(A) tail is added to the 3' end of the transcript (Edmonds et al., 1971, Darnell et al., 1971). The expression of specific regulators that bind directly to the sequence or to structural components in pre-mRNAs determines the poly(A) site used.

Alternative cleavage and polyadenylation produce different mRNA isoforms according to their coding sequence (CDS) or in their 3' UTRs. Over 80 proteins are involved in mRNA 3' end processing (Shi et al., 2009), with around 20 of these proteins being considered core proteins (Tian and Manley, 2017). The core pre-mRNA 3' end processing complex is formed of four subcomplexes, the cleavage stimulation factor (CSTF), cleavage and polyadenylation factor (CPSF), cleavage factor I (CFI) and CFII. Symplekin and poly(A) polymerase (PAP) are also part of the processing complex (Tian and Manley, 2017). The canonical poly(A) signal (PAS), a hexameric consensus motif AAUUAAA (Proudfoot and Brownlee, 1976), defines the site of pre-mRNA cleavage and polyadenylation. A co-transcriptionally recruited CPSF subcomplex (made up of CPSF1-CPSF4, WDR33 and FIP1L1) recognises the PAS site and cleaves the mRNA 3' end ~21 nucleotides downstream of the PAS and usually adjacent to an adenosine (Wang et al., 2018). CPSF3 performs endonucleolytic cleavage (Dominski et

al., 2005, Mandel et al., 2006) and poly(A) polymerase adds a poly(A) tail to the 5' cleavage product. Once the tail extends to ~250 nucleotides, PABPN1 binds to the poly(A) tail interfering with CPSF and poly(A) polymerase interactions thereby regulating poly(A) tail length (Eckmann et al., 2011).

1.4 Haematopoiesis

For many decades the haematopoietic system has served as a general model for studies of the molecular basis of cell fate decisions and gene regulation. Haematopoiesis is the process of blood cell development occurring in all invertebrates and vertebrates. In mammals, blood production takes place in the bone marrow. All red cells which allow the efficient carriage of oxygen, immune cells that protect against infection and disease and platelets essential for clotting are derived from a multipotent haematopoietic stem cell (HSC) which can self-renew in the right niche environment. HSC are defined as being multipotent, being able to self-renew and are capable of long-term engraftment and reconstitution of the entire haematopoietic system after transplantation into irradiated adult recipients (Majeti et al., 2007).

HSC lineage commitment to the discrete multipotent, oligopotent and unipotent progenitor cells is associated with loss of the self-renewal capacity of these cells. The development of these lineage restricted progenitor cells from the HSCs occurs through several sequential binary decisions, summarised by the classical hierarchical tree-like model of haematopoiesis (Figure 1.10) (Chao et al., 2008, Kondo et al., 1997, Akashi et al.,

2000, Notta et al., 2011). This model was formulated from analysis of fluorescence-activated cell sorting (FACS)-purified cell populations data and lineage contribution associated with each population determined by colony formation assays and/or transplantation assays. However, different assays and different markers used to classify cell types give rise to conflicting branching points within the model and thus alternative hierarchies have been proposed. More recent studies employing single-cell approaches have challenged this traditional stepwise hierarchical model, with data indicating that differentiation follows a continuous process along one trajectory from the HSC and that intermediate progenitors such as MPPs represent transitory states, rather than discreet cell types (Pellin et al., 2019, Naik et al., 2013).

In both humans and mice, mature HSCs found in the adult bone marrow derive from early immature HSCs which first emerge in the early embryo (Medvinsky et al., 1993). The process of HSC formation in embryonic development is a topic of intense study, as the ability to drive the emergence of large quantities of HSCs capable of giving rise to any of the constituents of blood on demand would be of extreme therapeutic and biotechnological value. Understanding the anatomical and cellular origin of HSCs including the signalling pathways and signalling responsive transcription factor regulatory networks which control their emergence is of high importance in order to be able to mimic *in vivo* conditions *in vitro*. The next sections of this chapter lay out our current understanding of mouse embryonic haematopoiesis and underpins the rational for using mouse embryonic haematopoiesis as a model process for the study of signalling responsive TF activity.

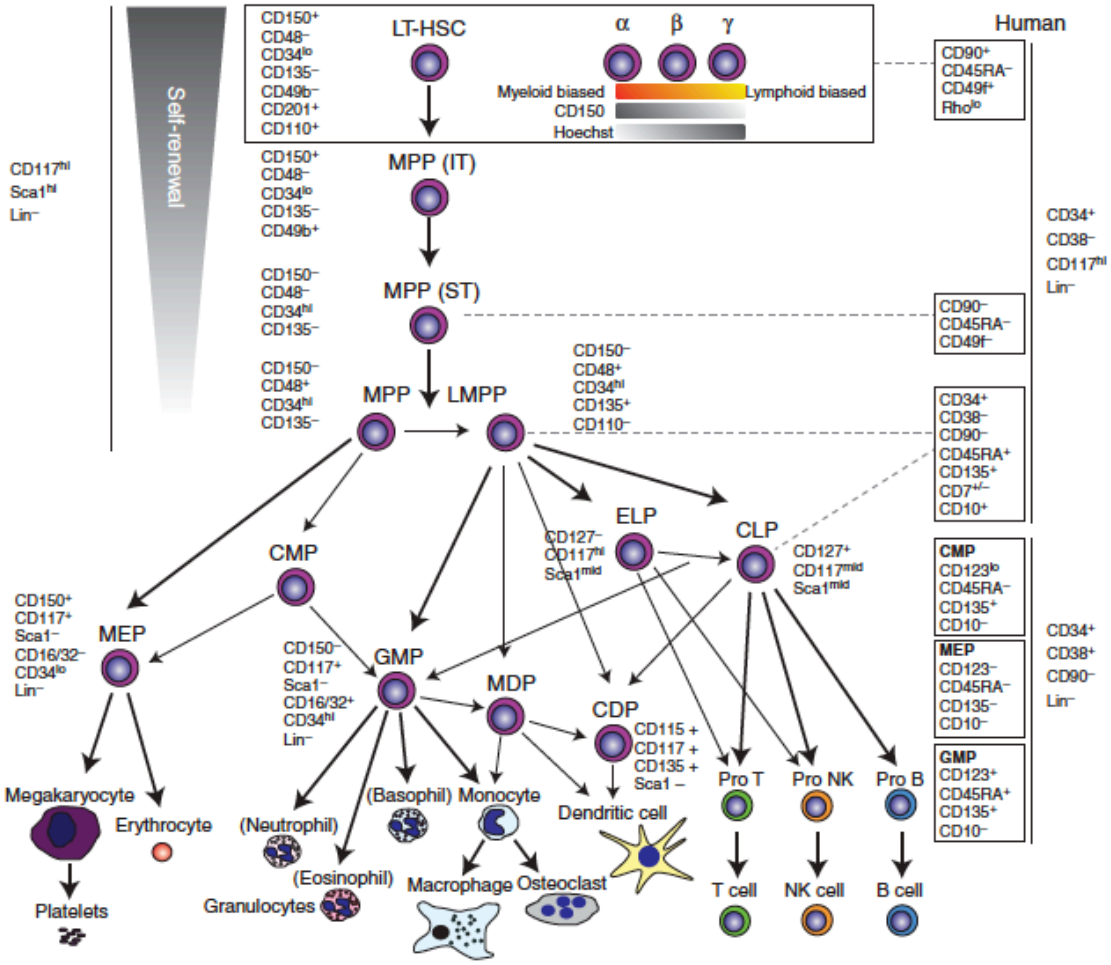


Figure 1.10- Murine adult haematopoiesis differentiation hierarchy.
At each stage of the hierarchy mouse markers define cell types are displayed (human markers are displayed on the right). Self-renewal capacity shown as decreasing from top to lower in the hierarchy. Long-term self-renewing HSCs are at the top of the hierarchy. HSC=haematopoietic stem cell; MPP=multipotent progenitor; LT-= long-term repopulating; IT-=Intermediate-term repopulating; ST-=short-term repopulating; LMPP= lymphoid-primed MPP; ELP= early lymphoid progenitor; CLP= common lymphoid progenitor; CMP= common myeloid progenitor; GMP= granulocyte-macrophage progenitor; MEP= megakaryocyte-erythrocyte progenitor; CDP= common dendritic progenitor; MDP= monocyte-dendritic cell progenitor; NK= natural killer cell. Image from (Rieger and Schroeder, 2012). (Permissions to reproduce this figure has been provided under CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).)

1.4.1 Anatomical origin of mouse embryonic haematopoiesis

The development of the embryonic haematopoietic system is an excellent model for studying how transcriptional mechanisms control cell fate decisions. The entire process is orchestrated by the complex interplay of stage specific TFs with chromatin. The location of the emergence of such cells is shared between mouse and human embryos and even in fish. Importantly, key transcriptional regulators and signalling pathways also appear to be conserved making the mouse a prime model for the study of early haematopoiesis in vertebrates (Sykes and Scadden, 2013).

During murine embryonic development, gastrulation gives rise to the three germ layers: ectoderm, mesoderm and the endoderm. These three layers form the basis of the body plan. During gastrulation the extra-embryonic and intra-embryonic territories are defined, constituting the yolk sac and the embryo, respectively (Bardot and Hadjantonakis, 2020). Haematopoiesis develops in successive steps in the early embryo. In murine embryos the first wave of primitive precursors emerges soon after the start of mesoderm formation around E7.5 within the blood islands of the yolk sac, generating primitive erythrocytes, megakaryocytes and macrophages (Palis et al., 1999, Xu et al., 2001) (Figure 1.11).

The macrophage progenitors which emerge from the first wave at around E7.5 are the source of early macrophages found in the embryo and are essential for the embryo's survival. A second wave of multipotent erythro-myeloid progenitors (EMPs) arise from the yolk sac at E8.5 (McGrath et al., 2015). These EMPs seed the skin or colonise the foetal liver and initiate myelopoiesis to form macrophages. These yolk sac derived macrophages go on

to form tissue resident macrophages across the body which fulfil important tissue-specific and niche-specific functions ranging from homeostatic functions to tissue immune surveillance and response to infection (Hoeffel et al., 2012, Schulz et al., 2012).

It was originally thought that the definitive blood cell lines also originated from the yolk sac, however, no long-term repopulating progenitor cells could be isolated in the yolk sac prior to the formation of the circulation. In 1975 a chimeric study involving the grafting of two-day old quail embryos on to chick yolk-sacs of developmental stages revealed that all the cells in the developing haematopoietic organs were of quail type. This result indicated an intra-embryonic source of HSCs, not yolk sac (Dieterlen-Lievre, 1975). These findings were later confirmed as an intra-embryonic site of HSC production (Medvinsky et al., 1993). Muller and colleagues (Muller et al., 1994) in 1994 demonstrated that transplantation of cells derived from a region called the aorta-gonad-mesonephros (AGM) region (which is derived from the mesodermal germ layer) to lethally irradiated mice resulted in long-term reconstitution of the haematopoietic system (Muller et al., 1994).

It is now thought that a second wave of definitive erythroid and myeloid progenitors is the source of the adult haematopoietic system and originates from endothelial cells found in the dorsal aorta (de Bruijn et al., 2000) of mid-gestational embryos in the AGM (Ivanovs et al., 2011, Medvinsky et al., 1993, Muller et al., 1994). These endothelial cells undergo a process called the endothelial to haematopoietic transition (EHT) producing the HSC cell type.

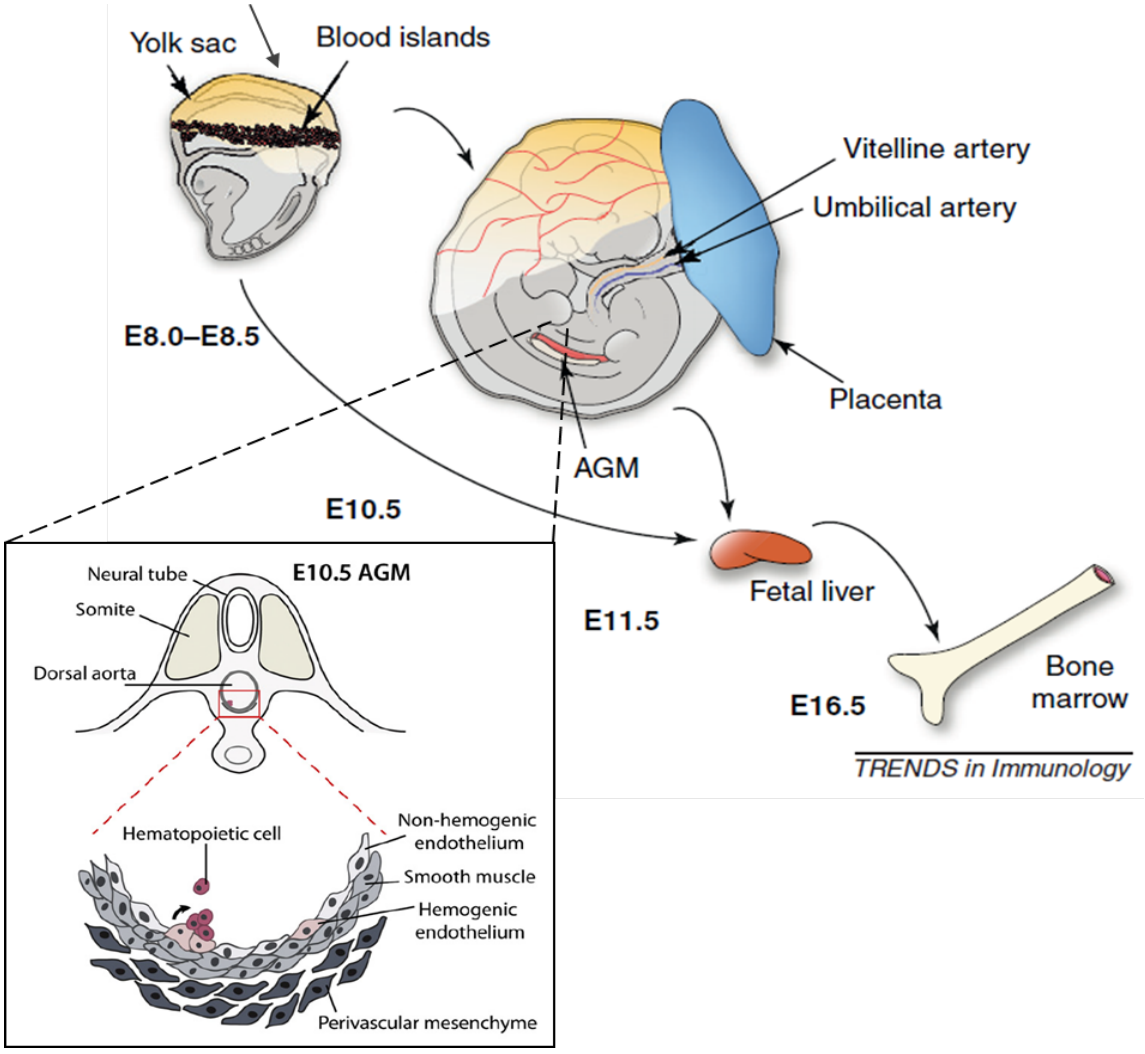


Figure 1.11- Embryonic haematopoiesis develops in distinct anatomical sites in a sequential manner.

Within the yolk sac of murine embryos primitive haematopoietic progenitors are formed within the blood islands at ~E8. At ~E10.5 within the AGM region at a site termed the dorsal aorta a specialised HE forms which gives rise to the HSCs. HSCs which bud off from the HE entre the circulation and colonise the foetal liver at ~E11.5 where they expand before populating the bone marrow prior to birth. The bone marrow remains the HSC niche during adult life. Adapted from (Costa et al., 2012, Swiers et al., 2013). (Permission to reproduce this figure has been granted by Elsevier.)

1.4.2 Cellular origin of embryonic haematopoiesis

The blood islands of the yolk sac form from an aggregation of mesoderm cells. Peripheral cells differentiate to form the endothelium and the central cells form embryonic haematopoietic cells. The similarities between the development of the endothelial and haematopoietic cells led to the hypothesis that they develop from a common precursor cell named the haemangioblast (HB) (Sabin, 1920).

The early concept of the HB gained support from the observation *in vitro* that the haematopoietic and endothelial lineages share expression of a number of genes and that, in response to vascular endothelial growth factor (VEGF) and c-kit ligand, these precursors give rise to blast colonies which express genes associated with haematopoietic precursors (Choi et al., 1998). The blast colonies also expressed β-H1 and β-major globin but lacked the mesodermal marker Brachyury (*Bry*). Kinetic analysis indicated that these colony-forming cells reflect a transient population prior to the production of the primitive erythroid and other lineage-restricted precursors (Choi et al., 1998). The HB expresses fibroblast growth factor receptor-1 (Fgfr1) and bFGF-mediated signalling is critical for the proliferation of the HB (Faloon et al., 2000).

Gene targeting experiments have demonstrated that a functional FLK-1 (*Kdr*) receptor tyrosine kinase (RTK) which is expressed on endothelial cells and their embryonic precursors the HB is required for the development of the blood islands, and that both the haematopoietic and endothelial lineages derive from a common precursor cell (Shalaby et

al., 1995). The ligand for the FLK-1 receptor is the vascular endothelial growth factor (VEGF). FLK-1 is essential for embryonic vasculogenesis and angiogenesis as mouse FLK-1 null embryos die *in utero* between E8.5 and E9.5 from failed development of endothelial and haematopoietic cells (Shalaby et al., 1995). However, these mice also displayed a failure of haematopoiesis demonstrating that this factor is essential for blood cell development. The ETS family transcription factor ETV2 is expressed in early mesoderm and FLK-1+ HB cells (Goode et al., 2016) and is essential for the development of the HB. ETV2 activity is modulated by VEGF induced MAPK signalling and acts to promote vascular endothelial differentiation (Casie Chetty et al., 2017). Knockout of *Etv2* in mice is embryonically lethal as blood islands in the yolk sack failing to form (Ishitobi et al., 2011).

ETV2 upregulates the expression of key TFs required for the next steps of differentiation such as FLI1, GATA2 and SCL/TAL1 (Wareing et al., 2012). Many TFs such as GATA1, GATA2, PU.1, MYB, IKAROS, SCL/TAL1 and RUNX1 have been investigated for their role in the development of this early HB precursor cell type from ES cells (Porcher et al., 1996). The basic-helix-loop-helix TF SCL/TAL-1 is essential for the differentiation of the HB to the specialised haemogenic endothelium. $SCL^{-/-}$ mouse ES cells fail to undergo myeloerythroid haematopoiesis in culture, $SCL^{-/-}$ mouse embryos die at the yolk sack stage, and introduction of SCL cDNA into SCL-/- cells has been shown to rescue myeloerythroid development (Porcher et al., 1996, Robb et al., 1996).

In previous studies difficulties in accessing the embryo prior to the establishment of blood made the characterisation of the putative HB challenging. However, differentiation of

embryonic stem cells into the haematopoietic lineages *in vitro* has offered an alternative approach to assessing this early precursor at the cellular and molecular level. ES cell *in vitro* differentiation cultures allow for the study of sequential steps from this early precursor to HSCs (Choi et al., 1998, Keller, 1995). Using such techniques, it was shown that ES cells differentiate into embryoid bodies which contain blood cell precursors with both primitive and definitive haematopoietic potential. Choi et al (Choi et al., 1998) demonstrated that transferring these EBs to liquid cultures containing growth factors known to be important in endothelial and haematopoietic cell growth leads to the generation of endothelial and haematopoietic cells, including colonies which gave rise to haematopoietic precursors (Choi et al., 1998). However, recent studies have questioned the role of the HB as a precursor and in fact indicate that primitive blood cells may also be derived from HECs located in the yolk sac (Padron-Barthe et al., 2014, Stefanska et al., 2017).

The HB differentiates to form a specialised haemogenic endothelium (HE) from which the HSCs later emerge. Early evidence for a HE was provided through histological observations in the pig embryo (Emmel, 1916, Jordan, 1916). Later, evidence was provided in chicken embryos through dye-labelling of vascular endothelial growth factor receptor 2 (FLK-1/KDR) expressing endothelial cells prior to haematopoietic development which was then passed onto emerging blood cells (Jaffredo et al., 1998). Transgenic markers in the mouse embryo were used to show that emerging HSCs localised to the endothelial layer of the dorsal aorta (de Bruijn et al., 2002). Lineage tracing studies using temporally restricted genetic labelling in mice confirmed that endothelial cells in the dorsal aorta give rise to emerging blood cells which then proceed to colonise the foetal liver and adult bone marrow

(Zovein et al., 2008). These findings have also been confirmed through *in vivo* imaging in mice (Boisset et al., 2010) and cell culture studies (Lancrin et al., 2009, Chen et al., 2011). HSCs were also shown by *in vivo* imaging to emerge from the aortic floor into the dorsal aortic lumen after an endothelium to haematopoietic transition in zebrafish (Bertrand et al., 2010, Kissa and Herbomel, 2010).

In culture, the HE can be split into an early HE cell type, haemogenic endothelium type-1 (HE1) defined as expressing cell markers KIT, TIE2 but not CD41 and a later HE cell type, haemogenic endothelium type-2 (HE2) defined as expressing cell markers KIT, TIE2 and CD41. Importantly, KIT is the receptor for stem cell factor (SCF), a molecule that is essential for haematopoietic progenitor (HP) growth (Huang et al., 1999). The *cadherin* 5 (CDH5) protein acts as a robust marker of haemogenic endothelium but is not required for the EHT process (Anderson et al., 2015). The HE1 cell type expresses transcription factors associated with an endothelial cell signature such as ETS, SOX, GATA, LMO2, MEIS, AP-1 and TEAD TFs. The HE2 cell type has upregulated RUNX1 which drives the expression of the CD41 surface marker (Mikkola et al., 2003).

1.4.3 Transcription factors driving haematopoietic specification

VEGF binding to the RTK FLK-1 results in signalling through the well characterised MAP kinase signalling pathways which terminate at the transcription factors ETS and SOX17, driving their activity. In response to VEGF directed signalling ETS and SOX17 factors establish and maintain NOTCH signalling and an overall endothelial signature is maintained.

The expression of *Sox17* marks the haemogenic endothelium and is required for the EHT to occur (Clarke et al., 2013). Loss of function assays in an endothelial-specific Cre recombinase mouse line crossed to a *Sox17* floxed line with a ROSA26Cre reporter demonstrated that *in vivo* endothelial genetic deletion of *Sox17* during EHT resulted in a decrease in the expression of *Notch1* (Lizama et al., 2015). Abrogating of *Sox17* expression has been shown to result in a decrease in the expression of *Notch4* and *Dll4* in freshly isolated endothelial cells from the AGM of mice (Corada et al., 2013). SOX17 directly binds *Runx1* and *Gata2* and represses haematopoietic cell fate (Lizama et al., 2015). Other members of the SoxF family (SOX7 and SOX18) also act to maintain an endothelial signature (Serrano et al., 2010). Increased NOTCH1 activation in E11.0 AGM explants has been shown to be capable of abrogating the observed EHT in *Sox17* mutants. *Notch1* loss of function in AGM explants resulted in an increase in the HE ratio and increased HSC production (Lizama et al., 2015).

TFs act to establish, maintain or repress cell type-specific gene expression programmes (Goode et al., 2016) and the activity of a large number of them is modulated by signalling. The Hippo-signalling responsive TEAD TFs are required for the efficient differentiation of the HB to HSCs. AP-1 factors, in response to MAPK signalling, modulate the balance between vascular smooth muscle and hemogenic cell fate. TEAD TFs, in response to Hippo signalling, interact with and are stabilised by YAP/TAZ proteins at target genes *cis*-regulatory elements. YAP/TAZ proteins are the transcriptional co-activators and major effectors of the Hippo signalling pathway (Goode et al., 2016, Obier et al., 2016) (MAPK, Hippo and NOTCH signalling pathways are discussed in detail in Introduction 1.4.6, 1.4.7

and 1.4.8).

Endothelial TFs in response to external signalling cues drive differentiation of HE1 to the HE2 cell type, resulting in the expression of GATA TFs such as *Gata2* and the low-level expression of the haematopoietic transcription factor *Runx1* (Figure 1.12). GATA TFs are highly conserved across vertebrates (Zaidan and Ottersbach, 2018). In vertebrates there are six TFs in the GATA family which are grouped into haematopoietic (GATA1, GATA2, GATA3) and endodermal (GATA4, GATA5, GATA6) subgroups (Katsumura et al., 2017). GATA2 is a master regulator of haematopoiesis. Deletion of *Gata2* in murine embryos results in embryonic death at E10.5 (Tsai et al., 1994). GATA2 drives the expression of *Runx1* through binding to its +23kb enhancer (Nottingham et al., 2007). GATA3 has also been shown to play an important role in HSC formation. Endothelial-specific deletion *Gata3* was found to significantly reduce HSC production (Zaidan et al., 2022).

RUNX1 is considered the master regulator of haematopoiesis and acts to establish a haematopoietic gene expression pattern while down regulating the endothelial programme. *Runx1* expression is considered essential for the endothelial-to-haematopoietic transition (EHT) (Lancrin et al., 2009, Eilken et al., 2009, Jaffredo et al., 1998, Chen et al., 2009) which is a highly conserved process that has been studied in *Xenopus*, zebrafish and mice and is relevant to human early haematopoiesis due to the endothelial origin of the human definitive blood system (Oatley et al., 2020).

Once expressed RUNX1 drives the EHT by binding to and repressing the transcription of

Sox17 and *Notch1* and leads to a relocation of SCL/TAL1 and FLI-1 to new binding sites (Lichtinger et al., 2012) while also upregulating the ETS family transcription factor PU.1 (Willcockson et al., 2019, Lichtinger et al., 2012, Lancrin et al., 2009). RUNX1 also upregulates the transcriptional repressor genes *Gfi1* and *Gfi1b* which are essential for the successful completion of the EHT and emergence of the HSC. They act by recruiting to endothelial target genes the histone demethylase LSD1, part of the coREST complex containing HDAC1 and HDAC2, thus epigenetically silencing the endothelial programme in the HE (Thambyrajah et al., 2016, Lancrin et al., 2012). *Runx1* knockout in the mouse embryo causes lethality at ~E12.5 as no haematopoietic clusters are formed in the dorsal aorta (North et al., 1999). *Runx1* knockout in ESCs resulted in a complete loss of the EHT *in vitro* (Lancrin et al., 2009).

During the EHT at ~E10.5, HE2 cells undergo morphological changes and begin to bud upwards. Non-invasive, high resolution live imaging of zebrafish embryos have confirmed that the HSCs emerge from the aortic floor, from endothelial cells that begin to egress from the aortic ventral wall into the sub-aortic space before emergence of the HSC (Kissa and Herbomel, 2010). Furthermore, these studies indicated that *Runx1* expression was essential for the exit of cells from this process as anti-*Runx1* morpholino experiments resulted in a failure of HSC emergence (Kissa and Herbomel, 2010). It is likely that as a result of these morphological changes and subsequent budding upwards that NOTCH cell to cell signalling contacts and Hippo signalling contacts are lost resulting in the further loss of an endothelial TF signature. Finally, haematopoietic progenitors (HP) bud off from the HE cluster and are considered to be fully committed to a blood cell fate (Boisset et al., 2010).

A three-stage HSC maturation hierarchy has been proposed based on phenotypic characteristics, developmental stage and cellular markers (Rybtsov et al., 2014). Pro-HSC are the least mature form of HSC precursor expressing CD41, but not CD43 or CD45. Type I pre-HSC follow expressing CD41 and CD43 but not CD45 and finally type II pre-HSCs express CD41, CD43 and CD45. These populations have been characterised based on both the expression of these cell surface markers and the developmental time point at which they were isolated from the embryos by plating into OP9 co-aggregates. Recent single cell transcriptomics and antibody screening has also identified CD44 as a marker of the EHT, allowing for a more accurate classification of these early progenitor populations (Oatley et al., 2020). Once HSCs bud off the specialised HE clusters they enter the bloodstream and populate the foetal liver. HSC numbers in the foetal liver increase between E12 and E15, before decreasing again. It is thought that the decrease in HSCs in the foetal liver occurs due to the mobilisation of these HSCs to the spleen or the bone marrow (Christensen et al., 2004).

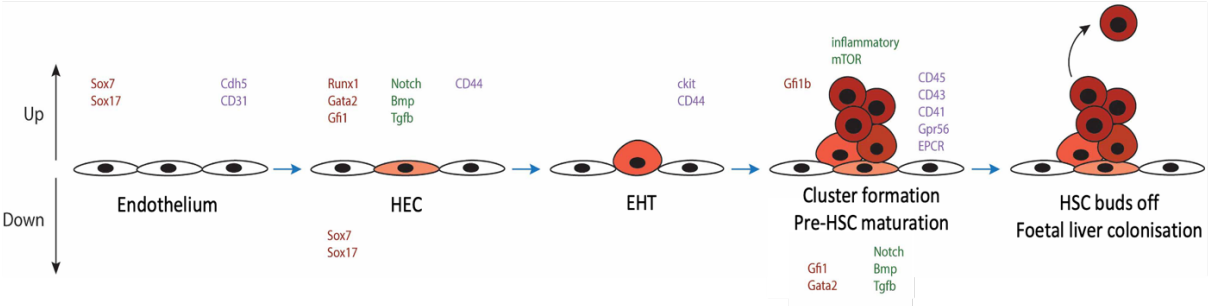


Figure 1.12- The regulation of the stepwise progression of endothelial cells to haematopoietic progenitors.
This diagram shows the stepwise progression of cell fate commitment from endothelial cells to haematopoietic progenitors in the dorsal aorta of the mouse embryo. The morphological changes which take place are shown. In response to upregulation of haematopoietic transcription factors in endothelial cells HECs form (marked red). As HECs

commit to the EHT they change shape becoming rounder and budding upwards into the aortic lumen resulting in the breaking of cell-to-cell tight junctions. HECs differentiate into pro-HSC and then Pre-HSC within the clusters before their maturation to HSCs. HSCs bud off and enter the circulation leading to foetal liver colonisation. Key transcription factors involved in the EHT are shown in red while key signalling pathways are shown in green. The surface markers used to classify each stage are shown in purple. The up and down arrow on the left of the image indicate whether a gene or signalling pathway is up or down regulated, respectively. Image from (Ottersbach, 2019). (Permission to reproduce this figure has been provided under CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).)

1.4.4 Signalling pathways direct haematopoietic specification

Early haematopoiesis is controlled by a complex interplay of stage specific TFs which act on their cognate *cis*-regulatory elements within a dynamic transcriptional network. The roles of the transcriptional regulators of haematopoiesis are beginning to be understood. It is still not clear how extrinsic signalling from growth factors direct transcription factor activity in a stage-specific manner (Obier et al., 2016). To have greater control over the differentiation of HSCs to blood precursors for downstream applications, a better understanding of how signalling factors and networks is required. Cytokines such as stem cell factor (SCF), bone morphogenic proteins (BMPs) and fibroblast growth factors (FGFs) regulate developmental haematopoiesis by binding their complementary receptors and inducing a signalling cascade which act on inducible TFs (Zhang and Lodish, 2008).

Signal transduction into the nucleus from the cell surface membrane requires surface molecules such as receptor tyrosine kinases (RTKs). Signals are conducted from RTKs through signalling cascades, such as a phosphorylation cascade, to a transcriptional factor.

These TFs can be expressed in multiple cell types and can colocalise with other inducible factors to alter the probability of transcriptional activity (Hill and Treisman, 1995, Pope and Medzhitov, 2018). Signalling pathways are challenging to study as they are highly cell type specific and there is significant cross talk between them. Down regulation or loss of signalling through one pathway can often be compensated by another. However, the signalling-responsive regulatory elements in the genome are invariant. Therefore, there is a significant need to study the function of signalling-responsive *cis*-regulatory elements and their interacting factors to better understand how stepwise lineage commitment is controlled during development.

Many signalling pathways transmit signals from cytokine-RTKs binding events, cell to cell signalling and sheer stress to TFs directing their activity during mouse ES cell differentiation to haematopoietic progenitors. BMP4, Notch, Wingless-type (Wnt) signalling, Sonic hedgehog (Shh) signalling, VEGF signalling, c-kit signalling, Hippo signalling and RAR/RXR mediated signalling all play a role in haematopoiesis. However, for the purpose of this study we will focus on how BMP4, MAPK, NOTCH and Hippo signalling act to regulate haematopoietic stem cell emergence from ES cells.

1.4.5 BMP4 signalling in early haematopoiesis

BMP4 signalling is part of the transforming growth factor-β (TGFβ) superfamily regulating essential cell fate decisions during early embryogenesis. BMPs signal through Type I and Type II receptors which form the only known family of transmembrane Ser-Thr kinases

(Figure 1.13). In response to ligand binding pre-formed dimers of the type II receptors and dimers of the type I receptors form heterotetrameric active receptor complexes (Schmierer and Hill, 2007). The type II receptor phosphorylates the type 1 receptor enabling the recruitment of and phosphorylation of the receptor regulated SMAD TF family. Phosphorylated SMADs then bind with cofactors to regulate gene expression (Schmierer and Hill, 2007). BMP is required during gastrulation and for mesoderm specification and has also been implicated in HSC emergence. BMP4 is expressed by the mesenchyme underlying intra-aortic haematopoietic clusters at the AGM within the dorsal aorta (Marshall et al., 2000, Durand et al., 2007). In zebrafish anti-sense knockdown of Bmp4 causes a loss of HSC emergence (Wilkinson et al., 2009). The mesenchyme signals in a paracrine fashion to the HE. BMP signalling acts on the TFs SMAD1 and SMAD5 which are expressed around the sites of HSC emergence (Wilkinson et al., 2009).

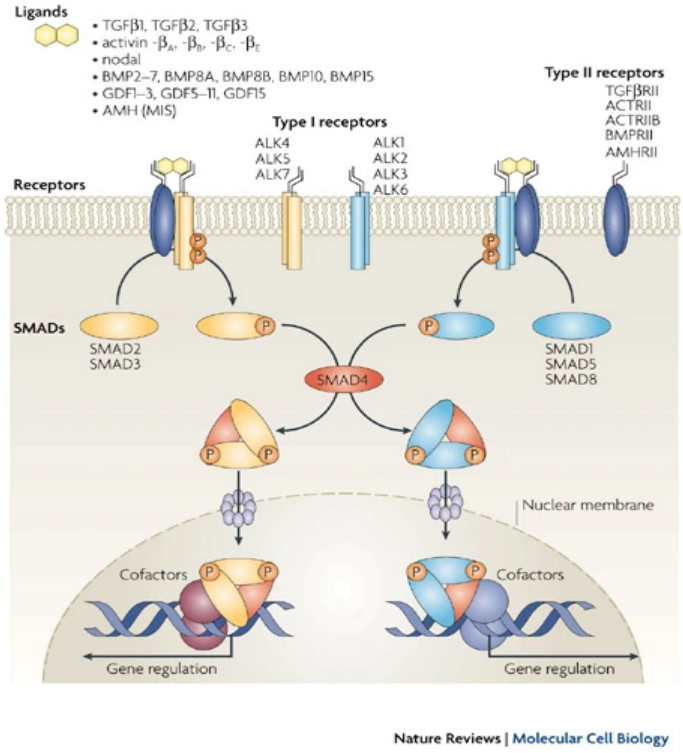


Figure 1.13- Core pathways of mammalian TGFβ-SMAD signalling.
Ligands and receptors of the TGFβ signalling pathway shown. Binding of ligands to the receptor causes Type I and Type II receptors to form heterotetrameric active receptor complexes and phosphorylation of SMAD proteins. SMAD proteins bind with cofactors to regulate gene expression. Image from (Schmierer and Hill, 2007). (Permission to reproduce this figure has been granted by Springer Nature.)

1.4.6 MAPK signalling in early haematopoiesis

VEGF signalling is required after gastrulation and axis formation. It is required by haemangioblasts and endothelial cells. The VEGF receptor FLK-1 (KDR) is expressed by haemangioblasts and endothelial cells in the dorsal aorta (Ziegler et al., 1999). VEGF regulates the activity of SoxF TFs (SOX17, SOX7, SOX18), FLI1, ETS, AP-1 and TEAD factors through the mitogen-activated protein kinases (MAPKs) signalling cascade.

MAPKs are a well-studied highly conserved set of enzymes that carry signals from RTKs to

intracellular targets (Figure 1.14). At least four MAPK families expressed in mammalian cells have been characterised; the classical MAPK/extracellular signal-regulated kinase known as ERK 1-2, C-JUN N-terminal kinase/stress-activated protein kinase (JNK/SAPK), the p38 kinase and ERK5 kinase (Chang and Karin, 2001, Widmann et al., 1999). Each member of the MAPKs signalling cascade contains three tiers consisting of the MAP kinase (MAPK), MAPK kinase (MAPKK) and the MAPKK kinase (MAPKKK). Signals from STE20 kinases or small GTP-binding proteins activate the MAPKs which in turn activate MAPK kinases. MEK1/2 activates ERK1/2, MK3/6 activates p38, MKK4/7 (JNKK1/2) activates JNKs and MEK5 activates ERK5. However, each MAPKK can be activated by more than one MAPKKK, increasing the complexity of MAPK signalling (Widmann et al., 1999). In mammalian cells MAPK signalling pathways relay information from the cell surface to the nucleus to induce a range of responses including cellular proliferation, differentiation, development, inflammatory responses and apoptosis (Chang and Karin, 2001).

Knockout of *Jun* in mice is lethal around the onset of HSC emergence around E13.0 with the lethality of *Jun* mutant foetuses reflecting the wide ranging roles of *Jun* in the development of key structures such as the neural crest and the maintenance of hepatic haematopoiesis (Eferl et al., 1999). Lee and colleagues demonstrated by overexpression of *junD* and c-*fos* and loss-of-function experiments using morpholino *junD* that the AP-1 TF family plays an important role in *Xenopus* haematopoiesis. They highlighted that JUND of AP-1$^{\text{JunD/c-FOS}}$ is required for BMP-4 induced haematopoiesis and that BMP4 regulated JunD at both the transcriptional and post-translational modification levels (Lee et al., 2012). In zebrafish, it was shown that the transcriptional co-repressor NCoR is required for haematopoietic stem

cell development, as *Ncor2* represses *fos-vegfd* signalling. Knockdown of *ncor2* resulted in an upregulation of *fos* transcription through altered acetylation in the *fos* promotor region, followed by enhanced VEGFD signalling. Increased Vegfd signalling induced Notch signalling and promoted arterial endothelial fate over haemogenic endothelial specification (Wei et al., 2014). Epoxyeicosatrienoic acids, a family of lipids, have been shown to enhance haematopoietic stem cell progenitor cell engraftment through activation of AP-1 and the R*unx1* transcription program (Li et al., 2015). Finally, FOS, along with GATA2, GFI1B and ETV6, was identified as an essential factor in the reprogramming of mouse embryonic fibroblasts (MEFs) to endothelial-like precursor cells and subsequently cells with an appearance of haematopoietic cells (Pereira et al., 2013).

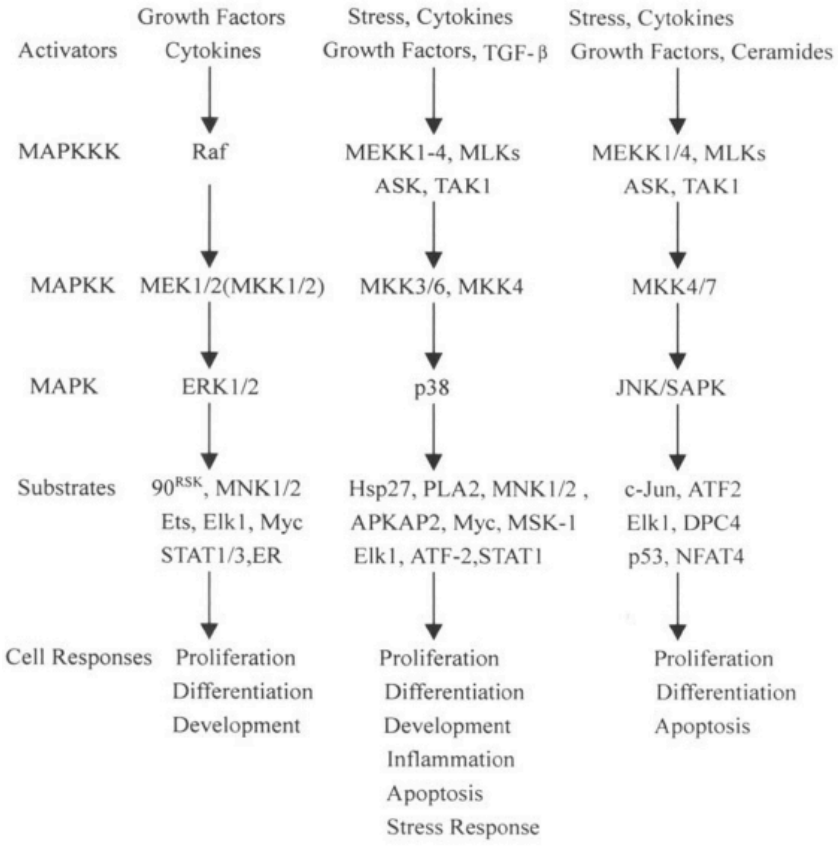


Figure 1.14- Diagram of the major MAP kinase cascades in mammalian cells.
Diagram shows the activators such as cytokines inducing the activity of three tiers of kinases which signal to substrates (including TFs) which in turn induce a cellular response.

Image from (Zhang and Liu, 2002). (Permission to reproduce this figure has been granted by Springer Nature.)

1.4.7 NOTCH signalling in early haematopoiesis

The Notch signalling pathway is conserved across species and is involved in a range of developmental processes and mediates cell to cell signalling. Canonical Notch signalling is at the heart of the complex mechanisms which drive definitive HSC production. In mammals there are four different Notch receptors: NOTCH1 to 4. Notch ligands are classified into two families, the Jagged and the Delta families. In mammals JAG1 and JAG2 are the two Jagged ligands while DLL1, DLL3 and DLL4 are the Delta ligands (Bray, 2016). These ligands are transmembrane proteins expressed by neighbouring cells. Both families of ligands induce Notch signalling, however, Notch1-Dll4 interactions induce high levels of Notch signalling while Notch1-Jag1 induce low levels of Notch signalling. Endothelial cell fate is determined through this high or low level of Notch signalling mechanism (Lomeli and Castillo-Castellanos, 2020).

In the HE, VEGF binding to the FLK-1 receptor induces Notch1 and Notch4 and Dll4 expression through ERK/MAPK. Interestingly, *Forkhead box c proteins* FOXC1 and 2 in response to VEGF-mediated PI3K and ERK/MAPK pathways have been shown to be transcriptional regulators of the repressive TF *Hey2* and induce the expression of *Dll4* (Hayashi and Kume, 2008, Seo et al., 2006). The TF SOX17 is a positive regulator of Notch signalling in HE (Clarke et al., 2013). After specification SOX17 mediates repression of *Runx1*

and *Gata2* thereby impeding the transition to hematopoietic progenitors. This highlights the arrangement of VEGF and Notch signalling into a VEGF-Notch signalling axis which establishes and maintains the haemogenic endothelium.

Gata2 is one of the downstream transcriptional targets of NOTCH1. Guiu et al (Guiu et al., 2013) presented that the human and mouse *Gata2* gene contains motifs for the Notch effector RBPJ and that mutation of these sites result in loss of *Gata2* reporter expression in transgenic embryos. Analysis of the mouse *Gata2* locus revealed motifs for the HES transcriptional repressors which are Notch signalling responsive (this family includes HES-1, HES-3, HES-5, HES-7, HRT-1 and HRT2) (Guiu et al., 2013). Mice embryos deficient for *Hes1* and *Hes5* have an intact arterial program but overproduce non-functional haematopoietic progenitors. Mutation of the HES binding sites results in an increase in *Gata2* expression in haematopoietic progenitors. They proposed a model by which Notch signalling activates *Gata2* and *Hes* genes, which continues until HES proteins reach a critical threshold resulting in the repression of the *Gata2* promoter (Guiu et al., 2013).

1.4.8 Hippo signalling in early haematopoiesis

Hippo signalling is a conserved signalling pathway which is required for the proper regulation of haematopoiesis. The effectors of hippo signalling are YES-Associated Protein (YAP) and transcriptional co-activator with PDZ-binding motif (TAZ). YAP and TAZ in response to Hippo signalling can bind with TEAD1-4 TFs responding to VEGF signalling, stabilising TEAD TFs resulting in target gene activation (Wang et al., 2017) (Figure 1.15).

Dysregulation of the Hippo pathway can lead to aberrant cell growth and neoplasia. The Hippo pathway contains a kinase cascade, transcription coactivators and DNA-binding proteins, in total the network contains >30 components. The pathway has a variety of inputs including cell-cell contact, cell polarity signalling, actin cytoskeleton and signals from cytokines which act via G-protein-coupled receptors (Meng et al., 2016). The core of the Hippo pathway involves a kinase cascade where the mammalian Ste20-like kinases 1/2 (MST1/2) activate through phosphorylation the large tumour suppressor 1/2 (LATS1/2). This action inhibits the action of the transcriptional coactivators YAP and TAZ (Meng et al., 2016). Once active YAP and TAZ translocate into the nucleus and bind to and stabilise the TEAD family of TFs, ultimately inducing genes involved in cell proliferation, survival and migration (Meng et al., 2016). The MST1/2 activation loop within the core of the Hippo pathway can be activated through phosphorylation by TAO kinases (TAOK1/2/3) which leads to the activation of MST1/2 (Boggiano et al., 2011).

Praskova and colleagues (Praskova et al., 2004) submitted evidence that MST1/2 loop activation can be achieved independent of TAO kinases via MST1/2 autophosphorylation (Praskova et al., 2004). In support of this notion the activation of loop phosphorylation has been shown to be enhanced by MST1/2 dimerisation (Glantschnig et al., 2002). Both SAV1 and MOB1A/B are phosphorylated by activated MST1/2 and act to stabilise the recruitment and phosphorylation of LATS1/2 at their hydrophobic motifs. NF2/Merlin directly interacts with LATS1/2 facilitating its phosphorylation by MST1/2-SAV1 complex (Yin et al., 2013). LATS1/2 subsequently undergoes autophosphorylation, becomes activated and

phosphorylate and inactivate YAP and TAZ (Chan et al., 2005). Two groups of the MAP4Ks (mitogen-activated protein kinase kinase kinase kinase), MAP4K1/2/3/5 and MAP4K4/6/7 directly phosphorylate LATS1/2 at their hydrophobic motifs which induces LATS1/2 activation (Zheng et al., 2015). Once YAP and TAZ translocate into the nucleus, they modulate gene expression through the transcription factors TEAD1-4 (Zhao et al., 2008). Within the nucleus TEAD can bind to VGLL4 which functions as a transcriptional repressor. YAP/TAZ interacts with TEAD1-4 dissociating VGLL4 from TEAD1-4 leading to TEAD driven gene transcription (Koontz et al., 2013).

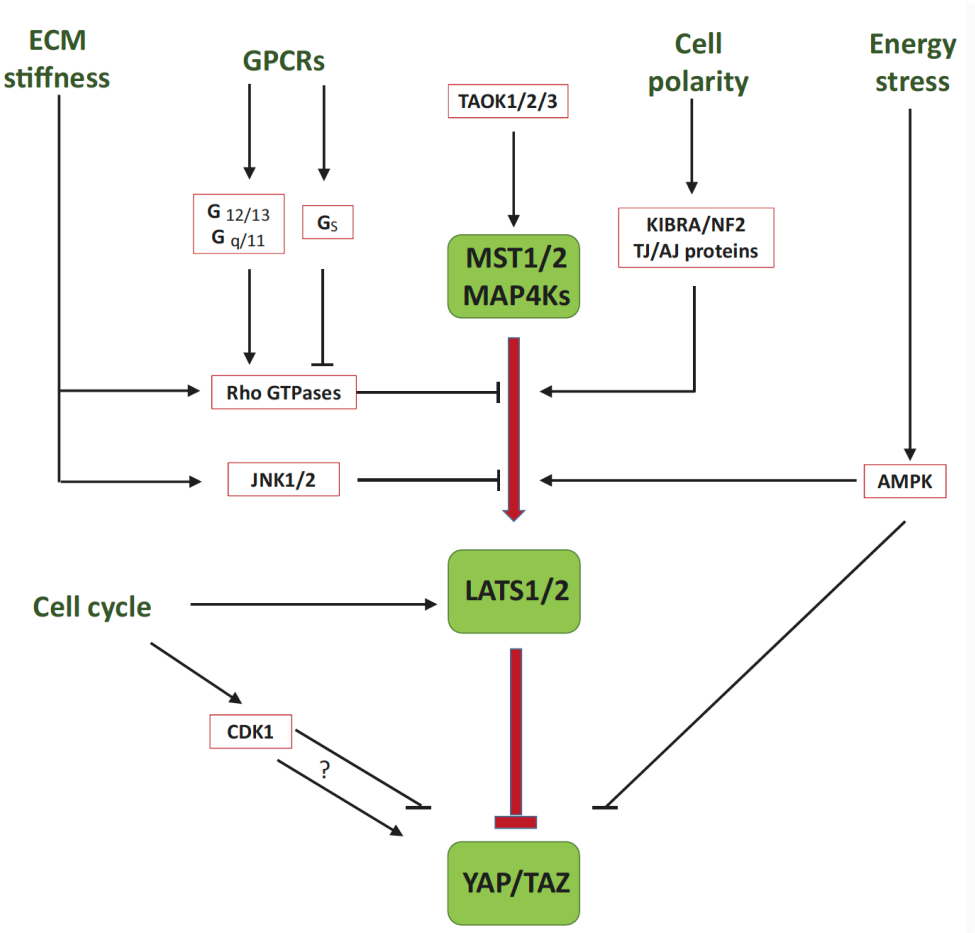


Figure 1.15- Diagram showing the regulation of the Hippo pathway.
Key signalling sources and mediators shown. Image from (Meng et al., 2016). (Permission

to reproduce this figure has been provided under CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).)

1.4.9 Embryonic stem cell *in vitro* differentiation

Embryonic stem (ES) cells are pluripotent cells isolated form the inner cell mass of blastocyst-stage embryos (Martin, 1981, Evans and Kaufman, 1981). These cells are remarkable for two important reasons. Firstly, ES cells can be maintained and expanded as pure populations of undifferentiated cells for long periods of time through well-established culture techniques. Secondly, they are pluripotent and so able to differentiate into every cell type in the body. The pluripotency of ES cells was highlighted in 1984 through chimeric studies in mice where their injection into host blastocysts resulted in live-born animals which were overtly chimeric (Bradley et al., 1984). ES cells also retain their capacity to differentiate into a range of cell types in culture (Keller, 1995, Smith, 2001). Human ES cells were isolated in 1998 and were shown to still have developmental potential to all three embryonic germ layers after 4 to 5 months of undifferentiated proliferation (Thomson et al., 1998).

The ability to culture and generate multiple lineages *in vitro*, including the haematopoietic lineages, from a single cell type is of great value for study. To conduct similar studies in the mouse embryo can be challenging and expensive and is impossible in the human embryo. ES cell differentiation systems may also provide an invaluable unlimited source of cells for cell and tissue transplantation for the treatment of a range of diseases. Both serum and

serum free *in vitro* differentiation systems have been developed to culture and differentiate ES cells and are discussed below.

1.4.10 Serum containing *in vitro* differentiation systems

To robustly study the mechanisms by which mouse ESCs commit to the haematopoietic lineage, cell culture techniques have been developed which utilise foetal calf serum (FCS). FCS contains a complex array of key constituents required for ESC differentiation to mature progenitor cells. (Choi et al., 1998, Nishikawa et al., 1998). Serum is rich in a large number of often unknown signalling factors and nutrient supplements, which make studying the effect of how each factor individually impacts on growth difficult. Mouse ES cells are cultured and maintained by co-culture with mouse embryonic fibroblast (MEFS) cells (Evans and Kaufman, 1981). Maintenance of the stem-cell phenotype *in* vitro requires a soluble factor differentiation inhibitory activity (DIA) such as leukaemia inhibitory factor (LIF). Recombinant LIF acts as a substitute for DIA and maintains ES cell lines totipotent potential (Williams et al., 1988). LIF acts through receptor complexes containing the signal transducer GP130 which acts on STAT3 (Niwa et al., 1998).

When mouse ES cells are cultured in the absence of factors that maintain their pluripotency and are plated into specific conditions, they differentiate to derivatives of the three embryonic germ layers: mesoderm, endoderm and ectoderm. Broadly three different approaches exist to differentiate ES cells to the haematopoietic lineages in serum each with their own advantages and disadvantages. (i) ES cells are plated into suspension and left to

aggregate and form three-dimensional colonies called embryoid bodies (EBs). (ii) ES cells are cultured on stromal cells such as OP9 cells and differentiation occurs through contacts with these cells. (iii) ES cells are differentiated in a single layer on an extracellular matrix of proteins such as type IV collagen (Keller, 2005, Nishikawa et al., 1998). Culturing of EBs in serum leads to the expression of *Flk-1* classifying them as HB. Sorting for the HB and then culturing them in FCS in the presence of additional VEGF gives rise to HE and HP cells (Keller, 2005).

1.4.11 Serum free *in vitro* differentiation systems

A number of serum free culturing techniques have been developed for the differentiation of human ESCs (Tian et al., 2004, Zambidis et al., 2006, Ng et al., 2008, Pick et al., 2013) and mouse ESCs to haematopoietic progenitors (Adelman et al., 2002, Pearson et al., 2008). These differentiation methods rely on the step wise addition of growth factors (cytokines) to drive lineage commitment from one cell type to another. Pearson and colleagues (Pearson et al., 2015) demonstrated that the sequential addition of bone morphogenic protein (BMP4), basic fibroblast growth factor (bFGF), Activin-A and vascular endothelial growth factor (VEGF) was enough drive the differentiation of mouse ESCs to haematopoietic progenitors. BMP4 induced the formation of the mesoderm, addition of Activin-A and bFGF was enough to drive the expression of *Flk-1* (*Kdr*) in HB. FLK-1+ HB cells were then sorted using anti-FLK-1 biotin beads. Finally, addition of VEGF to cultures of FLK-1+ HB cells resulted in the development of the adherent haemogenic endothelium and some floating haematopoietic progenitors (Pearson et al., 2015). The system therefore provides a controlled differentiation method with a reduced number of signalling factors

allowing for the study of how specific factors induce changes in gene expression and the chromatin landscape during the stepwise differentiation of ESCs to haematopoietic progenitor cells. As the cytokine factors present in the serum free I.V.D systems is controlled, these systems allow for the precise study of how specific cytokine signalling factors drive cell type-specific TF activity during ES cell differentiation to haematopoietic progenitors.

1.4.12 Dynamic gene regulatory networks direct haematopoietic specification

Cis-regulatory elements found across the genome encode the blueprint for the regulation of developmental gene expression (Yue et al., 2014). These elements are distributed across large distances and are controlled by TFs which bind in response to a multitude of intrinsic and extrinsic signals. TFs and TF encoding genes and their targets form gene regulatory networks (GRNs) which define the identity of a cell (Cockerill, 2011, Edginton-White and Bonifer, 2022). In response to signalling and developmental cues, TFs recruit chromatin remodellers/modifiers and interact with each other through looping to control gene expression (Field and Adelman, 2020). To understand how GRNs change from one developmental stage to another it is necessary to (i) identify cell type-specific *cis*-regulatory elements, (ii) identify the TFs which bind to them and (iii) to understand how they respond to signalling. The use of genome wide multi-omics approaches allows for the collection and then integration of data which elucidate these processes.

Previously, Goode et al., (Goode et al., 2016) generated global multi-omics data on TF

binding, gene expression, histone modifications and open chromatin regions during six sequential stages of haematopoietic differentiation from mouse ES cells to terminally differentiated macrophages. This work defined a dynamic TF GRNs driving gene expression at six sequential developmental stages (embryonic stem cells (ESCs), mesoderm, haemangioblasts (HB), haemogenic endothelium (HE), haematopoietic progenitors (HP) and macrophages) (Figure 1.16).

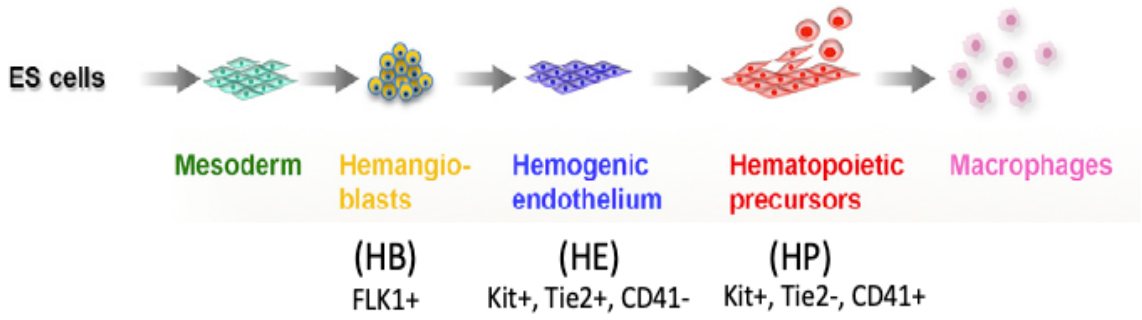


Figure 1.16- Differentiation pathway of the six sequential stages studied by Goode et al. ESCs to macrophages. ES cells initially differentiate into the mesoderm which contains haemangioblast (HB) cells expressing the cell marker *Flk1*. HB differentiate into the specialised haemogenic endothelium (HE) expressing the cell markers *Kit* and *Tie2*. HE differentiates into haematopoietic progenitor cells (HP) expressing *Kit* and *Cd41*. HP differentiates into macrophages. Adapted from (Goode et al, 2016). (Permission to reproduce this figure has been provided under CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).)

The work revealed the nature and activity of regulatory elements driving the differential gene expression required for haematopoietic lineage commitment in serum *in vitro* differentiation cultures and yielded a comprehensive dynamic core regulatory network model for haematopoietic specification (Figure 1.17).

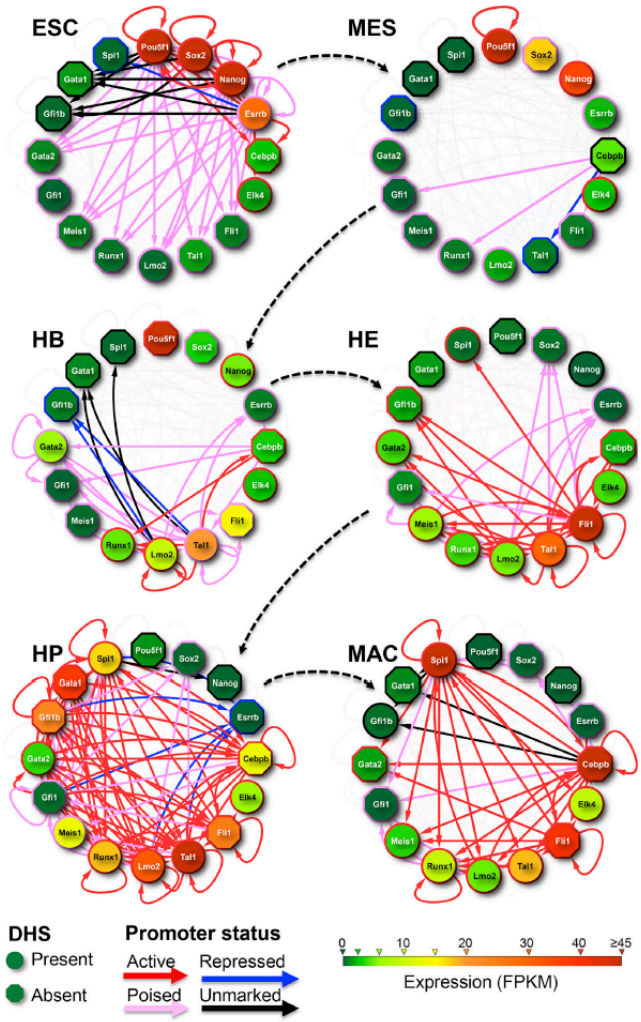


Figure 1.17- Model of dynamic gene regulatory network driving haematopoietic specification in the serum I.V.D system.

For each developmental stage the 16 TFs measured in ChIP-seq experiments are shown as nodes in a gene regulatory network. Chromatin accessibility at each promoter is shown as an open/circular (DHS presence) or closed/octagonal (DHS absence), and the border colour of the node corresponds to the promoter state. Arrows show binding event of a TF (source) at gene loci encoding all TFs (target). The arrow colour marks the promoter state of the target TF encoding gene as indicated. No arrow emerging from a node shows absence of ChIP data. Image from (Goode et al., 2016). (Permission to reproduce this figure has been provided under CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).)

The relative importance of TFs with respect to their importance for the maintenance of specific cell types is encoded in the binding motif composition of cell-type specific *cis*-regulatory elements (Neph et al., 2012). To identify transcriptional regulators during blood cell specification, chromatin accessibility data were used to perform a pairwise comparison of distal DNase hypersensitive sites (DHSs) from one cell type with all others as outlined in Figure 1.18a (Goode et al., 2016). For each set of DHSs unique to each cell type, a relative enrichment value was calculated for each sequence motif of interest and clustering analysis was performed against known motifs bound by TFs expressed in these cells (Figure 1.18b). The analysis confirmed known roles of TFs involved in blood specification at key developmental stages. For example, in ESCs pluripotency factor motifs cluster together, whilst the RUNX motif is enriched at the HP stage where the TF is critically required for the endothelial to haematopoietic transition (Lancrin et al., 2009).

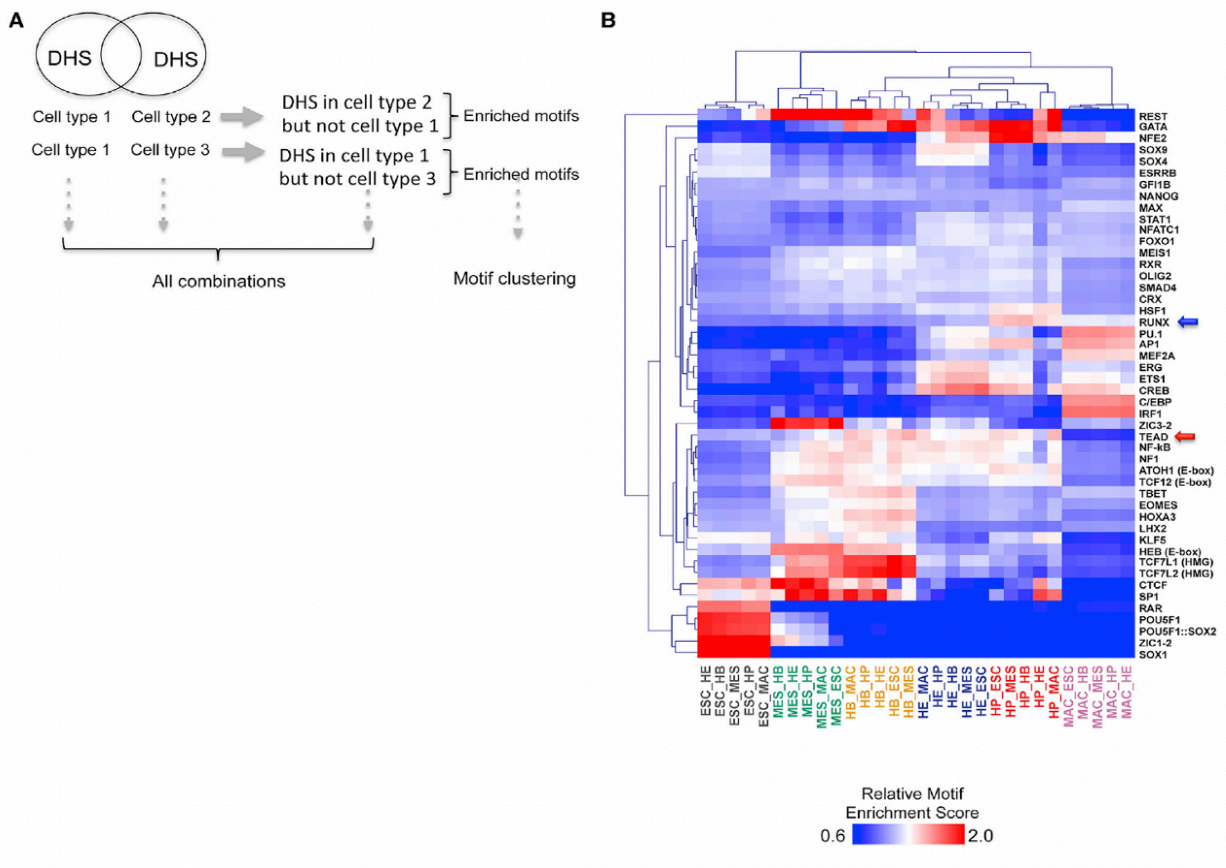


Figure 1.18- Chromatin and TF binding dynamics and identification of regulators in serum I.V.D

(a) Schematic of the methodology applied to generate pairwise motif clustering at distal DNase1 hypersensitive sites (DHS). (b) Heat map of relative motif enrichment (RMEA) scores for the motifs in DHS unique to a given cell type compared to of the other five cell types. RUNX and TEAD motifs are highlighted by blue and red arrows, respectively. Image from (Goode et al., 2016). (Permission to reproduce this figure has been provided under CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).)

The results were validated by comparing called enriched motifs with TF ChIP-seq peaks and thus form a valuable resource for determining transcriptional regulators driving stepwise lineage commitment from ESCs to terminally differentiated macrophages (Goode et al., 2016). Furthermore, functional studies motivated by the data identified stage-specific roles for factors such a TEAD/YAP that so far had not been known to be involved in the regulation of haematopoietic specification, such as TEAD/YAP or AP1. These studies

revealed that TEAD and AP-1 cooperatively bind to modulate the balance between vascular smooth muscle and haemogenic cell fate (Obier et al., 2016). The data also allowed for the bioinformatic prediction of a TEAD::AP-1 motif conformational arrangement separated by a 7 base spacer is *cis*-elements. The functional importance of spacing was shown by experiments employing a dominant negative FOS protein blocking all AP-1 binding to DNA as TEAD binding as measured by ChIP was also affected at such sites. This result demonstrates that the two factors co-operate, and such genomic elements integrate the two signalling pathways (Obier et al., 2016). The data also lead to studies investigating the role of the ubiquitous Sp1 and Sp3 TFs which showed that loss of these factors severely impacts haematopoietic progenitor production (Gilmour et al., 2019). These studies demonstrate the power of genome-wide multi-omics approaches for the study of GRNs during development.

1.4.13 Recapitulation of in vivo blood development in vitro by pluripotent stem cell differentiation

Embryonic stem cells are pluripotent cells isolated from the inner cell mass of blastocyst-stage embryos (Bradley et al., 1984, Martin, 1981, Evans and Kaufman, 1981). This cell type is of great interest because it can be maintained and expanded as pure populations of cells for long periods while maintaining the capacity to generate every cell in the body. This pluripotency was confirmed after their injection into host blastocysts which lead to production of all the tissues of the adult mouse (Bradley et al., 1984). The ability of ES cells to form different differentiated cell types *in vitro* has also been observed (Keller, 1995,

Smith, 2001) and as such these culture methods have been viewed not only as a model for the study of early development but as a potentially unlimited source of cells and tissues for transplantation for the treatment of a range of conditions. The development of blood from ES cells, for example, would be of great value. However, although transplantable progenitor cells were generated, so far *in vitro* systems have not generated HSCs capable of reconstituting the hematopoietic system long-term in transplantation assays (Cheng et al., 2013). The reason for this failure is thought to be a lack of the appropriate signals coming from a specific microenvironment.

Initial studies found that ES cells could be maintained on mouse embryonic feeder cells (Evans and Kaufman, 1981), which led to the identification of leukaemia inhibitory factor (LIF) as one of the feeder-cell-derived molecules which plays a role in the maintenance of these cells. Recombinant LIF was then shown to be able to replace feeder cell function and prevent the differentiation of ES cells in culture (Smith et al., 1988, Williams et al., 1988, Stewart et al., 1992).

Generally, all *in vitro* ES cell to HSC differentiation protocols start with ES cells being cultured in the presence of LIF maintaining their pluripotency. From here differentiation protocols can broadly be categorised into two groups (i) those being applied to study specific processes during ESC differentiation to blood (such as the serum I.V.D and restricted serum free I.V.D systems used in this project) and (ii) those which are truly attempting to produce HSCs for transplantation (Bruveris et al., 2020, Bruveris et al., 2021).

ES cells can differentiate into the three embryonic germ layers; mesoderm, endoderm and ectoderm (Keller, 1995). The blood lineages emerge through the mesoderm (Kinder et al., 1999). To initiate ES cell differentiation there are three general approaches. ES cells are plated into floating culture and left to aggregate to form floating three-dimensional colonies termed embryoid bodies (EBs) (Doetschman et al., 1985, Keller, 1995). The second method involves culturing ES cells on stromal cells (usually OP9 stromal cells) and differentiation takes place in contact with these cells (Nakano et al., 1994). The third method involves the culturing of ES cells in a monolayer on extracellular matrix proteins (Nishikawa et al., 1998).

Many culture systems involve the development of embryoid bodies which contain FLK-1+ (KDR) haemangioblast cells (such as the serum I.V.D and serum free I.V.D systems used in this project). As described previously it was believed that the HE and HSC emerges from a mesodermal progenitor cell called the haemangioblast (Haar and Ackerman, 1971). However, it is now believed that haemangioblast like cell types are only found *in vivo* in the yolk sac at around E7.5-E8.5 and does not form the specialised haemogenic endothelium found in the AGM of the dorsal aorta which produces the HSC (de Bruijn et al., 2000). In this regard HSCs which are derived *in vitro* from the haemangioblasts differ from HSCs produced in the mouse embryo. However, a number of key processes appear to be similar between the development of haemangioblast derived HSCs and those emerging from the AGM in the mouse around E10, mainly the occurrence of the endothelium-to-haematopoietic transition during which there is a switch from an endothelium type transcription factor gene regulatory network to a haematopoietic transcription factor

network through the upregulation of *Runx1* (Goode et al., 2016, Lancrin et al., 2012). It is this similarity with regards to the dependence on specific transcription factors which allows for *in vitro* differentiation systems to still be of value for studying the molecular details of the EHT.

The classification of cell types also differs slightly between *in vitro* and *in vivo* HSC development during the EHT. In culture, the specialised haemogenic endothelium is usually stained for the KIT receptor (KIT ligand receptor/stem cell factor receptor). However, in the mouse AGM only the HSC expresses the KIT receptor (Azzoni et al., 2018), indicating a slight difference between the expression of key surface markers *in vitro* compared to *in vivo. In vitro* systems also fail to fully mimic the extracellular matrix and the perivascular mesenchyme which underlies the specialised HE and forms the cellular niche within which HSCs develop and mature. Recent single cell transcriptomics and chromatin accessibility analyses of whole moue embryos and of the AGM of mouse embryos around the time of the EHT (Fadlullah et al., 2022, Argelaguet et al., 2019) will allow for the more accurate comparison of cell type, gene expression and chromatin accessibility seen *in vivo* compared to cells derived from *in vitro* culture systems.

2 AIMS AND OBJECTIVES

Previously our group generated global multi-omics data on TF binding, gene expression, histone modifications and open chromatin regions during six sequential stages of haematopoietic differentiation from mouse ES cells to terminally differentiated macrophages. This work defined a relay of dynamic GRNs driving gene expression at six sequential developmental stages (embryonic stem cells (ESCs), mesoderm, haemangioblasts (HB), haemogenic endothelium (HE), haematopoietic progenitors (HP) and macrophages) (Goode et al., 2016). The study revealed the nature and activity of regulatory elements driving the differential gene expression required for haematopoietic lineage commitment in serum *in vitro* differentiation cultures and yielded a comprehensive dynamic core regulatory network model for haematopoietic specification.

However, Goode et al (Goode et al., 2016) did not provide functional evidence for (i) which *cis*-regulatory elements have enhancer activity, (ii) how they are controlled at different developmental stages, (iii) which signalling factors (cytokines) control their activity and finally, (iv) which TFs mediate the signalling responsiveness. It is these questions which this study aims to answer by addressing the following objectives:

(i) Identify enhancer and promoter elements specifically active at defined stages of blood specification.

Based on the work by Dr Ben Edginton-White who has developed a high-throughput

method for enhancer identification at five stages of ES cell based haematopoietic differentiation, I will use bioinformatics methods to identify enhancer and promoter elements specifically active at defined stages of blood cell specification (ESCs, HB, HE1, HE2 and HP) capable of stimulating a minimal promoter (Edginton-White et al., 2022).

(ii) I will study how the enhancer elements are controlled at different developmental stages.

I will use bioinformatics methods to identify TFs controlling cell type-specific enhancer activity by identifying enriched transcription factor binding motifs and by investigating how such motifs co-localise. I will validate the developmental regulation of individual enhancer elements identified in our screen by creating ES cell lines expressing selected elements.

(iii) I will use a serum free *in vitro* differentiation system and modulated the cytokine composition to assess the signal responsiveness of identified enhancer elements in four cell types (HB, HE1, HE2 and HP) and identify cytokine responsive enhancers.

I will optimise the serum free system to drive the production of sufficient haematopoietic progenitors for downstream applications. I will modulate the cytokine composition of the culture system by omitting individual cytokines to assess the impact of omission on chromatin accessibility at enhancer elements, thus identifying cytokine specific responsive enhancer elements. I will used bioinformatics methods to identify transcription factor

binding motifs enriched or depleted in cytokine responsive enhancer elements.

(iv) For one cytokine, VEGF, I will study cytokine responsiveness of enhancers in more detail and identify the transcription factors mediating responsiveness.

I will validate cytokine responsiveness of individual enhancer elements identified in our screen by creating ES cell lines expressing selected elements. We will use normal and mutated version of cytokine-responsive enhancers to assess which key TFs involved in blood development respond to the presence and absence of cytokines.

3 MATERIALS AND METHODS

Many authors contributed to the success of this project. This section lays out the contributions of the authors. Dr Benjamin Edginton-White performed the genome wide enhancer screen. Dr Debbie Goode developed HM-1 ES cells containing fluorescence enhancer reporter constructs housing the wild type *Galnt1* enhancer and the *Galnt1* enhancer with mutated TEAD, GATA, EBOX and ETS sites. Dr Peter Keane processed the initial scRNA-seq data and assigned the clusters. Dr Salam Assi assigned enhancers to their gene by HiC analysis.

Alexander Maytum (AM), Dr Benjamin Edginton-White and Dr Jane Gilmour set up the serum free I.V.D system and AM optimised the serum I.V.D system used in this project. AM developed the HM-1 ES cell line containing a fluorescence enhancer reporter construct housing the *Galnt1* enhancer with a mutated AP-1 site. AM cultured the ES cell clones containing the fluorescence enhancer reporter constructs housing the WT *Galnt1* enhancer and the enhancers with mutated sites in the serum and serum free I.V.D systems with all cytokines and then with individual cytokines removed and assessed the enhancer activity by FACS. AM performed large serum free I.V.D cultures of ES cells and differentiated them into HB, HE1, HE2 and HP in the presence of all cytokines and then with one of BMP4, IL6, IL3 or VEGF withdrawn and assessed the effect on the proportion of cells gained from the cultures by FACS. AM stained these cell types with antibodies and then cells were sorted for by Dr Mary Clarke. Sorted cells were taken for ATAC-Seq and scRNA-Seq analysis. AM produced ATAC-Seq libraries and performed analysis of the subsequent ATAC-Seq data to generate lists of cytokine-responsive distal ATAC sites. AM intersected these lists with enhancer lists generated by Dr Benjamin Edginton-White to produce lists of signalling responsive enhancer elements. AM correlated enhancer activity against their corresponding gene expression pattern.

AM performed supervised transcription factor (TF) motif analysis and TF motif co-localisation analysis on differentially active enhancers from each cell stage in the differentiation pathway. AM performed *de novo* and supervised TF motif clustering analysis to identify the TFs mediating the stage-specific responsiveness of enhancers to cytokine signalling. AM selected enhancers assigned to *Sparc*, *Pxn*, *Hspg2*, *Cdh5*, *Dlk1* and two enhancers for *Runx1* which were predicted as being VEGF signalling responsive as a result of the previously mentioned cytokine responsive distal ATAC sites intersection with the results from the enhancer screen for validation as being VEGF signalling responsive. AM made ES cell lines containing the fluorescence enhancer reporter construct housing these enhancers and differentiated these cells in the serum free I.V.D system validating that they were VEGF signalling responsive. AM annotated these enhancer sequences highlighting a number of potentially signalling responsive TF motif sites. AM performed scRNA-Seq analysis of HE1, HE2 and HP cells derived from the serum free I.V.D system with and without VEGF and performed differential gene expression analysis, revealing the failure of *Runx1* upregulation and the failure of the downregulation of *Sox17* in HE2 and HP in the presence of VEGF. AM developed ES cell lines containing the fluorescence enhancer reporter construct housing the *Runx1* +23kb and +3.7kb enhancers wild-type sequences and then with mutated TF motifs and differentiated these ES cell clones in the serum I.V.D culture system and assessed the effect on enhancer activity by FACS.

3.1 Cell lines

Embryonic stem cell lines used in project were mouse HM-1 and A7 cells. HM-1 cells were previously isolated directly from HPRT-deficient strain 129 mice produced by blastocyst injection of E14tG2a cells (Magin et al., 1992). A7 cells from (Sroczynska et al., 2009) which contain a GFP reporter gene inserted into the *Runx1* locus by homologous recombination

under the control of the *Runx1* distal promoter and a truncated human CD4 cDNA introduced by homologous recombination downstream of the proximal 5' UTR. Embryonic stem cells were cultured on mouse embryonic fibroblast cells (MEFS) prior to differentiations (amsbio, #ASF-1201). Cells were periodically checked for mycoplasma infection and prior to genomic sequencing experiments by a luminescence-based assay (Lonza, #LT07-318).

3.2 Embryonic stem cell culture

Embryonic stem cells (ESCs) were cultivated on mitomycin C treated primary mouse embryonic fibroblasts (MEFs) in DMEM-ES (DMEM (Merck, #D5796), +15% heat-inactivated Foetal Bovine Serum (Hi-FBS) (Gibco), 1 x Penicillin-Streptomycin (Merch, #P4333), 1 x L-Glutamine (Merck, #G7513), 1mM Sodium pyruvate (Merck, #S8636), 1x non-essential amino acids(Merck, M7145), MTG 0.15mM (Merck), Hepes buffer 25mM (Merck, #H0887) and 1,000 U/ml ESGRO LIF (Merck, #ESG1107). $1x10^6$ MEFs were plated onto gelatinised (0.1% gelatine/PBS filter sterilized) 6-well plate (Falcon) and cultured for 24 hours, after which $1x10^5$ ESCs were plated per well in 3 ml DMEM-ES. The media was changed every 2 days and cells were split by brief trypsinisation (1x TrypLE^{TM} Express (ThermoFisher, #12605010)) at room temperature every 2/3 days depending on growth rate.

3.3 Serum I.V.D culture of ESCs

48 hours before embryoid body (EB) generation, ES cells were cultured on gelatinised (0.1% gelatine/PBS) 6 well plates at a density of $3x10^5$ cells per well in 3 ml DMEM-ES (1^{st} gelatine) to remove MEFS. 24 hours before EB generation, ES cells were split again onto gelatinised (0.1% gelatine/PBS) 6 well plates at a density of $3x10^5$ cells per well in 3 ml DMEM-ES (2^{nd} gelatine) to further remove MEFS.

To generate EBs, ES cells were cultured without LIF on bacteriological grade dishes selected for low adherence (TermoFisher, #501V). ES cells from the second gelatine treatment were trypsinised (TrypLE™ Express (ThermoFisher, #12605010)) and further cultivated in suspension at a concentration of 2.5 $x10^4$ cells per ml in I.V.D media (IMDM (Merck, #13390), 15% FCS, 1x Penicillin/Streptomycin (Merck, #P4333), 1 x L-Glutamine (Merck, #G7513), 0.15 mM MTG, 50μg/ml Ascorbic acid and 180 μg/ml Human transferrin (Merck, #T8158)) in a 15cm diameter dish (50ml media). The cultures were incubated for 3 days for EB development containing FLK-1+ haemangioblasts (HB).

To differentiate HB cells into haematopoietic progenitor cells, EBs were dissociated and Flk-1^+ cells were sorted. EBs were transferred to 50 ml tubes and cells allowed to settle. The supernatant was removed, and phosphate buffered saline (PBS)(Merck) was added to wash the pellet. 5 ml of trypsin (1x TrypLE™ Express (ThermoFisher, #12605010)) was added and left at room temperature for 3 minutes while the cells were gently pipetted up and down using a pipette to dissociate the pellet to a single cell suspension. Trypsin was inactivated

with an equal volume of IMDM +10% FCS, cells were then taken for counting before being centrifuged at 300xg for 5 minutes. Cells were resuspended to 10^7 cells per ml and 30ul was taken for FLK-1-PE staining, ISO-PE staining and for an unstained sample. Cells were stained with FLK1-PE antibody (1:200) (eBioscience) and ISO-PE antibody (1:200) (eBioscience). Cells were then incubated at 4°C for 15 minutes before being washed with 500ul MACS buffer (PBS (Merck), BSA (5%), EDTA (0.4%) (Invitrogen)), twice and resuspended in 500ul MACS buffer for flow cytometry. The percentage of FLK-1+ cells was assessed by flow cytometry over isotype stain. Remaining cells were incubated with mouse FLK-1 biotin antibody (1:200) (eBioscience, #13-5821-82) for 15 minutes at 4°C. Cells were washed with MACs buffer twice (centrifuged at 300xg for 5 minutes) and resuspended in MACS buffer at 90 µl/10^7 cells. The cells were then incubated with Anti-biotin Microbead (10 µl/10^7 cells) (Miltenyi Biotec, #130-090-485) for 15 minutes at 4°C before being washed with 10ml MACS buffer and spun down at 300xg for 5 minutes. The supernatant was then discarded, and the cells bound to the beads were then resuspended in 1ml of MACS buffer and pipetted into a sterile LS Column (Miltenyi Biotec, #130-042-401) attached to a magnet (Miltenyi Biotec). 500µl MACS buffer was added three times to wash the beads. The LS column was removed from the magnet, 500 µl MACS buffer was added to the LS column to collect FLK-1+ cells. The number of FLK-1+ cells were counted. FLK-1 positive cells were then plated in liquid Blast Media (IMDM (Merck, #13390), 10% FCS, 1 x Penicillin/Streptomycin (Merck, #P4333), 1 x L-Glutamine (Merck, G7513), 0.45mM MTG, 25µg/ml Ascorbic acid, 180 µg/ml Human transferrin (Merck, #T18158), 20% D4T Conditioned Media, 5ng/ml VEGF (PeproTech, #450-32), 10/ng/ml IL-6 (Peprotech, #216-16)) on gelatine coated 6 well plates (Falcon) at 1.0x10^5 cells per well. Cells were then left

to incubate at 37°C 5% CO_2 for 2.5 days before being harvested for cell sorting and FACS analysis.

Media containing floating progenitor cells was collected and spun down at 300xg for 5 minutes. Adherent blast cells were trypsinised (1x TrypLE^{TM} Express (ThermoFisher, #12605010)) for 5 minutes before an equal volume of Blast Media was added. Cells were spun down at 300xg for 5 minutes and the supernatant was discarded. The cells were then washed with 5ml MACS buffer and spun at 300xg for 5 minutes, the supernatant was discarded, and cells were resuspended in MACS buffer. For flow cytometry analysis and cell sorting, cells were divided into six FACS tubes: (1) an unstained sample, (2) a triple ISO stain containing ISO-APC (1:100) (Invitrogen), ISO-PE (1:200) (Invitrogen), ISO-APC (1:200) (Biolegend) antibodies, (3-5) Single stains for KIT-APC (1:200) (Invitrogen), Tie2-PE (1:100) (Invitrogen), CD41-PECY7 (1:200) (Invitrogen). (6) A triple stain containing KIT-APC (1:200) (Invitrogen), Tie2-PE (1:100) (Invitrogen), CD41-PECY7 (1:200) (Invitrogen).

Cells were incubated at 4°C for 15 minutes before being washed twice with 500µl MACS buffer. Cells were then assessed by FACS using CytoFLEX Flow Cytometer (Backman Coulter) and CytExpert Software (Beckman Coulter). For cell sorting, cells were sorted using BD FACS Area Sorter (BD) into 300µl MACS buffer and taken for downstream applications.

3.4 Serum free I.V.D culture

EBs were generated from ESCs by plating at 5.0x10^5 cells/ml in serum free (SF) media

(StemPro-34 Serum free media (Gibco), StemPro-24 Supplement (Gibco) (final 2mM), 30mg/ml Human Transferrin (Roche), ascorbic acid 0.5mM (Sigma), MTG (monothioglycerol) (Sigma) final $4.5x10^{-4}$ M) into petri-grade dishes (Sterilin). BMP4 (PeproTech) was added to a concentration of 5ng/ml. Cultures were left to incubate at 37°C and 5% CO_2 for 60 hours before bFGF (PeproTech) and Activin-A (PeproTech) were added to a concentration of 5ng/ml each. The cells were left to incubate for 16 hours at 37°C 5% CO_2 and then sorted for FLK-1+ cells as in Methods 3.3. Trypsin Inhibitor (ThermoFisher) was added for 5 minutes to neutralise Trypsin. FLK-1+ HB were plated in serum free media on 0.1% gelatine coated plates or for larger cultures flasks at $2.25x10^4$ cells per cm. BMP4 (5ng/ml) (PeproTech), Activin-A (5ng/ml) (PeproTech) and bFGF (5ng/ml) (PeproTech) were added to a concentration of 5ng/ml for 16 hours. SF media was then removed, and the blast culture was washed with PBS and fresh SF media was added containing BMP4 (5ng/ml) (PeproTech), VEGF (5ng/ml) (PeproTech), TPO (5ng/ml) (PeproTech), SCF (100ng/ml) (PeproTech), IL6 (10ng/ml) (PeproTech) and IL3 (1ng/ml) (PeproTech). For cytokine withdrawal experiments, one of BMP4, VEGF, IL6 or IL3 was not added at this stage. Blast cultures were left to incubate at 37°C and 5% CO_2 for 72 hours before cells were harvested for cell sorting and flow cytometry analysis as in Methods 3.5. Inhibition of trypsin was achieved using Trypsin Inhibitor (ThermoFisher) at manufacturer's instructions.

3.5 Flow cytometry

HB cells were centrifuged at 300g for 5 minutes and resuspended in 50µl MACS buffer (PBS (Sigma), BSA (5%), EDTA (0.4%) (Invitrogen)) before being stained with FLK-PE 1:200

(Invitrogen) and incubated for 15 minutes at 4°C. Cells were then washed twice with 500µl MACS buffer centrifuged at 300g for 5 minutes and finally resuspended in 50µl for FACS analysis. Flow cytometry was performed on a CyAn ADP flow cytometer (Beckman Coulter) to assess the number of FLK-1+ HB. Summit 4.3 (Beckman Coulter) was used to analyse the data. For enhancer activity screening assays FLK-1+ HB numbers were assessed on CytoFLEX (Beckman Coulter). Blast culture cell types, HE1, HE2 and HP were assessed on a CytoFLEX (Beckman Coulter) using CytExpert Software (Beckman Coulter). Blast culture was stained for using KIT-APC 1:200 (Invitrogen), TIE2-PE 1:100 (Invitrogen) and CD41-PECY7 1:200 (Invitrogen) with the same process described for staining HB for FLK-1 assessment.

Antibody	**Dilution**	**Manufacturer and Catalogue Number**
Anti-mouse FLK-PE	1:100	Invitrogen #4303076
Anti-mouse TIE2-PE	1:200	Invitrogen #12-5987-82
Anti-mouse CD41-PECY7	1:100	Invitrogen #17-1171-82
Anti-mouse KIT-APC	1:100	Invitrogen #17-1171-82
Anti-mouse KIT Super Bright 436	1:100	Invitrogen #67-1171-82
Rat ISO APC	1:100	Biolegend #400512
Rat ISO PE	1:200	Invitrogen #12-4321-82
Rat ISO PECY7	1:100	Invitrogen #25-4321-82
Anti-mouse FLK-Bio	1:100	Invitrogen #13-5821-85

Table 3.1- Antibodies used for FACS and cell sorting in this study.

3.6 Cell sorting

Cells for ATAC and sc-RNA-seq experiments were sorted by the Technology Hub Cell Sorting Facility University of Birmingham on a BD FACSAria II (BD) or by Dr Mary Clarke on a MoFlo (Beckman Coulter). First, cells were collected and centrifuged at 300g for 5 minutes. Cells were then filtered using a EASYstrainer filter (Greiner, #542070) and then resuspended in 1ml MACS buffer (PBS (Merck), BSA (5%), EDTA (0.4%) (Invitrogen)) for every $5x10^6$. Cells were then stained with FLK-PE 1:200 (Invitrogen) for HB and KIT-APC 1:200 (Invitrogen), TIE2-PE 1:100 (Invitrogen) and CD41-PECY7 1:200 (Invitrogen) for blast culture cells for 15 minutes at 4°C. Cells were then washed twice with 5ml MACS buffer centrifuged at 300g for 5 minutes before being resuspended in 3ml and taken for cell sorting and downstream applications.

3.7 Assay for Accessible Chromatin- ATAC

ATAC-Seq was performed as described in (Buenrostro et al., 2015) with slight modifications. 5000-50,000 HB, HE1, HE2 and HP cells were centrifuged at 300g for 5 minutes at 4°C. The supernatant was discarded, and the pellet was kept on ice. The cells were gently resuspended in transposition reaction mix (2xTD Buffer 25μL (1x), Transposase 2.5μL (100mM) (Illumina), PBS 16.5μL, 1% Digitonin (Sigma) (0.01%), Tween-20 0.5μL (0.1%), dH_2O 5μL) and incubated at 37°C for 30 minutes with gentle shaking. The samples were then immediately purified using a Qiagen Reaction Clean-up MinElute kit (Qiagen, #28206) as per manufacturers instruction. DNA was eluted in 20μL elution buffer (Qiagen). DNA was stored at -20°C or taken forward for PCR amplification. To amplify transposed DNA

fragments, the following was combined in a PCR tube: 10μL Transposed DNA, 10μL Nuclease Free (H_20) (ThermoFisher), 2.5μL 25μM Customised Nextera PCR primer 1, 2.5μL 25μM Customised Nextera PCR primer 2 (Barcoded) (Buenrostro et al., 2015), 25μL NEBNext High Fidelity 2xPCR Master Mix (New England Labs). The mixture was then taken for PCR amplification by thermocycler (T300 Thermocycler, Biometra) using the following cycles: (1) 72°C 5 minutes, (2) 98°C 30 seconds, (3) 98°C 10 seconds, (4) 63°C 30 seconds, (5) 72°C 1 minute, repeat steps 3-4 x4, hold at 4°C. To reduce GC and size bias in the PCR, the PCR reaction was monitored using qPCR to stop amplification prior to saturation. To achieve this 5μL of amplified DNA or water control was then added to 4.4μL nuclease free H_2O (ThermoFisher), 0.25μL 25μM customised Nextera PCR primer 1, 0.25μL 25μM customised Nextera PCR primer 2, 0.06μL 100x SYBR Green 1, 5μL NEBNext high-fidelity 2x PCR Master Mix (New England Biolabs). The reaction mix was then amplified by qPCR thermocycler (Applied Bio Systems, StepOnePlus™) under the following conditions for 25 cycles in total: (1) 96°C 30 seconds, (2) 98°C 10 seconds, (3) 63°C 30 seconds, (4) 72°C 60 seconds, (5) repeat steps 2-4 x24. The additional number of cycles for the remaining 45μL of PCR reaction was calculated by taking raw fluorescence values from the blue channel and calculating the number of cycles to achieve 25% amplification. The sample was run for the remaining number of cycles (X): (1) 98°C 30 seconds, (2) 98°C 10 seconds, (3) 63°C 30 seconds, (4) 72°C 60 seconds, (5) repeat steps 2-4 X times, (6) Hold at 4°C. Amplified libraries were then purified using Qiagen MinElute Reaction Clean-up Kit (Qiagen, #28206) as per manufacturer's instructions and the sample was eluted in nuclease free H_2O. For library clean-up, 1.2x volume of AMPure XP beads (Beckman Coulter) were added to the samples and mixed by gentle pipetting ten times. Samples were then placed on a magnet

until the liquid was clear, the supernatant was then removed, and samples were washed twice with 100μL fresh 80% EtOH. Samples were left to dry at room temperature with an open cap for 5 minutes. 20μL H_20 was added by pipetting 10x to elute the samples which were then incubated for 5 minutes at room temperature before being placed back on the magnetic rack, when the solution was clear the liquid proportion was transferred to fresh Eppendorf's and stored at -20°C. Quality control of libraries was performed using qPCR-based KAPA library Quantification Kit for Illumina Sequencing Platforms (Roche) as per manufacturer's instructions using a qPCR thermocycler (Applied Bio Systems, StepOnePlus™). 4μL of 1:1000, 1:2000 and 1:4000 diluted library was mixed with 6μL of KAPA master mix and qPCR run under the following conditions: 95°C for 5 minutes, 95°C for 30 seconds then 60°C for 45 seconds x35. DNA libraries and average library size were quantified using a High Sensitivity DNA assay (Agilent) on a Bioanalyser 2100 (Agilent). 1μL of library was mixed with 5μL high-sensitivity marker at room temperature and loaded onto the DNA bioanalyser chip. Gel-dye mix and DNA ladder were also added. Peak size was determined by comparing the DNA ladder and library quantification against the DNA marker.

3.8 Enhancer reporter library cloning

Conducted by Dr Benjamin Edginton-White. The enhancer reporter system designed by Wilkinson et al. (Wilkinson et al., 2013) was adapted to develop a genome-wide enhancer reporter assay. The enhancer reporter functions by inserting a fragment of interest upstream of a HSP68 minimal promoter and Venus-YFP reporter by Gateway cloning. The

reporter construct is then transfected into the HM-1 ES cell line which has a non-functional HPRT locus. Using HPRT homology arms the report cassette becomes integrated into the HPRT locus by homologous recombination, also repairing the locus, enabling selection of clones with successful recombination. This enhancer reporter system was modified and optimised for genome-wide screening as follows.

Genome-wide enhancer fragments for cloning into the reporter were generated using tn5 tagmented open-chromatin fragments based on the ATAC-Seq protocol (Buenrostro et al. 2015 & Corces et al. 2016). Cells were collected ($2.5x10^5$) from each differentiation stage (ES, HB, HE1, HE2, HP) as detailed in the serum I.V.D method (Methods 3.1). For each of the 5 populations of cells collected $50x10^3$ cells were then centrifuged at 300g for 5 minutes and following the ATAC-Seq protocol transposition mix was added (2x TD Buffer 25µl (Illumina), TN5 Transposase 2.5 µl (Illumina), PBS 16.5 µl, 1% Digitonin 0.5 µl (Promega), 10 % Tween-20 0.5 µl, H_2O 5 µl) and cells were resuspended by gentle pipetting.

The transposition reaction was incubated with agitation at 700 RPM, 37°C for 30 minutes. The 5 reactions were then combined and DNA was purified using the MinElute® Reaction Clean-up kit (Qiagen, #28206) as per manufacturer's instructions before being eluted in 26.5µl H_2O. 1/5 of the purified reaction from each cell stage was used to make ATAC-Seq libraries (see ATAC-Seq method). The remaining 4/5 had linkers added which incorporated AttB Gateway cloning sites which enables insertion into a Gateway doner vector (pDONR221) (ThermoFisher, #12536017). Custom designed linkers based on the tn5 transposase sequence were added by PCR in 2 steps. First, 5 cycles of PCR were performed

(Transposed DNA 20 µl, H2O 20 µl, 25 µM AttB tn5 Fwd primer 5 µl, 25 µM AttB tn5 Rev primer 5 µl, NEBNext Master Mix 50 µl (New England Biolabs)) under the conditions: 72°C 5 minutes, 98°C 30 seconds, 5 x 98°C 10 seconds, 63°C 30 seconds, 72°C 1 minute and hold at 4°C. The reaction was then divided and 5 µl was taken for a side qPCR reaction (0.125 µM AttB tn5 Fwd Primer, 0.125 µM AttB tn5 Rev Primer, 1 x SYBR Green I (New England Biolabs), NEBNext Master Mix (New England Biolabs)) under the conditions: 72°C 5 minutes, 98°C 30 seconds, 35 x 98°C 10 seconds, 63°C 30 seconds, 72°C 1 minute using a StepOnePlus™ Real-Time PCR System (ThermoFisher). The raw data from the cyan channel was used to calculate the number of cycles required for 25% amplification (X) and this number was then used to amplify the remaining material under the following conditions: 98°C 30 seconds, X cycles 98°C 10 seconds, 63°C 30 seconds, 72°C 1 minute and hold at 4°C.

To optimise for material from enhancer elements the resulting material was then size selected by electrophoresis on a 1.2% agarose gel with ethidium bromide, fragments between 350 and 650 bp were selected. Gel extraction was performed using the Qiagen MinElute® Gel Extraction Kit (Qiagen, #28604) following manufactures instructions. Next the size selected material was cloned into the pDONR™221 (ThermoFisher, #12536017) by BP Gateway reaction (100 fmol pDONR221, 100 fmol PCR material, 2 µl BP Clonase II (ThermoFisher, #11789020), made up to 20µl with TE buffer (pH 8.0)) and the reaction was incubated for 18 hours at 25°C. To end the reaction 0.2µg proteinase K (ThermoFisher, #EO0491) was added and the mix was incubated at 37°C for 10 minutes.

NEB 10-beta Electrocompetent (New England Biolabs) cells were transformed with the

plasmid by electroporation using a Bio-Rad GenePulser (Bio-rad) and program EC1 (2.0 kV, 200 Ohm, 25 µF). 500µl outgrowth media was added (New England Biolabs) and transformed cells were incubated 37°C, 250 rpm for 1 hour after which the bacteria were plated onto 150mm agar plates containing 50 µg/ml kanamycin and left for incubation overnight at 37°C. 10 ml of LB (ThermoFisher, #12795027) was added to each plate and the bacterial lawn was scraped off using a cell scraper. After Maxi-prep expansion of the cultures the plasmid DNA was extracted using the NucleoBond Xtra Midi EF kit (Machery-Nagel #740420.50) following Manufacturer's instructions and the resulting plasmid product was eluted in 500 µl H_2O. The chromatin fragments were then integrated into the enhancer reporter cassette in the pSKB-GW-Hsp68-Venus vector (Wilkinson et al., 2013) by Gateway LR reaction (150ng Entry clone (pDONR221) (ThermoFisher, #12536017), 150 ng pSKB, 2 µl LR Clonase II, made up to 10µl with TE buffer (pH 8.0)) and the reaction was incubated for 1 hour at 25°C. The reaction was terminated by addition of 0.2 µg proteinase K (ThermoFisher) and incubation at 37°C for 10 minutes. NEB 10-beta Electrocompetent cells were transformed with the material. 500µl outgrowth media was added (New England Biolabs) and transformed cells were incubated 37°C, 250 rpm for 1 hour. Transformed bacteria were then spread on agar plates containing 200 µg/ml ampicillin and left to incubate at 37°C overnight after which the bacteria were collected and the plasmid library extracted by Maxi-Prep using the NucleoBond Xtra Midiprep kit (Machery-Nagel #740420.50). The cloning was performed 10 times to maximise the diversity of fragments gained from each differentiation stage, the pDONR221 library material was combined before the LR clonase reaction.

3.9 Enhancer reporter library transfection

HM-1 cells were cultured as per Methods 3.2. HM-1 cells were treated for 1 week with 1x6-TG (6-Thioguanine (Merck, A4660)) to select for cells without a functional HPRT locus. Next the plasmid libraries were digested with PmeI (1μL) (New England Biolabs) to linearise the plasmids before transfection. Next CRISPR guides were designed to flank the region and cloned into the PX458 Cas9 and sgRNA expression vector (Ran et al., 2013).

$500x10^6$ HM-1 cells were transfected using a Nucleofector®-4D (Lonza) with the P3 Primary Cell X kit (Lonza, V4XP-3024) with $5x10^6$ per cuvette 10μg pSKB enhancer reporter plasmid and 10μg of PX458-HPRT CRISPR plasmid using program CG-104. The transfected cells were then plated onto gelatinised plates in DMEM-ES media and 50μM SCR7 pyrazine (Merck, #SML1546) was added (to inhibit non-homologous end joining and to promote homologous recombination after cutting by Cas9) and the cells were incubated at 37°C and 5% CO_2 for 12 hours. The media was then changed and 1 x HAT (Hypoxanthine-Aminopterin-Thymidine (Merck, #H0262)) was added to select for clones which had undergone successful recombination. The cells were cultured in DMEM-ES media containing HAT for 7 days and HT (hypoxanthine and thymidine (Merck, #H0137)) for a further 2 passages (4 days) to prevent the cells from dying after the withdrawal of HAT. The cells were then expanded and differentiated into HB, HE1, HE2 and HP as per the serum I.V.D method (Methods 3.3). In addition to the previously mentioned gating for obtaining the HB, HE1, HE2 and HP cell populations the cells were also gated by Venus-YFP and sorted into a negative, low, medium and high populations which represented the activity of an enhancer fragment

driven minimal promoter. A minimal promoter only control cell line was used to set the Venus-YFP gate threshold.

3.10 Enhancer reporter library preparation

After cell sorting the genomic DNA of populations was extruded using the Qiagen DNA Micro kit (Qiagen) following the manufacturer's instructions. The fragments in the reporter cassettes were then amplified and sequencing adaptors added by PCR using index primers compatible with Illumina Sequencing and targeted against the tn5 transposase sequence ((0.5µM Nextera PCR Primer i5, 0.5µM Nextera PCR Primer i7, 25µl NEBNext Master Mix (New England Biolabs)). Between 25 and 30 PCR cycles were then performed under the following conditions: 98°C 30 seconds, 25-30 cycles (dependent on amount of DNA gained from sorted cells) 98°C 10 seconds, 63°C 30 seconds, 72°C 1 minute and hold at 4°C. The libraries were then quantified using a Bioanalyzer 2100 (Agilent) with a High Sensitivity DNA Chip and pooled at equal concentration for sequencing. Kappa Library Quantification Kit (Roche) and Bioanalyzer was then used to validate the final pool concentration and fragments size. The pooled libraries were sequenced on a Next-Seq 550 (Illumina) as a paired end run with a 150 cycle High Output kit (Illumina) by Genomics Birmingham.

3.11 Generation of cell lines carrying individual reporter constructs

A putative enhancer for the *Galnt1* gene containing transcription factor binding sites of

interest was identified from DNaseI hypersensitive site data (Goode et al., 2016). Dr Debbie Goode developed HM-1 cell lines containing the *Galnt1* WT and *Galnt1* enhancer sequences containing mutated TEAD, ETS, EBOX and GATA sites. To test the activity of the enhancer throughout differentiation the enhancer reporter system used in the genome-wise screen was employed (Wilkinson et al., 2013). The genomic sequence for the enhancer was obtained and flanking attB1 and attB2 sites were added to enable Gateway® cloning. The wild-type enhancer sequence and mutated sequences were then synthesised as an Invitrogen GeneArt Strings DNA Fragment (ThermoFisher). The resulting fragments were then cloned by the one step Gateway protocol in which in a single reaction the fragment was inserted into pDONR221 (ThermoFisher, #12536017) generating a attL-flanked entry clone by BP clonase (ThermoFisher) and from these vectors into the pSKB GW-Hsp68-Venus reporter vector by LR clonase. To achieve this 100 ng of Enhancer DNA, 75 ng pDONR221 (ThermoFisher, #12536017), 75 ng pSKB GW-Hsp68-Venus, TE pH 8.0 to 6µl total volume, 1.5 µl LR clonase (ThermoFisher) and 0.5 µl BP clonase (ThermoFisher) was incubated at 25°C for 3 hours and then 0.2 µg proteinase K (ThermoFisher) was added and incubated for 10 minutes at 37°C to terminate the reaction. Competent DH5α bacteria (New England Biolabs) were transformed with 1µl of Gateway® reaction by incubation on ice for 30 minutes and heat shock at 42 °C for 45 seconds. Transformed bacteria were then incubated for 1 hour at 37°C in SOC (super optimal broth with catabolite repression) media (Sigma-Aldrich, #L3522) and then plated onto agar plates containing 100µg/ml ampicillin (ThermoFisher) and incubated overnight at 37 °C (Infor HT). Colonies were then selected for mini-prep cultures, mini-preps were performed using the Qiaprep® Miniprep kit (Qiagen).

The insert in the resulting plasmid was sanger sequenced (Source Biosciences) to check the sequence using primer EGFP_Nrev (pSKB) (Table 3.2). Following confirmation of the insert sequence the same bacterial colony was used to grow cultures for Maxi-prep which was performed using the EndoFree® Plasmid Maxi Kit (Qiagen, #12362). HM-1 cells were transformed as described in Methods 3.9. HM-1 cells were treated with 6-TG for 1 week after which $5x10^6$ HM-1 cells were transfected using a Nucleofector®-4D (Lonza) with the P3 Primary Cell X kit. The cells were then selected for those with successful integration of the reporter cassette by treatment for 5 days with 1 x HAT containing DMEM-ES. At 9-12 days post the appearance of colonies, clones were picked and replated on a gelatinised 96 well plate. The clones were then grown in DMEM-ES containing 1 x HT for 2 passages and then used for serum and serum free I.V.D and flow cytometry as previously described.

Name	**Sequence**
PDONR201_5 (pDONR221)	GCAGTTCCCTACTCTCGC
EGFP_Nrev (pSKB)	CGTCGCCGTCCAGCTCGACCAG

Table 3.2- Primers used for validation of inserts in pDONR221 and pSKB plasmids by sanger sequencing.

PDONR201_5 (pDONR221) used to validate sequences in pDONR221 plasmid and EGFP_Nrev (pSKB) used to validate sequences in pSKB plasmid.

In order to validate enhancer elements which were identified in our enhancer screen, enhancers which were associated by HiC to the *Sparc*, *Pxn*, *Hspg2*, *Cdh5, Dlk1* and two enhancers for *Runx1* were selected (the +3.7kb and the +23kb). These enhancers were also identified as being VEGF signalling responsive (taken as a 2-fold change in ATAC-Seq peak height with and without VEGF present in the culture). Primers were designed to amplify

out the 400bp enhancer sequence from genomic DNA (Table 3.3). To achieve this a PCR reaction was set up by mixing 50ng HM-1 genomic DNA, Phusion Buffer 5X (New England Biolabs) 10μl, Phusion (New England Biolabs) 0.5μl, dNTP, forward primer (10 μM) 2.5μl, reverse primer (10μM) 2.5μl and the reaction was made up to 50μl total with H_20. The mixture was run using the following PCR conditions: 98°C 10 seconds, 72°C for 30 seconds x30, 72°C for 30 seconds, 72°C for 5 minutes and the 4°C hold. The 10μl of the PCR product was then combined with 2μl gel loading dye and run on a 1.2% agarose gel with 100bp ladder to assess whether the correct product was amplified. The enhancer element was then inserted into HM-1 cells as described above for the *Galnt1* sequences. For analyses P-values were calculated using a two-tailed Student's t-test.

Gene	Forward (5'-3')	Reverse (5'-3')
Sparc	GGGGACAAGTTTGTACAAAAAAGCAGGCTACTTCTCAGTTCTCTGTTG	GGGGACCACTTTGTACAAGAAAGCTGGGTGTTTGAACCCTTTGAACCAC
Pxn	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGCAGGAGCCTTGGATCTCC	GGGGACCACTTTGTACAAGAAAGCTGGGTGACAGTGCCTCCCACAAAG
Hspg2	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAGCCTCAGTAAGGCCCAG	GGGGACCACTTTGTACAAGAAAGCTGGGTGCAGGAGAGGGAAAACAGAG
Cdh5	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGGTACTAAGGGATCGTGG	GGGGACCACTTTGTACAAGAAAGCTGGGTCCATGGGGTGGTTTTGTGTG
Dlk1	GGGGACAAGTTTGTACAAAAAAGCAGGCTGCAGTATGGTTCCTGGGAC	GGGGACCACTTTGTACAAGAAAGCTGGGTTGGCCGGATGTGGTTATCTG
Eif2b3	GGGGACAAGTTTGTACAAAAAAGCAGGCTACCAAGTGACAGATTTCCTG	GGGGACCACTTTGTACAAGAAAGCTGGGTCCATGGGGTGGTTTTGTGTG

Table 3.3- Primers used for the amplification of enhancer elements from genomic DNA.

To validate and study the +3.7kb and +23kb *Runx1* enhancers the 400bp wild type enhancer

sequences and each of the enhancer sequences with TF binding motifs mutated were synthesised by GeneArt String Synthesis (ThermoFisher) (Table 3.4). The sequences were then cloned into HM-1 cells as described previously.

Construct	Sequence
RUNX1-+23-WT	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGCAAGGAGCGATGGAGGGATGGTGTGAGGAGGAGACAG GAAGGGAGGCGGTGACACAGGCCTTCACAGGGCCGCTGCAGCTAGGAACTGCTGCAAAAGCAAGCTGCCCA CGTTATCAGTGGCGCGCAGGCCTGCGGTTTTCTCGCTCTTGCAACCCGGCTTCAACTGCCGGTTTATTTTTCGA CAAACAGGATGCCTCCATCTGAGGCTGTCAGCATCCCAGGCTCTTTGAGAAGAAAAGAGGTAGTGAGGGCCC CACCCTAGAGCCCAGCCGGGAGGGGTGGGTGAGAGGTCCTGTTTCTAGATGCTTCCCAGAGAAGTGAGAAC CTAGCAGGTGCAGTGGCCAGGGTTCAGGAGCGTGTCAGGAAACCCAGCTTTCTTGTACAAAGTGGTCCCC
RUNX1-+23-dRUNX1	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGCAAGGAGCGATGGAGGGATGGTGTGAGGAGGAGACAG GAAGGGAGGCGGTGACACAGGCCTTCACAGGGCCGCTGCAGCTAGGAACTGCTGCAAAAGCAAGCTGCCCA CGTTATCAGTGGCGCGCAGGCCTGGCCTTTTCTCGCTCTTGCAACCCGGCTTCAACTGCCGGTTTATTTTTCGA CAAACAGGATGCCTCCATCTGAGGCTGTCAGCATCCCAGGCTCTTTGAGAAGAAAAGAGGTAGTGAGGGCCC CACCCTAGAGCCCAGCCGGGAGGGGTGGGTGAGAGGTCCTGTTTCTAGATGCTTCCCAGAGAAGTGAGAAC CTAGCAGGTGCAGTGGCCAGGGTTCAGGAGCGTGTCAGGAAACCCAGCTTTCTTGTACAAAGTGGTCCCC
RUNX1-+23-dTEAD	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGCAAGGAGCGATGGAGGGATGGTGTGAGGAGGAGACAG GAAGGGAGGCGGTGACACAGGCCTTCACAGGGCCGCTGCAGCTAGGAACTGCTGCAAAAGCAAGCTGCCCA CGTTATCAGTGGCGCGCAGGCCTGCGGTTTTCTCGCTCTTGCAACCCGGCTTCAACTGCCGGTTTATTTTTCGA CAAACAGGATGCCTCCATCTGAGGCTGTCAGCGTCAAAGGCTCTTTGAGAAGAAAAGAGGTAGTGAGGGCC CCACCCTAGAGCCCAGCCGGGAGGGGTGGGTGAGAGGTCCTGTTTCTAGATGCTTCCCAGAGAAGTGAGAA CCTAGCAGGTGCAGTGGCCAGGGTTCAGGAGCGTGTCAGGAAACCCAGCTTTCTTGTACAAAGTGGTCCCC
RUNX1-+3.7 -WT	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGTTCCTGGGAGTCTCAGTCGAGTCTTATCGTCTTGCTTAGGC AGACTAAGTGGGAACCTGGTGTGGGGCGATAGAGGGGGAAAGGAGCTCCCTGTGGTAGACCACATTCTTTT TGCGTGCCAGTGTCTCAGATGCTGGCCCGACTCTCACGAGAAAGGGAACAATGGCTTGCTTGGTTCCCCATCT CTGTTTTCTCCCGTTTCCACAGTGGTATCAGGCTGAAAGTGAGTTACAGGTTTCCTCCTGCGAGGATTGCTGG GACTGATAAAGTACCATCCCAGGAGGAAAACAAGATGCTGAAATTACAGGCAGAAGCAGCCACCAGTGAGG AGCCCTAGGAGTGGCAGGCTGTGGTGTGGGGGGCCAGGAACCCAGCTTTCTTGTACAAAGTGGTCCCC
Galnt1-WT	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGAGTCACCCCGACTCAAGCAGAGCATCGTAAATACATTAAGA CACAAGACTGACATGTATATGACACACACCAGAGCAGCAGCAGAAACATCCTGCCATTTGAACCATGGATTG AGGGCTGCAGGACTGGGAATGACCCAGTGCAGCGTTGGCAAGGATTCAACACAGCGCATGCGCAGAAGTCT GCAAGTCCTGTAACACTGCGGTCCACGGGGCCCACTAGATTTCACCAGTGGGCGAAAGGATGGAAAGAGCA CTTTTCCAGTCTCTTCTTGCGGATCCTCAAGCACTCTGCCTTTCATGCAGGCCAGAGTTCTTGGGGGCTGTGGG TTGCTCCCACATGGAAAGTCCCAAGGCCGTTGCAGGATGTCCTCCACAGAGTCCCATCCTGTCCTTATCACAG CTGCAGGCAGCCTGGGCCACATTCCTGTCTCCTGAGTCACTGGTGTGGCCTGGGCTCCTCGCTCTGGAAACCA GGGCACTGCCTGCTCAGTCTTTGGCTTGTCTTTACTCCTCCATACATACCCAGCTTTCTTGTACAAAGTGGTCCC C
Galnt1-dTEAD	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGAGTCACCCCGACTCAAGCAGAGCATCGTAAATACATTAAGA CACAAGACTGACATGTATATGACACACACCAGAGCAGCAGCAGAAACATCCTGCCATTTGAACCATGGATTG AGGGCTGCAGGACTGGGAtcGACCCAGTGCAGCGTTGGCAAGGATTCAACACAGCGCATGCGCAGAAGTCT GCAAGTCCTGTAACACTGCGGTCCACGGGGCCCACTAGATTTCACCAGTGGGCGAAAGGATGGAAAGAGCA CTTTTCCAGTCTCTTCTTGCGGATCCTCAAGCACTCTGCCTTTCATGCAGGCCAGAGTTCTTGGGGGCTGTGGG TTGCTCCCACATGGAAAGTCCCAAGGCCGTTGCAGGATGTCCTCCACAGAGTCCCATCCTGTCCTTATCACAG CTGCAGGCAGCCTGGGCCACgaTCCTGTCTCCTGAGTCACTGGTGTGGCCTGGGCTCCTCGCTCTGGAAACCA GGGCACTGCCTGCTCAGTCTTTGGCTTGTCTTTACTCCTCCATACATACCCAGCTTTCTTGTACAAAGTGGTCCC C
Galnt1-dAP-1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGtGcgACCCCGACTCAAGCAGAGCATCGTAAATACATTAAGAC ACAAGACTGACATGTATATGACACACACCAGAGCAGCAGCAGAAACATCCTGCCATTTGAACCATGGATTGA GGGCTGCAGGACTGGGAATGACCCAGTGCAGCGTTGGCAAGGATTCAACACAGCGCATGCGCAGAAGTCTG CAAGTCCTGTAACACTGCGGTCCACGGGGCCCACTAGATTTCACCAGTGGGCGAAAGGATGGAAAGAGCACT TTTCCAGTCTCTTCTTGCGGATCCTCAAGCACTCTGCCTTTCATGCAGGCCAGAGTTCTTGGGGGCTGTGGGTT GCTCCCACATGGAAAGTCCCAAGGCCGTTGCAGGATGTCCTCCACAGAGTCCCATCCTGTCCTTATCACAGCT GCAGGCAGCCTGGGCCACATTCCTGTCTCCTcgGaCACTGGTGTGGCCTGGGCTCCTCGCTCTGGAAACCAGG GCACTGCCTGCTCAGTCTTTGGCTTGTCTTTACTCCTCCATACATACCCAGCTTTCTTGTACAAAGTGGTCCCC
Galnt1-dE-Box	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGAGTCACCCCGACTCAAGCAGAGCATCGTAAATACATTAAGA CACAAGACTGACATGTATATGACACACACCAGAGCAGCAGCAGAAACATCCTGCgAaTTcAACCATGGATTGA GGGCTGCAGGACTGGGAATGACCCAGTGCAGCGTTGGCAAGGATTCAACACAGCGCATGCGCAGAAGTCTG CAAGTCCTGTAACACTGCGGTCCACGGGGCCCACTAGATTTCACCAGTGGGCGAAAGGATGGAAAGAGCACT

	TTTCCAGTCTCTTCTTGCGGATCCTCAAGCACTCTGCCTTTCATGCAGGCCAGAGTTCTTGGGGGCTGTGGGTT GCTCCtACATGGAAAGTCCCAAGGCCGTTGCAGGATGTCCTCCACAGAGTCCCATCCTGTCCTTATCAgAGCTG CAGGCAGCCTGGGCCACATTCCTGTCTCCTGAGTCACTGGTGTGGCCTGGGCTCCTCGCTCTGGAAACCAGG GCACTGCCTGCTCAGTCTTTGGCTTGTCTTTACTCCTCCATACATACCCAGCTTTCTTGTACAAAGTGGTCCCC
Galnt1-dGATA	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGAGTCACCCCGACTCAAGCAGAGCATCGTAAATACATTAAGA CACAAGACTGACATGTATATGACACACACCAGAGCAGCAGCAGAAACATCCTGCCATTTGAACCATGGATTG AGGGCTGCAGGACTGGGAATGACCCAGTGCAGCGTTGGCAAGGATTCAACACAGCGCATGCGCAGAAGTCT GCAAGTCCTGTAACACTGCGGTCCACGGGGCCCACTAGATTTCACCAGTGGGCGAAAGGATGGAAAGAGCA CTTTTCCAGTCTCTTCTTGCGGATCCTCAAGCACTCTGCCTTTCATGCAGGCCAGAGTTCTTGGGGGCTGTGGG TTGCTCCCACATGGAAAGTCCCAAGGCCGTTGCAGGATGTCCTCCACAGAGTCCCATCCTGTCCTTAgCACAGC TGCAGGCAGCCTGGGCCACATTCCTGTCTCCTGAGTCACTGGTGTGGCCTGGGCTCCTCGCTCTGGAAACCAG GGCACTGCCTGCTCAGTCTTTGGCTTGTCTTTACTCCTCCATACATACCCAGCTTTCTTGTACAAAGTGGTCCCC
Galnt1-dETS	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGAGTCACCCCGACTCAAGCAGAGCATCGTAAATACATTAAGA CACAAGACTGACATGTATATGACACACACCAGAGCAGCAGCAGAAACcTCCTGCCATTTGAACCATCGATTGA GGGCTGCAGGACTGtGAATGACCCAGTGCAGCGTTGGCAAAGATTCAACACAGCGCATGCGCAGAAGTCTGC AAGTCCTGTAACACTGCGGTCCACGGGGCCCACTAGATTTCACCAGTGGGCGAAAaGATcGAAAGAGCACTT CTCCAGTCTCTTCTTGCGGcTCCTCAAGCACTCTGCCTTTCATGCAGGCCAGAGTTCTTGGGGGCTGTGGGTTG CTCCCACATGGAgAGTCCCAAGGCCGTTGCAGGtTGTCCTCCACAGAGTCCCcTCCTGTCCTTATCACAGCTGC AGGCAGCCTGGGCCACATTCgTGTCTCCTGAGTCACTGGTGTGGCCTGGGCTCCTCGCTCTcGAAACCAGGGC ACTGCCTGCTCAGTCTTTGGCTTGTCTTTACTCCTCCATACATACCCAGCTTTCTTGTACAAAGTGGTCCCC

Table 3.4- Table showing the Invitrogen GeneArt Strings synthesised DNA sequences used in this study for the *Runx1* +23kb, +3.7kb and *Galnt1* enhancer.

3.12 Single cell sequencing sort method

HM-1 cells were differentiated as described in Methods 3.4 using the serum free I.V.D system with and without VEGF. $7.0x10^6$ cells from the All-cytokines (+VEGF) condition blast culture and $2.8x10^6$ cells from the VEFG withdrawal (-VEGF) blast culture was collected for cell sorting and sc-RNA-Seq library preparation.

Both samples were stained with CD41-PECY7 (1:100) (Invitrogen), KIT-APC (1:100) (Invitrogen) and TIE2-PE (1:200) (Invitrogen) and then sorted into HE1 (KIT+, TIE2+, CD41-) and HE2 (KIT+, TIE2+, CD41+) populations. In total 758000 HE1 and 380000 HE2 cells were sorted from the All-cytokines sample and 71500 HE1 and 33300 HE2 cells were sorted from the VEGF withdrawal sample. Sorted populations were then resuspended in 80µL at a concentration of 100-1200 cells/µL for evaluation of cell viability. The viability of the All-cytokines populations was 73% for HE1 and 92% for HE2. The viability of the VEGF

withdrawal populations was 73% for HE1 and 75% for HE2. 10,000 single cells were loaded on a Chromium Single Cell Instrument (10x Genomics) for processing and sequencing on NovaSeq 6000 (Illumina) by Genomics Birmingham.

3.13 Miniprep

Plasmid DNA was extracted using the Qiagen spin miniprep kit (Qiagen #27104). Bacteria was grown in LB broth before being centrifuged at 4000g for 5 minutes. Pellets were resuspended in 250µl buffer P1. Buffer P2 was then added for 4 minutes to lyse bacterial cells before addition of 350µl buffer N3. The lysate was then centrifuged at 16000g for 10 minutes to pellet bacterial cell components leaving plasmid DNA in the supernatant. Supernatant was then pipetted into a QIAprep spin column which was centrifuged for 16000xg for 1 minute before being washed with buffer PE and eluted into 50µl buffer EB. The concentration of eluted plasmid was checked on a Nanodrop 2000c Spectrophotometer (ThermoFisher) using Nanodrop 2000c Software (ThermoFisher).

3.14 Maxiprep

Bacterial clones were grown in 250ml of LB broth with 100µg/ml kanamycin for clones containing pDONR and 100µg/ml ampicillin for clones containing pSKB overnight at 37°C, 200rpm. Bacterial culture was centrifuged at 4°C at 4000g for 20 minutes and the Nucleobond Xtra Midiprep kit (Machery-Nagel #740420.50) was used to extract plasmid DNA. The pellet was resuspended in 16ml buffer RES-EF. 16 ml buffer LYS-EF was added to lyse bacterial cells for 5 minutes before 16 ml of buffer NEU-EF was added to neutralise the reaction. Buffer EQU-ED was added to the Nucleobond column and then the lysed bacteria

suspension was loaded into the column and allowed to drain under gravity. The column filter was washed with buffer FIL-EF before being removed. The column was then washed with buffer ENDO-EF followed by buffer WASH-EF. 5ml ELU-EF was added to elute the material from the column by gravity. 3.5ml isopropanol was then added to the eluent before the mixture was centrifuged at 20000xg for 45 minutes. The supernatant was then discarded, and the pellet was washed with 70% ethanol before being resuspended with TE-EF buffer. The concentration of eluted plasmid was checked on a Nanodrop 2000c Spectrophotometer (ThermoFisher) using Nanodrop 2000c Software (ThermoFisher).

3.15 Enhancer sequencing data analysis

Conducted by Dr Benjamin Edginton-White. Initially paired end sequencing reads were trimmed using Trim Galore (V0.6.5) (Quinlan and Hall, 2010) with the options --nextera --length 70 –paired. Sequencing reads were then aligned to the mouse genome (version mm10) using bowtie2 (v1.3.0) (Langmead and Salzberg, 2012) using the options --very-sensitive --fr --no-discordant -X 600 --no-mixed. Aligned reads were then filtered for those which were properly paired and with a mapq score of over 40 using Samtools (v1.16) (Danecek et al., 2021) and output as a bedpe file. The file was then converted to a bed file of enhancer fragments by taking the first co-ordinate of read 1 and the final co-ordinate of read 2 for each pair. Duplicate fragments were filtered out using the uniq function (Bash) and fragments were further filtered using the intersect function on Bedtools (v2.30.0) (Quinlan and Hall, 2010) for fragments which had 100% overlap with an off-target background fragment list which was created by producing a library from un-transfected

cells.

The fragments were then filtered for those which were located in open chromatin, as defined by DNaseI-Seq from Goode et al. (Goode et al., 2016) and by a corresponding ATAC peak in the correct differentiation stage in HM-1 cells using the intersect function in bedtools. Finally, the annotatePeaks function of Homer (v4.11) (Heinz et al., 2010) was used to annotate enhancer fragments which were within 1.5 kb of a TSS (promoter fragments) and distal fragments taken as genuine enhancers. The resulting list of enhancer fragments was used for all further fragment analysis and by overlapping ATAC sites with enhancer fragments we defined full enhancers.

3.16 Gene annotation using HiC data

Gene annotation using HiC data was conducted by Dr Salam Assi. Promoter capture HiC data from embryonic stem cells were obtained from Novo et al. (Novo et al., 2018) (GEO accession numbers: GSM2753058, runs: SRR5972842, 1051 SRR5972842, SRR5972842, SRR5972842, SRR5972842) and from HPC7 cells from Comoglio et al. (Comoglio et al., 2018) (GEO accession numbers: GSM2702943 & GSM2702944, runs SRR5826938, SRR5826939). Chi-C paired-end sequencing reads were aligned using the HiCUP pipeline to the mouse genome (version mm10) (Wingett et al., 2015). Raw sequencing reads were separated and then mapped to the reference genome (mate pairs were mapped separately). Aligned reads were filtered for experimental artefacts and duplicate reads before being re-paired.

Statistically significant interactions were called using GOTHiC package (Mifsud et al., 2017) and HOMER software (v4.11) (Heinz et al., 2010). By using a background model of random interactions, this method uses a cumulative binomial test to identify interactions between distal genomic loci which have significantly more reads then expected by chance. This method assigns a p-value to each interaction which represents its significance. The union of all CHiC interactions from both ESC and HPC7 cells were used to annotate positive enhancer to their related promoter.

3.17 Motif co-localisation analysis

For each transcription factor (TF) the genomic co-ordinates of each of its binding motifs were retrieved from the sets of stage specific positive enhancer fragments and from within all distal ATAC-Seq peaks using HOMER (v4.11) (Heinz et al., 2010) using the annotatePeaks.pl function and exported as a BED file with the -mbed option. For ESC, HB, HE1, HE2 and HP the motif co-occurrence was then measured by counting the number of times a pair of TF motifs were found within 50bp of each other within the set of specific positive enhancer fragments. A re-sampling analysis was then performed to assess the significance of co-occurrence of motifs in enhancer fragments as follows. A set of ATAC-seq sites equal in number to the stage specific enhancer fragments was sampled at random from the complete set of distal ATAC-Seq peaks from that cell stage and the number of motif pairs were counted in these random peak sets. This process was repeated 1000 times resulting in a distribution of motif pair counts for each pair of TF binding motifs. Next a z-score was then calculated for each motif pair using the following equation:

$$Z = \frac{x - \mu}{\sigma}$$

Where x represents the number of motif pairs found in the specific positive enhancer fragments, μ is the mean number of motif pairs in 1000 random samples, and σ is the standard deviation of those motif pair counts. A z-score which is positive here suggests that the number of motif pairs found in the positive enhancers is greater than what would be expected by chance. The z-scores were then hierarchically clustered using complete linkage of the Euclidean distance in R (v3.6.1) and plotted as a heatmap.

3.18 ATAC-seq analysis

Illumina sequence read quality was assessed by MultiQC (Ewels et al., 2016). Illumina single-end reads from ATAC-seq experiments were processed using Trimmomatic (version 0.39) and then aligned to the mouse genome (version mm10) using Bowtie2 v2.4.4 (Langmead and Salzberg, 2012) using the options -very-sensitive-local. Open chromatin regions (peaks) were identified using MACS2 v2.2.7.1 (Zhang et al., 2008) using the options -B -trackline -nomodel. Next peaks were then filtered against the mm10 blacklist (Amemiya et al., 2019). Only aligned reads to the canonical set of chromosomes were taken forward to analysis, chromosomes annotated as chrUn, _random or as alternate haplotypes were removed. Peaks were then annotated as being either promoter-proximal if they were within 1.5kb of a transcription start site and as a distal element if not. For differential chromatin accessibility analysis, a peak union was first constructed by merging peaks from

two comparisons. To quantify peak height, the tag-density in a 400-bp range centred around the peak summits was calculated from bedGraph files produced by MACS2 peak calling using the annotatePeaks.pl function in Homer v4.11 (Heinz et al., 2010) with the options -size 400 -bedGraph. The tag-densities were then normalised as counts-per-million (CPM) using R v3.6.1 and then log_2- transformed as log_2(CPM+1). A peak was considered differentially accessible if it had a fold difference of at least 2 between conditions. HOMER was then used to perform a de-novo motif analysis on sets of gained and lost peaks using the findMotifsGenome.pl function with the options -size 200 -noknown. Tag density plots were constructed by retrieving the tag-density in a 2kb window centred on the peak summits with the annotatePeaks.pl function in Homer with the options -size 2000 -hist 10 -ghist -bedGraph. These were then plotted as a heat map using Java TreeView v1.1.6 (Saldanha, 2004). The intersect function in Bedtools (Quinlan and Hall, 2010) was then used to find open chromatin regions shared between two samples.

3.19 Relative motif enrichment analysis of ATAC-Seq data

To identify transcription factor binding motifs which were enriched in a set of peaks relative to another, we calculated a relative enrichment motif score as in Goode et al. (Goode et al., 2016). The relative enrichment motif score S_{ij} for each motif i in each peak set j as:

$$S_{ij} = \frac{n_{ij}/m_j}{\sum_j n_{ij} / \sum_j m_j},$$

where n_{ij} is the number of instances of motif i in peak set j and m_j is the total number of sites in peak set j. This score was calculated for each TF motif in each of the peak sets and a matrix of enrichment scores was produced which was then hierarchically clustered using complete linkage of the Euclidean distance in R and displayed as a heatmap with results scaled by either row or column using R (v3.6.1).

3.20 Single cell RNA-sequencing analysis

Fastq files from the sc-RNA-Seq experiments from HE1 and HE2 cultured in the serum free system with VEGF and with VEGF withdrawn were aligned to the mouse genome (version mm10) using the count function in CellRanger (V6.0.1) from 10X Genomics and using gene models from Ensemble (release 102) as the reference transcriptome. This produced the Unique Molecular Identifier (UMI) count matrices which were then processed using the Seurat package (v4.0.5) (Hao et al., 2021) in R (v4.1.2). Cell quality control was performed for each of the four samples which were sequenced individually in order to remove cells which had less than or higher than the expected number of detected genes or which had a high proportion of UMIs aligned to mitochondrial transcripts for each sample. Genes detected in fewer than 3 cells were excluded from the analysis.

The filtered data objects for each cell type were combined based on their VEGF status (+/-VEGF) and normalised using the LogNormalize method in Seurat. To calculate the cell cycle stage for each of the cells we used the in-built S-phase and G2M-phase marker gene list from Seurat. These lists were first converted from human gene symbols to the

corresponding mouse orthologs using the biomaRt package (v2.50.0) (Durinck et al., 2009) in R. The cell cycle stage was then determined using the CellCycleScoring function in Seurat and the possible effect of cell cycle stage on downstream analysis was then removed from the data by linear regression using the ScaleData function. Cells were then clustered using the first 20 principal components and visualised using the Uniform Manifold Approximation and Projection (UMAP) method. Cell marker genes for each of the clusters identified were calculated using the FindAllMakers function. A gene was considered a marker gene if it was expressed with a log2-fold-difference of 0.5 between the cluster being considered and all other cells as well as being detected in at least 50% of cells in that cluster. A Gene was taken as being statistically significant if it had a Bonferroni adjusted P-Values < 0.05. The cell types were then inferred for each cluster identified by manual inspection of the gene list and comparing these to the expression of known surface marker genes such as *Kit*, *Tie2* and *Cd41* for HE1, HE2 and HP.

To perform cell trajectory (pseudotime) analysis Monocle v3.1.0 (Trapnell et al., 2014) was used. The data from Seurat was exported to Monocle using the seuratwrappers package in R (https://github.com/satijalab/seurat-wrappers). The trajectories were then inferred using the learn graph function and then ordered along pseudotime by selecting a root node which corresponded to the earlier cell-type in the culture (HE1).

4 RESULTS

4.1 Functional identification of enhancer and promoter elements specifically active at defined stages of blood specification

4.1.1 Characterisation of the chromatin and transcription factor motif landscape at five stages of *in vitro* differentiation to blood progenitors in serum culture

To functionally identify cell-type specifically active enhancer elements in ESCs differentiating to HP we differentiated HM-1 mouse ES cells which carry an enhancer reporter system (Wilkinson et al., 2013) (Figure 4.1a-c). As this cell line is different compared to those used previously, we first validated the relative motif enrichment analysis findings reported in (Goode et al., 2016) as described in Methods 3.19. Briefly, mouse ESCs were cultured on mouse embryonic feeder cells with leukaemia inhibitory factor (LIF) before being split into suspension cultures for 72 hours. During this time floating embryoid bodies form containing FLK-1+ HB cells. These cells are then dissociated and sorted for expression of FLK-1 by magnetic cell sorting. Some FLK-1+ cells were used for ATAC-Seq analysis while the rest were plated into blast culture with Blast media. Blast cell colonies were allowed to form and at day 5 (day 2 of blast cell culture) HE1 cells expressing Kit, TIE2 but not CD41 and low levels of RUNX1 were sorted by FACS together with HE2 cells which upregulate RUNX1 and express KIT, TIE2, CD41 and HP expressing RUNX1, KIT, CD41 but not TIE2 which had undergone the endothelial-to-haematopoietic transition (EHT). Using the differential expression of these specific markers, the HB, HE1, HE2 and HP populations were purified to near homogeneity. These cell types were taken for ATAC-Seq

library preparation, libraries were sequenced, ATAC-Seq reads were aligned to the mouse mm10 genome and peaks were called as described in Methods 3.18.

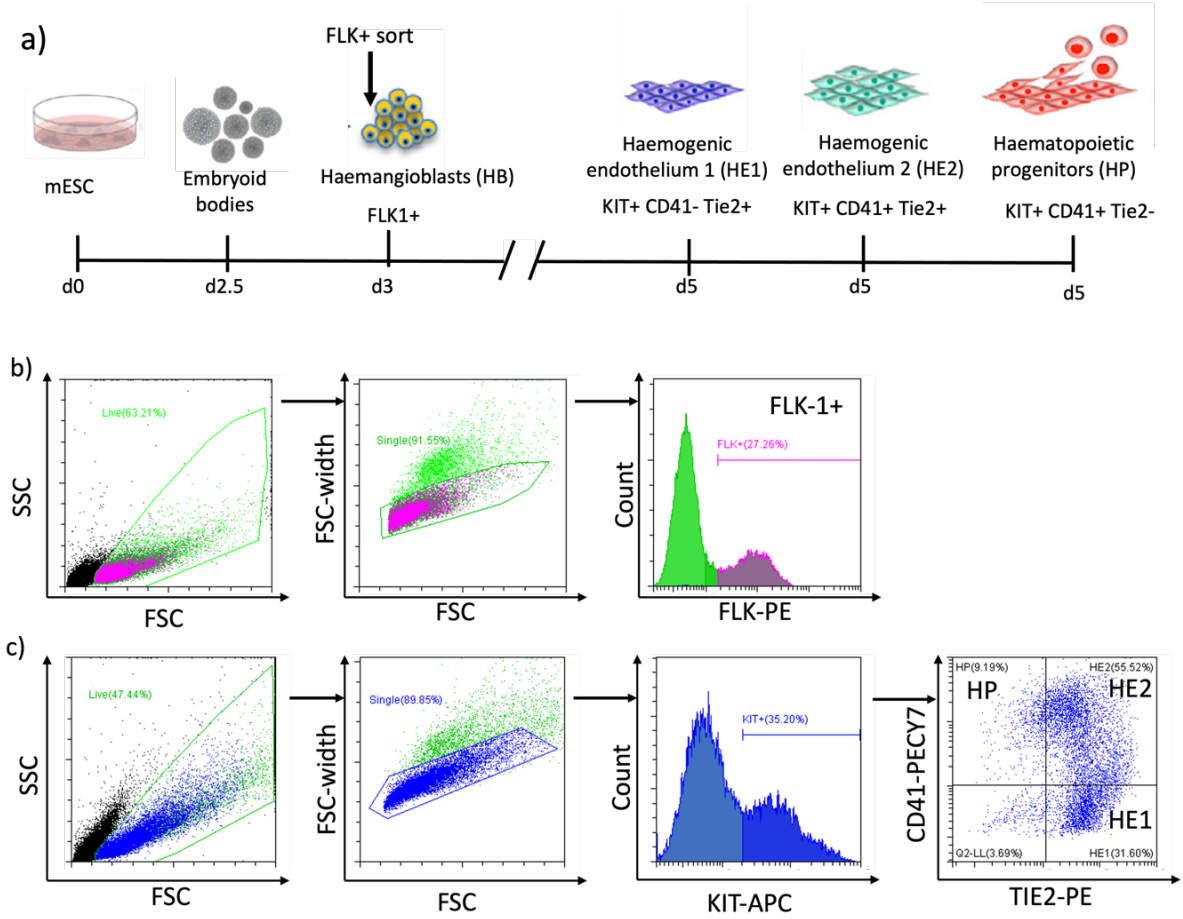


Figure 4.1- In vitro differentiation of HM-1 cells into blood progenitor cells (HP).
a) Schematic of the mouse serum ES cell differentiation system used in this study. Adapted from (Goode et al, 2016). b) FACS gating strategy for FLK-1+ HB cells. Live cells were gated, singlets were gated for and finally FLK-1+ cells (adapted from (Goode et al., 2016)). c) FACS gating strategy for day 2 blast cultures (d5 in schematic) shown. Live cells were gated for, then singlets, then KIT+ cells and then finally cells were gated on markers TIE2 and CD41.

4.1.2 Relative motif enrichment analysis of all distal ATAC-Seq sites in cells derived from serum *in vitro* differentiation

To identify TF binding motifs enriched in cell type-specific open chromatin regions, we used HM-1 chromatin accessibility data to perform a pairwise comparison of distal ATAC-Seq sites from one cell type with all other cell types as shown in Figure 4.2a Cell type specific ATAC-Seq peak sets were defined using the Bedtools Intersect function as described in Methods 3.18. These unique ATAC-Seq peak sets were taken for HOMER motif calling and then relative motif enrichment analysis as described in Methods 3.19 was conducted. Relative motif enrichment analysis scores were then hierarchically clustered and plotted as a heat map.

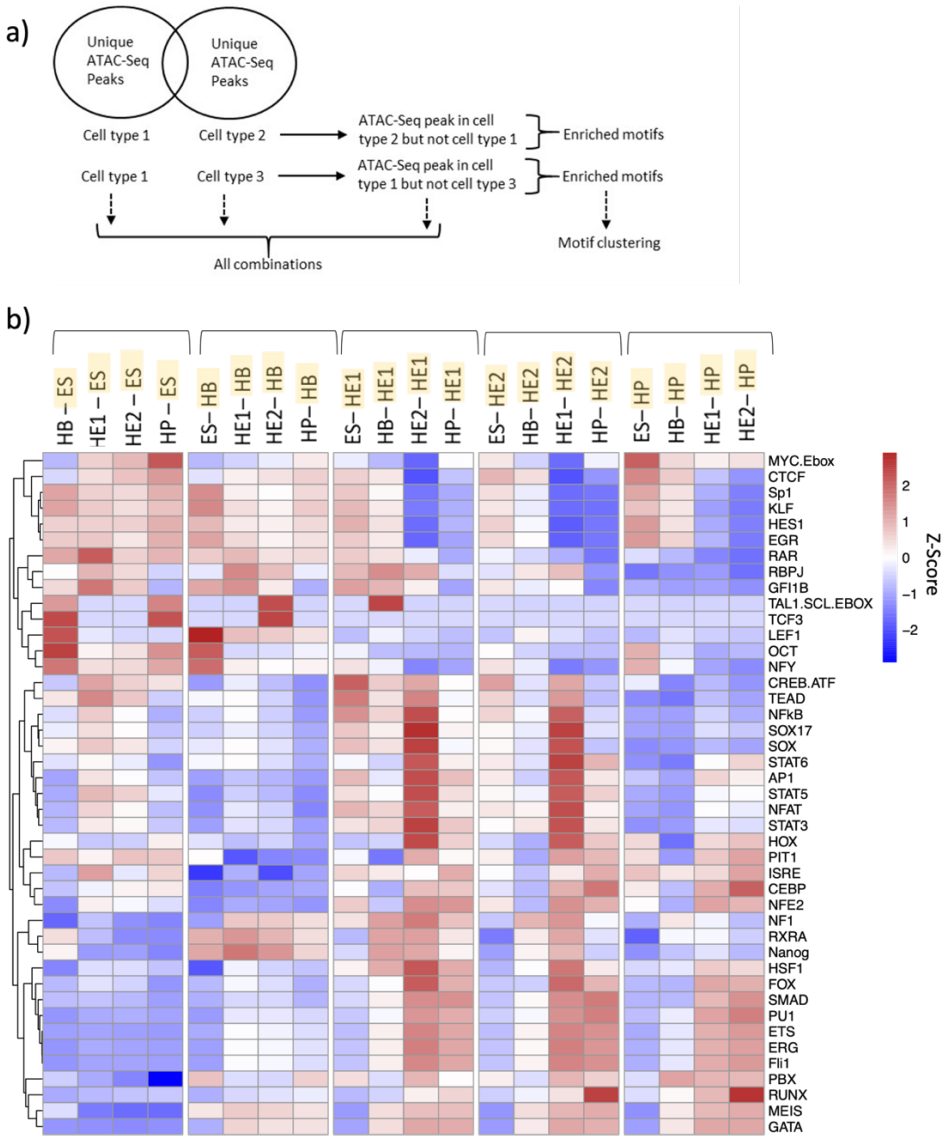


Figure 4.2- Hierarchical clustering of relative motifs enrichment analysis scores of unique distal ATAC-Seq peak sets from serum I.V.D.

a) Filtering strategy used for identifying specific TF binding motifs in cell type specific open chromatin region (distal elements only) as described in (Goode et al., 2016). Distal ATAC-Seq sites (at least 1500bp from the TSS) from each cell type were intersected to get lists of unique distal ATAC-Seq sets. HOMER was then used to call motifs for each peak set and a motif relative enrichment analysis score was calculated and taken for clustering. b) Heat map of motif relative enrichment scores hierarchically clustered by TF motif. The motif relative enrichment analysis scores were converted to a Z-score by columns. The top of the plot shows the cell type comparison and the motif relative enrichment analysis scores for peaks unique to the cell type highlighted in yellow. Cell types are grouped by brackets. The right of the plot shows the TF motifs used in the analysis. Dendrogram showing clustering of TF motifs on the left. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

Our analysis highlighted the known roles of specific factors in the relevant cell types and confirmed findings from (Goode et al., 2016) (Figure 4.2b). As expected, in ESCs the pluripotency factor motifs such as KLF, SOX and OCT (POU5F1) were enriched. In HBs peaks were enriched for motifs for the NOTCH signalling effectors HES1 and RBPJ. RAR, RXR, EGR, KLF, LEF1 and SP1 motifs were also enriched in HB specific peaks.

The HE1 specific peak sets were enriched for TF motifs associated with an overall endothelial signature such as motifs for the VEGF signalling responsive TF SOX, the Notch signalling responsive effectors HES1 and RBPJ as well as AP-1 and TEAD TFs. We also found an enrichment of NFκB motifs. The HE2 cell type comparisons display a similar TF motif enrichment profile with a slight increase in the enrichment of motifs for haematopoietic TFs. Motifs for the haematopoietic TF GATA2 and in particular RUNX1 were also enriched in the HE2 cell type compared to previous cell types, indicating that this cell type was on its way to become a blood progenitor.

Peak sets unique to HP were strongly enriched for haematopoietic TF motifs such as RUNX1 and its targets PU.1 and NFE2. Importantly motifs for endothelial TFs such as SOX, AP-1, TEAD, TAL-1 and the Notch signalling targets HES1 and RBPJ were depleted, indicating an overall switch from an endothelial to a haematopoietic TF signature as a result of the EHT.

4.1.3 Functional identification of cell type specifically active enhancer elements

To identify functional cell type specifically active enhancer elements we adapted the enhancer reporter system developed by (Wilkinson et al., 2013, Dickel et al., 2014) to allow for a high-throughput genome-wide enhancer screen as described in Methods 3.8-3.10. The genome-wide enhancer screen was developed and conducted by Dr Benjamin Edginton-White. Briefly, we first differentiated mouse ESCs in the serum I.V.D system (Figure 4.1) and sorted cells by FACS to isolate HB, HE1, HE2 and HP. These cell types including ESCs were taken for ATAC-Seq library generation. ATAC fragments were sequenced directly or cloned into a targeting vector to generate a fragment library which was then integrated into a defined target site in the *HPRT* locus carrying a minimal promoter to drive a reporter gene (Venus-YFP). By using this targeting system, we made sure that only one fragment and reporter construct was active in each cell and that the enhancer-reporter construct was in an accessible chromatin environment throughout differentiation of mESCs to HP. Bulk mouse ESCs were differentiated using the serum I.V.D system and cells from each stage of development (ESCs, HB, HE1, HE2 and HP) were purified to measure reporter activity by flow cytometry. Cells were sorted into a high, medium, low and negative YFP populations (Figure 4.3a-b).

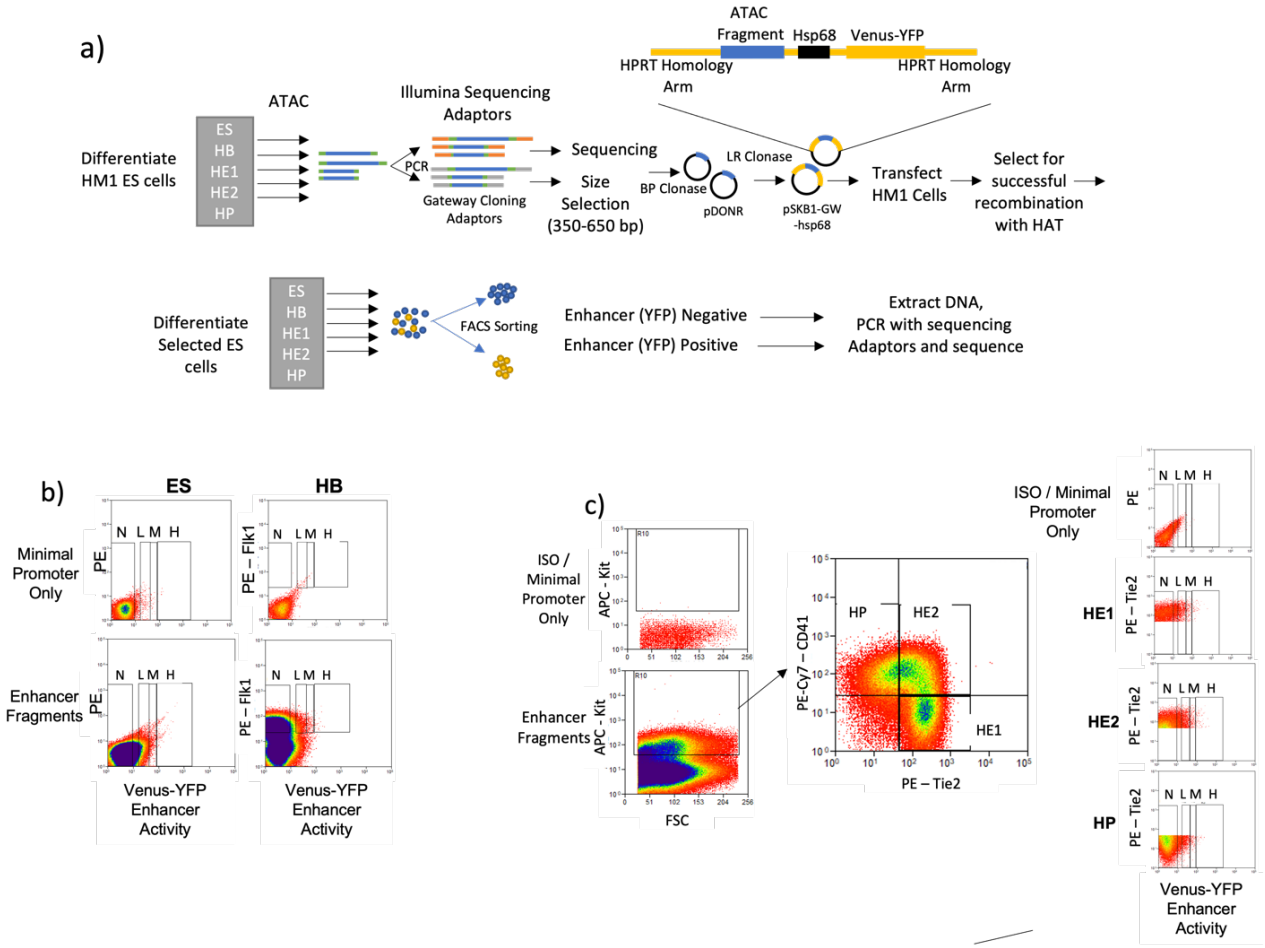
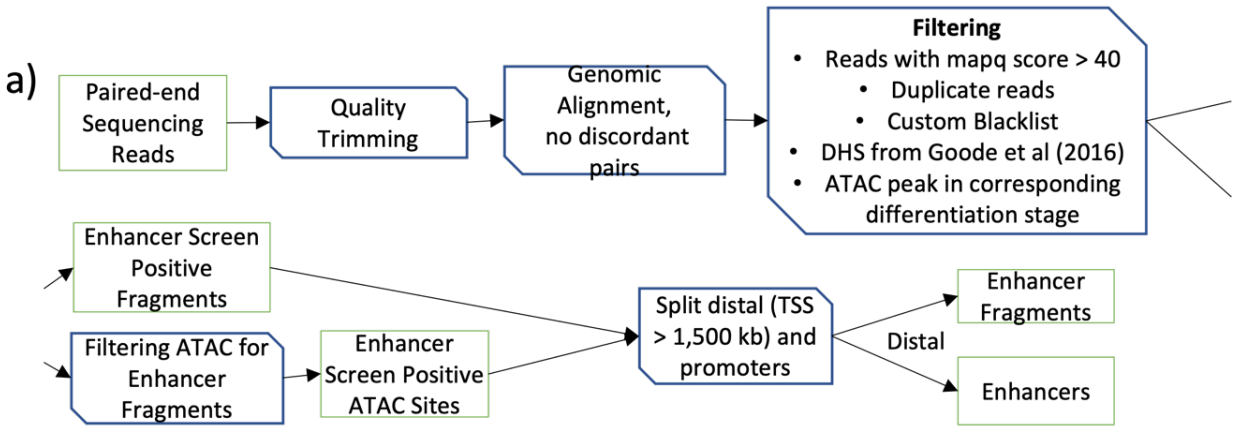


Figure 4.3- Enhancer screen strategy and cell sorting and analysis strategy for cells with activated reporter construct in the enhancer screen.

Enhancer screen conducted and figure produced by Dr Benjamin Edginton-White. a) Overview of the enhancer screen using ATAC-seq fragments from ESCs, HB, HE1, HE2 and HP. b) Example of gating strategy for the detection of enhancer activity compared to the minimal promoter control for ESCs and HB. c) Example of gating strategy for the detection of enhancer activity compared to the minimal promoter control for HE1, HE2 and HP. The low (L), medium (M) and high (H) gating strategy is shown.

Fragment inserts were then amplified by PCR using the barcoded primers recognising the ATAC linkers and then sequenced. Each unique read in the sequencing data is a fragment that was cloned into the reporter construct and is assigned activity based on which FACS (YFP) population it was sorted into. These aligned sequences then were rigorously filtered using the criteria detailed in Figure 4.4a.

a)



b)

Number of Enhancer Positive Distal ATAC Sites

	Total
ES	10,156
HB	12,062
HE1	18,180
HE2	16,891
HP	13,323

Number of Positive Promoter ATAC sites

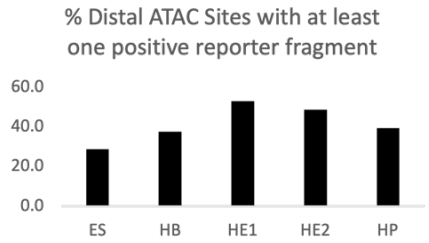
	Total
ES	11,444
HB	11,835
HE1	12,240
HE2	12,087
HP	11,633

Figure 4.4- The number of stage specific enhancer and promoter elements which scored positive in the enhancer screen.
Enhancer screen conducted and figure produced by Dr Benjamin Edginton-White. a) Overview of the filtering method used to filter positive (YFP+) fragments returned from the assay. b) The total number of distal and promoter sites which scored positive in the enhancer screen for each cell type.

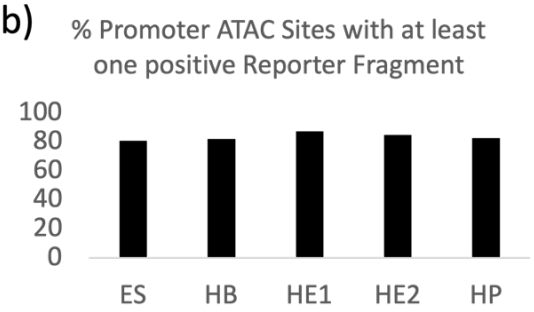
Our enhancer screen returned several hundred thousand fragments which could stimulate the enhancer reporter construct, of which 22% - 31% of fragments mapped to annotated distal elements and covered >70,000 enhancer positive ATAC sites across the five cell stages. The rest of the fragments were promoter sequences which were defined as being +/- 1.5kb from a transcription start site.

We found that between 30% - 50% of all distal ATAC-Seq sites (Figure 4.5a) and around 80% of all promoter sites (Figure 4.5b) in our serum derived ATAC-Seq library contained a fragment scoring in our assay with the medium number of fragments being between 3 and 6 (Figure 4.5c), suggesting that the majority of captured open chromatin regions have the capacity to influence transcription. To see, whether the same sequences were active *in vivo*, we overlapped our positively scoring ATAC fragments with open chromatin sites found in purified endothelial cells and haemogenic endothelium from mouse embryos at embryonic day 9.5 and embryonic day 13.5 (Zhu et al., 2020, Howell et al., 2021) and found that most of our enhancer positive ATAC fragments overlapped are with these sites (Figure 4.5d) and are likely to be functional.

a)



b)



c)

	Median number of unique positive fragments per distal ATAC site
ES	3
HB	4
HE1	6
HE2	4
HP	3

d) E9.5 ATAC Sites (Howell, et al. 2021) : 99,297

	No. of Enhancer Positive ATAC sites Overlapping E9.5	Percentage of Enhancer Positive ATAC sites found in E9.5
ES	7,956	79.8%
HB	10,088	84.9%
HE1	16,030	88.5%
HE2	14,749	87.4%
HP	11,345	85.5%

E13.5 ATAC Sites (Howell, et al. 2021) : 96,536

	No. of Enhancer Positive ATAC sites Overlapping E13.5	Percentage of Enhancer Positive ATAC sites found in E13.5
ES	7,494	75.2%
HB	9,453	79.5%
HE1	14,918	82.4%
HE2	13,906	82.4%
HP	10,895	82.1%

Figure 4.5- Characterising enhancer features of ATAC fragments scoring positive in the enhancer screen.

Enhancer screen was conducted by Dr Benjamin Edginton-White. a) Percentage of distal ATAC-Seq sites which contain at least one positive reporter fragment in our assay for ESCs to HP. b) Percentage of promoter ATAC-Seq sites with at least one positive reporter fragment in our assay from ESCs to HP. c) The medium number of unique positive fragments per distal ATAC-seq site from ESCs to HP. d) Overlap of distal fragments which scored positive in our enhancer screen with open chromatin sites in haemogenic endothelium of mouse embryos (Zhu et al., 2020, Howell et al., 2021), the number of enhancer positive ATAC sites overlapping and the percentage of enhancer positive ATAC sites found for E9.5 and E13.5 shown. g) Percentage of ATAC sites which scored positive and negative/unknown which overlap with RNA Pol II binding sites (Gilmour et al., 2018).

The next figures show individual examples of gene loci for which enhancer elements were previously characterised by others. Genomic locations of the sequences which scored positive in our assay were loaded onto the USCS genome browser for visualisation. An example of visualisation is shown in Figure 4.6, which shows the well-characterised *Pu.1* (*Spi1*) locus and its annotation with *cis*-regulatory elements scoring positive in our screen (enhancers) being indicated by vertical bars. Our enhancer screen returns all enhancer elements which have been previously reported and reports their known cell type specific activity (Leddin et al., 2011). The data show the promoter, 5' URE, 3' URE, -12kb enhancer, the anti-sense promoter from the *Pu.1* (*Spi1*) locus.

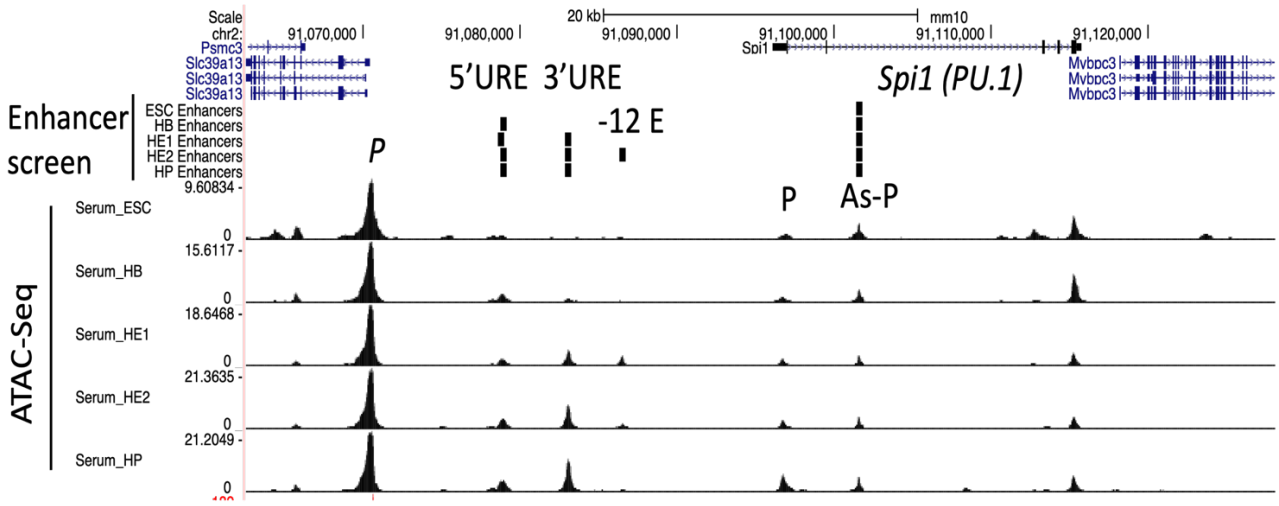


Figure 4.6- UCSC browser screenshot of the *Pu.1* (*Spi1*) locus showing cis-regulatory elements scoring positive in our enhancer screen.
Our enhancer screen returned the well characterised enhancers of this gene locus. The Promoter (P), 5' URE, 3'URE, -12kb enhancer (-12E), Promoter (P), and the antisense promoter (As-P) is shown. Distal *cis*-regulatory elements which scored positive in our enhancer screen are marked by vertical black bars (labelled Enhancer Screen). ATAC-Seq data for serum derived ESC, HB, HE1, HE2 and HP shown (labelled ATAC-Seq).

Figure 4.7 shows another example of the visualisation of the enhancers at the *Notch1* locus which scored positive in our enhancer screen assay. *Notch1* is required during blood cell specification and is expressed in HB, HE1 and HE2 cell types. Our analysis uncovered a highly complex regulatory region with multiple enhancer sequence which appears to be differentially active in development.

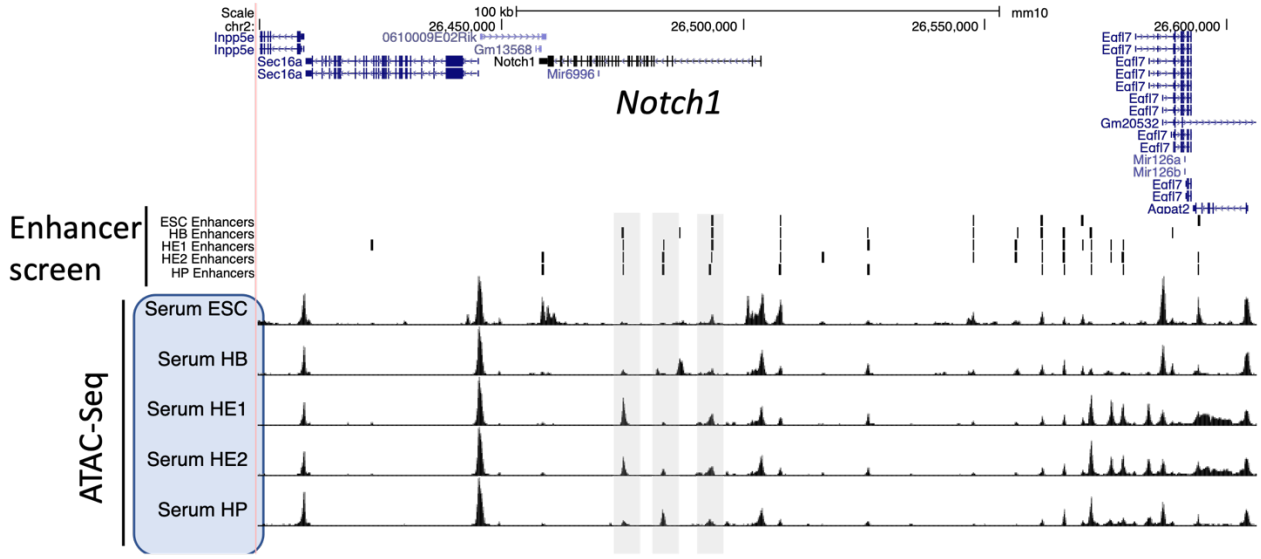


Figure 4.7- UCSC browser screenshot of the *Notch1* locus.
ATAC-Seq fragments which scored positive in our enhancer screen for each cell type are depicted by horizontal black bars (labelled Enhancer Screen). ATAC-Seq results for ESCs, HB, HE1, HE2 and HP cell types shown (labelled ATAC-Seq). Grey boxes highlight developmental-stage specific *cis*-elements.

4.1.4 Sequential stages of haematopoietic specification are defined by distinct enhancer sets which are associated with cell-type specific gene expression

We next correlated developmental stage-specific enhancer activity based on our enhancer screen data with the expression of their associated genes. Between 10 and 20% of all the distal sites and between 15% and 50% of all promoter sites displayed cell type specific activity in our screen as exemplified by the two examples shown above. We then used publicly available HiC and co-regulation data (Vijayabaskar et al., 2019, Novo et al., 2018, Wilson et al., 2016) to assign enhancer elements which scored positive in our assay to their rightful genes (as described in Methods 3.16). We then generated a matrix cataloguing enhancers as active (1) and inactive (0) at each of the five differentiation stages and plotted the expression level of associated genes in the different cell types as a heatmap (Figure 4.8). This heatmap shows that enhancer activity during development is continuous and that stage specific enhancer activity is strongly associated with stage specific gene expression.

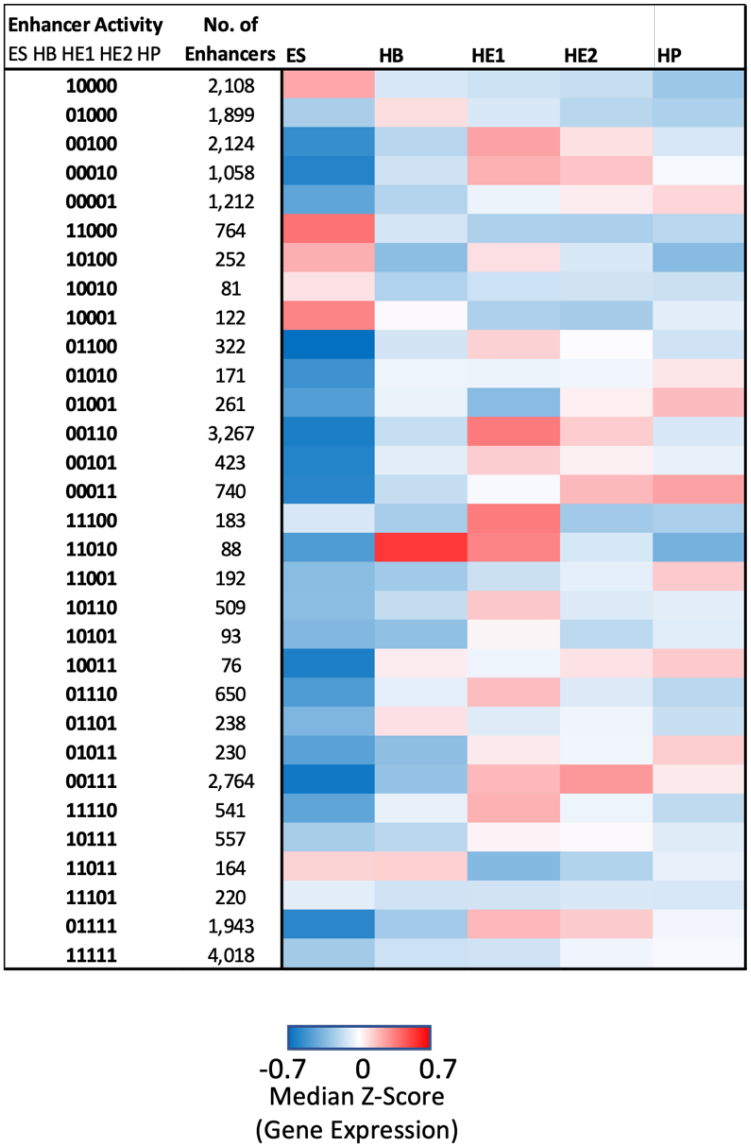


Figure 4.8- - Heatmap combining the presence or absence of enhancer activity with gene expression.

Presence or absence of enhancer activity in ESCs, HB, HE1, HE2 and HP expressed as a binary code (1= positive score, 0= negative score) and sorted by the activity pattern across the differentiation time course. The heatmap shows the activity of genes associated with these enhancer elements. Enhancer – promoter interactions were determined using the union of HiC data from ESCs and HPC7 (HP) cells (Wilson et al., 2016) and with co-regulation data defined by (Vijayabaskar et al., 2019) for a total of 21671 elements. 5599 promoters were not covered by these data and so were associated by being the nearest to the enhancer element.

4.2 Analysis of the control of enhancer elements at different developmental stages

4.2.1 Identification of transcription factors controlling cell type-specific enhancer activity

We next studied which transcription factors were involved in mediating cell stage specific activity of enhancer elements identified in our screen. We therefore performed a pair-wise relative motif enrichment analysis of enhancer elements which were unique to one cell type compared to another cell type (Filtering strategy shown in Figure 4.9a). We used HOMER to call motifs enriched in unique enhancer elements. A relative motif enrichment score was calculated and hierarchically plotted as a heatmap (Figure 4.9b).

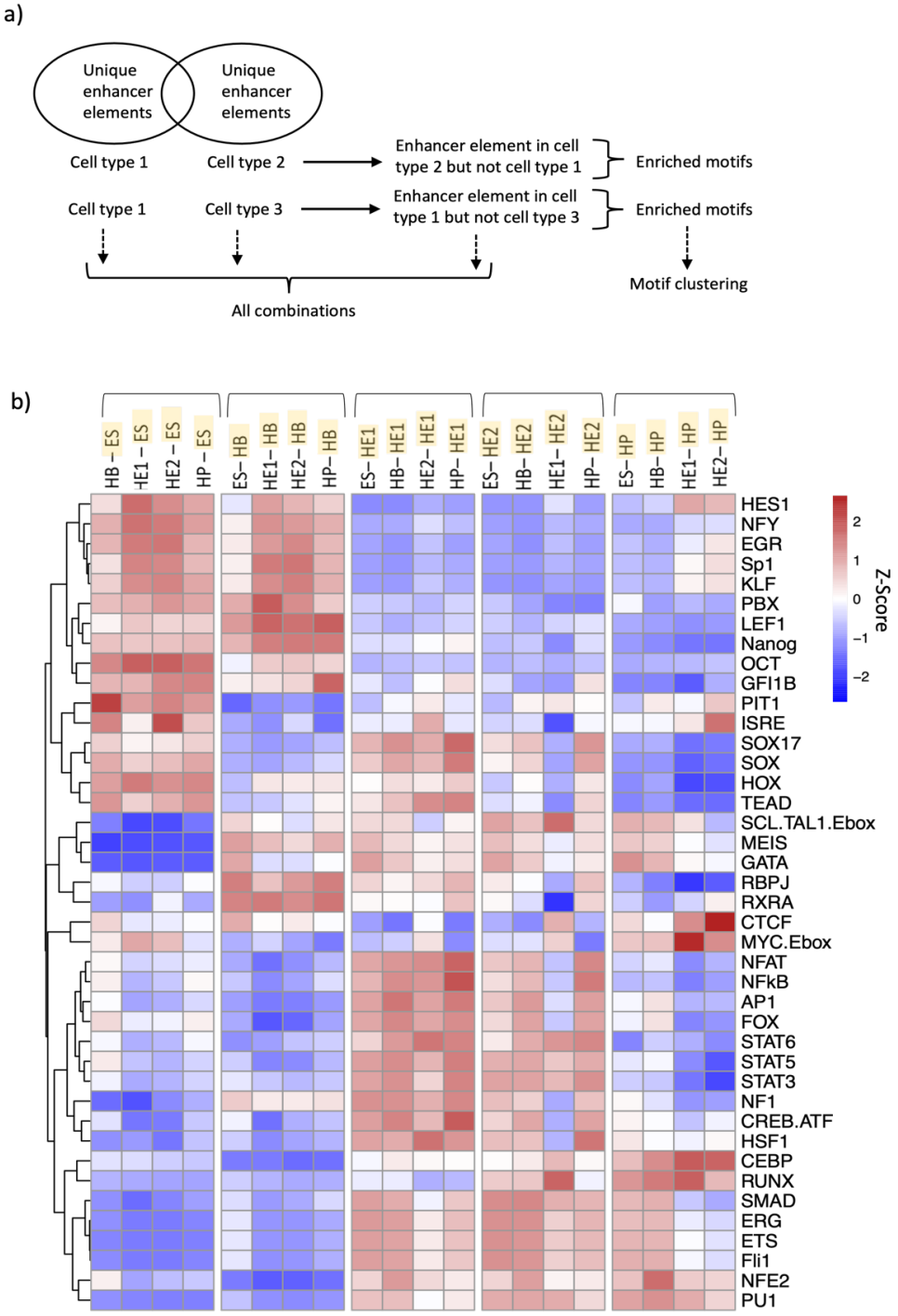


Figure 4.9- Hierarchical clustering of relative motifs enrichment analysis scores of enhancer elements identified in our screen.

a) Filtering strategy used for identifying enhancer elements uniquely active in each cell type in pair-wise manner. HOMER was then used to call motifs for each peak sets and a motif relative enrichment analysis score was calculated and taken for clustering. b) Heat map of relative motif enrichment scores from enhancer elements hierarchically clustered by TF motif. The relative motif enrichment analysis scores were converted to a Z-score by rows.

The top of the plot shows the cell type comparison and the relative motif enrichment analysis scores for peaks unique to the cell type highlighted in yellow. Cell types are grouped by brackets. The right of the plot shows the TF motifs used in the analysis. Dendrogram showing clustering of TF motifs on the left. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

Enhancer elements unique to ESCs were enriched for the pluripotency factors KLF, SOX and OCT (POU5F1) as well as HES and GFI1b motifs. Interestingly, in contrast to findings from relative motif enrichment analysis of all distal ATAC seq sites (Figure 4.2) CTCF motifs were depleted. This may indicate that CTCF motifs are depleted in functional enhancer elements compared to entire distal ATAC-seq sites, which is consistent with evidence that CTCFs mark TAD boundaries which gene regulatory elements sit within and may have a different function than transcription stimulation.

Enhancer elements which scored positive in HB cells were also enriched for KLF and OCT(POU5F1) pluripotency factors but showed an enrichment of RBPJ and HES motifs which are effectors of NOTCH signalling, together with GATA, SCL/TAL1 and MEIS motifs. This result confirms data from Goode et al (Goode et al., 2016) which show that expression and binding of these factors mark the onset of the haematopoietic program.

Enhancers elements scoring positive in HE1 were enriched for the endothelial TF motifs AP-1, TEAD, SOX17, ETS, NFKB STAT3,5,6 and FOX. Enhancers active in this cell type were depleted for pluripotency factors KLF and OCT (POU5F1). Enhancers which scored positive in HE2 maintained both an endothelial but also gained a haematopoietic TF motif signature.

We found an enrichment of endothelial TF motifs SOX, AP-1, ETS, FLI1 but TEAD TF motifs were not enriched compared to HE1 specific enhancers. Motifs for the haematopoietic TF motifs RUNX, PU.1 and NFE2 were starting to be enriched indicating that these enhancers may be activated during the EHT. After the EHT, enhancers uniquely active in HP cells compared to other cell types were depleted for the endothelial TF motifs SOX, AP-1, TEAD, STAT3,5,6 and the NOTCH signalling effectors RBPJ and HES1. Instead, these enhancers show strong TF motif enrichment for the haematopoietic TFs RUNX, CEBP, NFE2 and PU.1. Taken together, these analyses confirm the dynamic gene regulatory networks as described by the binding studies of Goode et al. (Goode et al., 2016) but identify upon which functional enhancer elements such factors act.

We next studied whether stage specific enhancer activity was regulated by cooperating transcription factors (Figure 4.10a-f). We therefore performed a motif co-localisation analysis for enhancers which were unique to each cell stage compared to all other stages. The motif colocalization analysis examined whether specific binding motif pairs located within 50bp of each other were enriched in cell type specific enhancer fragments as compared to all open chromatin sites. Motifs which showed an exact overlap were removed from the analysis (Analysis strategy shown in Figure 4.10a).

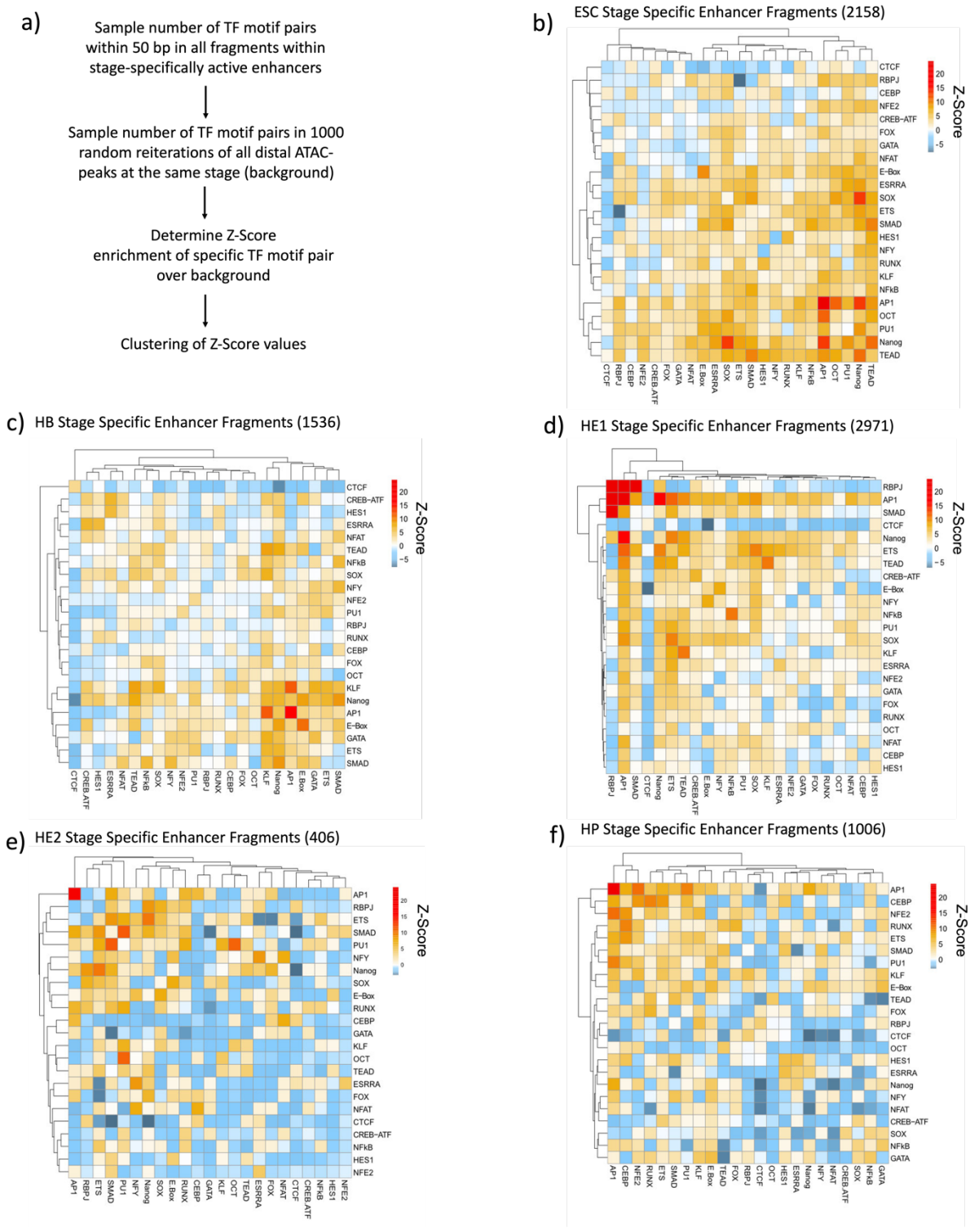


Figure 4.10- Cell type specifically active enhancer elements show a specific transcription factor motif co-localisation pattern.

a) Overview of motif co-localisation analysis pipeline for motif pairs co-localising within 50 base pairs in cell stage specific enhancer fragments. b-f) Transcription factor co-localisation analysis. Heatmaps of Z-score enrichments for pairs of TF binding motifs found with 50 base

pairs for ESC, HB, HE1, HE2 and HP. The number of fragments analysed indicated above each heatmap. The bottom and right of the heatmap label the binding motif combinations. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

Our data show that stage specific enhancer activity correlated with an enrichment of co-localisation of motifs for developmental-stage specific transcription factors. ESC stage specific enhancers showed the greatest number of transcription factors co-localising within 50bp which is in line with ES cells generally having a more open chromatin structure compared to later developmental stages (Figure 4.10b). Here motifs for transcription factors which maintain pluripotency such as NANOG, SOX2 and OCT4 showed a strong co-localisation. AP-1 motifs also co-localised with both itself and OCT and NANOG motifs potentially linking their activity to signalling processes.

The co-localisation of AP-1 motifs with themselves was also a feature of HB specific enhancers (Figure 4.10c). In addition, AP-1 motifs were co-localising with KLF motifs. NANOG motifs were co-localised with TEAD, SMAD, GATA and E-BOX motifs. Interestingly, there was a co-localisation of SMAD and GATA motifs. BMP4 directed SMAD signalling upregulates GATA in the HB and HE (Kirmizitas et al., 2017) and the co-localisation of these motifs at enhancers from these cell types may reflect a regulatory mechanism between SMAD and GATA TFs. We found no enrichment of co-localising haematopoietic TF motifs (i.e. RUNX/PU1/NFE2) in this cell type as expected.

HE1 specific enhancers showed a significant enrichment of binding motifs for the NOTCH signalling effector RBPJ with itself and AP-1 and SMAD transcription factors. This result is

consistent with findings that NOTCH signalling is established and active in this cell type (Uenishi et al., 2018). Motifs for the VEGF signalling responsive transcription factor family AP-1 also co-localised with HES1 motifs which is another effector of the NOTCH signalling pathway, suggesting that enhancers active in this cell type are co-regulated by VEGF directed AP-1 and NOTCH directed RBPJ/HES1 transcription factor activity. BMP4 signalling responsive SMAD motifs co-localising with NFκB, SOX, TEAD and ETS motifs were enriched as well. In this cell type AP-1 motifs were found to co-localise with TEAD motifs which was consistent with previous findings from our group (Obier et al., 2016).

In HE2 specific enhancers, co-localising AP-1 motifs were found to be significantly enriched. AP-1 and TEAD factors co-localisation was less pronounced in this cell type compared to in HE1, again consistent with previous findings from our group (Obier et al., 2016). Enhancers showed a marked reduction in the enrichment of motifs for the NOTCH signalling effectors RBPJ and HES1 co-localising with other TFs when compared to HE1, indicating a reduction in reliance of NOTCH signalling to regulate enhancers specifically active in this cell type. This finding is in concordance with literature suggesting a loss of NOTCH signalling activity in cells undergoing the EHT (Del Monte et al., 2007, Richard et al., 2013). Instead, we observed an enrichment in RUNX motifs co-localising with motifs for AP-1 and its target PU.1 in this cell type.

In HP cells, which have undergone the EHT and have switched from an endothelial to a haematopoietic TF identity, we observed an increase in enrichment of RUNX1 motifs co-localising with motifs for AP-1, GATA, TEAD, FOX, CEBP and its direct targets PU.1 and NFE2.

In contrast, motifs for endothelial TF motifs such as for RBPJ, HES, and SOX were lost.

Taken together these analyses highlight that the changes in the gene regulatory networks during haematopoietic specification are accompanied by a profound difference in the TF collaboration.

4.2.2 Validation of individual elements in serum *in vitro* differentiation cultures

In order to validate the function of selected functionally identified enhancer element candidates from our high-throughput screen we tested their individual activity in the enhancer reporter system using the serum I.V.D culture. Enhancers were selected using six criteria (i) the enhancer element scored as positive in HB, HE1, HE2 or HP cells, (ii) the enhancer element was assigned to a gene whose expression profile showed it was expressed in one of these cell types (iii) there was published literature about the genes function, (iv) their enhancer was expected to be VEGF signalling responsive (discussed in Results 4.3.8) and (v) primers could be designed to amplify the 400 base pair enhancer region from genomic DNA. 9 candidate enhancer elements were selected for serum I.V.D validation which were enhancers assigned to *Sparc*, *Eif2b3*, *Hspg2*, *Pxn*, *Cdh5*, *Dlk1*, *Mrpl15* and two for the gene for *Runx1.* ES cells were transfected with these constructs and reporter activity was measured by flow cytometry.

Sparc encodes a cysteine-rich acidic matrix-associated protein which modulates cell-extracellular matrix interactions and is involved in angiogenesis (Rivera et al., 2011). Its

importance in haematopoiesis has been studied in zebrafish using morpholino antisense oligonucleotide-based knockdown which resulted in specific erythroid progenitor cell differentiation defects (Ceinos et al., 2013). *Sparc* is expressed in HE1 and HE2 with ATAC-Seq sites present in these cell types for this enhancer (Chr11:55405590-55406032). The enhancer (indicated by a grey bar) scored as active in HE1, HE2 and HP in our enhancer screen. The enhancer was validated as able of driving the reporter gene activity with the highest activity seen in HE1 (Figure 4.11a-c).

Eif2b3 encodes a subunit for the initiation factor eIF2B and is expressed from ESC through to HP. Eif2b3 is one of the five subunits of the eIF2B pentameric complex (*Eif2b1*, *Eif2b2*, *Eif2b3*, *Eif2b4*, *Eif2b5*) (Smit et al., 2021). Serum ATAC-Seq revealed accessible chromatin in HE1, HE2 and HP overlapping with this enhancer (Chr4:117056629-117057073). The enhancer scored as positive in our assay in HE1, HE2 and HP and validation showed that the enhancer can drive the activity of the reporter gene in HE1 and HE2. These data indicate that this genes activity is regulated by this enhancer in HE1-HP but is regulated by other elements in ESCs and HB (Figure 4.11d-f).

Hspg2 encodes a perlecan protein which acts to bind and cross-link many extracellular matrix components and cell-surface molecules. *Hspg2* is expressed in HSPC niches by human and murine bone marrow stromal cells, however, it's direct role in haematopoiesis has not been reported (Lee-Thedieck et al., 2022). Perlecan has been shown to bind to granulocyte macrophage colony stimulating factor (GM-CSF) and present to HP cells in the niche (Klein et al., 1995). *Hspg2* is expressed in HE1 and HE2 and the enhancer element

overlaps with an ATAC-Seq site accessible in HE1 and HE2 (Chr4:137478251-137478690). The enhancer scored as positive in HE1 and HE2 and validation of the *Hspg2* enhancer showed that the element drives the highest activity of the reporter gene in HE1 and is then down-regulated in HE2 and HP which precisely follows the gene expression profile (Figure 4.12a-c).

Pxn encodes a cytoskeletal protein involved in actin-membrane attachment at sites of cell adhesion to the extracellular matrix and is important for many morphogenetic processes (Evans et al., 2003). *Pxn* expression increases from HB to HE1 through to HP and the enhancer element which scored positive in our assay in HE1, HE2 and HP (Chr5:115530006-115530483) overlaps with ATAC-Seq sites in those cell types (Figure 4.12d-f).

Dlk1 encodes a transmembrane protein which is a repressor of NOTCH signalling. *Dlk1* has been found to be a negative regulator of emerging haematopoietic stem and progenitor cells (Mirshekar-Syahkal et al., 2013) and is expressed in the smooth muscle layer of the dorsal aorta and the ventral sub-aortic mesenchyme. Dlk1 expression increases from HB peaking in HE1 before falling again in HE2 and HP. The enhancer scored as positive in ESCs, HB, HE1, HE2 and HP cells and overlaps with ATAC-Seq sites in these cell types (Chr12:109437601-109438085). The validation confirmed that the enhancer was able to stimulate the reporter gene activity in HB-HP cells (Figure 4.13a-c).

Mrpl15 is a gene encoding a 39s subunit protein that belongs to the EcoL15 ribosomal protein family and along with other mitochondrial environment genes beginning to be

recognised as important for various conditions (Giannos et al., 2022). Mrpl15 expression decreases from ESCs to HE1 before increasing slightly again in HE2 and then falling again in HP. The enhancer scored positive in HE2 in our screen and overlaps with ATAC-seq sites in HE2 (Chr1:4760408-4760960). The validation found that the enhancer was most active in HE1 and has the second highest value in HE2 (Figure 4.13d-f).

As discussed previously, *Runx1* encodes a master regulator of haematopoiesis which drives the EHT process (previously discussed in detail in Introduction 1.4.3). *Runx1* is upregulated in HE2 before being significantly upregulated in HP cells. A number of enhancer elements within the *Runx1* gene scored positive in our enhancer screen. Two enhancers were selected for validation and further interrogation, a +3.7kb *Runx1* enhancer (chr16:92822182-92822601) and the already well characterised +23kb *Runx1* (chr16:92,801,538-92,802,049) enhancer (Bee et al., 2009). The +3.7kb enhancer overlaps with an ATAC-Seq site present in HE1, HE2 and HP cells. Individual validation found that the enhancer could drive the activity of the reporter gene in HE1, HE2 and HP where it was strongly up-regulated (Figure 4.14a-c). The +23kb enhancer was active in HE1, HE2 and HP cells in concordance with ATAC-Seq sites overlapping the enhancer in these cell types. Validation in serum culture demonstrated that the enhancer could drive the reporter gene activity in HE1 and HE2 before its activity was reduced in HP (Figure 4.14a-c). The analysis of *Runx1* enhancers therefore revealed an interesting cooperation of enhancer elements which leads to the up-regulation of RUNX1 after the EHT.

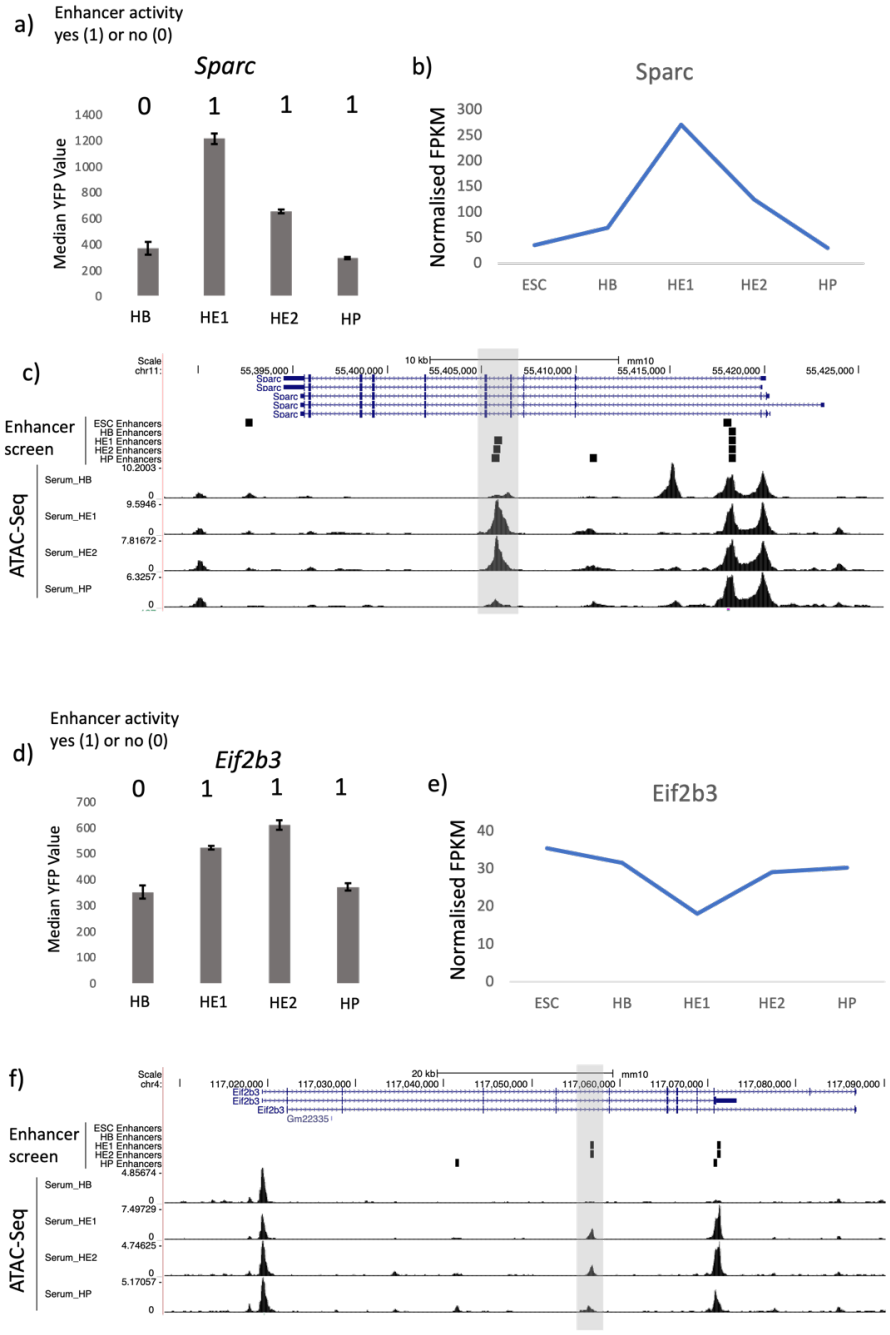


Figure 4.11- Individual in vitro validation of *Sparc* and *Eif2b3* enhancer elements in serum culture.

a) Validation of the indicated *Sparc* enhancer measured as median YFP value in HB, HE1, HE2 and HP. Binary code of enhancer scoring positive (1) or negative/unknown (0) in our enhancer screen shown above the bars for each cell type. Error bars show standard deviation, n=3. b) Expression profile of *Sparc* in ESCs, HB, HE1, HE2 and HP (RNA-Seq data from (Gilmour et al., 2018)). c) UCSC browser screenshot of the *Sparc* locus showing the enhancer scoring positive in our assay as vertical black bars highlighted in grey (Enhancer Screen) and serum derived ATAC-Seq for HB-HP. d) Validation of *Eif2b3* enhancer in serum

culture measured as median YFP value in HB, HE1, HE2 and HP. e) Expression profile of *Eif2b3* in ESCs, HB, HE1, HE2 and HP (RNA-Seq data from (Gilmour et al., 2018)). f) UCSC browser screenshot of the *Eif2b3* locus. All other features as in a).

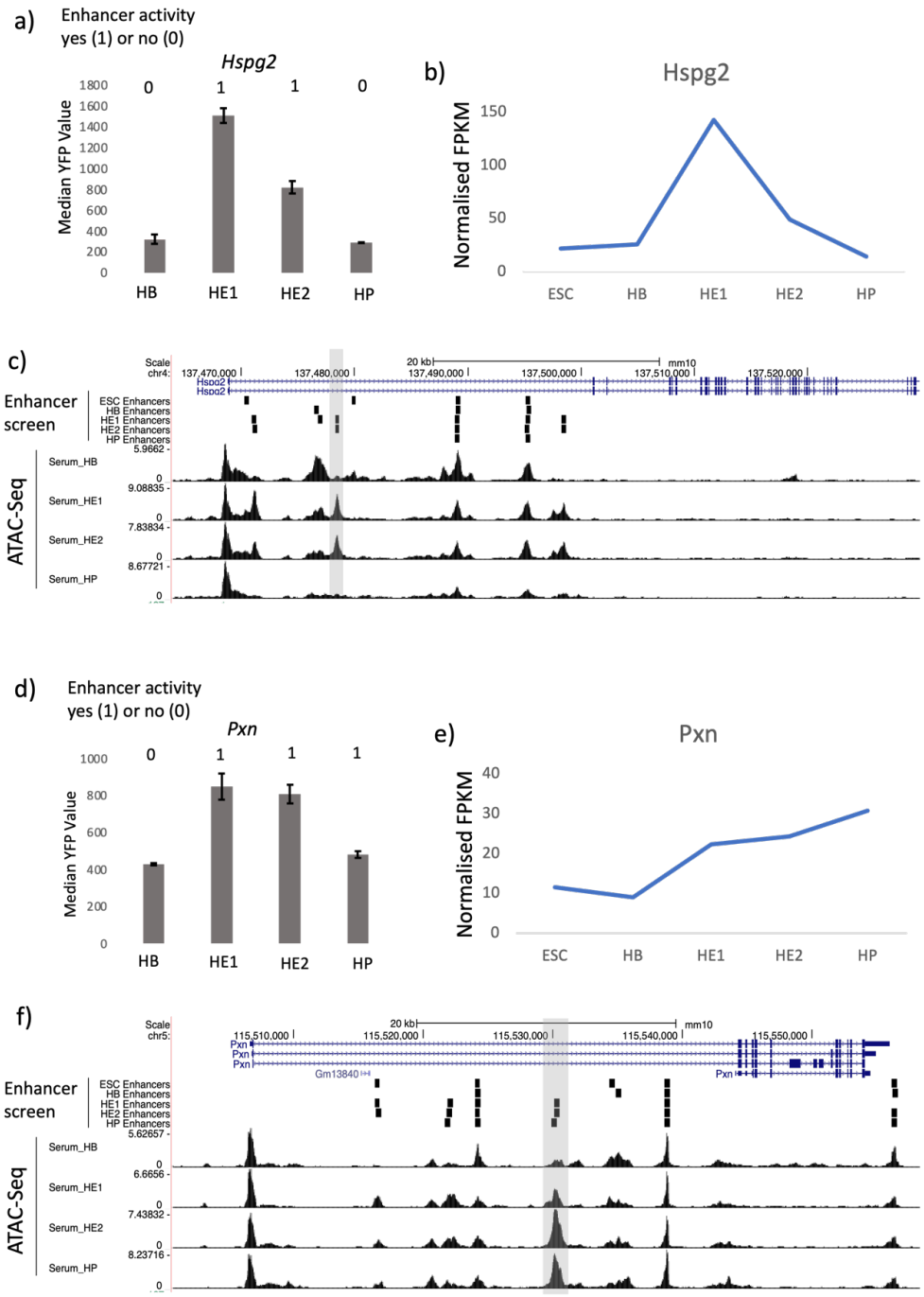


Figure 4.12- Serum in vitro validation of *Hspg2* and *Pxn* enhancer elements.
a) Serum I.V.D validation of *Hspg2* enhancer measured as median YFP value in HB, HE1, HE2 and HP. Binary code of enhancer scoring positive (1) or negative/unknown (0) in our enhancer screen shown above the bars for each cell type. Error bars show standard

deviation, n=3. b) Expression profile of *Hspg2* in ESCs, HB, HE1, HE2 and HP (RNA-Seq data from (Gilmour et al., 2018)). c) UCSC browser screenshot of the *Hspg2* locus showing the enhancer scoring positive in our assay as vertical black bars highlighted in grey (Enhancer Screen) and serum derived ATAC-Seq for HB-HP. d) Serum I.V.D validation of *Pxn* enhancer measured as median YFP value in HB, HE1, HE2 and HP. e) Expression profile of *Pxn* in ESCs, HB, HE1, HE2 and HP (RNA-Seq data from (Gilmour et al., 2018)). f) UCSC browser screenshot of the *Pxn* locus. All other features as in a).

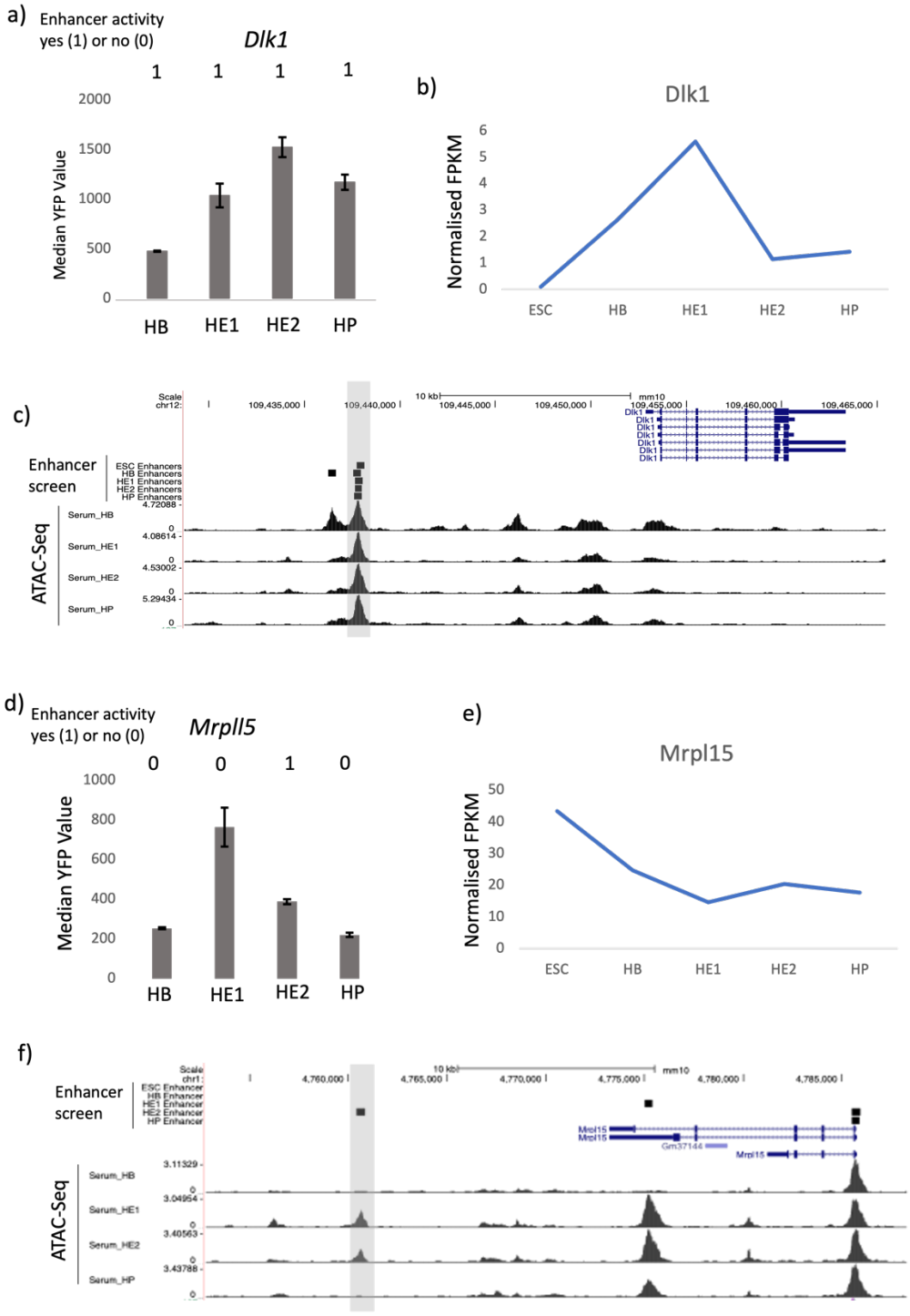


Figure 4.13- Serum in vitro validation of *Dlk1* and *Mrpl15* enhancer elements.
a) Serum I.V.D validation of *Dlk1* enhancer measured as median YFP value in HB, HE1, HE2 and HP. Binary code of enhancer scoring positive (1) or negative/unknown (0) in our enhancer screen shown above the bars for each cell type. Error bars show standard deviation, n=3. b) Expression profile of *Dlk1* in ESCs, HB, HE1, HE2 and HP (RNA-Seq data from (Gilmour et al., 2018)). c) UCSC browser screenshot of the *Mrpl15* locus showing the

enhancer scoring positive in our assay as vertical black bars highlighted in grey (Enhancer Screen) and serum derived ATAC-Seq for HB-HP. d) Serum I.V.D validation of *Mrpl15* enhancer measured as median YFP value in HB, HE1, HE2 and HP. e) Expression profile of *Mrpl15* in ESCs, HB, HE1, HE2 and HP (RNA-Seq data from (Gilmour et al., 2018)). f) UCSC browser screenshot of the *Mrpl15* locus. All other features as in a).

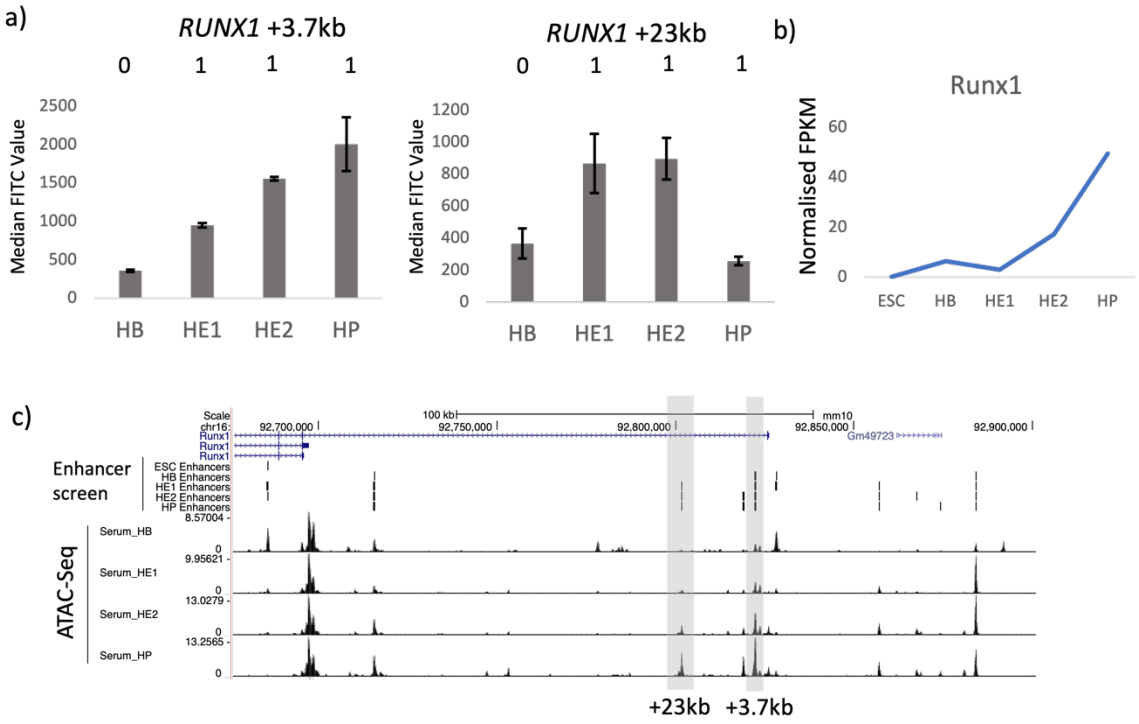


Figure 4.14- Serum in vitro validation of *Runx1* +3.7kb and +23kb enhancer elements.
a) Serum I.V.D validation of *Runx1 +3.7kb and +23kb* enhancer elements measured as median YFP value in HB, HE1, HE2 and HP. Binary code of enhancer scoring positive (1) or negative/unknown (0) in our enhancer screen shown above the bars for each cell type. Error bars show standard deviation, n=3. b) Expression profile of *Runx1* in ESCs, HB, HE1, HE2 and HP (RNA-Seq data from (Gilmour et al., 2018)). c) UCSC browser screenshot of the *Runx1* locus showing both enhancers scoring positive in our assay as vertical black bars highlighted in grey (Enhancer Screen) and serum derived ATAC-Seq for HB-HP.

4.3 Using serum free in vitro differentiation to identify cytokine responsive enhancers

Cell differentiation involves signalling factors which direct changes in growth and differentiation state by changing gene expression. Signalling components such as cytokines, their receptors, integrins and kinase molecules are well characterised contributors to blood cell development. However, less is known about how different signals are integrated at the level of the genome and how they regulate gene expression. In this study we aimed to investigate how cytokines drive signalling responsive transcription factor activity at their target enhancer elements during mouse ES cell differentiation to haematopoietic progenitors. To address this question, an *in vitro* differentiation culture system which allows for the tight control of signalling factors (cytokines) present in the system is required. Serum based *in vitro* differentiation systems contain foetal calf serum which contains a range of cytokine factors at undefined concentrations making the study of how specific signals drive specific transcription factor activity at their target regulatory elements extremely challenging. We therefore employed a previously published serum free *in vitro* differentiation system capable of driving mouse ES cells to HP (Pearson et al., 2015).

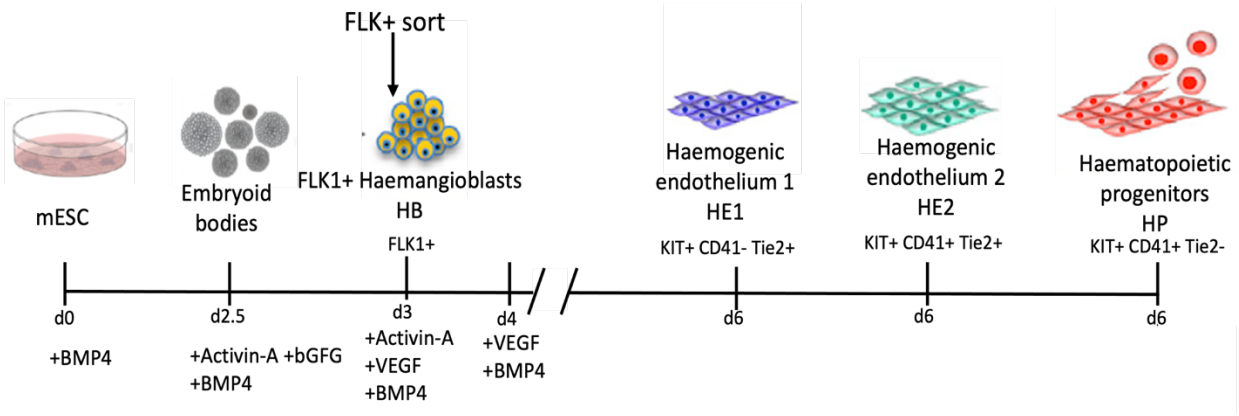


Figure 4.15- Schematic representation of experimental serum free culture strategy.
Serum free culture strategy from (Pearson et al., 2015). Mouse ESCs were differentiated to embryoid bodies (EB) in serum free culture by the addition of BMP4 at day 0. Activin-A and bFGF was then added at day 2.5 driving the expression of FLK-1 marking Haemangioblasts (HB). HB were then sorted for by FLK-1+ sort and plated on gelatinised plates in serum-free media in the presence of Activin-A, BMP4 and VEGF to form haemogenic endothelium (HE1, HE2). After 24 hours Activin-A was removed. By day 2 blast culture (day 6 culture) floating haematopoietic progenitors had formed. Adapted from (Goode et al, 2016).

Figure 4.15 gives a schematic overview over the serum free I.V.D culture system. Pearson et al (Pearson et al., 2015) demonstrated that this system drives the differentiation of mESCs to HP through the stepwise addition of cytokines. Here, the addition of bone morphogenic protein (BMP4), basic fibroblast growth factor (bFGF), Activin-A and vascular endothelial growth factor (VEGF) was enough to drive the differentiation of mouse ESCs to haematopoietic progenitors. BMP4 induced the formation of the mesoderm by day 2.5, addition of Activin-A and bFGF was enough to drive the development of FLK-1+ haemangioblasts enriched cells which were then sorted using anti-FLK-1 biotin beads. Plating of FLK-1+ haemangioblasts on gelatinised plates in the serum free media with the addition of Activin-A, BMP4 and VEGF (blast culture) resulted in the development of the haemogenic endothelium and also for the production of floating progenitor cells. The system therefore allows for a controlled differentiation method using a reduced number of

known signalling factors at a known concentration, thus allowing for the study of how specific factors drive specific transcription factor activity and at which regulatory elements.

4.3.1 Activin-A represses the emergence of haematopoietic progenitors in the serum free I.V.D system

We found that culturing our cell lines (HM-1 and A7) in the serum free system established by Pearson et al (Pearson et al., 2015) gave rise to very few (<2%) floating progenitors when following their exact timing of addition and withdrawal of cytokines used in this culture system. However, for this study we aimed to perform genome-wide omics analysis such as ATAC-Seq and transcriptomics of HB, HE1, HE2 and HP populations which required a sufficient number of cells from each cell type (>5x10^5 cells). Pearson et al (Pearson et al., 2015) reported that Activin-A withdrawal at 1 day (taken as 24 hours) was required to allow for the emergence of floating KIT+ TIE2- CD41+ haematopoietic progenitors and that there was an enrichment of haemogenic endothelium in the system (KIT+ TIE2+ CD41-) if Activin-A was not removed.

We therefore removed Activin A at different timepoints to determine when precisely to remove Activin-A (Figure 4.16a-c). We differentiated mouse ESCs to FLK-1 expressing haemangioblasts through addition of BMP4 at D0, Activin -A and bFGF at D2.5. FLK-1+ haemangioblasts were then sorted for and replated in serum free blast culture on gelatinised plates in the presence of BMP4 and VEGF. Activin-A was then withdrawn at 0h, 1h, 2h, 4h, 6h, 10h, 20h, 24h and finally 72h (left in culture). Blast cultures were then

collected for FACS analysis at day 1 (24 hours), day 2 (48 hours) and day 3 (72 hours) to assess the proportion of HE1 (expressing KIT and TIE2), HE2 (expressing KIT, TIE2 and CD41) and HP (expressing KIT and CD41) in the system. We found that maintaining Activin-A in the system up to 20 hours severely reduced the number of HP cells gained from the system at day 3 blast culture. The greatest number of gained HP gained occurred when Activin-A was removed after 10 hours (Figure 4.16a). Light microscopy revealed an increase in the number of floating progenitors in the system when Activin-A was removed at 10 hours compared to 24 hours (Figure 4.16c-d). We therefore amended our culture system protocol to remove Activin-A at 12 hours post blast culture plating.

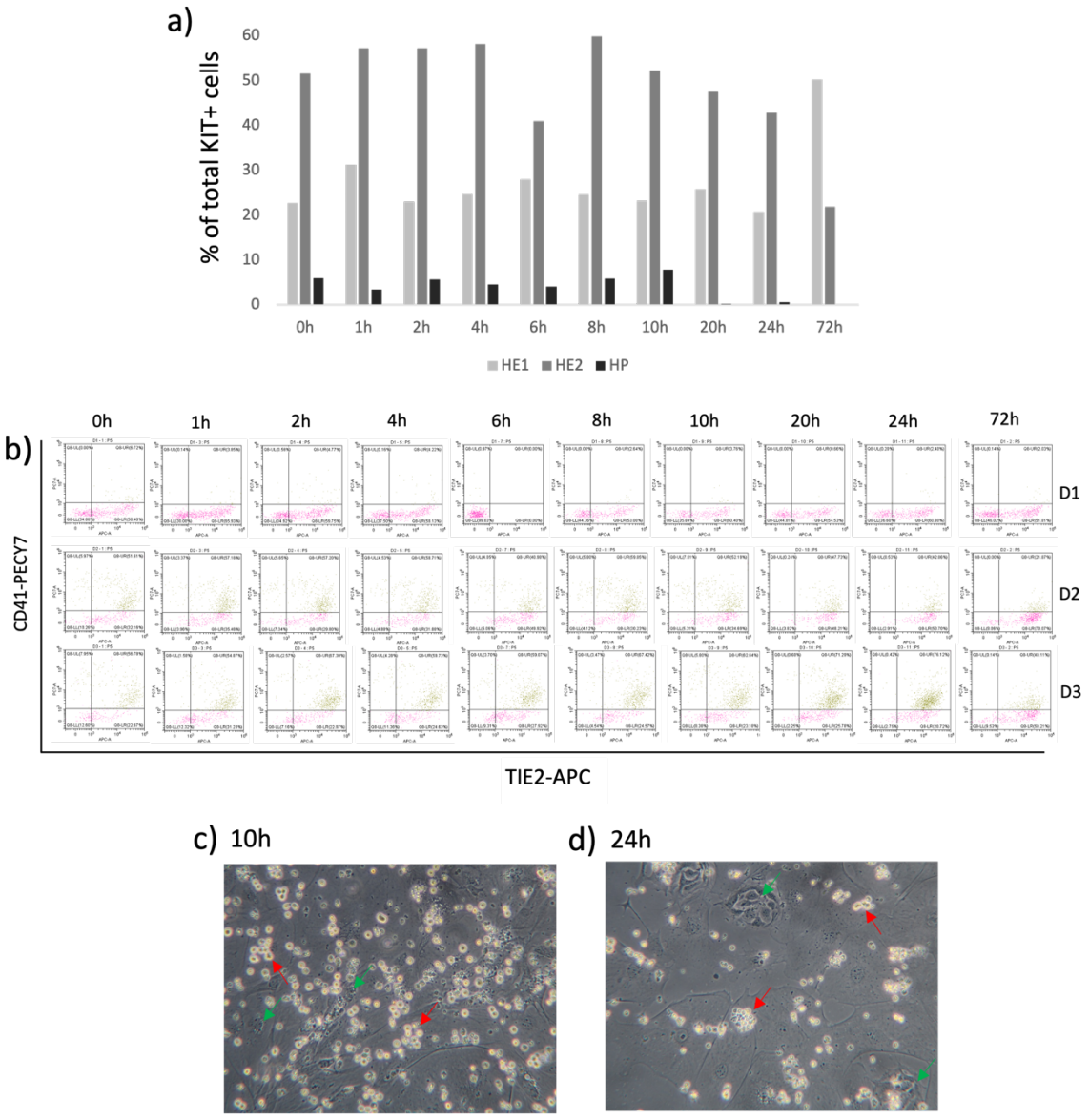


Figure 4.16- - Activin-A represses the emergence of the haematopoietic progenitor cell in serum free I.V.D.

a) Time course experiment of removal of Activin-A from serum free blast culture at 0h (not added), 1h, 2h, 4h, 6h, 8h, 10h, 20h, 24h and 72h post blast culture plating (n=1) (A7 cells). Percentage of total KIT expressing cells classified as HE1 (expressing KIT and TIE2), HE2 (expressing KIT, TIE2 and CD41) and HP (expressing KIT and CD41) from day 3 blast culture. b) Representative FACS plots of serum free blast cultures at day 1 (D1), day 2 (D2) and day 3 (D3) with Activin-A removed at the time points shown. c) Light microscope image of serum free blast culture at day 3 from cultures with Activin-A removed at 10 hours and d) at 24 hours (100x magnification). Green arrows show haemogenic endothelium, red arrows show floating haematopoietic progenitors.

4.3.2 Addition of SCF, IL6, IL3 and TPO to the serum free I.V.D system stimulates the differentiation of HE1 to HE2 and HP

Even with the altered timing of Activin-A withdrawal allowing for the emergence of haematopoietic progenitor cells, insufficient numbers of HE2 and HP cells were being generated for downstream applications. In order to increase the number of HE2 and HP populations additional cytokines were added with the aim of driving further lineage commitment of the HE1 cell type through to HE2 and then HP cells. To achieve this the haematopoietic cytokines SCF (KIT-ligand) (100ng/ml), IL6 (10ng/ml), IL3 (1ng/ml) and TPO (5ng/ml) were added to day 0.5 blast culture (after Activin-A removal).

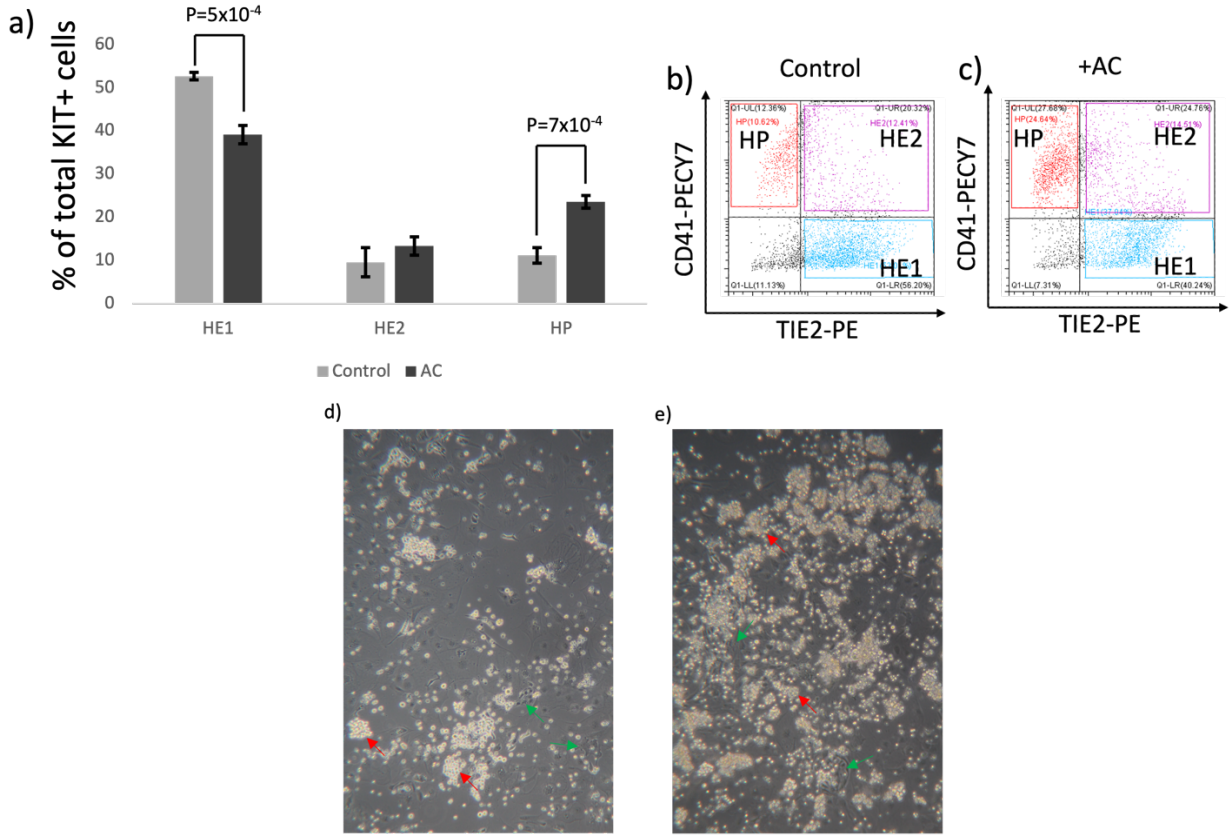


Figure 4.17- Optimisation of the serum free in vitro differentiation system.
a) Bar chart of haemogenic endothelium type 1 (HE1), haemogenic endothelium type 2

(HE2) and haematopoietic progenitors (HP) with the additional cytokines (+AC) SCF, IL6, IL3 and TPO and without (Control) day 3 blast culture (n=3, error bars indicate standard deviation) (A7 cells). b) FACS analysis of KIT expressing cells without additional cytokines (Control) gated by CD41 and TIE2 from day 3 blast culture. c) FACS plot of KIT expressing cells cultured with additional cytokines (+AC) gated by CD41 and TIE2 from day 3 blast culture. d) Light microscope image of blast culture without additional cytokines (Control), green arrows indicate haemogenic endothelium, red arrows indicate floating progenitor cells. e) Light microscope image of blast culture with additional cytokines (+AC), green arrows indicate haemogenic endothelium, red arrows indicate floating progenitor cells (100x magnification).

We found by FACS analysis that the addition of SCF, TPO, IL3 and IL6 to day 0.5 blast culture produced suitable numbers of HE1 cells gated on the cell markers KIT and TIE2 and HE2 cells gated on the cell markers KIT, TIE2 and CD41 at day 3 blast culture. FACS analysis also revealed that addition of SCF, IL6, IL3 and TPO resulted in a greater than 2-fold increase in the number of floating haematopoietic progenitors expressing KIT and CD41 over the control ($P=7x10^{-4}$) at day 3 blast culture. In parallel, we saw a decrease in the number of HE1 cells with the additional cytokines, indicating that these cells were differentiating to HE2 cells and on to HP cells. This process can also be seen in Figure 4.17b-c as the proportion of HE1 cells is greater in the control over the additional cytokines condition. Investigation by light microscopy showed a clear increase in the number of floating haematopoietic progenitor cells in the presence of the additional cytokines (Figure 4.17e) compared to the control (Figure 4.17d).

4.3.3 Identification of signalling responsive distal ATAC-seq sites

After optimising the serum free I.V.D system we investigated which enhancer elements were the genomic targets of signalling factors (in this case BMP4, VEGF, IL6 and IL3) during ESCs differentiation to haematopoietic progenitors and to elucidate the role of different cytokines on chromatin programming. Therefore, we employed the optimised serum free I.V.D system to differentiate ESCs to HP and withdrew either BMP4, VEGF, IL6 or IL3 from blast culture, meaning that the cytokines were not added to the blast culture after FLK-1+ cell sorting (Figure 4.18).

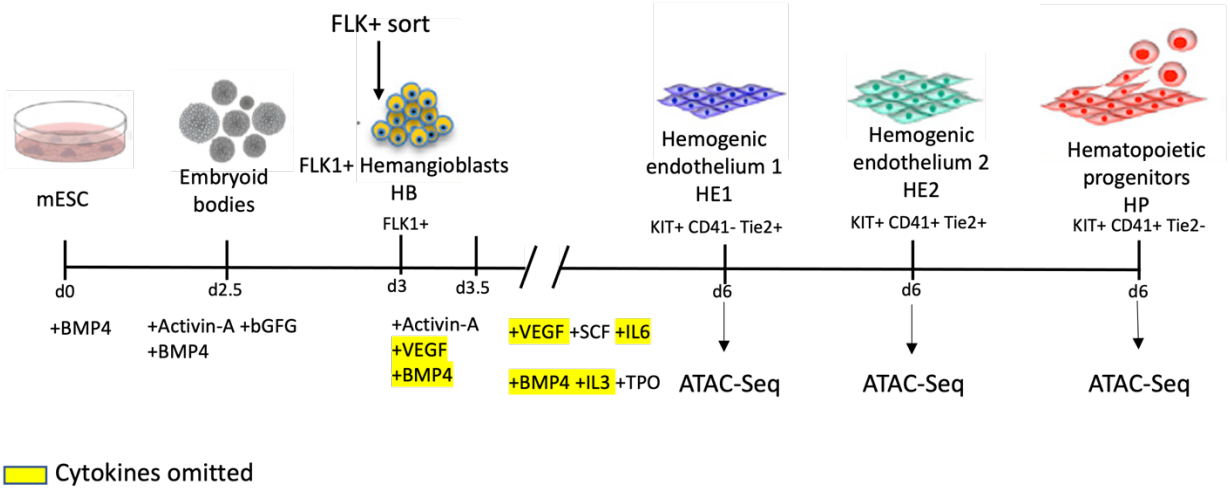


Figure 4.18- Identification of signalling responsive distal elements.
Overview of the optimised serum free *in vitro* differentiation system modified from (Pearson et al., 2015). Specific cytokines that were left out at the beginning of blast culture are highlighted in yellow. Adapted from (Goode et al, 2016).

Day 3 blast culture cells (day 6 within the differentiation protocol) were collected and cells were sorted for HE1 (expressing KIT and TIE2), HE2 (expressing KIT, TIE2 and CD41) and HP expressing (KIT and CD41). All sorted cells were taken for ATAC-sequencing. HOMER motif

analysis was then run on distal (defined as being +/- 1.5kb from a TSS) cytokine responsive ATAC-Seq peaks to examine which TF binding motifs were enriched in these peaks. Cytokine responsive peaks were defined as peaks which were at least 2-fold higher (enriched) or lower (depleted) in the presence of the cytokine. We defined these open chromatin regions as being signalling responsive. HB, HE1, HE2 and HP cells were collected for the All cytokines condition and after withdrawal of VEGF, IL6 or IL3. No HP cells were formed in the absence of BMP4.

4.3.4 Identification of enriched transcription factor binding motifs in cytokines responsive distal ATAC-seq sites

Next, we aimed to identify which transcription factors were responsive to the different cytokines in our system. To this end, we filtered for distal ATAC sites and performed HOMER motif enrichment analysis as described in Methods 3.18 on our signalling responsive distal ATAC sites. ATAC peaks from the control (all cytokines) and withdrawal cytokine condition were ranked by fold change and heat maps were plotted for each comparison.

4.3.4.1 Open chromatin regions dependent on BMP4 signalling are enriched for GATA TF motifs

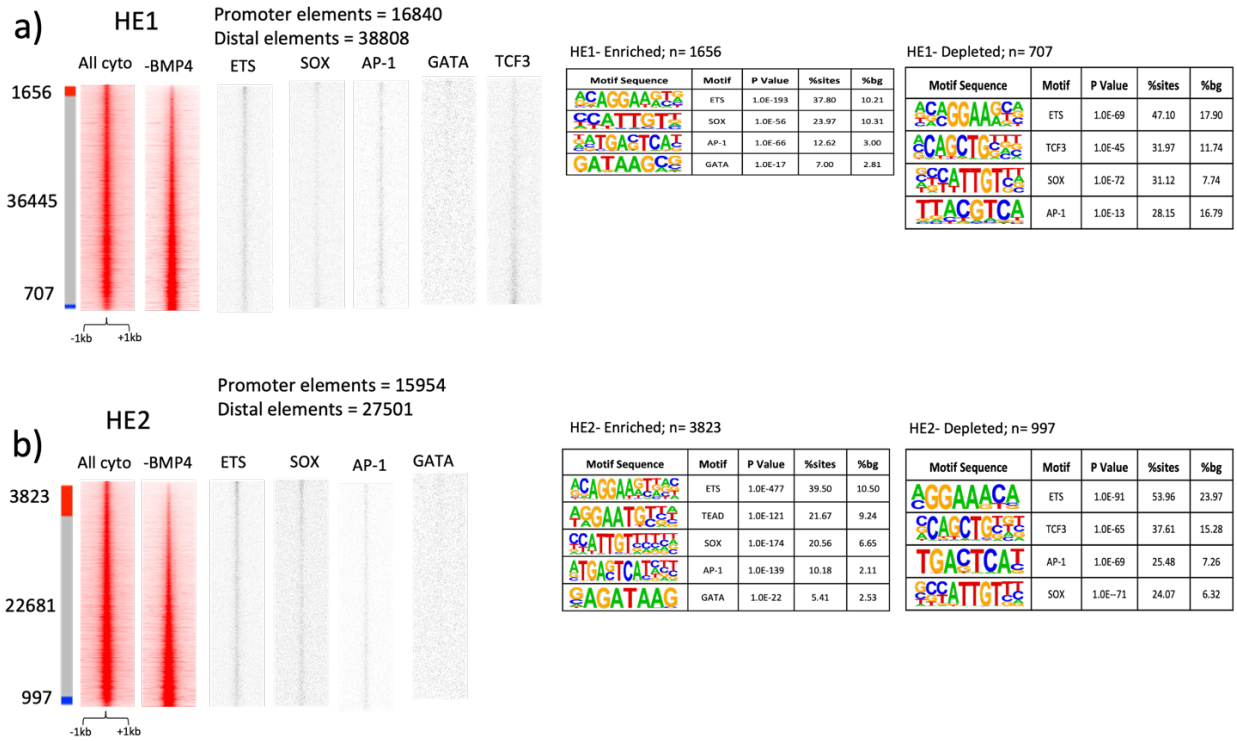


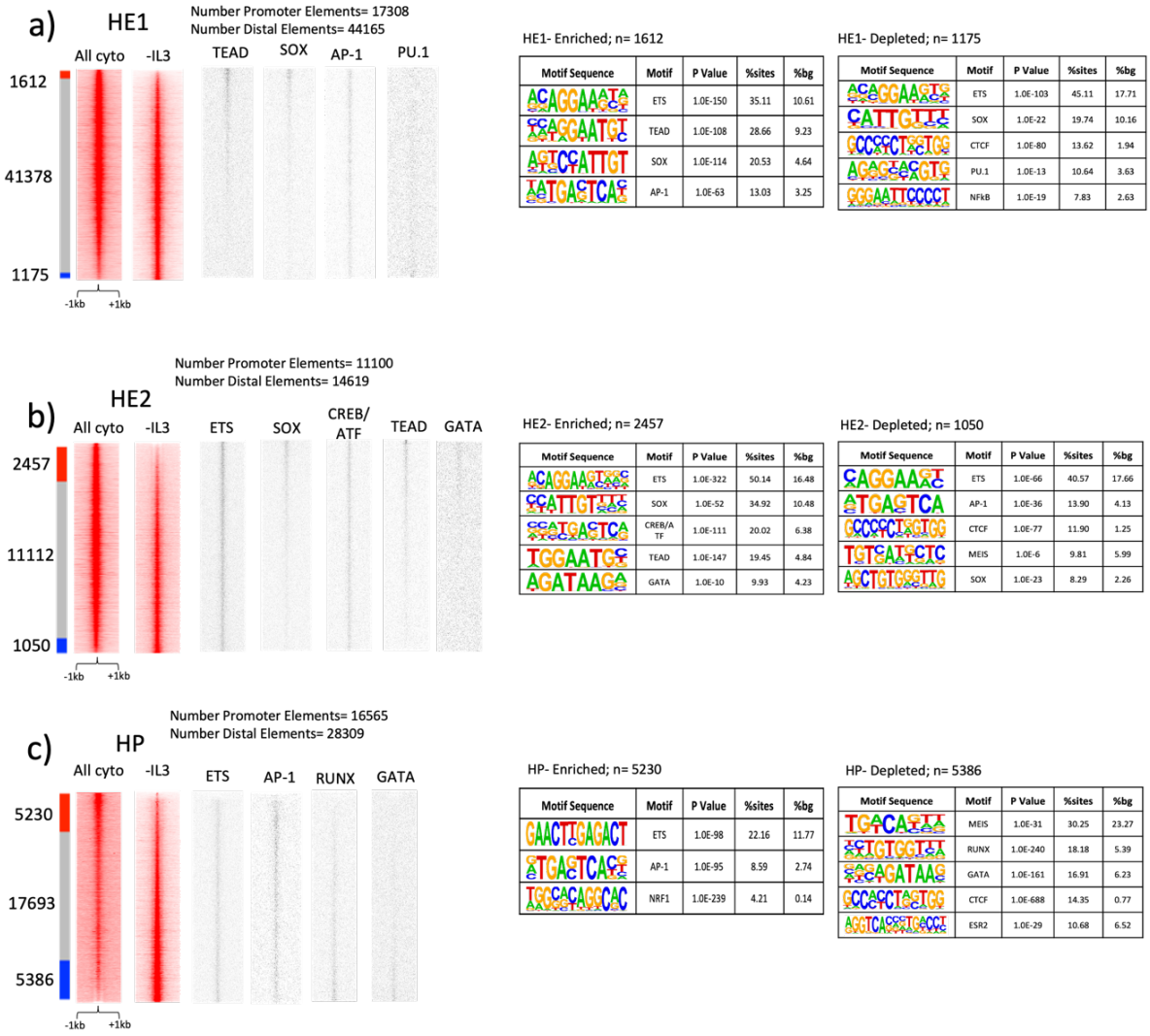
Figure 4.19- Comparison of ATAC-Seq profiles of cells derived from cultures containing BMP4 (All cyto) and without BMP4 (-BMP4) for HE1 and HE2.
For both conditions the two-fold difference in peak height was calculated for all peaks and all plots were centred to the peak summit. Peaks were then plotted according to the fold difference to indicate All cytokine and –BMP4 specific peaks. Individual motifs enrichment within peak sets are plotted alongside. Right panels: HOMER TF motif enrichment analysis in specific peak sets was performed with all motifs displayed passing P<0.01 threshold significance. %bg: percentage of background sites enriched for motif. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

Withdrawal of BMP4 had a pronounced effect on the number of cells gained from the serum free I.V.D system at day 3 blast culture and was incompatible with the emergence of HP. The ratio of distal elements to promoter elements for HE1 (38808 to 16840) and HE2 (27501 to 15954) reflect the quality of the ATAC data being compared. Interestingly, even

though BMP4 withdrawal had a profound effect on the overall number of cells gained from the system very few peaks were 2-fold increased (1656) or decreased (707) in peak height in the presence of the cytokine for HE1. This may reflect that BMP4 signalling is essential for the commitment of HB to HE and that its withdrawal severely impacts on HE emergence. The low number of HB cells which did commit to HE may have done so due to cross signalling and therefore pushed through to HE.

HOMER motif enrichment analysis of peaks which were both 2-fold higher and lower in peak height revealed that they were both enriched for ETS, SOX and AP-1 motifs marking this cell type as having an endothelial TF signature. Peaks with a 2-fold increase in height in the presence of BMP4 were also enriched for GATA motifs which is consistent with data indicating that BMP4 signalling upregulates GATA in the HE (Kirmizitas et al., 2017). Peaks derived from HE2 cells which were 2-fold increased and peaks which were decreased in height with BMP4 were also enriched for endothelial TFs SOX ETS and AP-1. Peaks which were 2-fold higher in the presence of BMP4 were also enriched for TEAD and GATA sites.

4.3.4.2 Open chromatin regions dependent on IL3 are depleted for haematopoietic TFs RUNX and GATA



HE1- Enriched; n= 1612

Motif Sequence	Motif	P Value	%sites	%bg
	ETS	1.0E-150	35.11	10.61
	TEAD	1.0E-108	28.66	9.23
	SOX	1.0E-114	20.53	4.64
	AP-1	1.0E-63	13.03	3.25

HE1- Depleted; n= 1175

Motif Sequence	Motif	P Value	%sites	%bg
	ETS	1.0E-103	45.11	17.71
	SOX	1.0E-22	19.74	10.16
	CTCF	1.0E-80	13.62	1.94
	PU.1	1.0E-13	10.64	3.63
	NFkB	1.0E-19	7.83	2.63

HE2- Enriched; n= 2457

Motif Sequence	Motif	P Value	%sites	%bg
	ETS	1.0E-322	50.14	16.48
	SOX	1.0E-52	34.92	10.48
	CREB/A TF	1.0E-111	20.02	6.38
	TEAD	1.0E-147	19.45	4.84
	GATA	1.0E-10	9.93	4.23

HE2- Depleted; n= 1050

Motif Sequence	Motif	P Value	%sites	%bg
	ETS	1.0E-66	40.57	17.66
	AP-1	1.0E-36	13.90	4.13
	CTCF	1.0E-77	11.90	1.25
	MEIS	1.0E-6	9.81	5.99
	SOX	1.0E-23	8.29	2.26

HP- Enriched; n= 5230

Motif Sequence	Motif	P Value	%sites	%bg
	ETS	1.0E-98	22.16	11.77
	AP-1	1.0E-95	8.59	2.74
	NRF1	1.0E-239	4.21	0.14

HP- Depleted; n= 5386

Motif Sequence	Motif	P Value	%sites	%bg
	MEIS	1.0E-31	30.25	23.27
	RUNX	1.0E-240	18.18	5.39
	GATA	1.0E-161	16.91	6.23
	CTCF	1.0E-688	14.35	0.77
	ESR2	1.0E-29	10.68	6.52

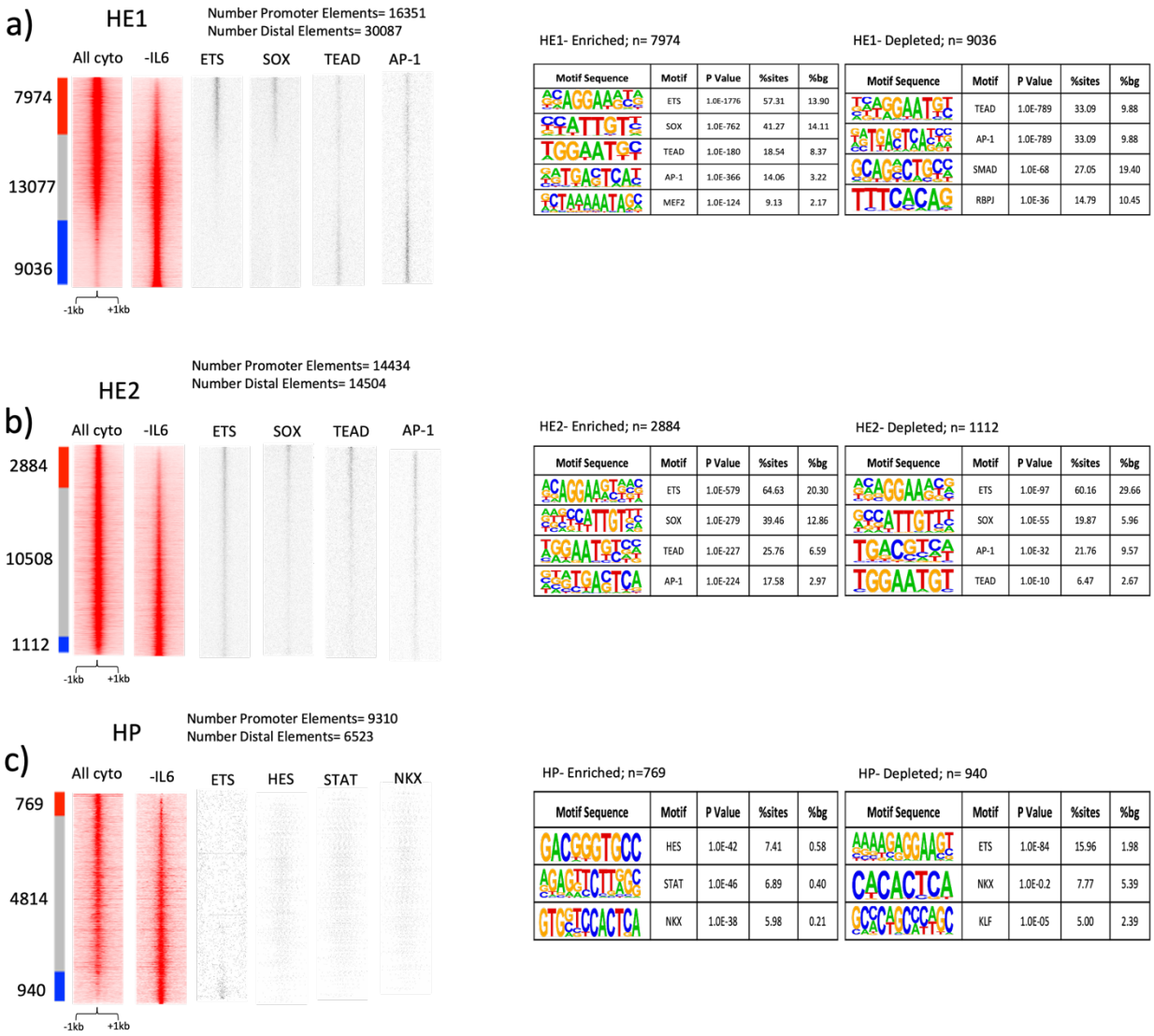
Figure 4.20- Comparison of ATAC-Seq profiles of cells derived from cultures containing IL3 (All cyto) and without IL3 (-IL3).

For both conditions the two-fold difference in peak height was calculated for all peaks and all plots were centred to the peak summit. Peaks were then plotted according to the fold difference to indicate All cytokine and –IL3 specific peaks. Individual motifs enrichment within peak sets are plotted alongside. Right panels: HOMER TF motif enrichment analysis in specific peak sets was performed with all motifs displayed passing P<0.01 threshold significance. %bg: percentage of background sites enriched for motif. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

In the presence of IL3 few peaks were 2-fold increased (1612) or decreased (1175) in HE1 (Figure 4.20a). Both peak sets were enriched for motifs for endothelial TF motifs SOX, TEAD and AP-1. Peaks which were 2-fold lower with the cytokine were also enriched for PU.1 and NFKB motifs. HE2 cytokine responsive peaks were also enriched for endothelial motifs SOX and AP-1. However, peaks which were increased in peak height were enriched for TEAD and GATA motifs suggesting that these cells had achieved a haematopoietic TF signature and were undergoing the EHT as a result of IL3 signalling (Figure 4.20b).

In the HP cell type comparison, a large number of peaks were 2-fold increased (5230) and decreased (5386) in the presence of IL3 indicating that in this cell type the cytokine exerted a significant effect on the chromatin landscape (Figure 4.20c). Interestingly, peaks which were decreased in the presence of IL3 were enriched for the haematopoietic TF motifs RUNX and GATA. The IL3 receptor (*Il3ra*) is expressed at low levels in HE and HP but is upregulated in macrophages (Goode et al., 2016), it is possible that IL3 acts to drive the differentiation of HP to macrophages and in turn negatively regulate haematopoietic TF activity.

4.3.4.3 Open chromatin regions dependent on IL6 are depleted for SMAD motifs



Motif Sequence	Motif	P Value	%sites	%bg
	ETS	1.0E-1776	57.31	13.90
	SOX	1.0E-762	41.27	14.11
	TEAD	1.0E-180	18.54	8.37
	AP-1	1.0E-366	14.06	3.22
	MEF2	1.0E-124	9.13	2.17

Motif Sequence	Motif	P Value	%sites	%bg
	TEAD	1.0E-789	33.09	9.88
	AP-1	1.0E-789	33.09	9.88
	SMAD	1.0E-68	27.05	19.40
	RBPJ	1.0E-36	14.79	10.45

Motif Sequence	Motif	P Value	%sites	%bg
	ETS	1.0E-579	64.63	20.30
	SOX	1.0E-279	39.46	12.86
	TEAD	1.0E-227	25.76	6.59
	AP-1	1.0E-224	17.58	2.97

Motif Sequence	Motif	P Value	%sites	%bg
	ETS	1.0E-97	60.16	29.66
	SOX	1.0E-55	19.87	5.96
	AP-1	1.0E-32	21.76	9.57
	TEAD	1.0E-10	6.47	2.67

Motif Sequence	Motif	P Value	%sites	%bg
	HES	1.0E-42	7.41	0.58
	STAT	1.0E-46	6.89	0.40
	NKX	1.0E-38	5.98	0.21

Motif Sequence	Motif	P Value	%sites	%bg
	ETS	1.0E-84	15.96	1.98
	NKX	1.0E-0.2	7.77	5.39
	KLF	1.0E-05	5.00	2.39

Figure 4.21- Comparison of ATAC-Seq profiles of cells derived from cultures containing IL6 (All cyto) and without IL6 (-IL6).

For both conditions the two-fold difference in peak height was calculated for all peaks and all plots were centred to the peak summit. Peaks were then plotted according to the fold difference to indicate All cytokine and –IL6 specific peaks. Individual motifs enrichment within peak sets are plotted alongside. Right panels: HOMER TF motif enrichment analysis in specific peak sets was performed with all motifs displayed passing P<0.01 threshold significance. %bg: percentage of background sites enriched for motif. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

In HE1 there was a large number of peaks showing a 2-fold increase or decrease in accessibility in the presence of IL6 (Figure 4.21a). In both enriched and depleted peak sets endothelial TF motifs SOX, TEAD and AP-1 were enriched. Interestingly NOTCH signalling effect RBPJ TF motifs were depleted in peaks in the presence of IL6 in HE1. This suggests that IL6 signalling may impact on NOTCH signalling processes.

In HE2 there was an enrichment of SOX, TEAD and AP-1 TF motifs (Figure 4.21b). Ranked density plots reveal that TEAD and SOX enrichment was greater in the peaks high in the presence of IL6, indicating that IL6 signalling may impact the activity of these TFs.

In HP there were few peaks 2-fold increased (769) or decreased (940) in height between the All cytokine (+IL6) and -IL6 conditions (Figure 4.21c). In peaks which were 2-fold increased in height there was an enrichment of NOTCH signalling responsive TF HES motifs suggesting that in the presence of IL6 in these cell types IL6 signal to HES TFs. In peaks which were 2-fold decreased in height there was an enrichment of ETS TF motifs which indicates a more endothelial TF signature in these peaks.

4.3.4.4 VEGF dependent peaks are enriched for endothelial TF motifs and depleted for haematopoietic TF motifs

VEGF withdrawal had the most notable effect at the chromatin level. In HE1 6697 peaks saw a 2-fold increase in peak height while 5264 peaks saw a 2-fold decrease in peak height (Figure 4.22a). The majority of peaks (13404) were shared. HOMER motif analysis of peaks derived from HE1 cells revealed that ATAC sites which were 2-fold higher with VEGF were enriched for EBOX (SCL/TAL1) and SOX motifs. In peaks which were 2-fold depleted in the presence of VEGF we found an enrichment of TEAD and GATA motifs.

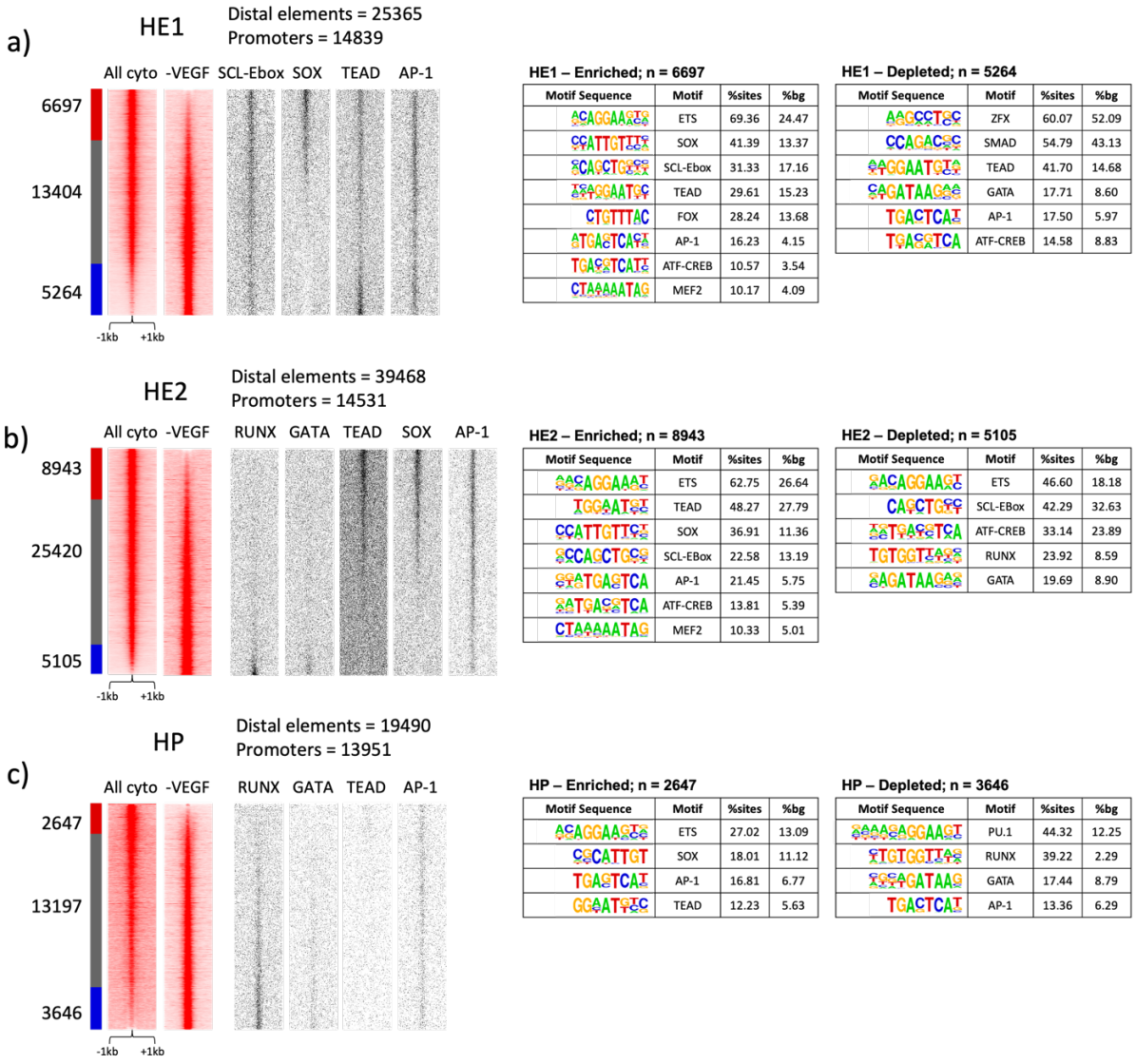


Figure 4.22- Comparison of ATAC-Seq profiles of cells derived from cultures containing VEGF (All cyto) and without VEGF (-VEGF).
For both conditions the two-fold difference in peak height was calculated for all peaks and all plots were centred to the peak summit. Peaks were then plotted according to the fold difference to indicate All cytokine and –VEGF specific peaks. Individual motifs enrichment within peak sets are plotted alongside. Right panels: HOMER TF motif enrichment analysis in specific peak sets was performed with all motifs displayed passing P<0.01 threshold significance. %bg: percentage of background sites enriched for motif. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

In HE2 8943 sites were 2-fold increased in the presence of VEGF and 5105 were 2-fold decreased in the presence of VEGF, the majority of sites were shared (25420) (Figure 4.22b). Motif analysis revealed that ATAC sites which were 2-fold increased in peak height derived from HE2 cells cultured in the presence of VEGF were enriched for motifs for endothelial transcription factors such as SOX, TEAD and AP-1, indicating that these cells had maintained their endothelial signature. However, peaks which were 2-fold decreased in peak height in the presence of VEGF were enriched for the haematopoietic transcription factors GATA and RUNX.

In HP, 2647 sites were 2-fold increased while 3646 sites were 2-fold decreased in peak heigh in the presence of VEGF (Figure 4.22c). The cells should have undergone the EHT and should be marked by a haematopoietic TF motif enrichment signature at distal ATAC sites. However, in the presence of VEGF sites which were 2-fold increase in peak height were enriched for endothelial TF motifs SOX and TEAD. Sites which were 2-fold decrease in peak height were enriched for the haematopoietic transcription factors GATA, RUNX and the RUNX1 target PU.1 (SPI1). These findings suggest at the level of the genome, that VEGF is

repressing the transition from an endothelial to haematopoietic chromatin and TF motif landscape within these cells.

4.3.5 Identification of cytokine responsive binding motif signatures at the genome-wide level

To identify transcription factor motifs enriched in open chromatin regions which were signalling responsive to VEGF, BMP4, IL6 or IL3 we performed supervised HOMER motif analysis on peak sets that showed a 2-fold increase or decrease in peak height in the presence of the cytokine. For each TF motif in each peak set a relative motif enrichment value was calculated and then hierarchically clustered and plotted as a heat map. The analysis strategy is depicted in Figure 4.23a.

Peaks which were 2-fold higher in HE1 and HE2 cells cultured in the presence of BMP4 were enriched for MYC-EBOX, OCT, CTCF, and PBX transcription factor motifs (Figure 4.23b left panel). In the presence of BMP4 we found an enrichment of GATA motifs in these sites, which is consistent with literature indicating that the BMP4 directed SMAD activity drives *Gata2* expression (Kirmizitas et al., 2017). Peaks which were 2-fold decreased in height (Figure 4.23b right panel) displayed an enrichment of TFs motifs which were associated with an endothelial signature such as SOX and RBPJ together with an enrichment of RAR and FOX motifs.

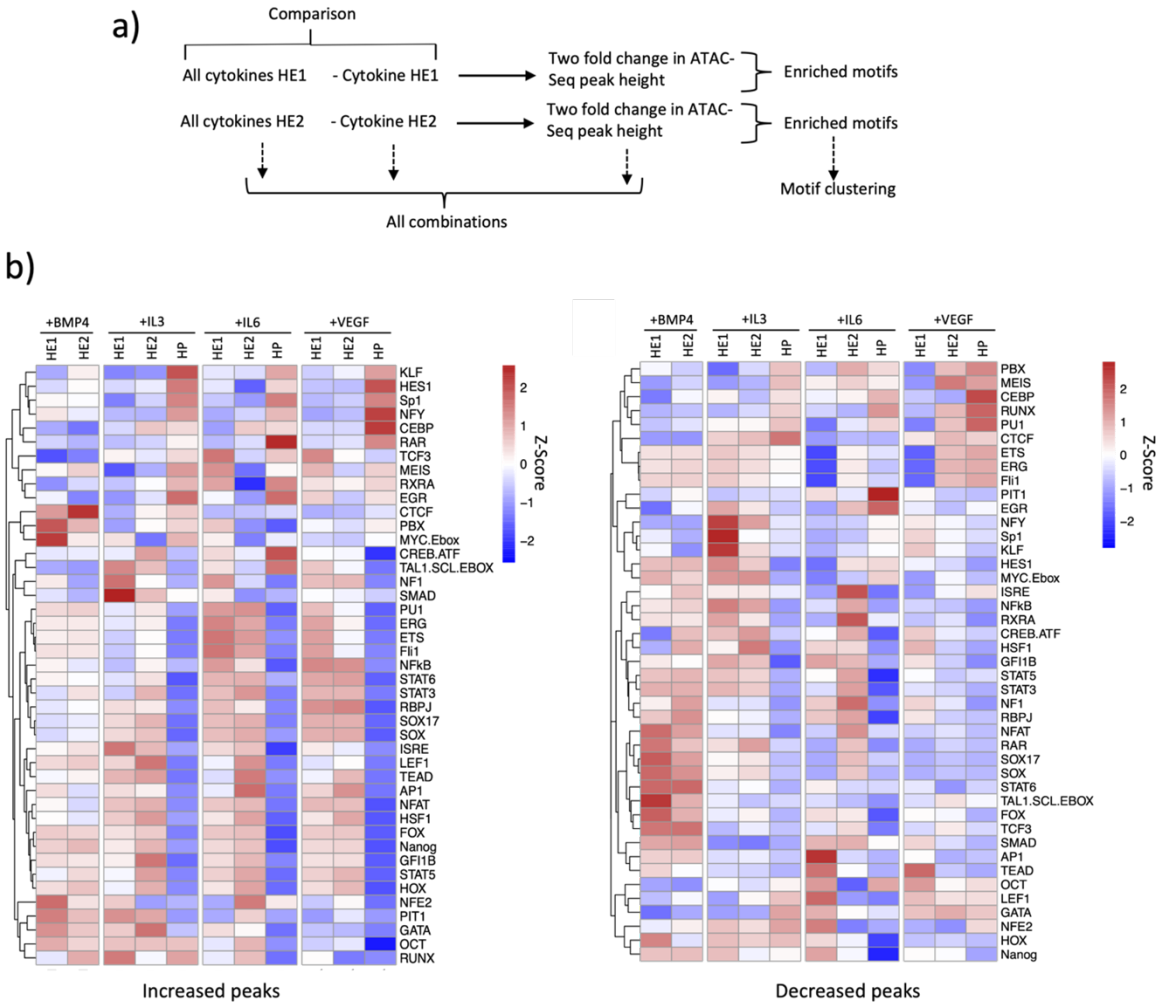


Figure 4.23- Relative motif enrichment analysis of cytokine responsive distal ATAC sites. a) Overview of motif enrichment analysis strategy. Relative motif enrichment analysis of distal ATAC peaks (b) that are increased (left) or decreased (right) in the presence of the indicated cytokines (direction of peak height change indicated below the heatmap). Cell type grouping and cytokine condition shown at the top. TF motifs on the right. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

Peaks which were 2-fold increased in height in the presence of IL3 in HE1 cells and to a lesser extend in HE2 cells were enriched in motifs for the SMAD family of TFs. These peaks were also enriched for the haematopoietic TF motifs GATA, RUNX and PU.1, potentially highlighting the notion that IL3 is a 'haematopoietic' cytokine which drives the emergence

and/or maintenance of the HSC from the HE. IL3 has been shown to be a Runx1 target and is able to drive the development of HSCs from the E11 AGM. Furthermore, *IL3* mutant embryos fail to form HSCs indicating that IL3 is required prior to the emergence of the HSC acting as a proliferation or survival factor for the earliest HSC in the embryo (Robin et al., 2006). Coaggregation cultures of caudal parts of E9.5 mouse embryos with OP9 cells revealed that HSC-precursors and type I pre-HSCs can mature into definitive HSCs in response to stem cell factor (SCF) but not IL3. However, for type II pre-HSCs, SCF and IL3 are both equally potent allowing for the identification of different haematopoietically committed cell types (Rybtsov et al., 2014). LEF1, TEAD and IRSE motifs were also enriched in this peak set. In peaks which were 2-fold decreased in the presence of IL3 there was a depletion of the haematopoietic TF motifs GATA, RUNX and PU.1.

Interpreting the IL6 results was challenging as no clear signature emerged. Peaks from HE cells cultured in the presence of IL6 showing a 2-fold increase in peak height were enriched for endothelial motif TFs FLI1, ETS, ERG, RBPJ and STAT. There was a slight enrichment of RUNX seen in HE2. Peaks which were 2-fold decrease in peak height were depleted for endothelial TF motifs. However, not that these cultures contained VEGF and the motif pattern for enriched motifs appeared to be dominated by this cytokine (see also below). Decreased peaks displayed the respective motif signature for each cell type: AP-1/TEAD/SOX for HE1, RBPJ/HES/SOX for HE2 and a weak enrichment of motifs for haematopoietic TFs such as RUNX and PU.1 in HP.

As outlined above, VEGF had the most drastic influence on the chromatin landscape. Peaks

which were 2-fold increased in peak height in HE1 and HE2 cell types derived from cultures containing VEGF were enriched for endothelial motifs such as SOX17/SOX, FLI1, AP-1, TEAD and the NOTCH signalling effector RBPJ. Peaks which were increased in peak height in HE2 and HP cultured with VEGF saw a depletion of haematopoietic TF motifs such as for RUNX, PU.1 and NFE2, suggesting that these cells in the presence of VEGF had failed to undergo the EHT. In contrast peaks from HE cells which where 2-fold decreased in peak height from cultures without VEGF showed a depletion of endothelial motifs such as SOX17/SOX, FLI1, ETS and EGR. 2-fold decreased peaks from HP cells cultured with VEGF were significantly enriched for the haematopoietic TF motifs RUNX, PU.1, CEBP and NFE2 indicating that these cells had undergone the EHT and adopted a more haematopoietic signature at the level of the genome. Results from supervised relative motif enrichment analysis therefore is concordant with the results from the pair-wise unsupervised de novo HOMER motif analysis previously described, indicating at the level of the genome that VEGF signalling maintains an endothelial TF signature and represses the emergence of a haematopoietic TF signature.

4.3.6 Identification of cytokine responsive enhancer elements

In spite of a number of efforts looking at how specific genomic regions are controlled by signalling (Simeonov et al., 2017, Hewitt et al., 2017), we have limited global information about which *cis*-regulatory elements can respond to signals. Our global enhancer collection combined with our optimised serum free *in vitro* different system allows us to answer this question. Previously we employed a serum free *in vitro* differentiation system to identify

signalling responsive cis ATAC sites in HB, HE1, HE2 and HP cell types. Peaks which were 2-fold increased or decreased in peak height in the presence of each cytokine (VEGF, BMP4, IL3 and IL6) were defined as being signalling responsive. Enhancer elements which were identified in our enhancer screen sit within a subset of distal ATAC sites found in cell types derived from our serum free I.V.D system with and without cytokines. We therefore asked the question whether we could identify signalling responsive enhancer elements.

Initially we compared the chromatin signature of cells differentiated in serum and under serum-free conditions (in the presence of all cytokines) which showed that around 80%-90% of promoters seen in cells from serum free culture overlapped in both conditions, indicating the reproducibility of our differentiation system (Figure 4.24a). Distal elements for HB and HE1 cells were high (>70%) the overlap was reduced in HE2 and HP cells (~50%). This result suggests that while the overall cellular identity of the cell types seemed to be largely preserved in sorted cells expressing the correct surface markers, the difference in the signalling environment exerted a strong effect on the chromatin landscape.

To identify cytokine responsive enhancer elements in our enhancer collection we overlapped our cytokine responsive ATAC sites with our enhancer elements for each developmental stage which generated developmental stage specific signalling responsive enhancer sets. Figure 4.24b summarises the numbers of enhancers which were responsive to each cytokine. In total we found over 10,000 responsive enhancer and promoter elements and that the greatest number of enhancers were signalling responsive to VEGF.

a)

Percentage of ATAC peaks from serum free culture overlapping with ATAC peaks from serum culture

	All	Promoters	Distal
HB	74.10	80.96	70.83
HE1	78.20	83.49	74.53
HE2	63.27	79.07	56.66
HP	77.21	90.48	49.65

b)

Cytokine	Cell type	#Distal elements	#Distal elements identified as enhancers	**#Two-fold increase enhancers**	**#Two-fold decrease enhancers**	#Promoter elements	**#Two-fold increase promoters**	**#Two-fold decrease promoters**
VEGF	HE1	14739	7197	**2426**	**1357**	14585	**1405**	**353**
	HE2	13964	7720	**1733**	**2008**	15894	**1661**	**311**
	HP	18256	7834	**622**	**1331**	12896	**1204**	**506**
IL6	HE1	25437	9746	**2993**	**1139**	15404	**497**	**248**
	HE2	16991	8275	**303**	**247**	15735	**162**	**55**
	HP	5864	1639	**144**	**253**	9047	**572**	**168**
IL3	HE1	34042	13667	**248**	**283**	15992	**141**	**97**
	HE2	12040	6584	**349**	**334**	15411	**322**	**111**
	HP	35691	10906	**794**	**1946**	14862	**2138**	**514**
BMP4	HE1	32645	12911	**405**	**99**	15887	**109**	**38**
	HE2	27583	10544	**577**	**141**	16008	**205**	**67**

Figure 4.24- Identification of cytokine responsive enhancer elements identified from our enhancer screen.

a) Percentage of ATAC sites derived from serum free I.V.D culture overlapping with ATAC peaks derived from serum IVD culture (n=2). b) Table showing the number of enhancer fragments identified in our assay which were cytokine-responsive (signalling-responsive) by intersecting elements scoring positive in our screen with the ATAC sites from cytokine withdrawal experiments. For each cell type and cytokine withdrawal condition studied the table shows the number of distal elements (defined as +/-1.5kb from a transcription start site) present in each cell type (#Distal elements), the number of distal elements which overlapped with an ATAC fragment scoring positive in our enhancer screen (#Distal elements identified as enhancers) and the number of promoter elements which scored positive in our screen (#Promoter elements). The table also shows the number of enhancer elements in distal sites which were two-fold increased in peak height (#Two-fold increase enhancers) or two-fold decrease in peak height (#Two-fold decrease enhancers) and the number of promoter elements which were two-fold increased (#Two-fold increase promoters) or decreased (#Two-fold decrease promoters) in peak height in the presence of each cytokine.

4.3.7 Global analysis of relative motif enrichment within cytokine responsive enhancer elements

Next, we studied the motif enrichment profiles of functionally identified enhancer sets which were signalling responsive to VEGF, IL6, IL3 and BMP4 to identify TFs responding to signalling. To this end we performed HOMER motif enrichment analysis on peak sets which were 2-fold increased or decreased in peak height when in the presence of each cytokine compared to the all cytokine control (Figure 4.25a).

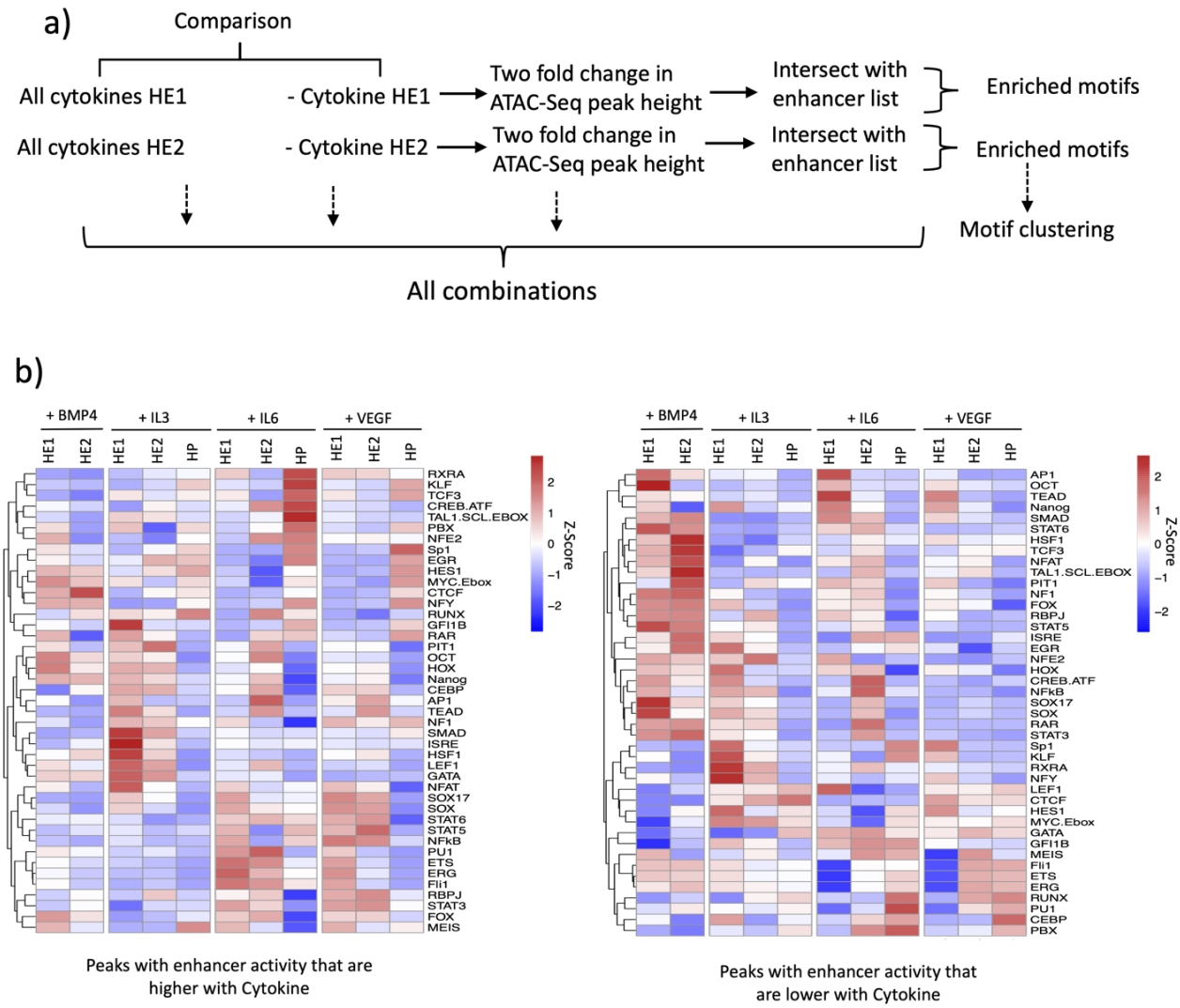


Figure 4.25- Relative motif enrichment analysis of cytokine responsive enhancer elements.

a) Overview of filtering strategy for the identification of cytokine responsive enhancer elements. Relative motif enrichment analysis of enhancer elements (b) that are increased (left) or decreased (right) in the presence of the indicated cytokines (direction of peak height change indicated below the heatmap). Cell type grouping and cytokine condition shown at the top. TF motifs on the right. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

Figure 4.25b left panel show peaks with enhancer activity that were 2-fold increased in the presence of each cytokine while the right panel shows peaks which were decreased in the presence of the cytokine. In enhancers from HE which were 2-fold increased in the

presence of BMP4 there was an enrichment of GATA motifs and those of the NOTCH signalling effector HES1. Consistent with SMADs being mediators of BMP signalling (Kirmizitas et al., 2017), we observed an overall depletion of SMAD motifs in these enhancers. HE1 and HE2 enhancers which were 2-fold increased in height in the presence of IL3 were enriched for GATA, SMAD and NFAT motifs. OCT, HOX and NANOG motifs were also enriched. In HP enhancers with a 2-fold increase in height in the presence of IL3 were enriched for GATA, LEF and RUNX TF motifs. Peaks which were 2-fold increased in height in the presence of IL6 were enriched for the endothelial TFs ETS and AP-1. Peaks with a 2-fold decrease in peak height were enriched for NFKB, GATA and SMAD motifs. In enhancers from HE1 and HE2 with a 2-fold increase in height in the presence of VEGF we found an enrichment of endothelial TF motifs such as SOX, STAT, RBPJ, NFKB and TEAD. In HE we found a depletion of haematopoietic TF motifs GATA, RUNX and PU.1. HP enhancers which were 2-fold decreased in the presence of VEGF were enriched for haematopoietic TF motifs RUNX, CEBP and PU.1. These findings were consistent with findings from direct pairwise *de novo* HOMER motifs enrichment analysis and relative motif enrichment analysis of signalling responsive distal ATAC sites and confirm that in the presence of VEGF cells were stuck in the HE.

4.3.8 Validation of VEGF responsiveness of individual enhancer elements identified in our screen

Our enhancer screen combined with our serum free I.V.D system identified 7396 enhancer elements as being VEGF signalling responsive. We therefore validated the VEGF signalling responsiveness of enhancer elements identified in our screen. Previously we validated the enhancers for *Sparc*, *Eif2b3*, *Hspg2*, *Pxn*, *Cdh5*, *Dlk1* and *Mrpl15* as enhancers using our serum I.V.D system. These enhancers were selected as they overlapped with an ATAC site which was 2-fold increased or decreased in the presence of VEGF. We next generated ESC lines carrying fluorescent reporter versions of these elements and clones were then cultured in the serum free I.V.D system with VEGF withdrawn at blast culture plating and left for 3 days before the enhancer activity was assessed in HE1, HE2 and HP by FACS.

The *Sparc* enhancer showed a more than 2-fold decrease in peak height in the absence of VEGF signalling in HE1, HE2 and HP cells (Figure 4.26a-c). Annotation of the enhancer sequence revealed motifs for endothelial TFs SOX17, AP-1, TAL-1, FLI1 and ETS and a motif for the haematopoietic TF RUNX1. The enhancer was active in the presence of VEGF in HE2 cells, the activity was lost in response to VEGF withdrawal (P=0.0007). The *Pxn* enhancer was decreased in peak height in HE1 when VEGF was withdrawn (Figure 4.27a-c). Annotation of the enhancer sequence revealed endothelial TF motifs ETS, RBPJ and FLI1 and haematopoietic TF motifs. Removal of VEGF resulted in a decrease in enhancer activity in HE1 (P=0.046) and HE2 (P=0.0007).

The *Hspg2* enhancer was a decreased in peak height in HE2 cells after VEGF withdrawal

(Figure 4.28a-c). Annotation of the enhancer sequence revealed endothelial TF motifs for SOX17, ETS and FLI1 (Figure 4.28c). The absence of haematopoietic TF motifs coincides with this enhancer lacking an ATAC site and enhancer activity in HP. VEGF withdrawal resulted in a reduction of enhancer activity in HE2 (P= 0.01).

The *Dlk1* enhancer showed an increase in peak height in HE1 and HE2 cells when VEGF was withdrawn (Figure 4.29a-c). Annotation of the enhancer sequence revealed endothelial (ETS, E-BOX, TEAD) and haematopoietic (GATA and RUNX) TFs (Figure 4.29c). Withdrawal of VEGF resulted in a higher level of activation in HE1 (P=0.0008) but a lower level of activation in HE2 (P=0.0018) when compared to all cytokine cultures.

The *Cdh5* enhancer displayed an increase in peak height in the presence of VEGF in HE1 (Figure 4.30a-c). The *cadherin* 5 (CDH5) protein acts as a robust marker of haemogenic endothelium but is not required for the EHT process (Anderson et al., 2015). Annotation of the enhancer sequence revealed endothelial TF motifs TAL1 and SOX17 and an overlapping RUNX/TEAD site (Figure 4.30c). VEGF withdrawal resulted in a loss of enhancer activity in HE1 (P=0.0064) and HE2 (P=0.0018).

Globally we have seen that withdrawal of cytokines from the serum free I.V.D system results in changes of a motif enrichment signature at distal ATAC sites and at enhancers which overlap with these sites. Withdrawal of BMP4 results in the of GATA motif enrichment consistent with finds that BMP4 signalling regulates *Gata2* expression (Kirmizitas et al., 2017). We also observe that VEGF withdrawal results in a loss of ETS, FLI

and SOX motifs supporting the known role of VEGF signalling for the maintenance of an endothelial programme. Withdrawal of VEGF results in an enrichment of haematopoietic TFs RUNX and PU.1 supporting finds that VEGF represses the expression of these factors by maintaining NOTCH signalling. We have combined the serum free I.V.D system with our enhancer reporter system to identify signalling responsive enhancer elements, and for VEGF we have validated a number of those enhancers as being responsive.

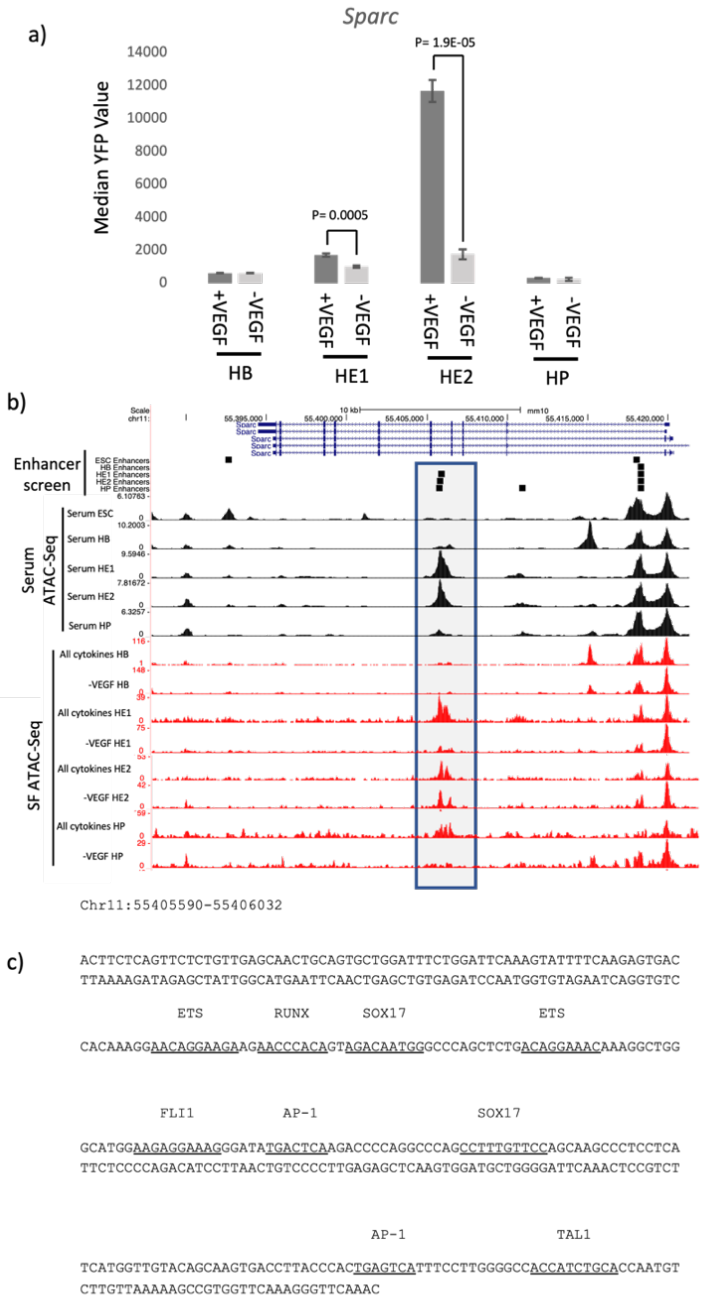


Figure 4.26- Validation of the *Sparc* enhancer as being VEGF signalling responsive.
a) Bar plot showing the Median YFP value for the enhancer in the serum free I.V.D system with (+VEGF) and without (-VEGF) for FLK-1+ HB (HB) and HE1, HE2 and HP at day 3 blast culture (n=3). Error bars represent the standard deviation, P value calculated using Student's two tailed t-test. b) UCSC browser screenshot showing the enhancer scoring as positive in our assay represented by black bars. ATAC-Seq tracks from serum (Serum ATAC-Seq) and serum free I.V.D (SF ATAC-Seq) highlighted with overlapping enhancers by the grey box. c) Enhancer sequence with endothelial and haematopoietic TF motifs possibly regulating the signalling responsive of the enhancer element to VEGF annotated.

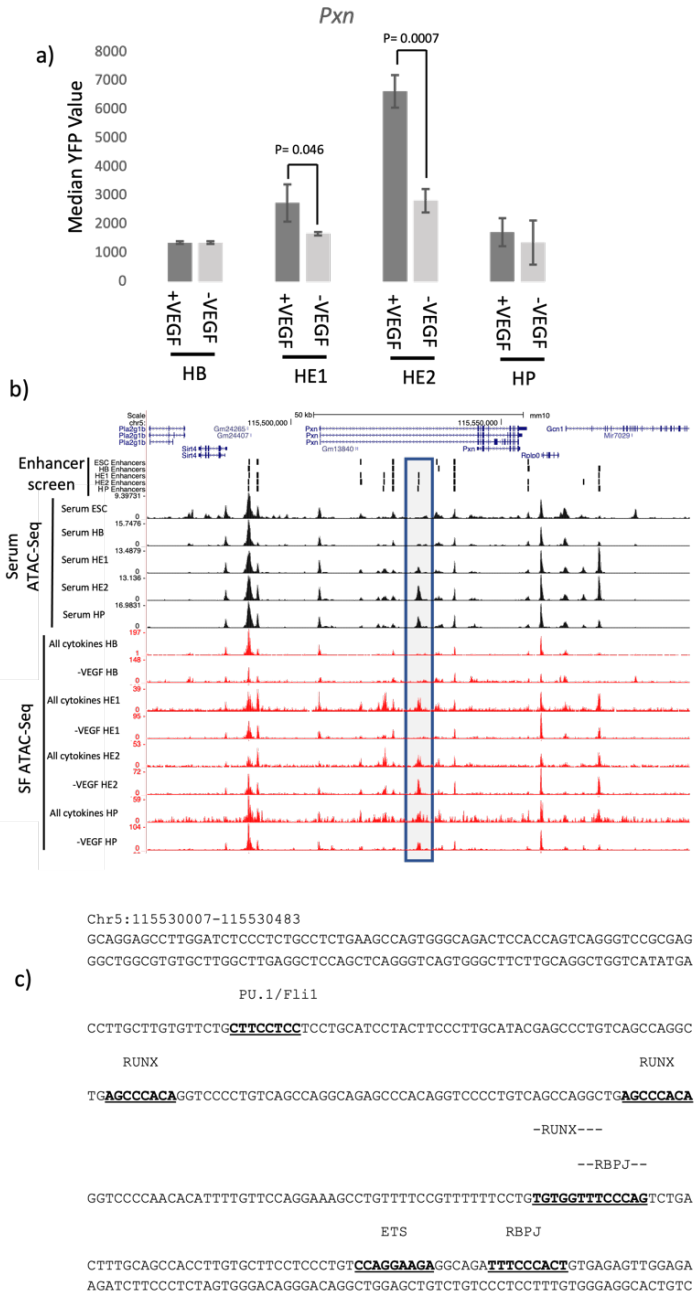


Figure 4.27- Validation of the *Pxn* enhancer as being VEGF signalling responsive.
a) Bar plot showing the Median YFP value for the enhancer in the serum free I.V.D system with (+VEGF) and without (-VEGF) for FLK-1+ HB (HB) and HE1, HE2 and HP at day 3 blast culture (n=3). Error bars represent the standard deviation, P value calculated using Student's two tailed t-test. b) UCSC browser screenshot showing the enhancer scoring as positive in our assay represented by black bars. ATAC-Seq tracks from serum (Serum ATAC-Seq) and serum free I.V.D (SF ATAC-Seq) highlighted with overlapping enhancers by the grey box. c) Enhancer sequence with endothelial and haematopoietic TF motifs possibly regulating the signalling responsive of the enhancer element to VEGF annotated.

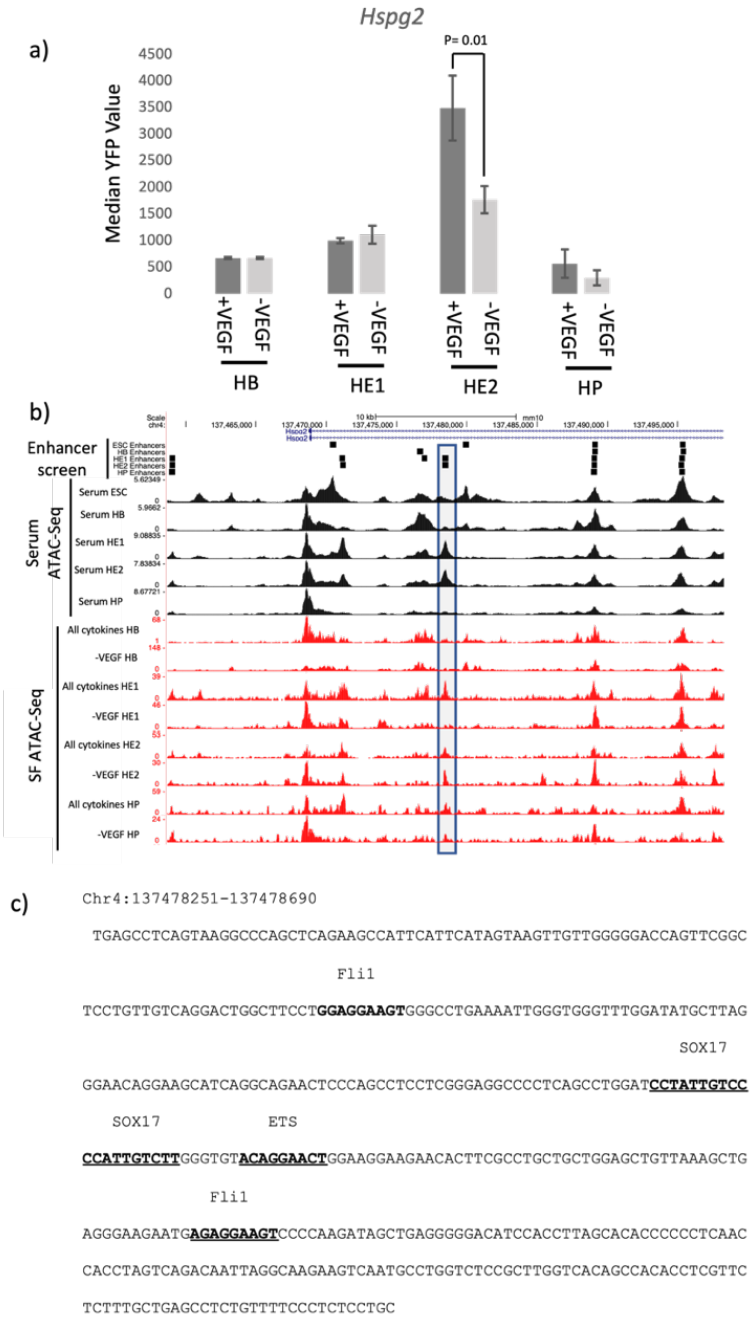


Figure 4.28- Validation of the *Hspg2* enhancer as being VEGF signalling responsive.
a) Bar plot showing the Median YFP value for the enhancer in the serum free I.V.D system with (+VEGF) and without (-VEGF) for FLK-1+ HB (HB) and HE1, HE2 and HP at day 3 blast culture (n=3). Error bars represent the standard deviation, P value calculated using Student's two tailed t-test. b) UCSC browser screenshot showing the enhancer scoring as positive in our assay represented by black bars. ATAC-Seq tracks from serum (Serum ATAC-Seq) and serum free I.V.D (SF ATAC-Seq) highlighted with overlapping enhancers by the grey box. c) Enhancer sequence with endothelial and haematopoietic TF motifs possibly regulating the signalling responsive of the enhancer element to VEGF annotated.

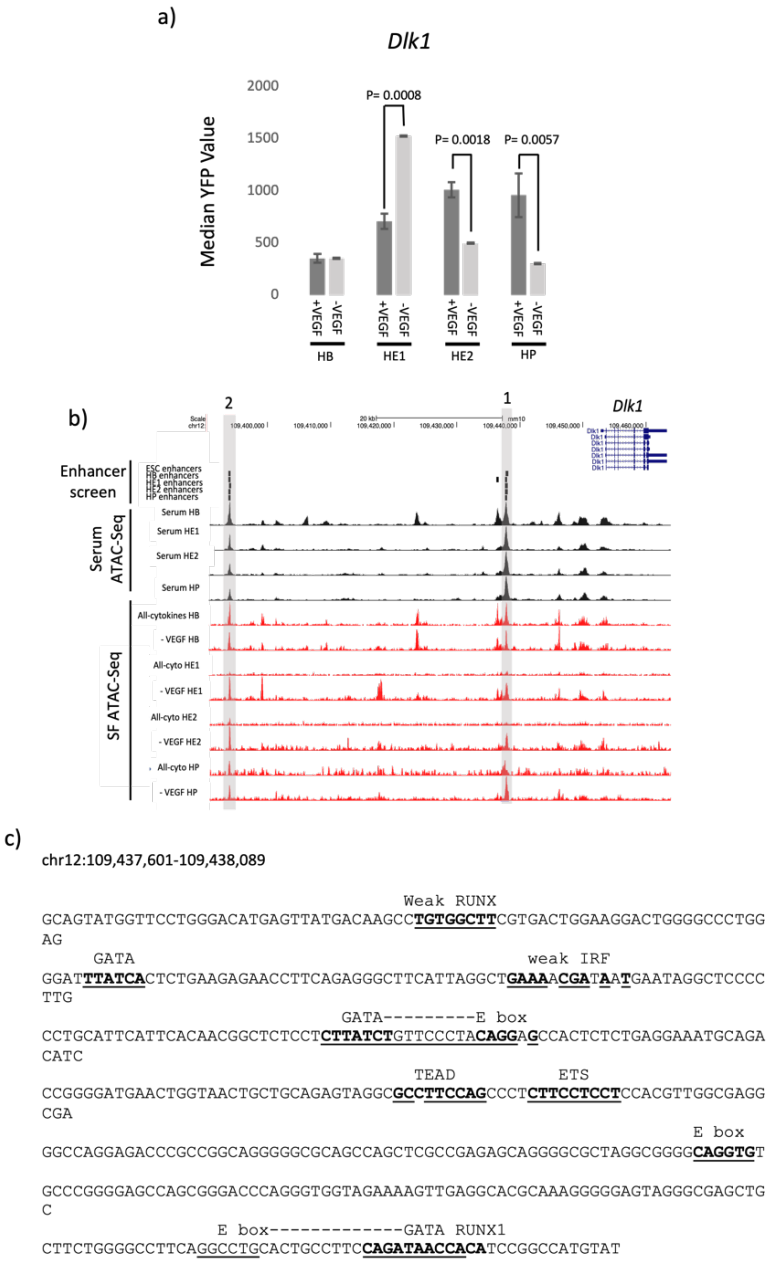


Figure 4.29- Validation of the *Dlk1* enhancer as being VEGF signalling responsive.
a) Bar plot showing the Median YFP value for the enhancer in the serum free I.V.D system with (+VEGF) and without (-VEGF) for FLK-1+ HB (HB) and HE1, HE2 and HP at day 3 blast culture (n=3). Error bars represent the standard deviation, P value calculated using Student's two tailed t-test. b) UCSC browser screenshot showing the enhancer scoring as positive in our assay represented by black bars (labelled 1). ATAC-Seq tracks from serum (Serum ATAC-Seq) and serum free I.V.D (SF ATAC-Seq) highlighted with overlapping enhancers by the grey box. c) Enhancer sequence with endothelial and haematopoietic TF motifs possibly regulating the signalling responsive of the enhancer element to VEGF annotated.

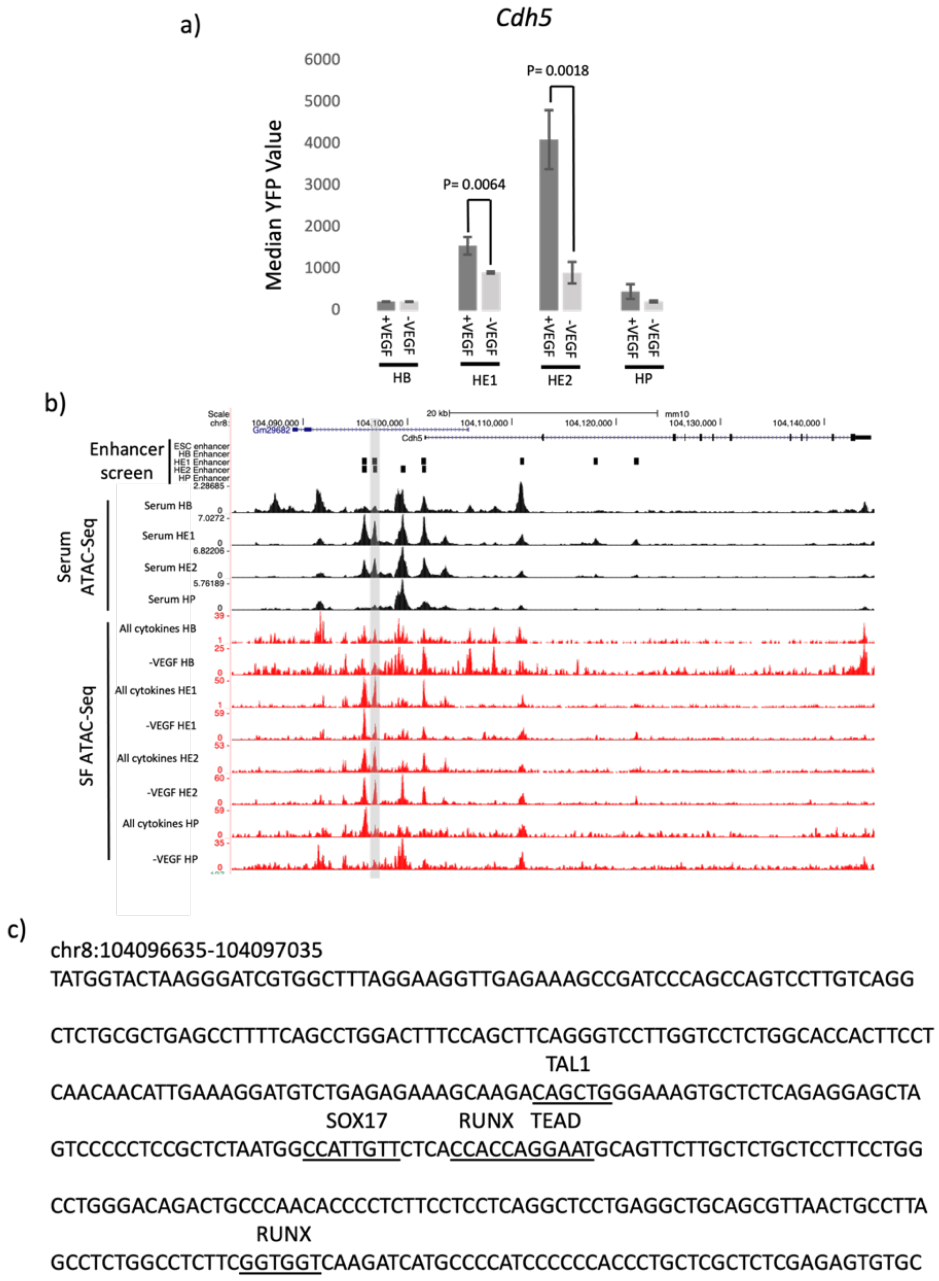


Figure 4.30-Validation of the *Cdh5* enhancer as being VEGF signalling responsive.
a) Bar plot showing the Median YFP value for the enhancer in the serum free I.V.D system with (+VEGF) and without (-VEGF) for FLK-1+ HB (HB) and HE1, HE2 and HP at day 3 blast culture (n=3). Error bars represent the standard deviation, P value calculated using Student's two tailed t-test. b) UCSC browser screenshot showing the enhancer scoring as positive in our assay represented by black bars. ATAC-Seq tracks from serum (Serum ATAC-Seq) and serum free I.V.D (SF ATAC-Seq) highlighted with overlapping enhancers by the grey box. c) Enhancer sequence with endothelial and haematopoietic TF motifs possibly regulating the signalling responsive of the enhancer element to VEGF annotated.

4.4 VEGF blocks HP development by repressing the endothelial-to-haematopoietic transition

4.4.1 VEGF withdrawal results in a decrease in the proportion of HE2 cells and an increase in the proportion of HP cells in serum free I.V.D

Next, we assessed the effect of cytokine withdrawal on the differentiation of mouse ESCs to HP cells in the serum free I.V.D system (Figure 4.31a-c). Withdrawal of IL3, IL6 and BMP4 had little effect on the relative proportion of HE1, HE2 and HP cells as measured by Flow Cytometry. We found that in the absence of BMP4 the overall number of cells gained from the serum free I.V.D culture was significantly reduced and that BMP4 withdrawal was incompatible with the emergence of large numbers of HP (data not shown). This result supports findings that BMP4 is required for the emergence of the HE from the HB. Withdrawal of VEGF also resulted in an overall reduction in the number of HE1, HE2 and HP cells gained from the serum free I.V.D culture. However, the presence of VEGF strongly supressed HP formation. VEGF withdrawal resulted in a significant increase in the number of KIT+ cells being classified as HP (CD41+, TIE2-) (P=4.34E-05) (Figure 4.31a). In VEGF withdrawal cultures there was also a marked decrease in the number of HE2 cells (P=0.049) indicating that these cells had differentiated to HP.

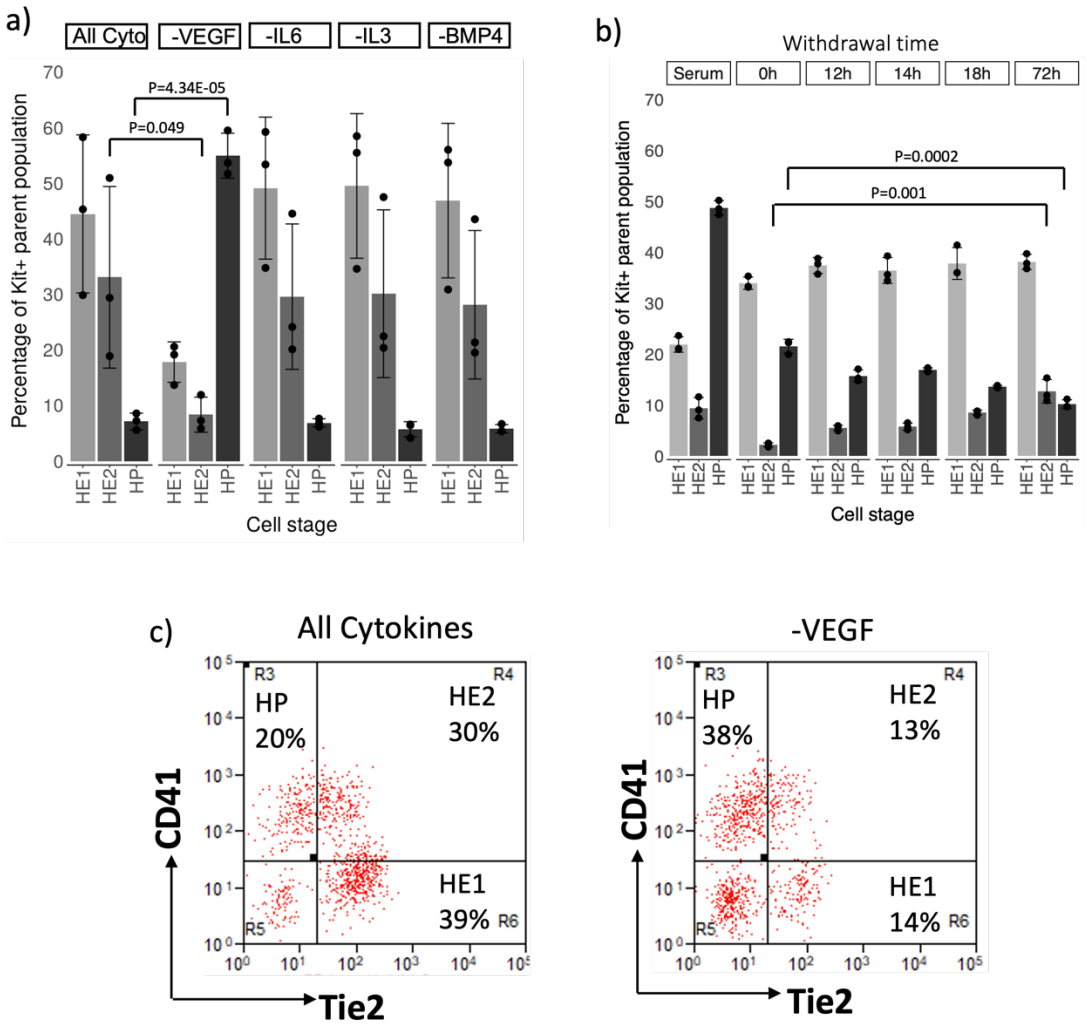


Figure 4.31- VEGF induces HE2 development but represses HP emergence.
a) Graph showing the proportion of HE1/HE and HP cells gained from serum free differentiation cultures in the absence of each cytokine measured by FACS. b) Proportion of HE1, HE2 and HP cells gained from serum free culture after VEGF was withdrawn at different times during differentiation culture. VEGF withdrawal started at the beginning of blast culture (0h). The proportion of each cell type gained from serum IVD also shown. Error bars represent the standard deviation (n=3) with the spread of data indicated by dots. P values were calculated using two tailed Student's t-test. c) Representative FACS analysis of cKIT+ cells stained against Tie2 and CD41 derived from cultures +/- VEGF indicating the percentage of HE1, HE2 and HP cells in each culture condition.

We then performed a time course experiment where VEGF was removed from blast culture at 0 hours (not added), 12 hours, 14 hours, 18 hours and at 72 hours (left on) and measured

the proportion of KIT+ HE1, HE2 and HP cells gained by FACS (Figure 4.31b). Withdrawal of VEGF at 0h returned the greatest proportion of HP cells while withdrawal at 72h returned the lowest proportion. However, the absolute number of HE cells was reduced. The optimal time for VEGF withdrawal or manipulation appeared to be at 12-14 hours where there is a trade-off between the amount of HE2 being produced and HP emergence. These results are consistent with previous observations that the receptor for VEGF, FLK-1, is essential for the formation of the HE from HB, but once formed, it needs to be removed for the EHT to occur. HP cells then become dependent on haematopoietic cytokines.

4.4.2 Motif enrichment analysis clustering reveals VEGF signalling responsive TFs families acting at cis-regulatory elements

Withdrawal of VEGF results in an increase in the proportion of HP at day 3 blast culture. The molecular basis for this finding, i.e., which genomic events are responsible for this phenomenon has so far been unclear. Our ATAC-Seq analysis identified 7814 chromatin regions carrying enhancer elements which responded to VEGF. To identify VEGF responsive TFs, we repeated pair-wise supervised relative motif enrichment clustering analysis with peaks which were signalling responsive to VEGF (the strategy is depicted in Figure 4.32a).

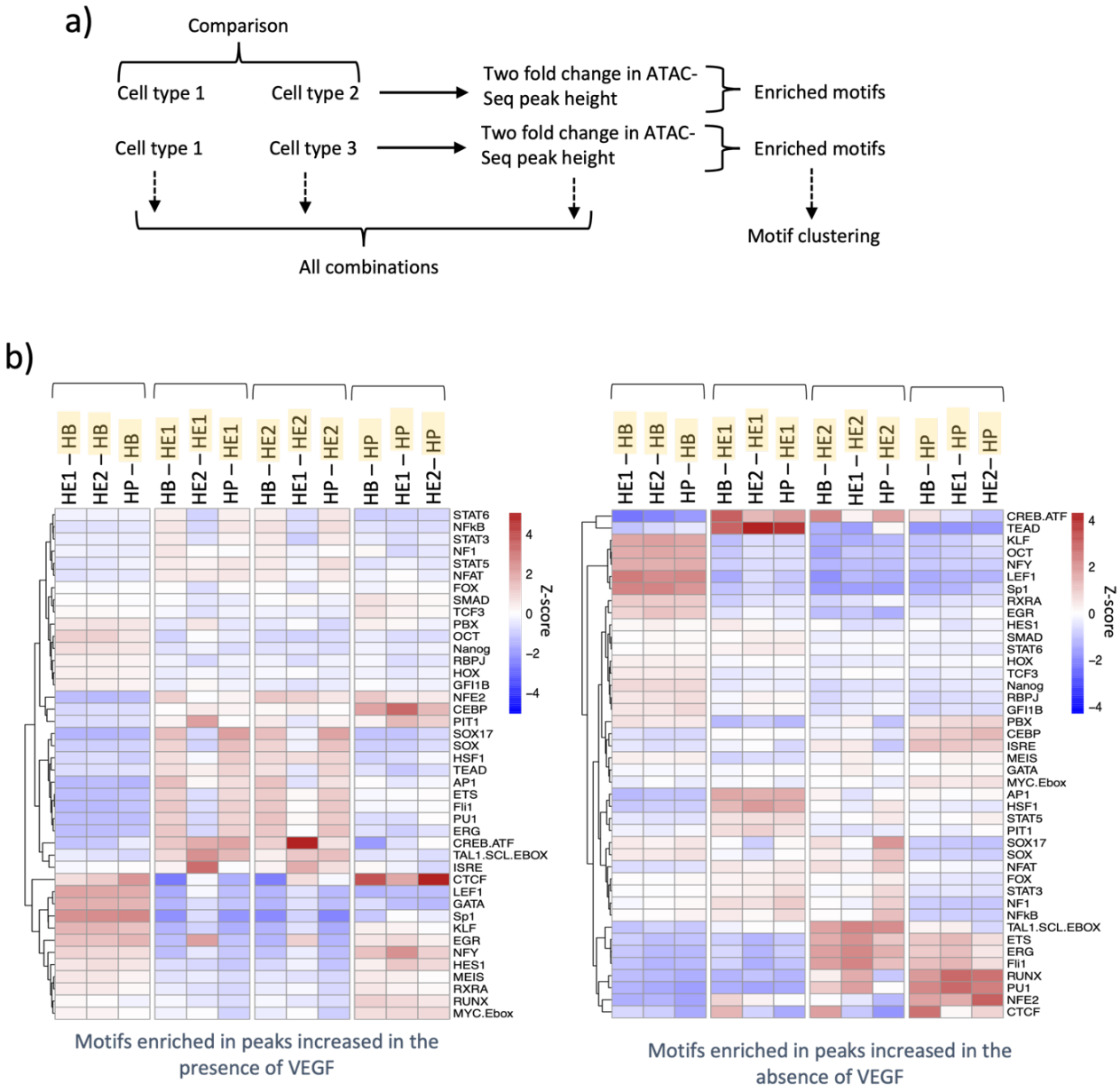


Figure 4.32- VEGF suppresses haematopoietic progenitor cell development.
a) Overview of the relative motif enrichment strategy. b) Clustering of HOMER TF relative enrichment analysis in distal ATAC-seq peaks that scored as positive in our enhancer screen that were 2-fold increase in the presence of (left panel) and in the absence of (right panel) VEGF. Motifs enriched in peaks of specific cell type (highlighted yellow). TF binding motifs are listed on the right on the panel. Relative motif enrichment Z-scores were calculated by columns. Motif position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Data.

In the presence of VEGF peaks with a 2-fold increase in height in HE1 and HE2 were enriched for endothelial TFs AP-1, TEAD, ETS, FLI1 and for SOX17, indicating that VEGF is directly or indirectly regulating the expression and/or binding activity of these factors (Figure 4.32b left panel). In HE2, we found no enrichment of haematopoietic TFs GATA and RUNX indicating that sustained VEGF signalling maintained an endothelial TF signature in these cells. Peaks which were higher in HP cells cultured with VEGF displayed only a slight enrichment of haematopoietic TF motifs.

When VEGF was withdrawn, peaks with a 2-fold increase in height in HE1 and HE2 were depleted for endothelial TF motifs SOX17, ERG, ETS and FLI1 indicating that the loss of VEGF signalling led to the activation of open chromatin regions lacking an endothelial motif signature (Figure 4.32b right panel). In contrast, in HE2 and HP cultured without VEGF we found an enrichment in motifs for the haematopoietic TFs RUNX, PU.1 and NFE2 indicating that these cells now reflect a haematopoietic TF motif signature. These findings are in keeping with our observations that VEGF withdrawal results in an increase in the overall number of HP cells and a depletion in the number of HE2 cells developing in our serum free I.V.D system.

4.4.3 Single cell RNA-Seq reveals a reduced progression of HE to HP cells in the presence of VEGF signalling

The development of the HE to HP cells is accompanied by a change in the gene expression program. To examine the transcriptional response to VEGF withdrawal at the level of individual cell types, we performed sc-RNA-sequencing on HE1 and HE2 from the serum free system cultured with and without VEGF to investigate whether VEGF withdrawal induced a haematopoietic transcription profile in these cells and whether the overall developmental trajectory changes. HE1 and HE2 cell types were collected at day 3 blast cultures and taken for SC-RNA-seq library preparation and sequencing.

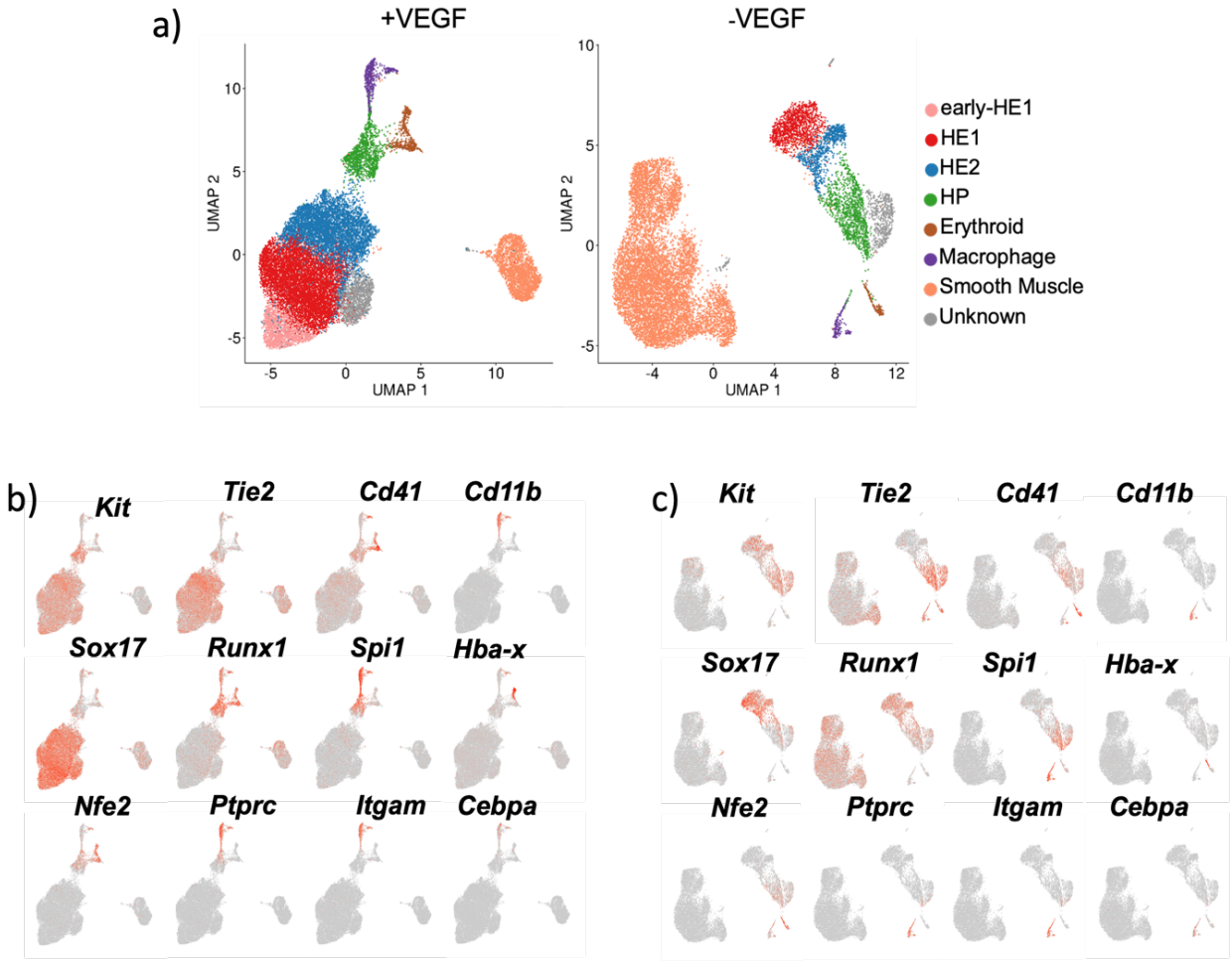


Figure 4.33- Withdrawal of VEGF results in a shift of the proportion of endothelial/smooth muscle and haemogenic endothelial/blood progenitor cells by blocking *Runx1* upregulation.

a) Single cell gene expression analysis of FACS sorted HE1 and HE2 populations in the presence (+VEGF) and absence (-VEGF) of VEGF. After HE1 and HE2 cell type data were pooled for each condition UMAP clustering was performed. b,c) gene expression levels of selected genes projected onto clusters with b) and without c) VEGF.

Single cell data were clustered by UMAP for cell types derived from All cytokines (+VEGF) and VEGF withdrawal (-VEGF) conditions (Figure 4.33a). In both the presence and absence of VEGF the full repertoire of cell types from HE1 to macrophages developed, demonstrating that the overall developmental trajectory was unaffected by the cytokine.

In the absence of VEGF the overall number of HE1, HE2 and HP cells were reduced and the proportion of smooth muscle cells was increased. However, the population showed an increased proportion of HP cells, suggesting that VEGF controls the frequency of the EHT. HE-specific *cis*-regulatory elements are enriched in AP-1 and TEAD motifs. This result is therefore consistent with findings from our group that VEGF signalling responsive AP-1 along with Hippo signalling responsive TEAD mediate the balance between smooth muscle and blood lineage commitment (Obier et al., 2016).

Sc-RNA-Seq differential gene expression analysis of all cells cultured in the presence of VEGF saw an upregulation of endothelial genes in HE1 such as *Sox17* (Avg log2FC 1.069, P Val Adj <1.38E-303) and *Cdh5* (Avg log2FC 1.982, P Val Adj <1.38E-303). In HE2 endothelial genes *Sox17* and Sox18 were also upregulated, however, haematopoietic genes such as *Runx1* (Avg log2FC -0.276, Adj P Val 2.6E-251), *Pu.1* (*Spi1*) (Avg log2FC 0.-0.271, Adj P Val 1.64E-144) were downregulated (For full lists of differentially expressed genes in HE1, HE2 and HP clusters see Supplementary table 6.2, Supplementary table 6.3 and Supplementary table 6.4, respectively). Analysis of marker expression across clusters also revealed a lack of upregulation of *Runx1* in HE2 and HP cells cultured with VEGF (Figure 4.33b-c).

Developmental trajectory analysis confirmed that the full haematopoietic lineage development occurred both in the presence and absence of VEGF (Figure 4.34a-d) and also confirmed that the smooth muscle population clustered separately from the haematopoietic cell types.

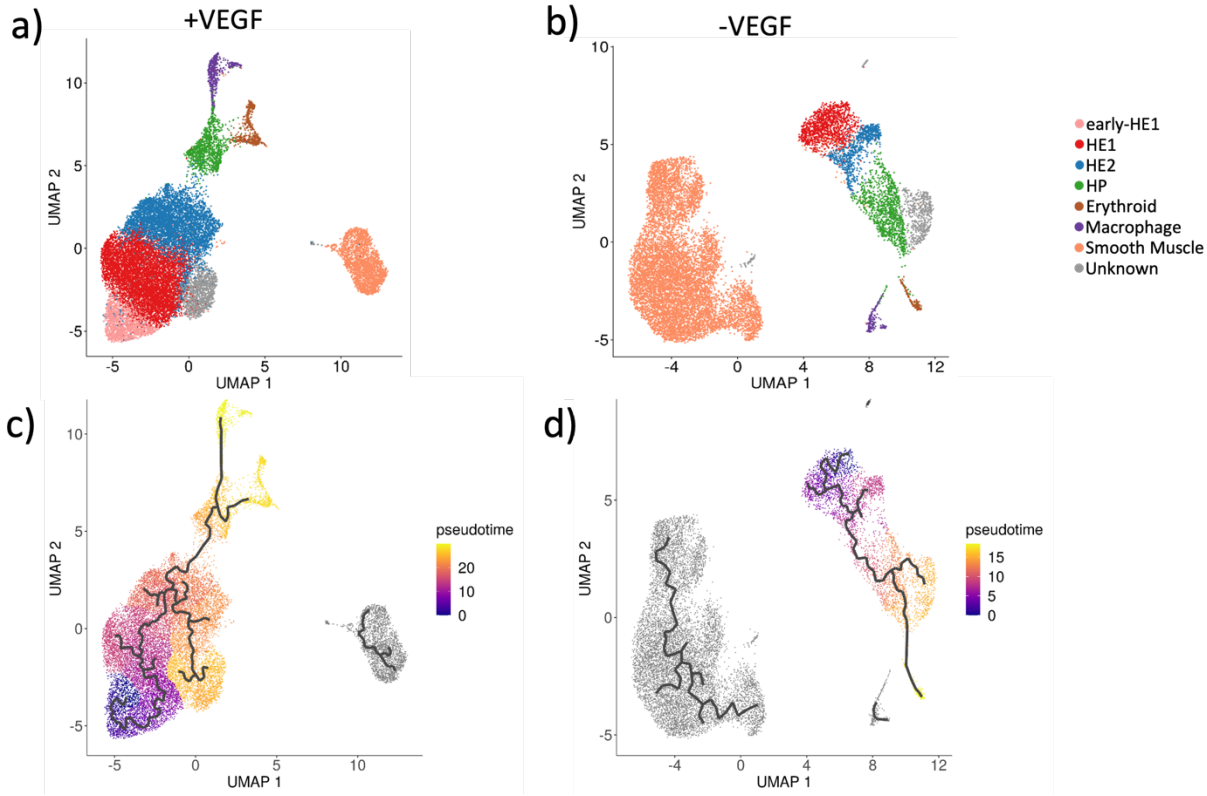


Figure 4.34- Single cell RNA seq analysis reveals altered developmental trajectories but full haematopoietic lineage development +/- VEGF.
a,b) Single cell gene expression analysis of FACS sorted HE1 and HE2 populations in the presence (+VEGF) and absence (-VEGF) of VEGF. After HE1 and HE2 cell type data was pooled for each condition UMAP clustering was performed. c,d) Pseudotime trajectory of cells derived from All cytokine (c, +VEGF) and VEGF omission (d, -VEGF) cultures of the clusters shown in Figure 4.33a.

We next performed differential gene expression analysis between each cluster of HE1, HE2 and HP from cultures with and without VEGF (+VEGF HE1 vs -VEGF HE1) to examine how VEGF signalling affects the TF expression within the individual cell types (Figure 4.35a-b).

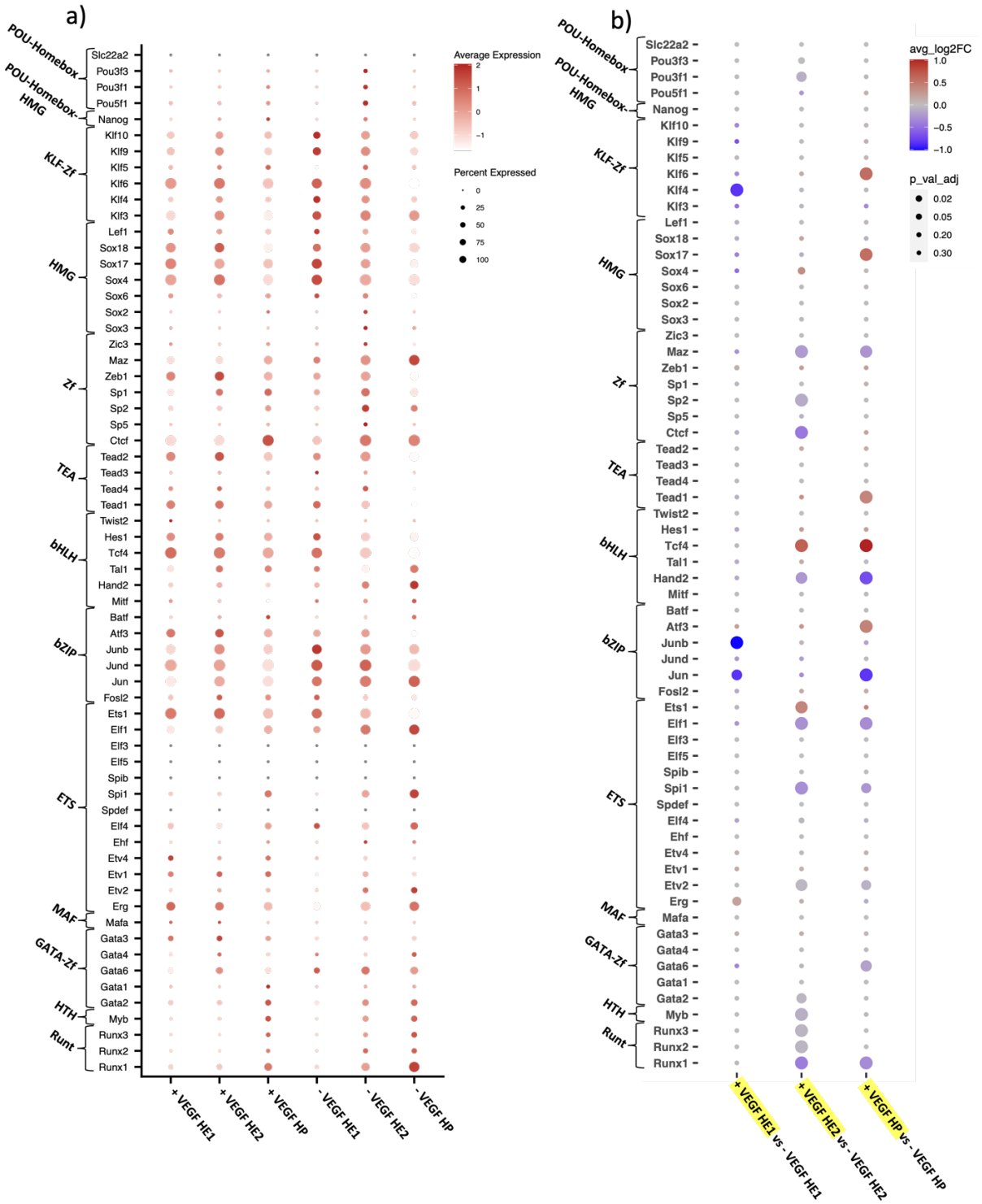


Figure 4.35- VEGF signalling alters transcription factor gene expression patterns as seen by sc-RNA-Seq.

a) Dot plot showing the average expression value (Normalised average UMI counts) and percentage of cells expressing each transcription factor for HE1, HE2 and HP clusters from Figure 4.33a with (+VEGF) and without (-VEGF) VEGF. b) Dot plot showing results of

differential gene expression analysis between HE1, HE2 and HP clusters with (+VEGF) and without (-VEGF) VEGF. The average log2 fold change (avg_log2FC) is shown in respect to the cell type highlighted yellow in the label. P_val_adj shows Bonferroni corrected P-Value. TFs grouped by TF family.

In HE1 SoxF factors were expressed in a greater percentage of cells in clusters cultured with VEGF (60.9% of +VEGF HE1 cells expressed *Sox18* compared to 34.1% in HE1 from - VEGF) (Figure 4.35a). However, SoxF factors were not differentially expressed in the HE1 or HE2 comparison indicating that VEGF does not regulate the expression of SoxF factors at the transcript level (Figure 4.35b). This result is consistent with findings from (Kim et al., 2016) showing that VEGF regulates SoxF factor expression not at the transcript level but at the protein level through mTOR signalling, and that SoxF factors are positive feedback regulators of VEGF signalling. The ETS family TF ERG was not differentially expressed between HE1 clusters, however, 46.1% of HE1 from the +VEGF cluster expressed *Erg* compared to 21.0% in HE1 without VEGF. In the presence of VEGF *Klf4* was differentially expressed in HE1 (Av log2FC -0.842 P Val Adj 1.0E-8).

JunB was significantly down regulated in HE1 in the presence of VEGF (Av log2FC -1.009 P Val Adj 0.004) (Figure 4.35b). Recent work from Chen and colleagues (Chen et al., 2022) demonstrated the requirement of *JUNB* expression and activity for the development of HE2 (haemogenic endothelial cell/HEC) and stated that *JUNB* is an essential regulator of the EHT. Our single cell transcriptomics data indicate that VEGF signalling directly represses *JunB* expression in HE1 cells (a pre-HEC endothelial cell type). Differential gene expression analysis of HE2 and HP clusters from +/-VEGF cultures reveal that *JunB* is no longer differentially expressed and that the percentage of cells from each cluster expressing *JunB*

is equal (in HE2 clusters: +VEGF 70%, -VEGF 72%, in HP clusters: +VEGF 54%, -VEGF 63%). These data support the findings of (Chen et al., 2022), and indicate (i) that *JunB* expression facilitates the transition of HE1 to HE2 (HECs) and (ii) that cells which have adopted a haematopoietic transcriptomic and TF motif signature (HE2 and HP) have previously reached a threshold level of *JunB* expression.

Runx1 is absolutely essential for the EHT, and the dosage of RUNX1 determines the frequency of cells undergoing the transition (Lie et al., 2018). *Runx1* and its target *Pu.1* (*Spi1*) were differentially expressed across the HE2 clusters (Figure 4.35b) with the expression of these genes being lower in VEGF containing cultures (Av log2FC -0.407 P Val Adj 3.2E-12 and Av log2FC -0.303 P Val Adj 1.9E-23 respectively). These findings support our previous observations at the level of the chromatin, where VEGF prevented the activation of *cis*-regulatory elements with a haematopoietic TF signature, and our FACS data showing that VEGF repressed the emergence of HP. Differential gene expression analysis between HP clusters similarly found that VEGF led to a down-regulation of the expression of *Runx1* (Av log2FC -0.323 P Val Adj 1.15E-05). The AP-1 TF JUN was also differentially expressed between HP clusters (Av log2FC -0.834 P Val Adj 4.09E-06). *Sox17* was differentially expressed between HP clusters with higher expression seen in HP derived from +VEGF culture (Av log2FC -0.572 P Val Adj 0.001) indicating that it had not been downregulated by RUNX1 (Lichtinger et al., 2012). Our data therefore suggest that VEGF regulates the EHT by controlling RUNX1 levels.

4.4.4 VEGF withdrawal alters the expression pattern of NOTCH signalling factors

We next investigated which other signalling factors were regulated by VEGF signalling. This cytokine directs the establishment of NOTCH signalling which maintains the endothelial cell type. Therefore, we studied the effect of VEGF withdrawal on the expression of NOTCH signalling factors (Figure 4.36a-c).

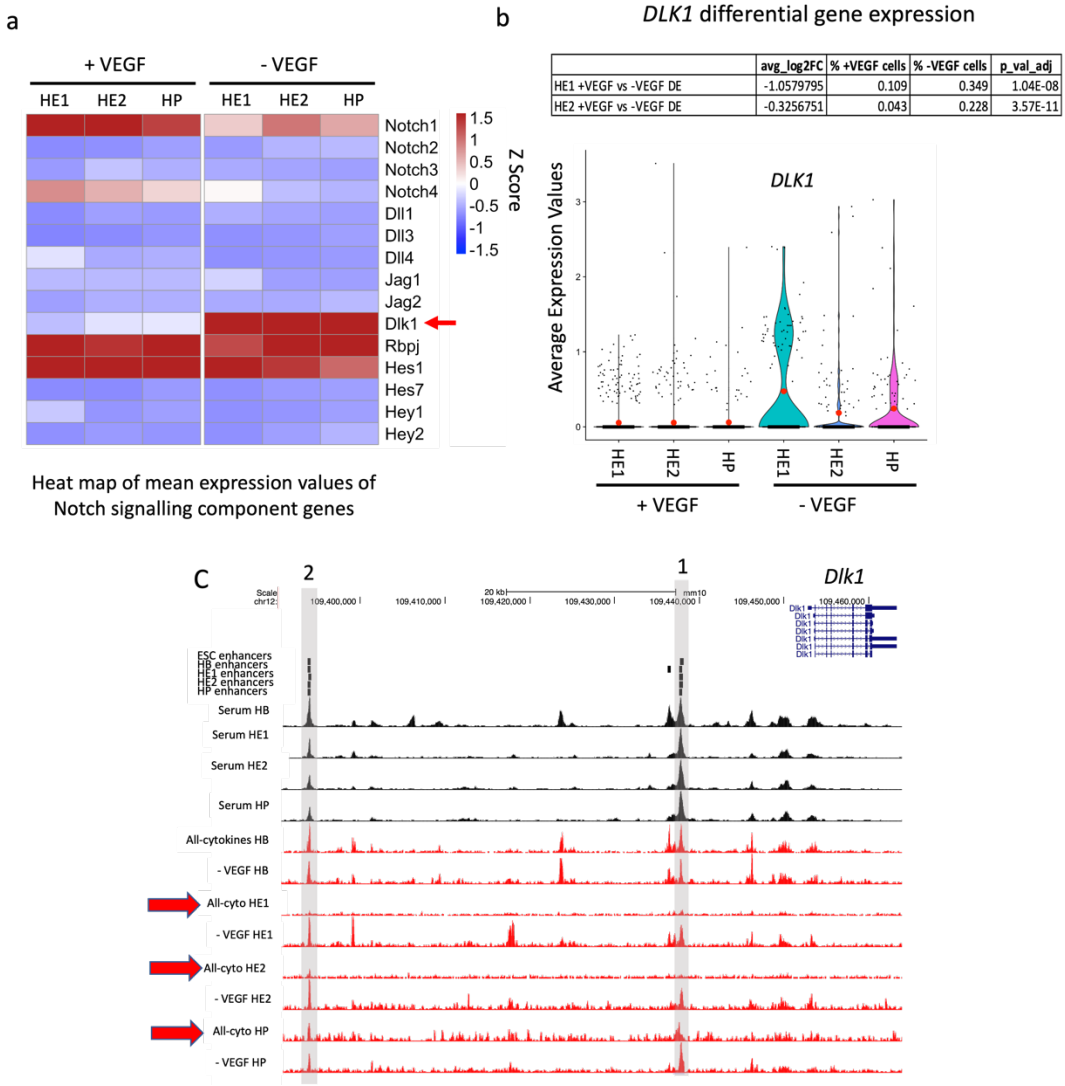


Figure 4.36- VEGF regulates the balance between endothelial and haemogenic development by altering NOTCH1 activity.
a) Heatmap showing the expression of key NOTCH signalling pathway genes with (+VEGF)

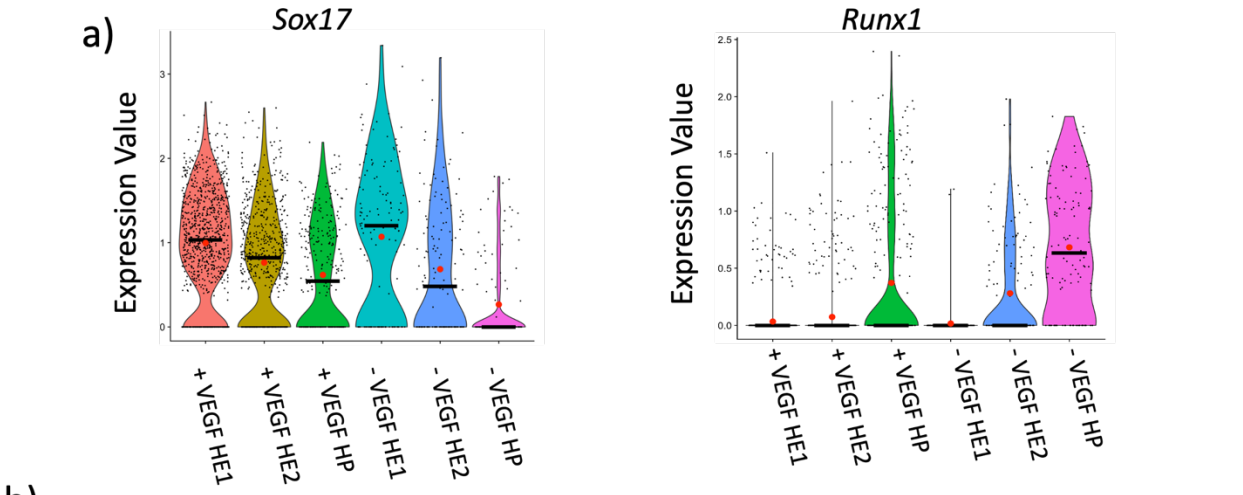
and in the absence of (-VEGF) VEGF from single cell expression data. b) Table showing differential gene expression analysis results from single cell data showing that *Dlk1* is differentially expressed between +VEGF and –VEGF containing cultures. Differential gene expression taken as +/-0.25 average log2 fold change and adjusted P-value<0.01. c) Violin plot showing the average expression values of *Dlk1* in HE1, HE2 and HP cell types cultured with (+VEGF) and without (-VEGF) VEGF. Black dots represent the spread of the data, red dots show the mean, black bar the median. c) Screenshot of the *Dlk1* gene locus with enhancers scoring positive in our screen marked as black vertical black bars. Serum and serum free +/-VEGF ATAC-Seq tracks shown with red arrows highlighting loss of open chromatin sites for the *Dlk1* enhancers in the presence of VEGF.

The analysis of our sc-RNA-Seq data showed that withdrawal of VEGF resulted in a significant decrease in the average mRNA expression values of *Notch1* in HE1, HE2 and HP (Figure 4.36a). In contrast, the *Dlk1* gene which encodes a membrane bound repressor of NOTCH1 activity (Mirshekar-Syahkal et al., 2013) was strongly upregulated in the absence of VEGF and differential gene expression analysis revealed this gene was differentially expressed between clusters (Figure 4.36b). Previously we showed that an enhancer which scored positive in our screen and which was linked to *Dlk1* by HiC was VEGF signalling responsive (Figure 4.29a-c). We also identified enhancers for *Notch1* and our ChIP data confirms occupancy of VEGF signalling responsive TFs at this element (Goode et al., 2016) (Supplementary data Figure 6.3).

Taken together, our data show that VEGF has a profound influence on the balance of expression of endothelial and haematopoietic transcription factors, resulting in an alteration of both the gene regulatory and signalling network.

4.5 VEGF blocks the upregulation of RUNX1 at the chromatin and gene expression level

The data described above showed that VEGF withdrawal resulted in an increase in the number of HP cells in our system and a decrease in the number of HE2 cells, suggesting that HE2 cells had undergone the EHT. Sc-RNA-Seq revealed that withdrawal of VEGF resulted in an upregulation of the haematopoietic TF *Runx1* in HE2 and HP cells and that in the presence of continued VEGF signalling the endothelial TF *Sox17* was not downregulated by RUNX1 in HP (Figure 4.37a-b).



b)

	+ VEGF HE1	+ VEGF HE2	+ VEGF HP	- VEGF HE1	- VEGF HE2	- VEGF HP
Sox17	2.26	1.66	1.27	3.23	1.82	0.52

Mean Normalised Expression Value

	+ VEGF HE1	+ VEGF HE2	+ VEGF HP	- VEGF HE1	- VEGF HE2	- VEGF HP
Runx1	0.051	0.125	0.838	0.032	0.492	1.300

Mean Normalised Expression Value

Figure 4.37- VEGF prevents the upregulation of *Runx1* and the downregulation of *Sox17* in HE2 and HP.
a) Violin plots of *Sox17* and *Runx1* average expression values for HE1, HE2 and HP cells from sc-RNA-Seq data with (+VEGF) and without (-VEGF). Spread of data represented by dots, red dot shows the mean, black bar the median. b) Tables showing mean normalised (Seurat log normalised) gene expression values for *Sox17* and *Runx1* for the indicated cell types in

the presence (+VEGF) and absence (-VEGF) of VEGF calculated from single cell expression data.

Taken together these data indicate that maintained VEGF signalling impairs HE1 committing to HE2 and represses the EHT process via a block in activation of *Runx1*. These data support previous findings that VEGF establishes and maintains NOTCH signalling and that VEGF signalling maintains the HE via ETV2 and SOX17 (Liu et al., 2015, Casie Chetty et al., 2017).

We therefore decided to study the regulatory enhancer elements of *Runx1* to elucidate how VEGF-NOTCH signalling acts at them. First, we studied the *Runx1* locus alongside our enhancer screen data and our ATAC-seq data for HB, HE1, HE2 and HP derived from serum cultures and serum free cultures with and without VEGF (Figure 4.38). Our enhancer screen faithfully captured previously identified and validated *Runx1* enhancer elements (Figure 4.38): the +23kb enhancer (Bee et al., 2009), the +3.7kb enhancer (Zhu et al., 2020), a -371 kb enhancer (Zhu et al., 2020), the +204kb enhancer (Owens et al., 2022), the +110kb enhancer (Owens et al., 2022) and others (Owens et al., 2022). Strikingly, in the presence of VEGF signalling open chromatin sites overlapping these enhancer elements failed to form.

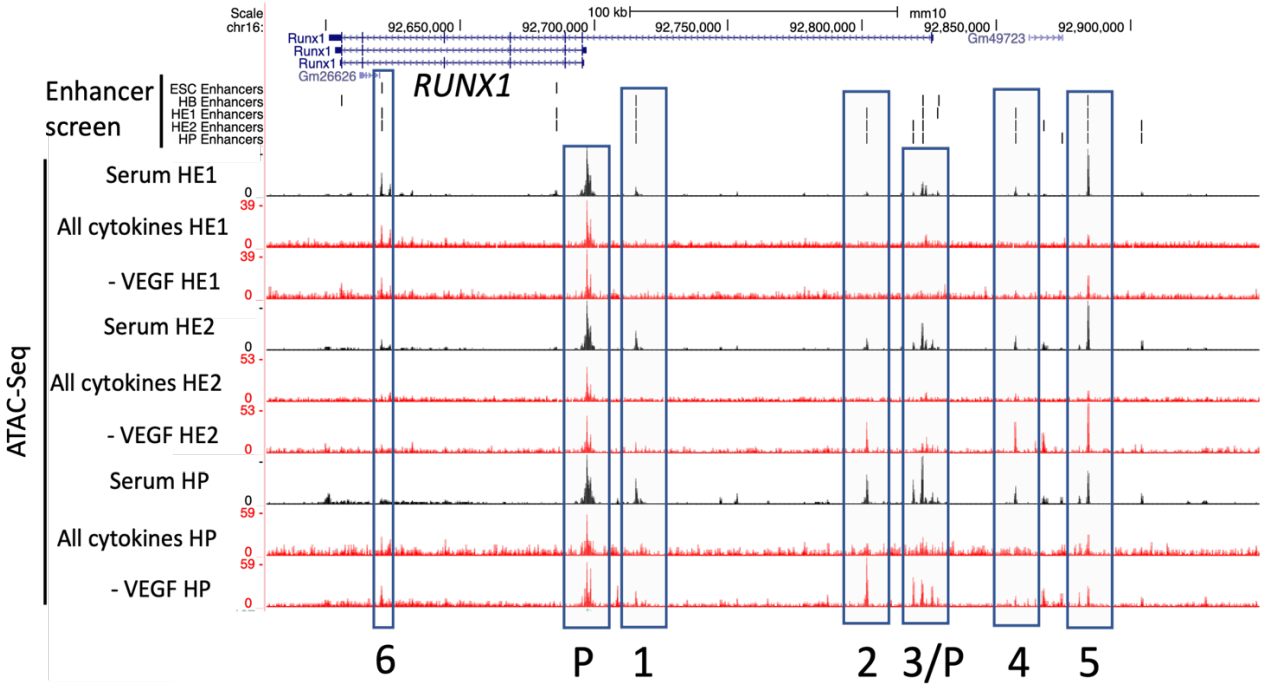


Figure 4.38- VEGF signalling results in altered chromatin accessibility at the *Runx1* locus distal regulatory elements.
UCSC genome browser screenshot showing open chromatin regions and the locations of distal enhancer elements scoring positive in our enhancer screen (vertical bars) at the *Runx1* locus under the indicated culture conditions (serum IVD, serum free IVD with VEGF (All cytokines), serum free IVD without VEGF (-VEGF)) at the different developmental stages (HE1, HE2 and HP). Enhancers are labelled 1 to 5 and promoter elements are marked P, both are indicated by blue boxes. Enhancer 1 is located at +110kb, Enhancer 2 is located at +23kb and enhancer 3 is at +3.7kb, enhancer 4 is at -34kb, enhancer 5 is at -58kb and enhancer 6 is at +204kb.

ChIP data from our group (Goode et al., 2016) showed that a number of these enhancer elements were bound by endothelial and haematopoietic TFs in HB, HE and HP (Supplementary data Figure 6.4). Annotation of the +23kb, +3.7kb, -31kb and -58kb enhancer elements revealed TF motifs for endothelial and haematopoietic TFs including SOX17, TAL1, Notch signalling effectors RBPJ and haematopoietic TFs GATA, RUNX1 and PU.1 (Figure 4.39a-b). ChIP data also revealed that endothelial TFs TEAD, AP-1 and haematopoietic TFs such as GATA, RUNX1 and PU.1 bind to the +23kb and +3.7kb

enhancers in HE and HP (Figure 4.39c) We selected the +23kb enhancer and the +3.7kb enhancer for validation and further interrogation.

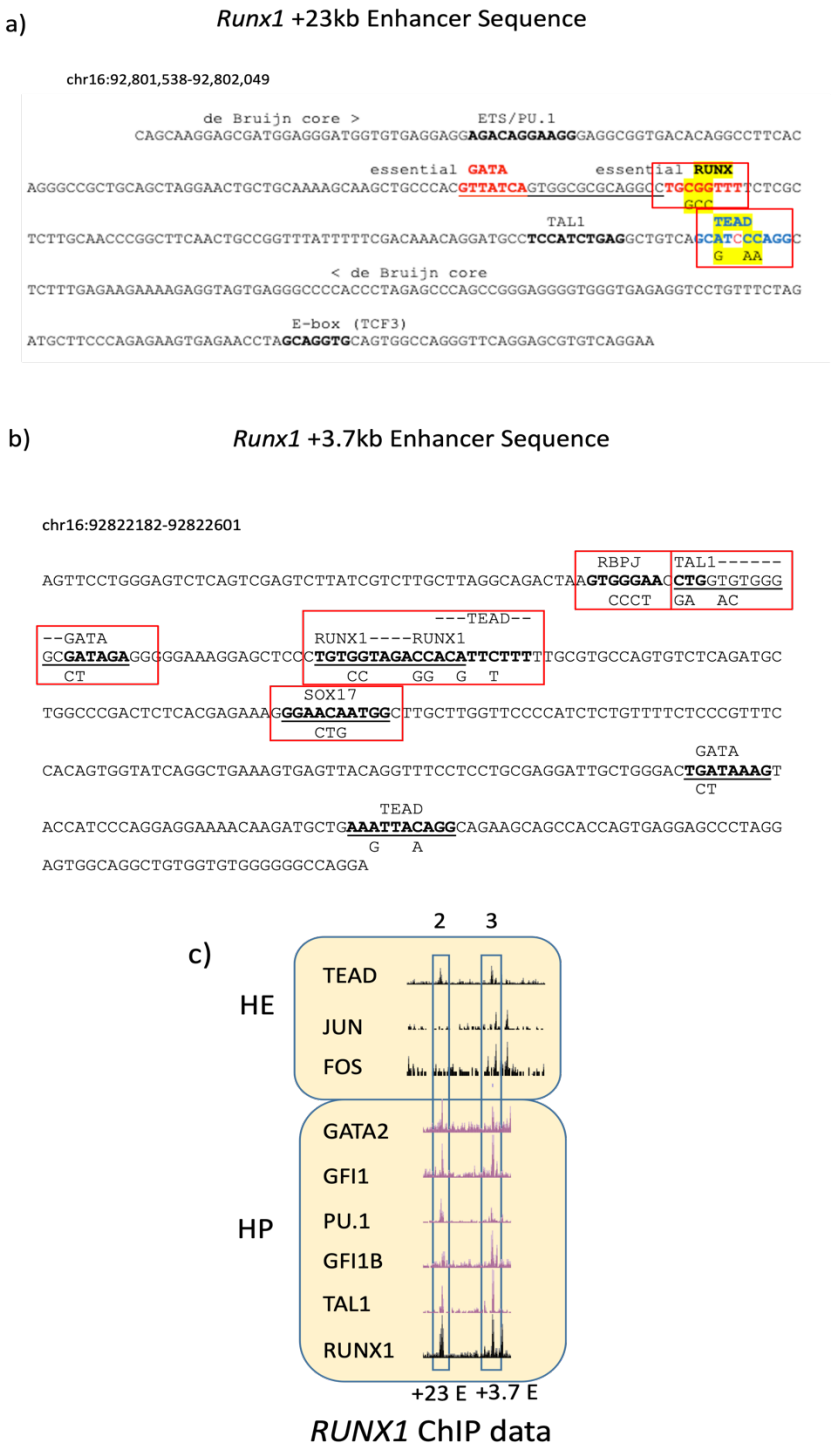


Figure 4.39-Wild type *Runx1* enhancer sequences and mutant elements.
Sequence of the +23kb enhancer (a) and the +3.7kb enhancer (b) with annotated binding

motifs. The position of motifs are highlighted by the red box and the nature of the mutations in both enhancers are depicted under the motifs. c) ChIP data showing the binding of indicated transcription factors to *Runx1* +23kb (2) and +3.7kb (3) enhancers in HE and HP cells (Goode et al., 2016).

We then validated the enhancers in our serum I.V.D system using Blast Media which is supplemented with additional VEGF. In the serum I.V.D system both enhancers displayed activity in HE1 and HE2 with the +3.7kb enhancer also being active in HP (Figure 4.14a and Figure 4.40a-b). We then tested which TF motifs in the enhancer elements could be regulating the activity of the enhancer elements. Our ChIP data (Goode et al., 2016) highlighted that the +23kb enhancer is bound by TEAD in HB and HE and RUNX1 in HP (Figure 4.39c). Previously, the GATA and RUNX1 TF motifs in this enhancer element had been demonstrated as being essential for its activity (Bee et al., 2009) but the role of TEAD is unknown. To address this issue, we mutated the RUNX1 and TEAD TF motif sites to abrogate TF binding and cloned the mutated enhancer sequences into the enhancer reporter system. Mutation of the essential RUNX1 site resulted in a reduction of enhancer activity in HE1 (P=0.028) and HE2 (P=3.95E-6) (Figure 4.40a), confirming that the RUNX1 site is essential for the activity of this enhancer. Mutation of the TEAD site resulted in a significant increase of reporter activity in HE2 (P=1.961E-5) and HP (P=2.148E-6), indicating that TEAD binding to this motif in the presence of VEGF signalling represses enhancer activity.

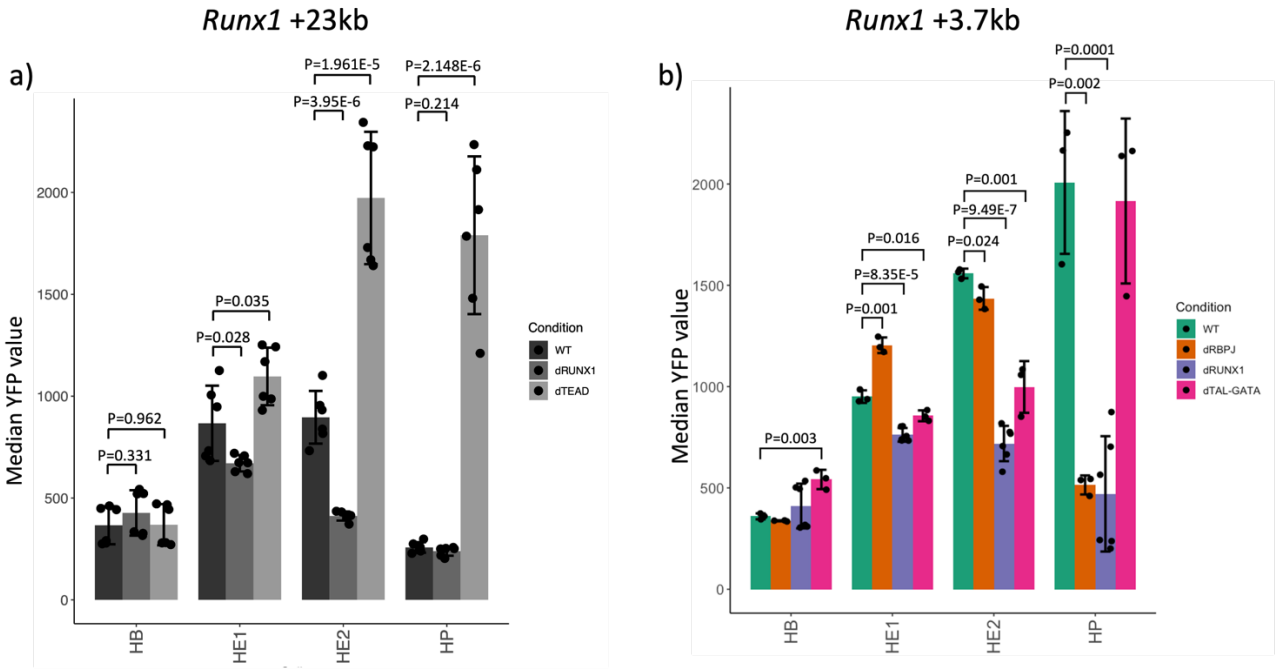


Figure 4.40- Specific transcription factors regulate the activity of the +23kb and +3.7kb *Runx1* enhancers.

a) Effect of mutations of the RUNX1 (dRUNX) and TEAD (dTEAD) binding sites on the activity of the *Runx1* +23kb enhancer in serum culture in HB, HE1, HE2 and HP compared to the wildtype (WT) enhancer. Error bars show standard deviation, P-value calculated from Student's two tailed t-test (n=6). b) Effect of mutations of the RBPJ (dRBPJ), RUNX1 (dRUNX) and TAL1-GATA (dTAL-GATA) binding sites on the activity of the wildtype (WT) *Runx1* 3.7kb enhancer in serum culture. Error bars show standard deviation, P-value calculated from Student's two tailed t-test (WT, dRBPJ, dTAL-GATA n=3. dRUNX n=6).

We then investigated the role of TF binding motifs found in the +3.7kb enhancer element. To this end we mutated the RUNX1, TAL1-GATA and RBPJ/IKAROS motifs in the element and differentiated the ESCs in the serum I.V.D system. Mutation of the RUNX motif resulted in the reduction of enhancer activity in HE2 (P=9.19E-7) and HP (P=0.0001) (Figure 4.40b). Mutation of the TAL1-GATA site resulted in a reduction of enhancer activity in HE1 (P=0.016) and HE2 (P=0.001), however, enhancer activity in HP cells was not significantly different to the activity seen in the wildtype. This suggests that the enhancer activity is driven by TAL1 and GATA in the HE but these factors are no longer required for enhancer

activity in HP.

Mutation of the RBPJ motif resulted in a slight reduction of activity in HE2 (P=0.024) followed by a significant reduction of activity in HP (P=0.002). The NOTCH signalling effector RBPJ shares the same TF motif as the haematopoietic TF IKAROS (IKZF1). There is evidence that RBPJ and IKAROS factors compete for the same motif (TGGGA) from studies in T cells (Geimer Le Lay et al., 2014) and IKAROS expression increases from HE2 to HP (Gilmour et al., 2018, Goode et al., 2016) before it is reduced again in macrophages. It is therefore possible that this motif is not bound by RBPJ in HE1 and HE2 but it is bound by IKAROS in HE2 and HP which drives its activity.

4.5.1 Identification of transcription factors mediating VEGF responsiveness using the GALNT1 enhancer reporter

The final experiments of this study examined the question of which TFs were connected to endothelium-specific gene expression and VEGF signalling responsiveness. It is known that HIPPO signalling is being crucial for HE development (Goode et al., 2016) and others have indeed shown that shear stress activating the TEAD partner YAP induces the EHT via activation of *Runx1 (Lundin et al., 2020)*. From our ChIP data it is clear that TEAD binds to its motifs in HE (Obier et al., 2016) but binding is lost in HP cells (Goode et al., 2016). TEAD and AP-1 often show a specific 7 base spacing between binding motifs (Obier et al., 2016). The same study demonstrated that the expression of a dominant negative (dn) FOS peptide

which blocks all AP-1 DNA binding caused a reduction of TEAD binding at such sites. This result suggests that HIPPO and MAP kinase signalling converge at such sites. RUNX1 has been shown to repress the endothelial program in cooperation with the TF GFI1 (Lancrin et al., 2012, Thambyrajah et al., 2016, Hirai et al., 2005). We were therefore also interested in understanding the role of RUNX1 in regulating developmental-stage specific enhancer activity. To examine the role of these TFs in more detail we selected enhancer elements which were (i) specifically active in the HE1 and HE2 cell types, (ii) were bound by several of these factors as seen by ChIP and (iii) were VEGF responsive. The *Galnt1* enhancer fitted these criteria (Figure 4.41a-b) as it contains a compositive TEAD-AP-1 site (Figure 4.41b highlighted in blue) overlapping a RUNX1 binding site together with motifs for ETS, TAL1 and GATA. The combination of these motifs enabled the study of the interplay between these factors in response to VEGF. ChIP data analysis revealed that the enhancer is bound by these TFs during the different developmental stages (Figure 4.41c).

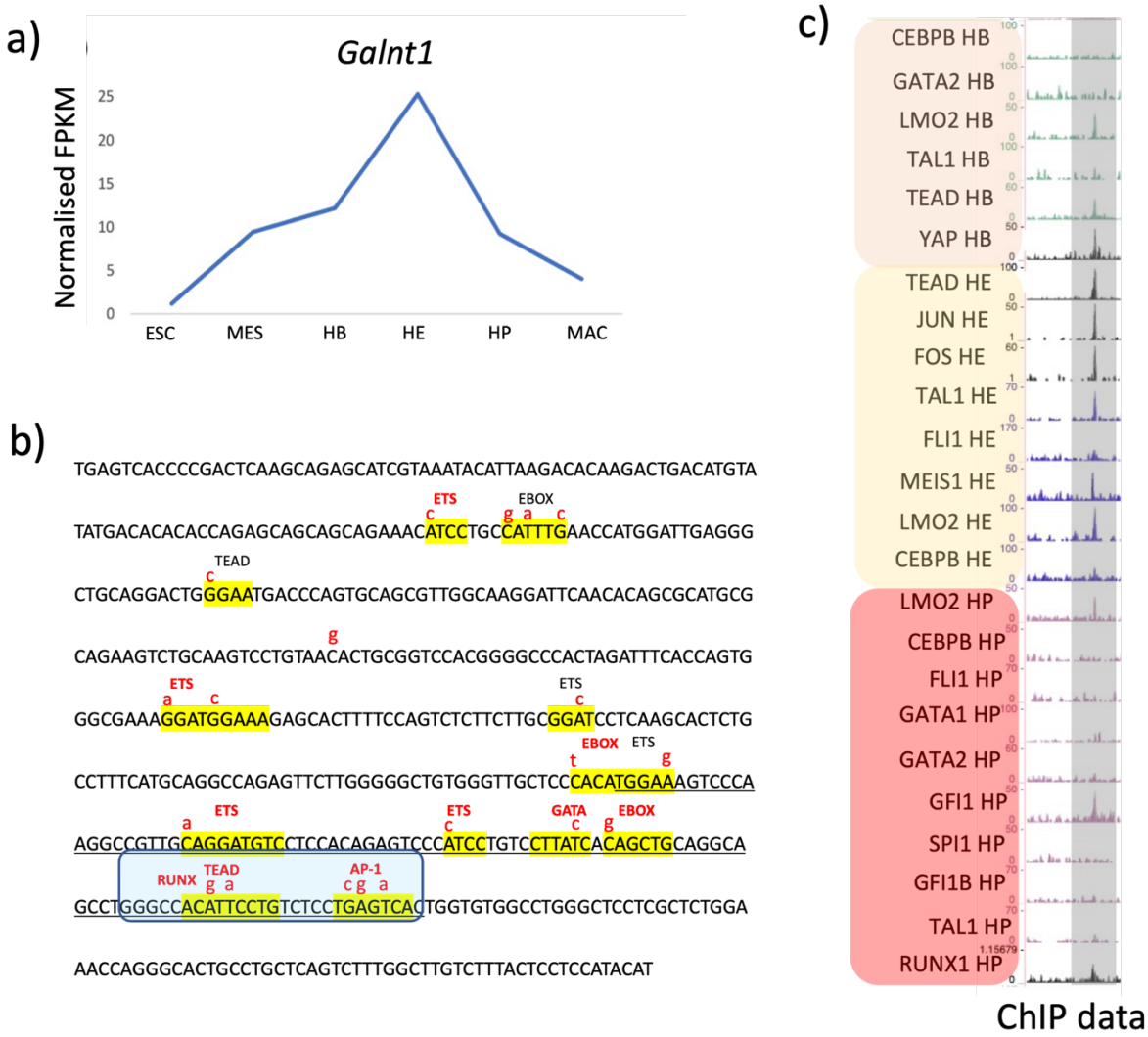


Figure 4.41- The *Galnt1* enhancer as a reporter for developmental-stage specific TF activity.

a) Plot showing the HE-specific expression of *Galnt1* (Goode et al., 2016). b) Sequence of the *Galnt1* enhancer with TF binding motifs highlighted and the mutations made indicated in red. Strong consensus sequences are highlighted in red. The TEAD-AP-1 conformational arrangement (Obier et al., 2016) is highlighted by the blue box. c) Genome Browser screenshot depicting binging of the indicated transcription factors at the different developmental stages to the *Galnt1* enhancer (Goode et al., 2016).

dnFOS resulted in a 2-fold decrease in the *Galtn1* enhancer ATAC peak height in HE (Figure 4.42a) and resulted in a greater than 2-fold decrease in the expression of *Galnt1* (Figure

4.42b), indicating that AP-1 activates the element and is required for the binding of TEAD factors within this composite module. Analysis of our Sc-RNA-seq data reveals that VEGF withdrawal results in a slight decrease of *Galnt1* expression in HP (Figure 4.43a-b).

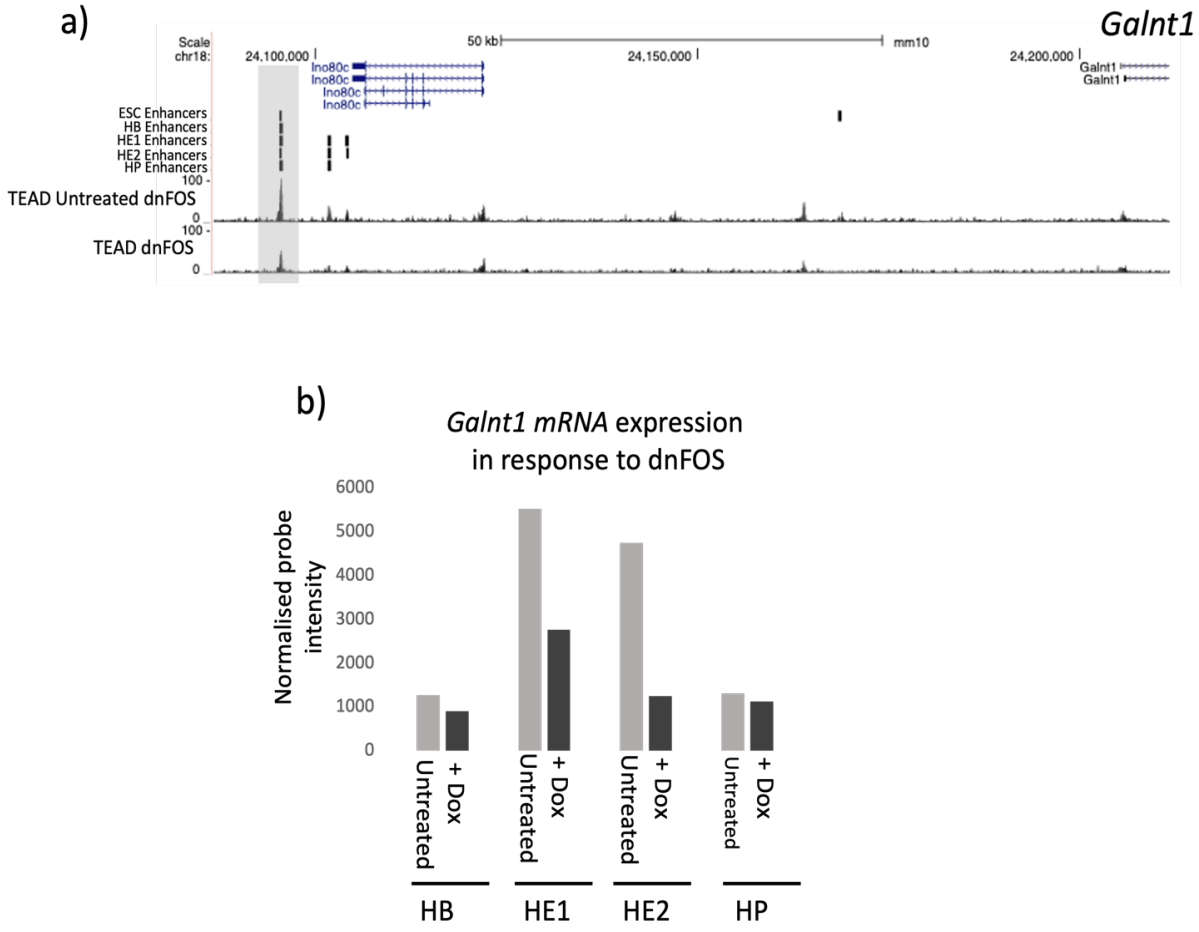


Figure 4.42- Cooperation of the TEAD4 and AP-1 transcription factors are required at the *Galnt1* enhancer- AP-1 is required for TEAD4 binding and enhancer activity.
a) Genome Browser screenshot of ChIP data showing the binding of TEAD4 at the *Galnt1* enhancer (highlighted in grey) with and without expression of a doxycycline inducible dominant negative FOS (dnFOS) peptide. Expression of dnFOS resulted in a two-fold decrease in the peak height at the *Galnt1* enhancer. b) *Galnt1* mRNA expression (Normalised probe intensity) in HB, HE1, HE2 and HP cells before (untreated) and after expression (+Dox) of the dnFOS peptide (merged duplicate microarray data from (Obier et al., 2016)).

a) Mean Normalised Expression Value

	+VEGF HE1	+ VEGF HE2	+ VEGF HP	- VEGF HE1	- VEGF HE2	- VEGF HP
Galnt1	1.970	2.496	1.846	2.320	1.978	1.389

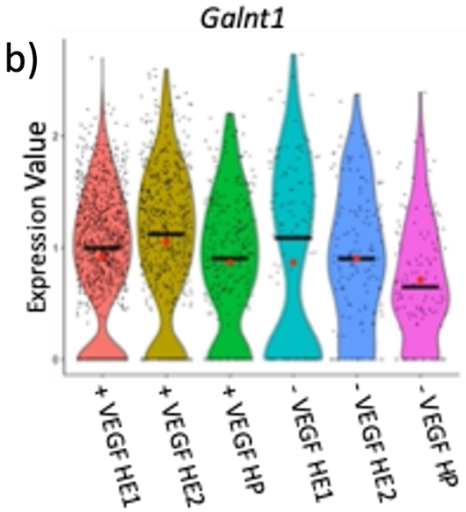


Figure 4.43- *Galnt1* expression in response to VEGF.
a) Average Expression Value (Seurat log normalised average UMI counts) *Galnt1* mRNA expression in HE1, HE2 and HP cells cultured with (+VEGF) and without (-VEGF) VEGF as measured by sc-RNA-seq b) Violin plots showing expression value of *Ganlt1* from sc-RNA-seq data. Black dots show the spread of the data. Black line represents the median and red dot represents the mean.

We therefore studied the role of these factors in more detail by mutating the different binding sites and culturing ESCs harbouring the mutated enhancer sequences in the serum free system with and without each of VEGF, BMP4, IL6 or IL3. First, we measured the activity of the WT enhancer (Figure 4.44) and compared it to the promoter control by FACS. We assayed the median YFP value (Figure 4.44) and the number of YFP+ cells (Figure 4.45) for each cell type. The activity of the enhancer (Figure 4.44) mirrored the gene expression profile (Figure 4.41a) in the presence of all cytokines.

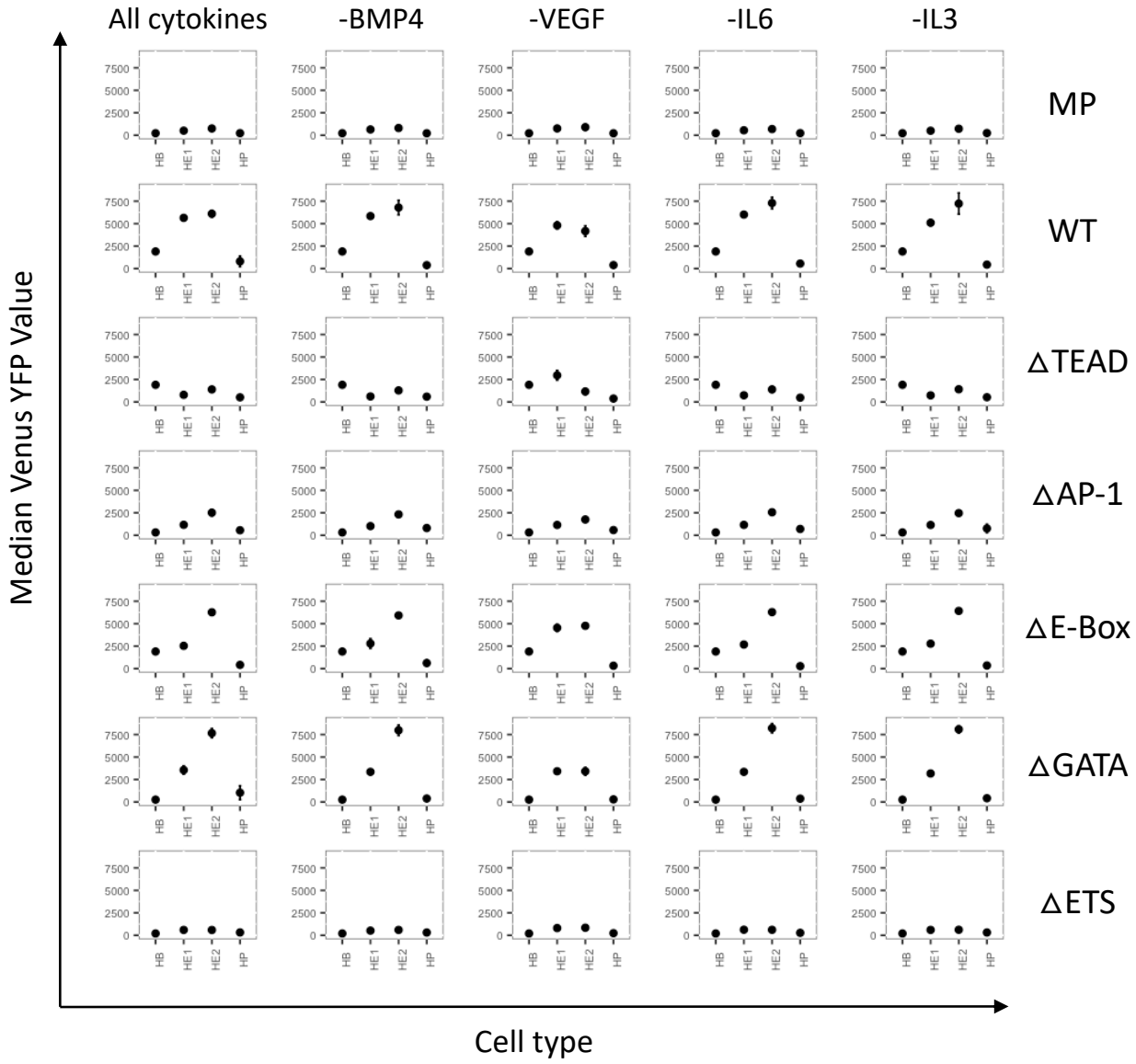


Figure 4.44- Reporter activity (median Venus YFP value) of the minimal promoter control (MP) wild type (WT) and mutated *Galnt1* enhancer elements in the presence and absence of the indicated cytokines.

Cytokine condition shown at the top of the grid, cell type condition shown on the right (ΔTEAD=enhancer sequence with TEAD site mutated). Error bars represent the standard deviation from n=3.

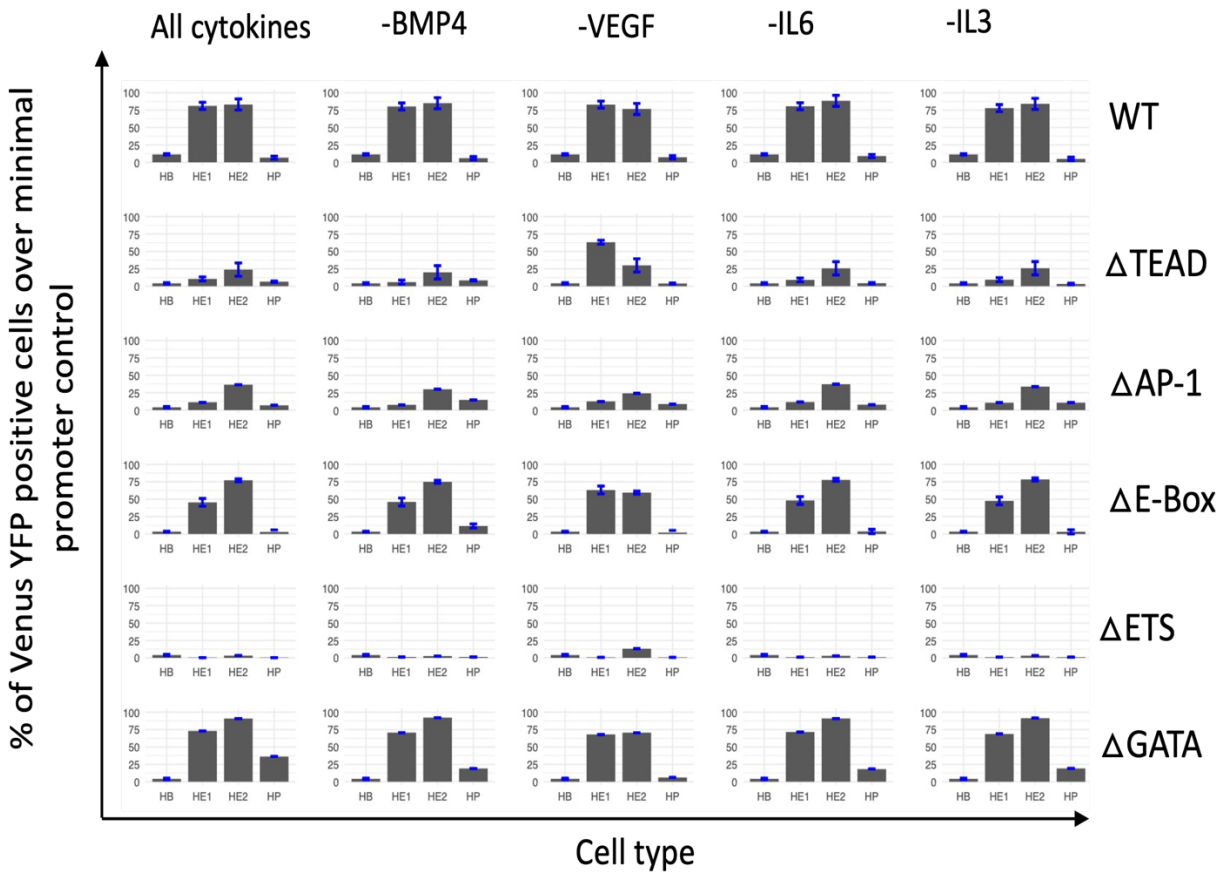


Figure 4.45- Reporter activity (% of Venus YFP positive cells over the minimal promoter control) of the wild type (WT) and mutated *Galnt1* enhancer elements in the presence and absence of the indicated cytokines.
Cytokine condition shown at the top of the grid, cell type condition shown on the right (ΔTEAD=enhancer sequence with TEAD site mutated). Error bars represent the standard deviation from n=3.

Analysis of the mutant enhancer elements revealed that ETTS, TEAD and AP-1 motifs were essential for the activity of the enhancer. Mutation of either the TEAD or AP-1 motifs resulted in a loss of activity, confirming the cooperative activity of these TF in this element (Figure 4.44). Mutation of the TAL1 and GATA sites resulted in a loss of activity in HE1 in the presence of all cytokines (P=3.5E-05 and P=0.009, respectively). However, activity was restored in HE2 indicating that TAL1 or GATA are required for enhancer activity in HE1 but not in HE2. Our ChIP data reveal that this enhancer is bound by RUNX1 and GFI1 (Figure

4.41c) at a site overlapping the TEAD motif suggesting that a factor exchange occurs here and a repressive RUNX1 complex is formed.

The analysis of the effects of cytokine withdrawal showed that withdrawal of BMP4, IL6 or IL3 had no effect on enhancer activity. However, withdrawal of VEGF from the cultures resulted in a decrease in WT enhancer activity HE2 (P=0.017). In addition, the mutation of the TEAD and overlapping RUNX site resulted in an increase in the percentage of YFP+ cells in HE1 in the absence of VEGF (Figure 4.45), suggesting that TEAD and RUNX1 are in balance to activate or restrict enhancer activity. Taken together our data indicate that VEGF regulates enhancer activity in the HE through regulation of AP-1, TEAD and RUNX1 TFs at this site. We suggest that sites like this are genomic regions where different signals converge and drive differential gene expression.

5 DISCUSSION AND FUTURE EXPERIMENTS

In this study, we (i) identified thousands of differentially active enhancer elements at five stages of development from ESCs to HP. (ii) We optimised a serum free I.V.D system to drive the differentiation of ESCs to HP and (iii) modulated the cytokine environment to study the signalling responsive open chromatin landscape of the system at each developmental stage. (iv) We overlapped signalling responsive open chromatin sites with our functionally identified enhancer list to identify signalling responsive enhancer elements and (v) defined the TF landscape acting at each set of signalling responsive enhancer elements. (vi) For one cytokine, VEGF, we validated a selection of enhancers as being signalling responsive. (vii) We studied the effects of VEGF on the differentiation from HB to HP cells at the cellular, transcriptomic and chromatin level, recovering the known role of VEGF driving the differentiation of HB to HE. (viii) We added to the literature by demonstrating that VEGF signalling represses the EHT and HP emergence through a VEGF-NOTCH signalling axis which represses *Runx1* expression. Finally, we Interrogate two *Runx1* enhancer elements mediating the responsiveness to VEGF and used the *Ganlt1* enhancer as a reporter to study which TFs are responding to which cytokines.

5.1 Our enhancer screen identified thousands of elements which are bound by specific transcription factors

Our high-throughput genome-wide enhancer screen allows for the identification of thousands of enhancer elements in cell types derived from ESCs. A previous study using open chromatin region-based libraries used a lentiviral integration method and not a single integration site. The study also only studied elements active in a single cell type (Murtha et al., 2014). Another study employed the same integration system we used but only studied limited genomic regions (Dickel et al., 2014). Through the integration of a reporter cassette into a HPRT locus in ESCs we make sure that the reporter is in open chromatin throughout differentiation. This is an important distinction from other methods because it allows for a sequence housed in the open chromatin region to be assayed for enhancer activity during different stages of differentiation from ESCs to more committed cell types. This method allows for the study of how different transcription factor and signalling landscapes can influence the enhancer activity of a given sequence. This methodology could be applied to any differentiation pathway which originates from ESCs using both serum and serum free cultures (discussed below). Our method returned previously well characterised enhancers such as for *Pu.1* (*Spi1*) (Leddin et al., 2011) and *Runx1* (Zhu et al., 2020, Bee et al., 2009).

One limitation of our method however is that it does not recapitulate the topology of the endogenous locus for a particular enhancer region. Furthermore, our method does not assess the role and impact of two or more enhancer elements acting to stimulate a promoter element through chromatin looping. These are of course limitations that also

apply to most other high throughput enhancer screen methods including the various STARR-Seq techniques (Arnold et al., 2013).

There are also considerations to be had about the TN5 cutting of open chromatin regions into sub-fragments of a putative enhancer. If the enhancer is larger than the fragment, and aspects of the enhancer required for its activity is not included in the fragment, then it is possible that we miss the reporting of such an enhancer element. However, from the size of DNaseI hypersensitive sites mapped in Goode et al., (Goode et al., 2016), most enhancer elements do not cover more than one or two nucleosomal regions, and the isolation of ATAC-seq fragments which are cloned into the enhancer library includes a size selection step (350-650bp). It is therefore unlikely that we miss large fragments. Furthermore, if a portion of an enhancer containing binding sites for specific factors active in a specific cell type which are repressive are not included in a fragment, then the enhancer could be incorrectly classified as active in that cell type. While the integration of enhancer sub fragments and not the full putative element may be considered a limitation, it could also provide a valuable resource. Studying the negative and positive enhancers sub-fragments may reveal information about which parts of an enhancer are required for its activity or repression. Our data are therefore a valuable resource that should be mined in future studies assessing enhancer sub-structures.

Importantly our enhancer list contains many enhancer elements found in conserved sequences and more than 20% of enhancer elements were also found in the human genome (data not shown). Interestingly, we found that the majority of open chromatin

regions can stimulate the minimal promoter. Previous work by Zuin, et al. (Zuin et al., 2022) demonstrated that the proximity of an element to a promoter has a significant effect on the level of enhancer activity. However, this work examined enhancer sequences by positioning them at different distances up to several kilobases away from a promoter and measured the promoter output and interactions. These distances are much larger than what is used in our reporter assay and in STARR-Seq techniques. It is possible that at increased distances from the promoter some elements may not have scored positive.

Our method identifies regulatory elements which are capable of stimulating a minimal promoter similar to what many transient reporter assays have done in the past. However, it is clear that in an endogenous locus the developmental control by enhancers is more stringent and a combination of elements may be needed to stimulate a specific promoter at the right level and in the right cell type. Individual elements captured by our method may form groups of regulatory elements or super-enhancers (SEs) to regulate a promoter. SEs are a class of compound regulatory elements which control the expression of developmental genes (Blayney et al., 2022). However, it is unclear whether they are groups of independent enhancer elements or if a SEs architecture contains certain properties which are essential for their function and therefore should be taken as a distinct class of element.

The question was addressed by Blayney et al. (Blayney et al., 2022). Their study employed synthetic biology and genome editing techniques to interrogate the constituent enhancer elements of the well characterised erythroid α-globin SE at its endogenous location. Initially

they removed all the SE elements in a mouse embryonic stem cell line. They then rebuilt the SE by individual and combinatorial reinsertion of its five elements (R1, R2, R3, Rm and R4) and assessed the importance of each sequence and its location. They found a high degree of functional interdependence between the five elements and that the classical enhancers R1 and R2 acted sub-optimally on their own and in combination. The R3, Rm and R4 elements provided no enhancer activity alone, however, they each exerted a major positive effect in facilitating the activity of the classical enhancers R1 and R2. Their findings suggest that as well as the actual sequence the position and combination of sequences influence the activity of the classical enhancers. They propose that these elements are a novel form of regulatory elements termed "facilitators". Our enhancer screen captured the R1 and R2 enhancers in HE2 and HP and the Rm enhancer in HE2 only, supporting their findings that the R3 and R4 elements had low enhancer activity alone. Our enhancer dataset could therefore be mined to study which components of SE elements are capable of driving a minimal promoter and which elements may play a complementary role.

5.2 Developmental stage-specifically active enhancer elements are regulated by developmental stage-specific transcription factors

Relative motif enrichment analysis of distal ATAC sites from HB, HE1, HE2 and HP confirmed the findings from Goode et al. (Goode et al., 2016). Open chromatin regions were enriched for motifs of TFs which were active in each developmental cell type and played a role in regulating cell stage specific *cis*-regulatory elements. The same analysis of functional

enhancer elements revealed a similar but not identical motif enrichment pattern compared to the global motif enrichment profile of distal sites. For example, CTCF binding motifs were enriched in distal sites but depleted in enhancer elements. This result is consistent with findings that CTCFs binding marks TAD and loop boundaries within which gene regulatory elements are located and are not commonly located in elements impacting on transcription activity (Dixon et al., 2012). We noted that the overall enrichment signal was much clearer when analysing enhancer elements alone compared to all distal sites.

Our motif co-localisation analysis of enhancer elements revealed that numerous stage-specific TFs co-localise within a distance of 50 bp in stage specifically active enhancer elements. Notably, AP-1 binding sites co-localised with many TFs across every cell type. Several studies have highlighted that AP-1 family TFs often co-operate with other TFs to bind to DNA and regulate gene expression. For example, BATF-3 requires AP-1 cofactors to dimerise as BATF family members lack a transcriptional activation domain (Schwering et al., 2003, Verma et al., 1995) and has been found to directly interact with JUN and JUNB as measured by co-immunoprecipitation and mass spectrometry (Lollies et al., 2018, Schleussner et al., 2018). Our group also showed that at many genomic sites TEAD and AP-1 cooperatively bind in the HE, and these factor motifs are enriched together in enhancers active in this cell type (Obier et al., 2016). AP-1 motifs also co-localise with other AP-1 motifs in all cell types. Depending on their compositions, AP-1 dimers can bind to different types of palindromic sequences. For example, FOS:JUN and JUN:JUN dimers preferentially bind to DNA motifs known as 12-O-tetradecanoylphorbol-13-acetate (TPA)-response element (TGATCA) and have a reduced affinity for cAMP-responsive element (CRE) which

has another base-pair in the middle (TGACGTCA). ATF-containing dimers preferentially bind to CRE and have a weaker affinity for TPA-response elements (Eferl et al., 2008).

AP-1 factors have been associated with glucocorticoid receptor (GR) dependent chromatin remodelling. GR and AP-1 regulate gene expression by interacting with chromatin whereby AP-1 recruits a chromatin remodelling complex which acts to create a more accessible chromatin state. This allows for additional TFs to bind and regulate gene expression. AP-1 can also transiently bind its motif and induce chromatin remodelling events through a fast "hit and run" mechanism (Voss et al., 2011). These findings indicate that the large AP-1 TF family plays an important role in regulating gene expression across many cell types and cooperates with numerous TF families.

As outlined in the introduction, many accessible chromatin regions are formed before the expression of its associated gene is activated, and thus serve as 'priming elements' (reviewed in (Bonifer and Cockerill, 2017)). Our enhancer screen enables the tracking of enhancer element activity through differentiation of ESCs to HP. It is therefore possible to identify which enhancer elements may act as priming element in a gene locus. Filtering our enhancer list for enhancers which are active in the cell type prior to gene expression of the gene they are associated with may yield a list of potential priming elements. By studying which stage specific TFs bind to these elements or which motif elements are found in these elements it may be possible to further classify such elements. CRISPR-Cas9 knock-out validation of these predicted priming elements could be performed to study the impact of their deletion on the kinetics of expression of their associated genes.

Future bioinformatic studies could also focus on studying the co-localisation of three or more TF motifs in these elements and determine the most commonly occurring spacing between these elements as compared to a background of sites. If identified TF motifs patterns are functional at enhancers, then it is possible that they are functionally conserved in these elements compared to a background of sites. This method could be used to predict TF binding complexes or TF binding arrangements which could highlight binding patterns which activate or repress an enhancers activity.

5.3 Identification of thousands of signalling responsive enhancer elements

The presence of signalling-responsive factors such as AP-1 co-localising with multiple other transcription factor families suggested that they could confer signalling responsiveness to specific enhancer elements. Our work has indeed demonstrated throughout this work that our enhancer screen could identify thousands of signalling responsive chromatin elements at each developmental state of mouse ESCs differentiating to HP most of which overlap with functional enhancers. We found that the greatest number of enhancer elements were responsive to VEGF in HB, HE and HP. This finding is consistent with literature showing the importance of VEGF signalling for the development and maintenance of these cell types (Liu et al., 2015). We validated five enhancer elements as being VEGF signalling responsive in the serum free I.V.D system including an enhancer linked to the *Cdh5* gene, which is a

known marker of the HE. From our data it may also be possible to identify signalling responsive priming elements.

As discussed previously our enhancer screen method can be applied to any cell type which can be derived *in vitro* from ESCs and through combining this system with a serum free I.V.D system we can identify signalling responsive enhancers. Other serum free I.V.D culture systems have been developed to drive the differentiation of ESCs to specific cell types through the three germ layers. For example serum free I.V.D systems exist for the differentiation of ESCs to cardiac cells (Gissel et al., 2006) pancreatic lineages (Blyszczuk and Wobus, 2006, Micallef et al., 2007), alveolar epithelial cells (Winkler et al., 2008) and osteoblasts (Alfred et al., 2010) among others. This method used in this study could be applied to identify enhancers active in these cell types and the signalling factors regulating their activity through signalling responsive TFs.

5.4 VEGF is required for the development of the HE from HB

As previously described, HB cells express the VEGF receptor tyrosine kinase cell surface receptor FLK-1. Studies have demonstrated that the FLK-1 receptor is essential for the development of endothelial and haematopoietic cells *in vivo* and that VEGF binding to FLK-1 induces MAPK signalling to the ETS transcription factor ETV2 which acts to promote endothelial differentiation (Liu et al., 2015, Casie Chetty et al., 2017). Our motif analysis revealed that VEGF signalling induced an enrichment of endothelial TF motifs ETS, AP-1, TEAD and SOX17, suggesting that VEGF signalling acts through these factors. Previously our

group demonstrated that the VEGF signalling responsive TF AP-1 and the Hippo signalling responsive TF TEAD cooperatively act at sites to drive the commitment of HB to HE over smooth muscle (Obier et al., 2016). Our scRNA-seq analysis revealed that in the absence of VEGF some cells do differentiate from HB to HE1 and commit to the hematopoietic lineage. However, withdrawal of VEGF resulted in an expansion of smooth muscle cells which is consistent with studies showing that VEGF signalling is required for the development of endothelial cells which thereafter give rise to the haematopoietic cells via the EHT.

The analysis of enhancers which were signalling responsive to VEGF revealed that they were enriched in NOTCH signalling effectors RBPJ in HE1 and HE2 and HES1 in HP. When VEGF was withdrawn these TF binding motifs were depleted in enhancer elements. sc-RNA-Seq analysis revealed that *Notch4* expression was higher in HE1 cultured with VEGF than without. *Notch4* expression is highest in the HB and drops significantly to the HE (Goode et al., 2016). Continued VEGF signalling maintains low levels of *Notch4* expression closer to what is seen in the HB. *Notch1* was also more expressed in HE and HP from cultures with VEGF. *Notch1* is the main NOTCH receptor expressed in the HE (Goode et al., 2016) and its activation through interacting with the NOTCH ligand DLL1 on neighbouring endothelial cell surfaces leads to intracellular NOTCH signalling. Therefore, our data indicates that VEGF signalling upregulates both *Notch4* and *Notch1* and maintains a high NOTCH signalling environment, which is consistent with the literature (Liu et al., 2015, Casie Chetty et al., 2017). Analysis of the expression of the NOTCH family ligands revealed that both *Dll1* and *Jagged* were not differentially expressed between HE and HP cell types cultured with and without VEGF, indicating that VEGF signalling does not impact on the expression of these

ligands and only acts to regulate their receptors. The *Dlk1* repressive NOTCH ligand which is expressed by the ventral sub-aortic mesenchyme (Mirshekar-Syahkal et al., 2013) in the dorsal aorta showed an increase in expression in HE and HP when VEGF was removed. In the presence of VEGF, enhancer elements linked to *Dlk1* by HiC displayed reduced chromatin accessibility and for one enhancer we validated it as being VEGF signalling responsive.

Our data supports previous studies indicating that VEGF signalling establishes and maintains the NOTCH signalling pathway which drives the differentiation of HB to HE and maintains an endothelial cell state (Liu et al., 2015). VEGF should therefore be maintained in early-stage serum and serum free I.V.D cultures to push HB cells to differentiate into HE over smooth muscle and in large quantities.

5.5 VEGF signalling represses the EHT and emergence of HP

During the EHT haemogenic endothelial cells begin to form clusters and finally bud off to form floating progenitors. Understanding the EHT process is therefore essential for the development of large quantities of HP for downstream applications. During the EHT the transcriptomic signature of the cells switches from an endothelial to haematopoietic pattern. Most notably, the expression of *Runx1* in the HE is essential for the EHT and emergence of the HP (Chen et al., 2009, Lancrin et al., 2009).

We observed by FACS that the withdrawal of VEGF resulted in a decrease in HE2 and a

significant increase in the emergence of HP cells. We also observed that in the absence of VEGF, fewer HB cells committed to the blood lineages. Our VEGF withdrawal timepoint study indicated that the trade-off timepoint for the withdrawal was around 12h-14h. Our motif enrichment analysis of distal ATAC sites from HE1 and HE2 revealed that the withdrawal of VEGF resulted in peaks being depleted in endothelial TF motifs such as SOX17, AP-1 and ETS. In contrast, distal ATAC sites from HE2 and HP were enriched for the haematopoietic TF motifs RUNX, PU.1 (SPI1) and NFE2 indicating that these cells had adopted a haematopoietic TF motif signature. Sc-RNA-Seq analysis of HE and HP from cultures where VEGF had been withdrawn revealed a significant increase in *Runx1* expression in HE2 and HP. There was also a significant increase in the expression of RUNX1 target genes such as *Pu.1* (*Spi1*), *Ikaros* and *Meis1*. In HP derived from cultures without VEGF there was also a downregulation of *Sox17* which maintains an endothelial cell state which is consistent with findings from out group (Lichtinger et al., 2012).

To investigate further we studied the ATAC profiles of the *Runx1* locus from HE1, HE2 and HP cultured without VEGF and found that in the presence of VEGF multiple previously well characterised enhancers failed to be activated and form open chromatin, most notably the +23kb enhancer (Bee et al., 2009) and also the +3.7kb enhancer (Zhu et al., 2020). We therefore mutated a number of binding motifs in these enhancer elements and found that the mutation of the TEAD motif in the +23kb enhancer resulted in the de-repression of the enhancer element in cultures supplemented with VEGF.

Recently Fadlullah et al. (Fadlullah et al., 2022) performed a full-length sc-RNA-seq of the

EHT process in mouse embryos focusing on the HE and dorsal aorta niche cells developing into intra-aortic haematopoietic clusters (IAHCs) which HP buds off from. Their data reveals a detailed HE differentiation continuum, which spanned the pre-HE and HE stages. By mining this resource, we can determine the timing and sequence of cytokine and receptor expression by the mesenchyme, endothelial and haematopoietic cells. We can then aim to mimic these conditions *in vitro* to maximise the development of HSCs from ESCs. Mining Fadlullah et al. data (Fadlullah et al., 2022) reveals that *Kdr* (*Flk-1*) expression is highest in the endothelium and pre-EHT HE before reducing in expression dramatically in HE undergoing the EHT and IAHCs. *Vegfa*, which is endogenously expressed by the endothelium, also follows the same expression pattern. Conversely, *Runx1* expression is low in endothelium and pre-EHT HE before being upregulated in HE undergoing the EHT and IAHC. Previously, *in vitro* RUNX1 has been demonstrated to downregulate *Kdr* (*Flk-1*) in endothelial cells undergoing the EHT by binding directly to the *Kdr* promoter (Hirai et al., 2005). NOTCH genes such as *Notch1*, *Dll1*, *Dll4, Jag1* and J*ag2* were also downregulated in HE undergoing the EHT and IAHC compared to the endothelium. These findings support our observations that *Vegfa* and *Kdr* expression is highest when *Runx1* expression is low and this state switches in cells undergoing the EHT and build on previous studies revealing the regulatory role VEGF signalling has on the EHT and *Runx1* expression.

Fadlullah et al. (Fadlullah et al., 2022) also highlighted that a sub-aortic mesenchymal population of smooth muscle and PDGFRa^{+} support haematopoiesis. Our VEGF withdrawal cultures showed an expansion of smooth muscle cells expressing PFGFRa and therefore may have further aided the development of greater number of HP. The surrounding niche

of the dorsal aorta which houses cell types that support the specialised HE and the EHT process must also be closely studied and recapitulated *in vitro*. Our serum free culture system used in this report both involve a FLK-1+ cell sort step, thereby removing mesodermal cells which may give rise to important supporting cells. Other culture systems do not isolate FLK-1+ cells and allow for the development of a number of cells found in the niche (Nafria et al., 2020). Further work needs to be conducted into determining the optimal cell type conditions which support the development of large quantities of reconstituting HP.

Recently published serum free I.V.D protocols induce the differentiation of HB to HE by adding VEGF to the culture system (Nafria et al., 2020). However, these protocols do not reduce the concentration of VEGF, possibly resulting in a decreased yield of HP gained from cultures. It was shown that endogenously expressed VEGF is insufficient to drive HP development in serum free cultures (Bruveris et al., 2021) meaning additional VEGF must be added but our data indicates that VEGF withdrawal increases the overall number of HP. Other protocols embed cytokines such as BMP4 and Activin-A and extra-cellular matrix (ECM) components into a Matrigel layer on top of which EBs differentiate (Zhang et al., 2012). This method simulates the directional expression of certain cytokines from the mesenchyme and provide a cytokine concentration gradient which determines the polarity of endothelial cells and the direction for budding. Future work could study if embedding VEGF into a Matrigel ECM matrix would allow for large numbers of HE development but also provide a concentration gradient allowing for the emergence of HP.

5.6 VEGF signals through AP-1 and TEAD transcription factors

An important aspect of our enhancer assay is the ability of studying individual mutant enhancer elements during cell differentiation to obtain a first idea of the transcription factors regulating its activity at different developmental stages. In this study we therefore employed a previously identified enhancer for *Galnt1* to act as a tracer to report which cytokines signal to which transcription factors in ESCs differentiating to HP. This enhancer sequence was selected because it contained key TF binding motifs for TF expressed in these cell types and our ChIP data had previously confirmed the binding of these factors in HB, HE and HP cells where *Galnt1* is expressed. We found that the ETS motifs were essential for enhancer activity. Interestingly, the mutation of the GATA and E-BOX motifs both resulted in a loss of enhancer activity in HE1 only suggesting that either GATA or E-BOX factors bind and drive the activity in HE1. GATA4 TF acts to drive the differentiation of HB to HE1 (Canete et al., 2017).

When assaying cytokine responsiveness, we found that VEGF withdrawal had the greatest effect on *Galnt1* enhancer activity. VEGF responsiveness is mediated by a composite TEAD-AP-1 binding site (Obier et al., 2016), where the mutation of either TEAD or AP-1 caused a loss of activity in HE1 and HE2. Abrogating the global binding of AP-1 using a dnFOS peptide resulted in a reduction of *Galnt1* expression seen by microarray analysis (Obier et al., 2016). This site also overlaps with a RUNX1 binding site which binds a repressive RUNX1/GFI1 once it is upregulated. Mutation of this site leads to an upregulation of the enhancer in HE in the presence of VEGF. The opposite interplay between TEAD and RUNX1 was seen at the *Runx1*

+23kb enhancer which its activity was strongly upregulated when the TEAD site was mutated, and downregulated once the essential RUNX1 site (Bee et al., 2009) was eliminated.

The 3.7kb enhancer required the RUNX1 binding to drive the enhancer activity in HP, suggesting that RUNX1 may autoregulate itself through this enhancer. Enhancer activity was also lost when the RBPJ/Ikaros motif was mutated in HP. Interestingly, the enhancer contains a SOX motif. SOX2 and SOX11 have been found to act as pioneer transcription factor which are involved in maintaining pluripotency and self-renewal of embryonic stem cells (Dodonova et al., 2020). SOX factors are HMG proteins which have been shown to cooperate with other transcription factors such as OCT4, Nanog and PRX1 and with ATP-consuming chromatin remodellers to increase chromatin accessibility and allow for additional factors to bind (Dodonova et al., 2020, Michael et al., 2020, Remenyi et al., 2003). Mutation of this SOX site may impact on the ability of transcription factors such as TAL1, GATA and RUNX1 to bind to this enhancer alongside SOX.

Our data confirm that a gene regulatory network containing TEAD, AP-1 and RUNX1 determine the developmental activity of endothelial and haematopoietic enhancer elements. However, each enhancer shows a different factor binding architecture of these factors determining whether it is activated or repressed during development. Future studies could focus on using alternative enhancer elements to further study how specific cytokines signal to TFs.

5.7 Defining a signalling responsive gene regulatory network regulating the balance between endothelial and haematopoietic development

Our study provides important mechanistic insights into the core regulatory and signalling network which regulates haematopoietic specification (Figure 5.1a-c). In particular, it links extracellular signalling to the regulation of transcription factor activity acting on specific *cis*-regulatory elements and show that signalling has a profound impact on genomic events.

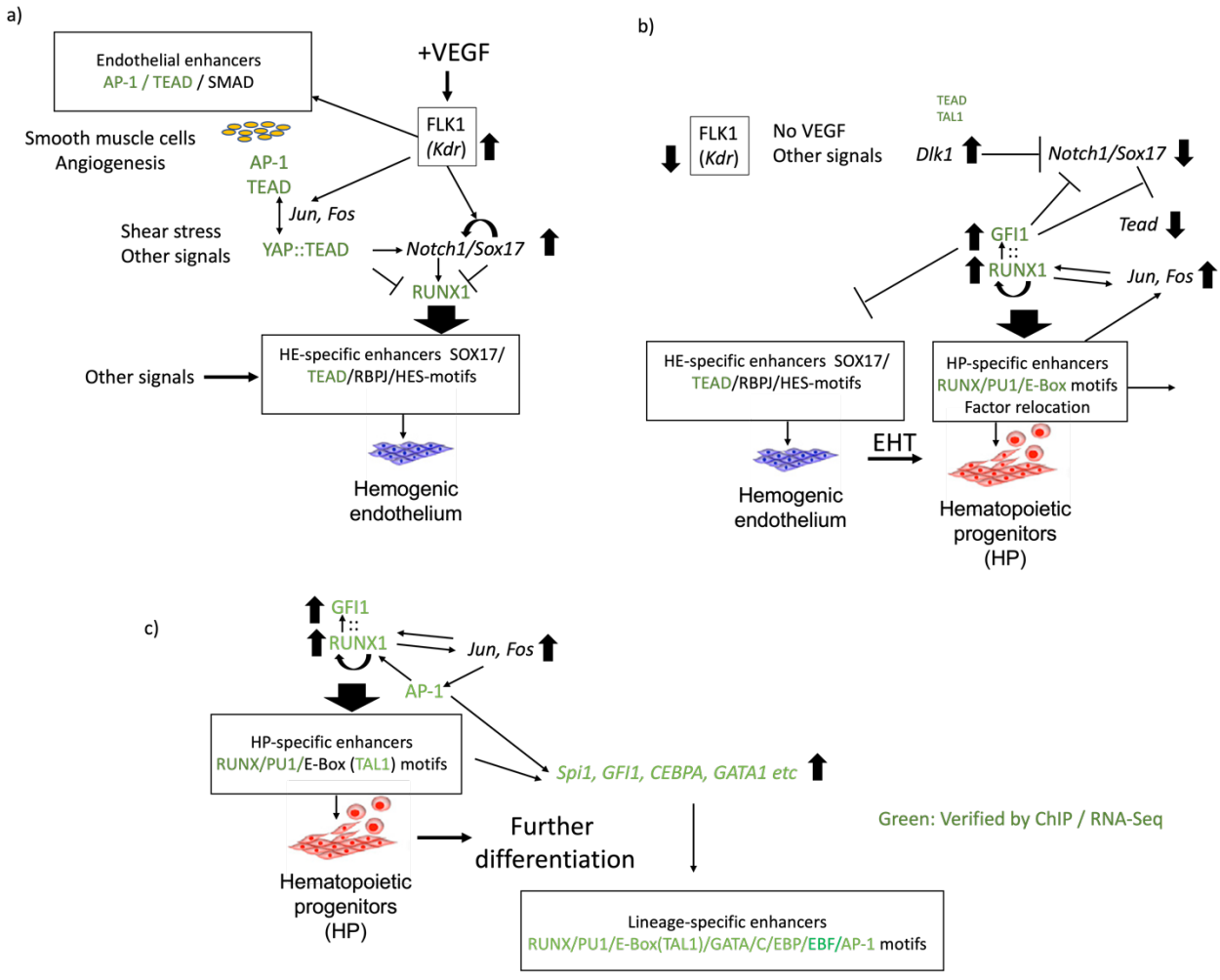


Figure 5.1- VEGF regulates the balance between endothelial and haemogenic development directly and also by regulating NOTCH signalling.
a-c) Core gene regulatory and signalling network regulating blood stem cell emergence. a)

Gene and signalling regulatory network controlling the development of the HE from the HB, b) the gene and signalling regulatory network controlling the EHT and c) the regulatory network maintaining a haematopoietic state and driving further differentiation. Arrows represent positive/activator interaction, repressive activity is shown by a T bar, factors highlighted in green are verified by ChIP and RNA-Seq data from (Goode et al., 2016). Adapted from (Goode et al, 2016).

As previously described in this work, VEGF bindings to its receptor FLK-1 (KDR) inducing a MAPK signalling cascade terminating at the ETS transcription factor ETV2 and at SOX17 (Figure 5.1a). These TFs act to orchestrate differential enhancer and promoter activity which establishes NOTCH signalling through the upregulation of the *Notch1* receptor and its ligands *Dll1 and Dll4*. It regulates the NOTCH1/SOX17 signalling axis which establishes and maintains an endothelial cell signature. VEGF signals to AP-1 TFs (Armesilla et al., 1999, Jia et al., 2016) which regulate the balance between smooth muscle and endothelial cells (Obier et al., 2016). MAPK signalling activated AP-1 also interphases with HIPPO signalling via TEAD TFs at a subset of enhancers in this cell type. Cell to cell interactions induce HIPPO signalling causing YAP/TAZ to translocate into the nucleus stabilising TEAD at their target genes (Lundin et al., 2020). SOX17 directly represses *Runx1* expression (Lizama et al., 2015) while NOTCH signalling upregulates *Gata2* expression (Robert-Moreno et al., 2005, Butko et al., 2015) which in turn starts to upregulate *Runx1* expression.

Upon the upregulation of *Runx1* by GATA2, RUNX1 binds to the *Kdr* (*Flk-1*) promoter directly repressing its activity (Figure 5.1b), thereby inducing a downregulation in the FLK-1 receptor (Hirai et al., 2005). These observations have been confirmed *in vitro* and reflect what is observed from recent *in vivo* single cell studies of the AGM during the EHT of mouse

embryos (Fadlullah et al., 2022). Upon the downregulation of *Kdr* (*Flk-1*) VEGF-MAPK the VEGF-MAPK-NOTCH signalling axis is lost and the repression of *Runx1* through SOX17 and NOTCH signalling effectors is lost as well, leading to a significant upregulation of *Runx1* in HE2. As cells begin to form clusters and bud upwards, endothelial cell to cell contacts which signal through HIPPO signalling (Jansson and Larsson, 2012) to TEAD TFs are lost, leading to the loss of TEAD TF binding at target enhancer sites. TEAD repression of the +23kb enhancer is thus lost, allowing for the upregulation of *Runx1.* RUNX1 autoregulates itself by binding to its own enhancers (Bee et al., 2009) as shown in this study. RUNX1 begins to upregulate its target genes *Pu.1* (*Spi1*), *Nfe2* and along with the repressive TF *Gfi1* which acts to repress endothelial genes such as *Sox17* (Lichtinger et al., 2012) shutting down the endothelial transcriptomic signature in these cells.

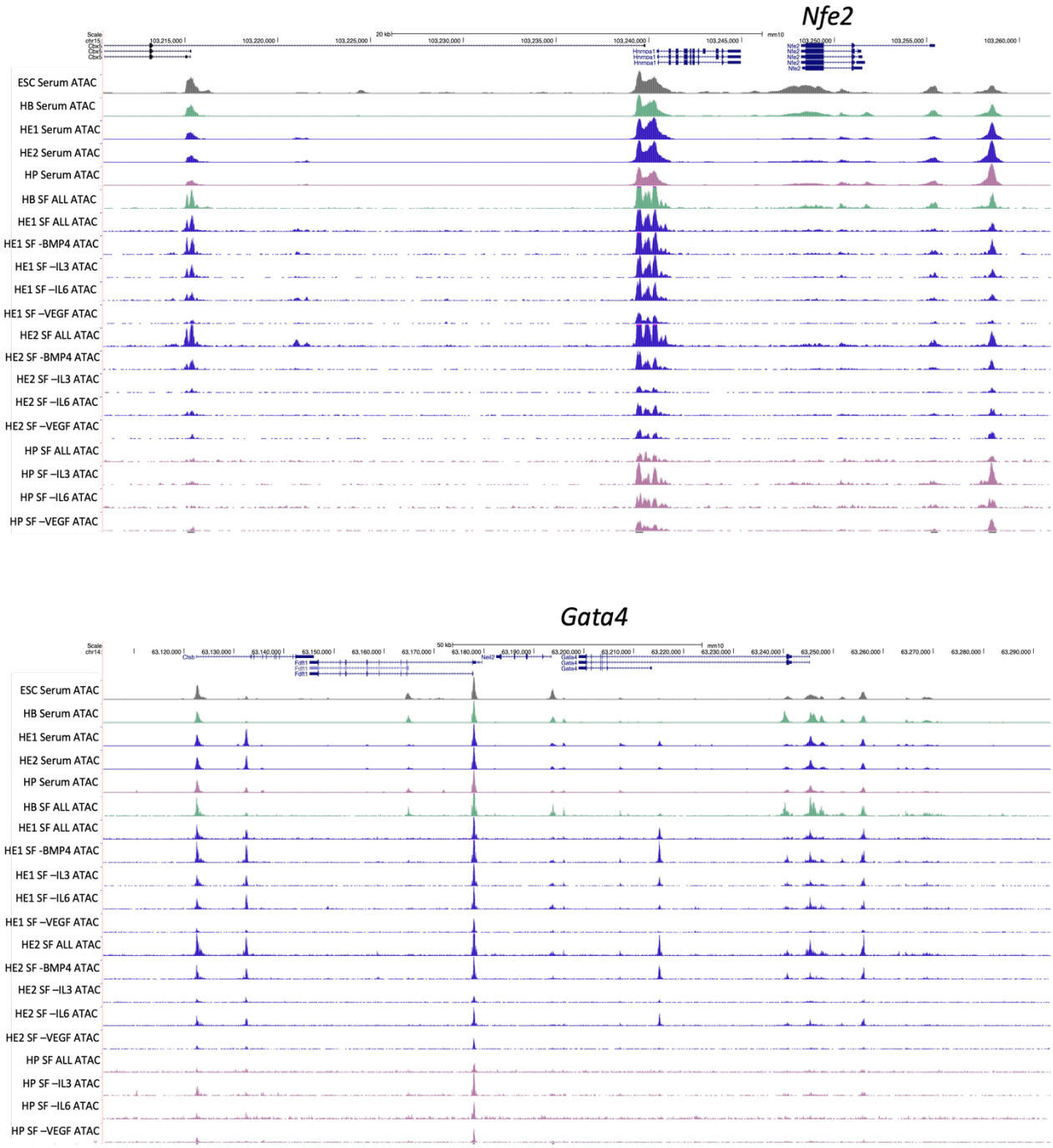
Finally, RUNX1 relocates haematopoietic transcription factors such as TAL1/SCL and FLI-1 to haematopoietic genes (Lichtinger et al., 2012)(Figure 5.1c). RUNX1 continues to upregulate haematopoietic transcription factors such as *Cebpa*, *Gfi1* and *Pu.1* (*Spi1*) which along with RUNX1 drive the commitment of HP through the HSC hierarchy while also repressing endothelial genes.

Taken together, our enhancer identification, our single-cell gene expression and reporter gene studies combined with our previously published ChIP data show that the interplay between ETS (ETV2), SOX17, RUNX1, VEGF-MAPK (AP-1), NOTCH (RBPJ, HES) and Hippo-signalling (TEAD) effectors acting at specific enhancer elements is at the heart of the decision between endothelial or haematopoietic cell fate commitment. Although most of

the regulatory network constituents and their role are known from many studies including perturbation and knock-out experiments (Ottersbach, 2019), it was unclear how and where signals and transcription factors act at specific enhancer elements in the genome and how they are interconnected within cell stage specific gene regulatory networks. This study has allowed us to address these issues.

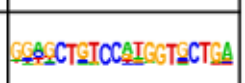
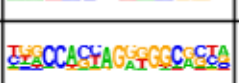
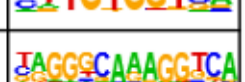
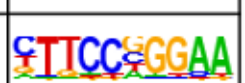
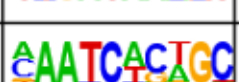
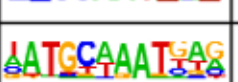
In conclusion, we developed a high throughput enhancer screen which identified thousands of differentially active enhancers at five stages of development from ESCs to HP and characterised the TF motif landscape of these enhancers at each developmental stage. We generated lists of signalling responsive enhancers for four cytokines; BMP4, IL6, IL3 and VEGF and validated a selection of enhancers as being VEGF signalling responsive. We have contributed important findings relating to the role of VEGF signalling in establishing the HE and its repressive activity on the EHT. We identified that key enhancers for the *Runx1* gene were responsive to VEGF signalling. By studying the *Galnt1* enhancer, we determined which factors were mediating this responsiveness by mutating individual TF binding motifs. As numerous other studies use serum free systems and differentiate ESCs into various cell types, we have highlighted that the methodology presented here could easily be adapted and applied in these contexts. As such we believe this work should be of interest to investigators in the field and more broadly (Edginton-White et al., 2023).

6 SUPPLEMENTARY DATA

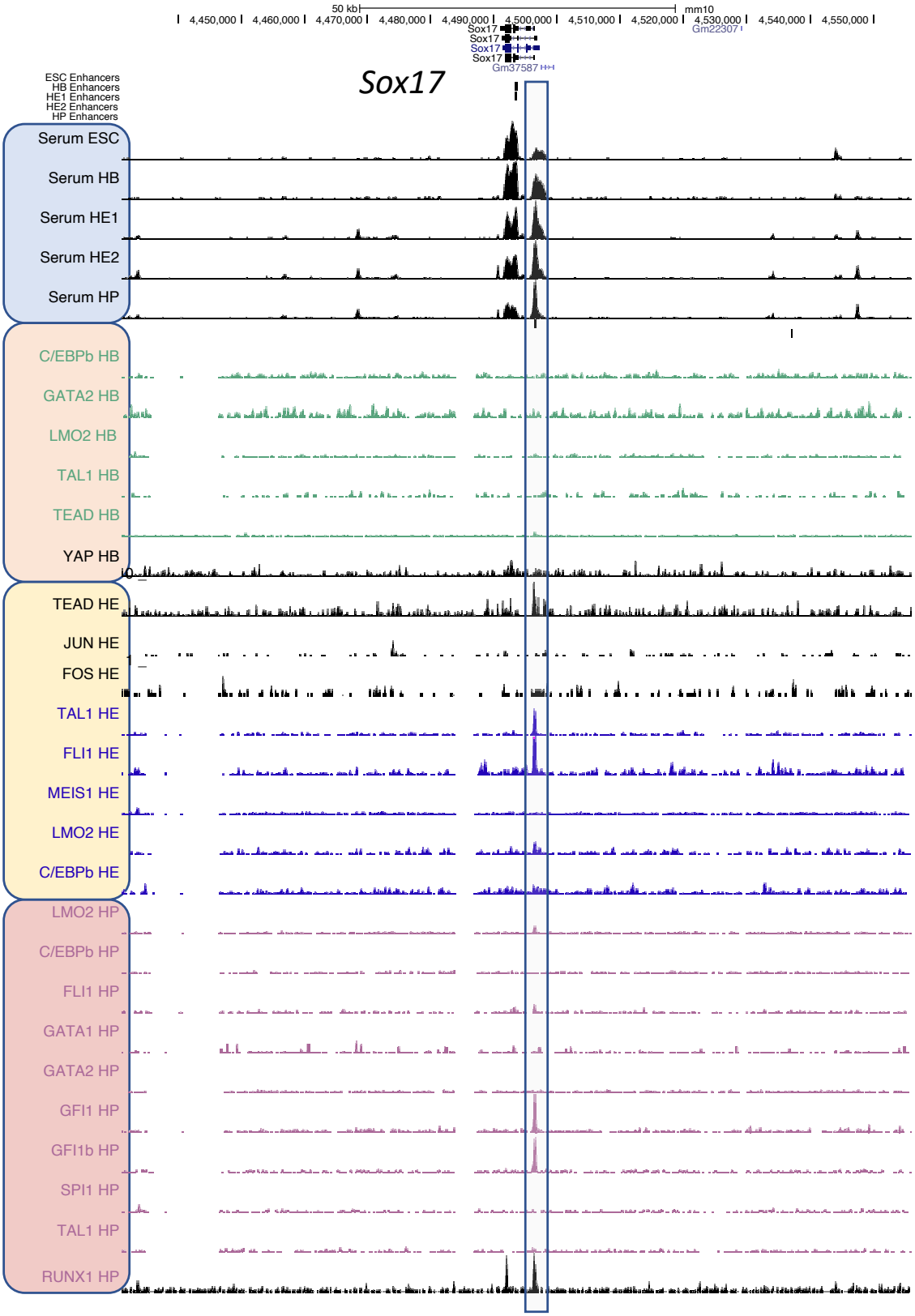


Supplementary Figure 6.1- UCSC genome browser screenshots showing the *Nfe2* and *Gata4* gene locus and the serum and serum free ATAC-Seq data used in this study.
ATAC tracks from ESCs, HB, HE1, HE2 and HP from serum I.V.D (Serum) shown. ATAC tracks from HB, HE1, HE2 and HP from serum free (SF) I.V.D shown with All cytokines (All), withdrawal of BMP4 (-BMP4), withdrawal of IL6 (-IL6), withdrawal of IL3 (-IL3) and

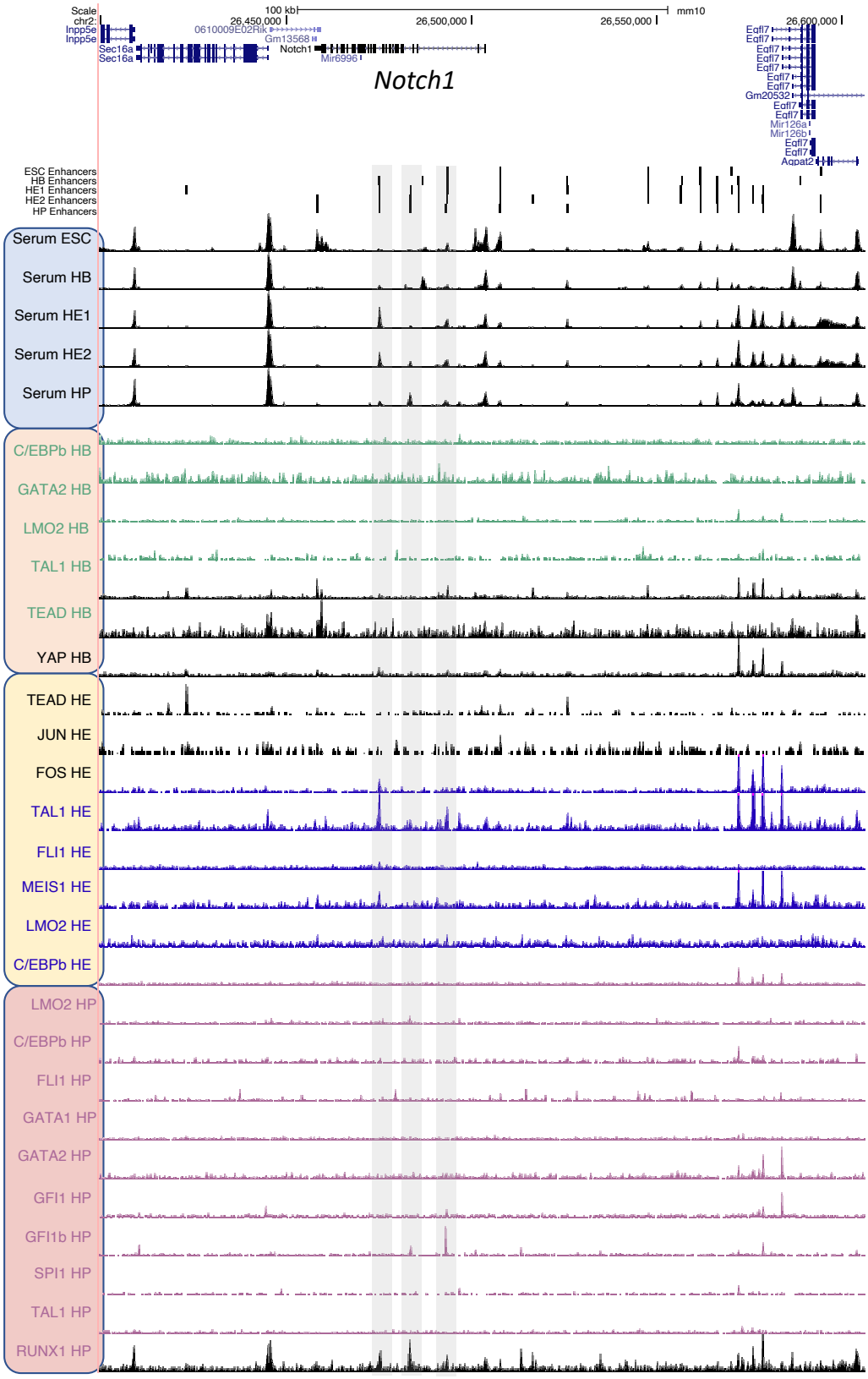
withdrawal of VEGF (-VEGF). The figure demonstrates the high quality of the ATAC data used in this study.

motif	logo	motif	logo	motif	logo
AP-1		ISRE		RARG	
CEBP		KLF		RBPJ	
CREB.ATF		LEF1		REST	
CTCF		MEIS		RUNX	
EGR		MYC.EBOX		RXR	
ERG		NANOG		TCF3	
ESRRA		NF1		SMAD	
ETS		NFAT		SOX	
Fli1		NFE2		SOX17	
FOX		NFκB		Sp1	
GATA		NFY		STAT3	
GFI1b		OCT		STAT5	
HES1		PBX		STAT6	
HOX		PIT1		SCL.TAL1.EBOX	
HSF1		PU.1		TEAD	

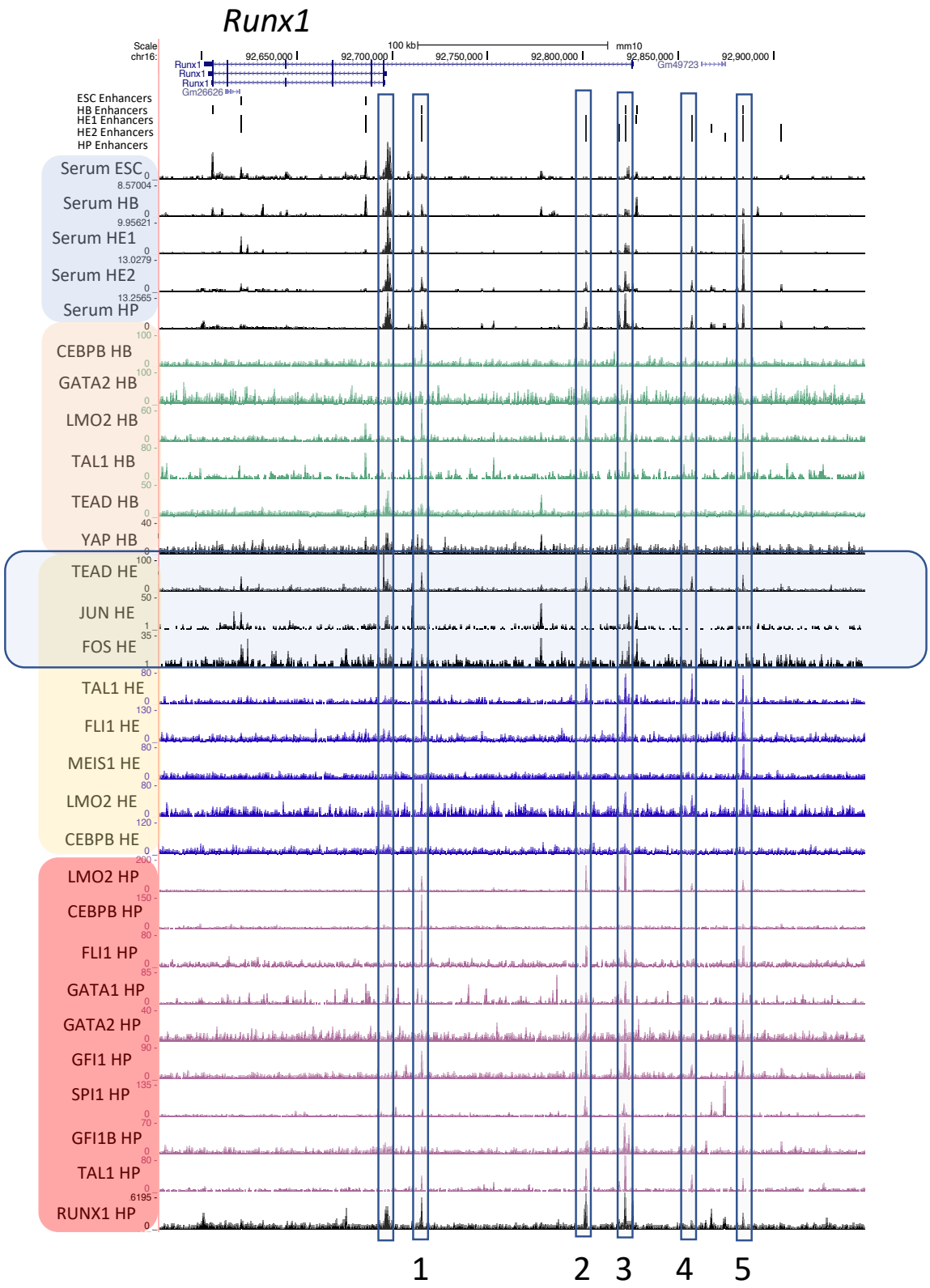
Supplementary Table 6.1- Table of the logos showing position weight matrices of motifs used for motif enrichment and motif colocalization analysis in this study.



Supplementary Figure 6.2- UCSC browser screen shot of the *Sox17* locus.
Enhancer activity, chromatin structure and developmentally regulated transcription factor binding at the *Sox17* locus (Goode et al., 2016). RUNX1, GFI1, GFI1b TAL1 and FLI1 binding highlighted by blue box.



Supplementary Figure 6.3- UCSC browser screen shot of the *Notch1* locus.
Enhancer activity, chromatin structure and developmentally regulated transcription factor binding at the *Notch1* locus (Goode et al., 2016). *Notch1* enhancers highlighted in grey.



Supplementary Figure 6.4- UCSC browser screen shot of the *Runx1* locus.
Enhancer activity, chromatin structure and developmentally regulated transcription factor binding at the *Runx1* locus (Goode et al., 2016). *Runx1* enhancers highlighted by blue box. TEAD, JUN and FOS TF tracks in HE highlighted by blue box.

Supplementary Table 6.2- List of differentially expressed genes from sc-RNA-Seq HE1 clusters.

List of differentially expressed genes from All cytokines (+VEGF) HE1 (pct.1) cluster vs withdrawal VEGF (-VEGF) HE1 cluster (pct.2).

Gene	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
Cotl1	5.10E-41	1.30587741	0.887	0.232	1.03E-36
Ccnd1	8.55E-41	1.64092418	0.879	0.254	1.73E-36
mt-Cytb	1.17E-40	0.70769356	1	0.986	2.37E-36
Tmsb10	1.79E-38	0.76031357	1	0.986	3.63E-34
Lgals1	2.62E-37	1.26710198	0.99	0.819	5.30E-33
mt-Atp6	7.48E-35	0.61006899	1	0.986	1.51E-30
Actb	7.24E-34	0.74151034	1	1	1.46E-29
Rpl41	3.13E-31	0.51804302	1	0.986	6.34E-27
Actg1	8.53E-31	0.68521418	1	0.986	1.73E-26
Nes	1.77E-27	1.02169175	0.801	0.246	3.57E-23
Tmsb4x	7.77E-27	0.61472735	1	1	1.57E-22
Ppia	1.64E-26	0.39848324	1	1	3.32E-22
Arhgap18	5.80E-26	1.05506099	0.738	0.203	1.17E-21
Tspan13	1.03E-25	0.92895643	0.619	0.087	2.08E-21
mt-Co2	4.01E-25	0.48095274	0.999	0.986	8.12E-21
Dlk1	2.26E-24	-0.9481755	0.086	0.362	4.57E-20
Pdlim1	3.68E-24	-1.2287008	0.442	0.703	7.44E-20
Ednrb	1.18E-22	-0.7353133	0.038	0.254	2.40E-18
Adam19	3.68E-22	0.67177128	0.653	0.13	7.45E-18
Tagln	3.78E-22	1.56032892	0.871	0.558	7.66E-18
Tpt1	1.10E-21	-0.4249589	1	0.993	2.22E-17
Hmgn5	1.41E-21	0.80842583	0.668	0.167	2.86E-17
Cryab	3.91E-21	-1.4172047	0.382	0.659	7.92E-17
Gng2	9.27E-21	0.7265852	0.718	0.217	1.88E-16
Malat1	1.25E-20	-0.9427602	1	0.993	2.53E-16
Col1a1	1.84E-20	-0.9361344	0.103	0.362	3.72E-16
Gm10076	2.76E-20	1.12768013	0.966	0.717	5.58E-16
H3f3b	4.23E-20	-0.6017638	1	0.993	8.56E-16
Tpm4	6.65E-19	0.75314647	0.986	0.739	1.35E-14
Myl6	8.96E-19	0.53500001	1	0.964	1.81E-14
Casp3	1.74E-18	0.57369087	0.789	0.297	3.53E-14
Pfn1	2.50E-18	0.45977943	1	0.986	5.05E-14
Gnb4	4.79E-18	0.60930182	0.674	0.196	9.69E-14
Txn1	5.43E-18	0.52479174	0.998	0.906	1.10E-13

Stmn1	1.96E-17	0.71746502	0.974	0.71	3.96E-13
Ethe1	2.25E-17	0.67720095	0.574	0.138	4.55E-13
Esm1	2.67E-17	1.05771387	0.381	0.007	5.40E-13
Map1b	5.39E-17	0.68734109	0.841	0.399	1.09E-12
Fosb	7.87E-17	-0.3456173	0.005	0.109	1.59E-12
Maged2	1.19E-16	0.6797892	0.933	0.536	2.41E-12
Ptpn18	1.52E-16	0.57606479	0.797	0.319	3.07E-12
Csrp1	2.96E-16	0.72075049	0.827	0.384	6.00E-12
Lcp1	3.05E-16	0.61054496	0.492	0.087	6.17E-12
Meg3	4.02E-16	-1.462907	0.47	0.688	8.14E-12
Tagln2	8.12E-16	0.58119088	0.995	0.833	1.64E-11
Rps2	9.14E-16	0.39633566	1	0.986	1.85E-11
Tubb2b	2.11E-15	0.5266713	0.812	0.355	4.27E-11
Ndufa9	2.35E-15	0.5031447	0.695	0.239	4.76E-11
Fabp5	9.81E-15	0.49461932	0.722	0.304	1.99E-10
Fdps	1.09E-14	0.50503125	0.711	0.254	2.22E-10
Mrps6	1.11E-14	0.51812784	0.767	0.319	2.24E-10
Arhgdib	1.35E-14	0.57466466	0.828	0.399	2.74E-10
Crmp1	1.78E-14	0.46138167	0.564	0.145	3.60E-10
Hnrnpa1	2.62E-14	0.55456419	0.973	0.688	5.29E-10
Gapdh	2.68E-14	0.45825289	0.996	0.957	5.42E-10
Rps10	2.91E-14	-0.6833921	0.94	0.935	5.90E-10
Ddx5	3.56E-14	-0.5514004	0.992	0.971	7.20E-10
Dusp6	5.42E-14	0.41861464	0.446	0.08	1.10E-09
Tmem176b	8.98E-14	-1.0282246	0.515	0.659	1.82E-09
Slc25a4	8.99E-14	0.39791965	1	0.949	1.82E-09
Mecom	9.70E-14	0.37731074	0.478	0.094	1.96E-09
Fos	1.10E-13	-1.1469672	0.075	0.261	2.24E-09
Selenoh	1.26E-13	0.63575717	0.823	0.435	2.56E-09
Tnfaip2	1.71E-13	0.47996517	0.541	0.152	3.46E-09
Ccl2	2.39E-13	-1.3740518	0.103	0.312	4.84E-09
Ccdc88a	3.35E-13	0.53452739	0.822	0.391	6.78E-09
H2az1	3.62E-13	0.64382212	0.979	0.71	7.32E-09
Klf4	5.09E-13	-0.8427888	0.064	0.239	1.03E-08
Fth1	7.34E-13	0.51865656	1	0.993	1.49E-08
Ran	8.26E-13	0.47877386	0.99	0.833	1.67E-08
Atp5j2	8.30E-13	0.47807422	0.992	0.826	1.68E-08
Nt5dc2	1.63E-12	0.49526211	0.694	0.283	3.29E-08
Ubqln2	1.75E-12	0.46247563	0.542	0.174	3.54E-08
Hmgb3	2.00E-12	0.48936011	0.806	0.37	4.05E-08

Gmfg	2.04E-12	0.57714887	0.825	0.413	4.13E-08
Pttg1	2.05E-12	0.77806788	0.37	0.058	4.16E-08
Exoc3l4	2.30E-12	0.41659169	0.413	0.08	4.66E-08
Rps29	2.32E-12	0.27757595	1	1	4.70E-08
Lrrfip1	2.46E-12	0.46238571	0.74	0.326	4.98E-08
Tnfaip8l1	2.55E-12	0.42325959	0.453	0.101	5.15E-08
Bin1	3.58E-12	0.38484963	0.618	0.225	7.24E-08
Tpm1	3.65E-12	0.83296241	0.973	0.899	7.40E-08
Ybx1	3.83E-12	0.38798863	0.998	0.957	7.76E-08
Col1a2	4.58E-12	-0.5188857	0.069	0.239	9.27E-08
Cnn2	5.12E-12	0.51840048	0.885	0.522	1.04E-07
Gng11	6.86E-12	-0.688571	0.987	0.971	1.39E-07
Pdia6	8.05E-12	0.4924314	0.945	0.623	1.63E-07
Nme1	8.76E-12	0.52910566	0.985	0.797	1.77E-07
Efna1	1.12E-11	0.4667668	0.415	0.094	2.26E-07
Arpc3	1.14E-11	0.39408949	0.991	0.891	2.31E-07
Lmo2	1.16E-11	0.48093535	0.65	0.254	2.34E-07
Sec61g	1.20E-11	0.37235482	0.999	0.957	2.42E-07
Acta2	1.43E-11	-0.4228696	0.494	0.717	2.90E-07
Ucn	1.65E-11	-0.4955887	0.023	0.138	3.34E-07
Gchfr	1.78E-11	0.44009935	0.496	0.152	3.59E-07
Pfdn1	1.82E-11	0.49124615	0.938	0.616	3.68E-07
Pmp22	2.06E-11	-0.8878283	0.27	0.464	4.17E-07
Prdx4	2.16E-11	0.44702087	0.824	0.428	4.36E-07
Cisd1	2.81E-11	0.39908866	0.674	0.254	5.69E-07
Igfbp4	2.89E-11	0.49121607	0.917	0.522	5.84E-07
Bex2	2.94E-11	0.41062335	0.608	0.225	5.95E-07
Atp5k	3.30E-11	0.39090856	0.985	0.79	6.68E-07
Pkm	3.98E-11	0.43330365	0.907	0.543	8.05E-07
Rell1	4.56E-11	0.37506138	0.493	0.145	9.23E-07
Smarcd1	5.29E-11	0.34180527	0.718	0.312	1.07E-06
Taldo1	6.19E-11	0.42664692	0.898	0.493	1.25E-06
Egfl7	6.20E-11	-0.6360057	0.963	0.92	1.25E-06
Rspo3	6.56E-11	-0.8269676	0.638	0.703	1.33E-06
Bloc1s2	6.59E-11	0.3929228	0.682	0.283	1.33E-06
Ctnnbip1	7.00E-11	0.40648275	0.738	0.333	1.42E-06
Srsf3	7.58E-11	0.4633622	0.923	0.594	1.53E-06
Myl9	8.78E-11	0.72739817	0.441	0.123	1.78E-06
Actr3	9.33E-11	0.41891738	0.898	0.507	1.89E-06
Ehd4	1.02E-10	0.47882635	0.798	0.413	2.07E-06

Tusc3	1.11E-10	0.30438739	0.508	0.174	2.24E-06
Apln	1.22E-10	0.47628448	0.482	0.152	2.48E-06
Acot7	1.26E-10	0.37530346	0.409	0.101	2.56E-06
Hmcn1	1.47E-10	0.31240259	0.608	0.225	2.98E-06
Lcp2	1.70E-10	0.39670117	0.253	0.007	3.45E-06
Calm3	1.72E-10	0.4053932	0.785	0.355	3.48E-06
Rpl35	1.74E-10	0.37077712	1	0.971	3.53E-06
Mrps36	1.77E-10	0.30402974	0.555	0.188	3.59E-06
mt-Nd4l	1.82E-10	0.49082345	0.975	0.768	3.68E-06
Dstn	1.82E-10	0.51981678	0.984	0.884	3.69E-06
Gm1673	1.84E-10	0.35813722	0.504	0.167	3.73E-06
Snrpc	1.98E-10	0.35272547	0.815	0.391	4.00E-06
Ifitm2	1.98E-10	-0.5831514	0.951	0.884	4.01E-06
Hmgn3	2.07E-10	0.31236214	0.561	0.203	4.18E-06
Ccnb2	2.10E-10	0.5482585	0.354	0.072	4.25E-06
Tubb6	2.20E-10	0.47210403	0.634	0.254	4.46E-06
Cdc42ep3	2.69E-10	0.47341567	0.334	0.065	5.45E-06
Ebpl	3.18E-10	0.2911422	0.575	0.203	6.44E-06
Pcdh12	3.47E-10	0.32401897	0.435	0.116	7.02E-06
Hmox1	4.14E-10	0.34899302	0.404	0.109	8.39E-06
Snrpg	4.15E-10	0.35997259	0.992	0.899	8.40E-06
Septin1	4.15E-10	0.34699082	0.287	0.029	8.40E-06
Eef1a1	4.17E-10	-0.253394	1	1	8.43E-06
Cp	4.67E-10	-0.2724038	0.005	0.072	9.45E-06
Slc5a3	4.76E-10	0.33924641	0.28	0.029	9.63E-06
Reep5	5.87E-10	0.34469004	0.911	0.551	1.19E-05
Sh3bp5	6.09E-10	0.48321361	0.875	0.529	1.23E-05
Ldb2	7.20E-10	0.31702792	0.493	0.152	1.46E-05
Rcbtb2	7.21E-10	0.33733185	0.523	0.181	1.46E-05
Anxa5	7.62E-10	0.38019684	0.965	0.71	1.54E-05
Twf2	7.71E-10	0.32260415	0.527	0.181	1.56E-05
Id3	8.20E-10	0.49083445	0.887	0.536	1.66E-05
Gfer	9.48E-10	0.30869197	0.486	0.159	1.92E-05
Plac1	1.22E-09	0.28319339	0.305	0.043	2.46E-05
Cox5a	1.22E-09	0.3647214	0.989	0.833	2.47E-05
Gpx8	1.34E-09	0.382702	0.502	0.181	2.72E-05
Txnl4a	1.38E-09	0.36555316	0.796	0.377	2.79E-05
Postn	1.40E-09	-0.3783529	0.023	0.123	2.84E-05
Esam	1.41E-09	0.4250577	0.862	0.529	2.85E-05
Pdlim5	1.43E-09	0.40048066	0.65	0.268	2.89E-05

Prkar2b	1.60E-09	0.36042052	0.282	0.036	3.23E-05
Cd93	1.65E-09	0.41069141	0.857	0.464	3.33E-05
Eif1	1.92E-09	0.28080593	1	0.978	3.89E-05
Hmgb2	1.94E-09	0.50474052	0.914	0.601	3.92E-05
Lims1	2.25E-09	0.31007579	0.545	0.196	4.55E-05
Selenop	2.48E-09	0.4503064	0.518	0.196	5.03E-05
Bzw2	2.58E-09	0.32570891	0.772	0.391	5.23E-05
Mtpn	2.60E-09	0.31018034	0.683	0.283	5.27E-05
Plp2	2.69E-09	0.35295274	0.664	0.297	5.44E-05
H3f3a	2.81E-09	-0.3899447	0.996	0.964	5.68E-05
Ap2s1	3.26E-09	0.40179645	0.921	0.587	6.60E-05
Birc5	3.28E-09	0.69242966	0.428	0.145	6.64E-05
Tmem101	3.46E-09	0.28241635	0.4	0.116	6.99E-05
Mgmt	3.80E-09	0.32998049	0.572	0.239	7.70E-05
Gtf2h5	3.89E-09	0.29591546	0.818	0.399	7.88E-05
mt-Nd1	4.14E-09	0.34000571	0.999	0.964	8.38E-05
Fam89a	4.16E-09	0.28833389	0.301	0.051	8.42E-05
Ptbp1	4.38E-09	0.31657354	0.597	0.254	8.87E-05
S100a16	4.83E-09	0.30595316	0.474	0.159	9.79E-05
Hmgn2	4.87E-09	0.35593994	0.58	0.246	9.87E-05
Klf7	4.99E-09	0.37271764	0.623	0.283	0.00010103
Uqcc2	5.09E-09	0.2602307	0.863	0.428	0.0001031
Dda1	5.22E-09	0.37791175	0.7	0.341	0.00010574
Cks1b	5.39E-09	0.37747579	0.674	0.312	0.00010901
Cmc2	5.91E-09	0.29777058	0.392	0.109	0.00011957
Cfl1	5.96E-09	0.3272426	0.998	0.92	0.0001206
Snrpe	6.18E-09	0.29510034	0.987	0.819	0.000125
Lsm4	6.72E-09	0.40425898	0.866	0.514	0.0001361
Rpn2	7.33E-09	0.31476116	0.622	0.254	0.00014842
Rasgrp2	8.59E-09	0.25662424	0.458	0.159	0.00017382
Acat1	8.62E-09	0.30773814	0.574	0.232	0.00017456
Eno1	8.65E-09	0.31580056	0.644	0.283	0.00017502
Rpl23a	8.70E-09	0.28484802	1	0.986	0.0001762
Morrbid	9.61E-09	0.31380074	0.236	0.022	0.00019461
Thyn1	9.66E-09	0.27719868	0.52	0.203	0.00019551
Serpina3n	9.82E-09	-0.263448	0.001	0.043	0.00019878
Tmem204	1.01E-08	0.31265246	0.392	0.116	0.00020374
Coq7	1.07E-08	0.26675985	0.395	0.116	0.0002159
Mtarc2	1.15E-08	0.39833386	0.779	0.42	0.00023213
Luc7l3	1.19E-08	0.36978066	0.751	0.37	0.00024034

Cks2	1.22E-08	0.32013281	0.548	0.232	0.00024672
Banf1	1.38E-08	0.2795984	0.856	0.442	0.00027868
Ifnar1	1.44E-08	0.31845565	0.541	0.225	0.00029214
Tmeff1	1.45E-08	0.32513707	0.276	0.051	0.00029381
Slc39a4	1.48E-08	0.29850836	0.236	0.022	0.00029908
Clec1a	1.57E-08	0.3669088	0.583	0.246	0.000317
Atp5b	1.58E-08	0.3202777	0.995	0.891	0.0003196
Akap12	1.59E-08	0.35993569	0.888	0.507	0.00032192
Dbi	1.65E-08	0.37006279	0.963	0.71	0.00033426
Glrx3	1.77E-08	0.36639475	0.961	0.696	0.00035885
Rps27l	1.86E-08	0.29775007	0.999	0.935	0.00037553
Rasgrp3	1.88E-08	0.25805441	0.664	0.283	0.00038035
Ptp4a3	1.91E-08	0.50996831	0.439	0.159	0.0003876
Cldn5	2.01E-08	0.32642335	0.917	0.609	0.00040631
Rsu1	2.11E-08	0.37748414	0.838	0.478	0.0004278
Ei24	2.22E-08	0.27563475	0.567	0.232	0.00044978
Col5a1	2.28E-08	-0.3110426	0.021	0.109	0.00046063
Cenpw	2.30E-08	0.29438976	0.376	0.109	0.00046629
Gnpda2	2.35E-08	0.27101613	0.428	0.152	0.00047486
Junb	2.36E-08	-1.0095451	0.543	0.601	0.00047701
Upp1	2.38E-08	0.26956365	0.643	0.297	0.00048128
Scarb1	2.40E-08	0.28722492	0.252	0.036	0.00048515
Sorbs2	2.46E-08	0.30740375	0.304	0.065	0.00049818
Ppp1ca	2.50E-08	0.37141137	0.97	0.732	0.00050697
Rnpep	2.63E-08	0.27358927	0.444	0.159	0.00053186
Prnp	2.80E-08	-0.827609	0.651	0.638	0.00056732
Lmnb1	2.82E-08	0.37385027	0.552	0.232	0.00056992
Flna	2.93E-08	0.37511627	0.857	0.507	0.00059331
Map4k3	2.96E-08	0.30297863	0.303	0.065	0.00059922
mt-Nd4	3.69E-08	0.27815222	0.999	0.971	0.00074786
Rhoc	3.71E-08	0.40203318	0.978	0.812	0.00075135
Sh3bgrl3	3.75E-08	0.31674266	0.987	0.826	0.00075888
Degs1	3.79E-08	0.29278828	0.6	0.254	0.00076795
Atp5mpl	3.89E-08	0.2998117	0.958	0.725	0.00078743
Mocs2	3.95E-08	0.27437847	0.495	0.188	0.00080009
Mapre1	4.02E-08	0.36292174	0.815	0.413	0.00081475
Cd63	4.08E-08	-0.6498438	0.863	0.775	0.00082638
Traf4	4.27E-08	0.29368025	0.584	0.254	0.00086368
Arhgap29	4.34E-08	0.35457611	0.876	0.471	0.00087797
Tmem256	4.58E-08	0.40262879	0.975	0.696	0.00092811

Flrt2	4.80E-08	-0.7899358	0.334	0.471	0.00097232
Mtdh	4.81E-08	-0.6358162	0.87	0.783	0.00097386
Cwf19l2	5.22E-08	0.28403027	0.418	0.138	0.0010558
Icam2	5.62E-08	0.35665617	0.965	0.746	0.00113665
Cenpa	5.68E-08	0.51830432	0.345	0.109	0.00115011
Ppp1r14b	5.74E-08	0.32022719	0.98	0.87	0.00116249
Gpatch4	6.27E-08	0.2839022	0.505	0.203	0.00126911
Tpx2	6.35E-08	0.50801527	0.368	0.116	0.00128576
Gcsh	8.14E-08	0.27872043	0.762	0.384	0.00164677
Ctsh	8.55E-08	-0.7377391	0.144	0.29	0.00173094
Rap2a	9.35E-08	0.27465795	0.587	0.261	0.00189353
Smdt1	9.56E-08	0.27721382	0.964	0.717	0.00193575
Kdr	9.62E-08	0.38867396	0.902	0.565	0.0019466
Ushbp1	1.05E-07	0.26437375	0.305	0.072	0.00213305
Prelid1	1.11E-07	0.34099623	0.917	0.565	0.00224367
Tuba1a	1.11E-07	0.37670125	0.99	0.877	0.00224997
Manf	1.12E-07	0.3533486	0.875	0.529	0.00227191
Stmn2	1.15E-07	-0.8178435	0.154	0.312	0.00232964
Vasp	1.31E-07	0.29786168	0.693	0.333	0.00264269
Mif	1.40E-07	0.3854217	0.946	0.688	0.00283695
Col18a1	1.42E-07	0.34717233	0.927	0.58	0.00286458
Lpp	1.62E-07	0.38122425	0.718	0.399	0.00328903
Gstk1	1.66E-07	0.294816	0.251	0.043	0.00336623
Mrpl28	1.74E-07	0.26768369	0.674	0.304	0.00352778
Ankrd1	1.75E-07	1.12975185	0.473	0.225	0.00353925
Slc31a2	1.86E-07	0.25134623	0.459	0.181	0.00376032
Mrpl58	1.96E-07	0.36435857	0.85	0.471	0.00397249
Ccng1	2.28E-07	0.2571855	0.779	0.413	0.00462403
Prrg3	2.39E-07	0.26444868	0.269	0.058	0.00483399
Scarf1	2.42E-07	0.25061564	0.412	0.145	0.00490419
Acadl	2.43E-07	0.32592227	0.75	0.391	0.00491589
Eif4ebp1	2.55E-07	0.36630271	0.927	0.681	0.00516096
Nudcd2	3.25E-07	0.32234482	0.909	0.565	0.00657725
Glud1	3.29E-07	0.26842134	0.934	0.616	0.00665935
Tpi1	3.32E-07	0.2844029	0.716	0.362	0.0067133
Arpc2	3.34E-07	0.29628117	0.987	0.826	0.00676354
Igf2bp1	3.42E-07	0.27679939	0.723	0.355	0.00692549
Eif5	3.64E-07	0.27627054	0.961	0.71	0.00737784
Fhl2	3.81E-07	0.25312527	0.306	0.08	0.00770934
Mrps14	3.83E-07	0.26152643	0.745	0.377	0.00774713

Oaf	3.93E-07	0.26464655	0.243	0.043	0.00795042
Specc1	3.94E-07	0.30406941	0.591	0.275	0.00796864
Snrpa	3.96E-07	0.25754088	0.713	0.355	0.00802524
Serpinb6a	4.02E-07	0.28302266	0.89	0.529	0.00814396
Slfn5	4.21E-07	0.27616692	0.34	0.109	0.00851388
Utrn	4.23E-07	0.26356168	0.593	0.283	0.00856242
Nabp2	4.32E-07	0.30392835	0.51	0.217	0.00873531
Bud31	4.41E-07	0.26090251	0.595	0.261	0.00892716
Cox8a	4.47E-07	0.28665696	0.995	0.899	0.00904662
Cox7b	4.53E-07	0.36568278	0.965	0.71	0.00916594
Sec61b	4.73E-07	0.30521191	0.99	0.855	0.00957487
Calr	4.94E-07	0.2934573	0.961	0.746	0.00999594
H2ax	4.95E-07	0.30328127	0.578	0.275	0.01001384
Eif5a	5.04E-07	0.30072706	0.998	0.949	0.01019275
Tmem176a	5.10E-07	-0.7496088	0.455	0.543	0.0103299
Ckap2	5.28E-07	0.4415242	0.428	0.174	0.01068436
Bambi	5.53E-07	0.26826616	0.444	0.174	0.01120173
Mrpl18	5.55E-07	0.3108314	0.782	0.406	0.01123432
Clec1b	5.92E-07	0.33132781	0.212	0.029	0.01197723
Nudt4	5.92E-07	0.2828167	0.689	0.348	0.01199065
Lgals9	6.21E-07	-0.7146055	0.185	0.333	0.01257091
Cox5b	6.53E-07	0.25901172	0.995	0.899	0.0132088
S100a13	6.84E-07	0.25034155	0.846	0.514	0.0138432
Flt1	7.11E-07	0.27125652	0.685	0.348	0.01440208
Ube2c	7.26E-07	0.63174074	0.376	0.145	0.01469539
Tkt	7.40E-07	0.27859624	0.884	0.587	0.01497669
Fscn1	8.15E-07	0.34608005	0.925	0.652	0.01649781
Eif1ax	8.20E-07	0.28739672	0.851	0.478	0.01659967
Hsp90aa1	8.45E-07	0.30740259	0.983	0.848	0.01710832
Prdx1	8.63E-07	0.28106523	0.998	0.899	0.01747294
Rab13	8.83E-07	0.26197311	0.623	0.326	0.01787178
Scd2	9.28E-07	0.27008097	0.667	0.341	0.01879096
Cebpb	9.87E-07	-0.5475855	0.034	0.123	0.01998785
Col15a1	9.99E-07	0.29422276	0.337	0.109	0.02021607
Socs3	1.04E-06	-0.5189493	0.05	0.152	0.02113224
Glo1	1.20E-06	0.27533158	0.841	0.486	0.02429155
Pdgfa	1.25E-06	-0.500231	0.068	0.181	0.02520334
Sub1	1.32E-06	0.29576873	0.985	0.79	0.02664891
Alyref	1.33E-06	0.32933897	0.808	0.457	0.02697196
Gas6	1.38E-06	-0.7671484	0.307	0.428	0.02789846

Ndufs6	1.44E-06	0.27126051	0.881	0.543	0.0292128
Dpysl3	1.46E-06	0.31742166	0.879	0.601	0.02954781
Cdkn2a	1.49E-06	0.30154088	0.368	0.138	0.03008312
Gpc3	1.53E-06	-0.3378262	0.021	0.094	0.03099609
Naa38	1.54E-06	0.31287791	0.872	0.551	0.03120439
Eif4a1	1.54E-06	0.26739416	0.999	0.928	0.03123116
Mrpl42	1.62E-06	0.25828718	0.686	0.355	0.03284961
Klf2	1.64E-06	-1.1064806	0.528	0.609	0.03324476
Serpinh1	1.75E-06	0.27584175	0.971	0.79	0.03543333
Mrps21	1.80E-06	0.27096758	0.841	0.493	0.03640117
Siva1	2.07E-06	0.29022036	0.684	0.362	0.04181257
Zwint	2.21E-06	0.25875394	0.828	0.478	0.04480122
Clic4	2.30E-06	0.31685046	0.691	0.384	0.0464862
Cdc20	2.35E-06	0.40969138	0.301	0.101	0.04757965
Exosc8	2.41E-06	0.26958393	0.51	0.239	0.04874082

Supplementary Table 6.3- List of differentially expressed genes from sc-RNA-Seq HE2 clusters.

List of differentially expressed genes from All cytokines (+VEGF) HE2 (pct.1) cluster vs withdrawal VEGF (-VEGF) HE2 cluster (pct.2).

Gene	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
Rps29	3.63E-67	1.093573268	1	1	7.34E-63
Rps10	1.14E-65	-1.461802743	0.933	0.993	2.31E-61
mt-Atp6	3.76E-63	1.170183567	1	1	7.62E-59
mt-Co1	2.16E-59	1.062200288	0.998	1	4.37E-55
Rpl35	4.24E-59	1.279078367	0.995	0.986	8.59E-55
Gm8797	3.63E-58	-1.422455759	0.562	0.973	7.36E-54
H2ac4	2.22E-57	-0.985566479	0.014	0.438	4.49E-53
H2ac20	7.56E-55	-0.817177659	0.02	0.445	1.53E-50
Rpl36al	1.60E-54	-1.177321041	0.851	0.986	3.24E-50
H2ac10	3.24E-53	-0.53759429	0.005	0.37	6.56E-49
mt-Co2	4.59E-53	0.834152953	1	1	9.29E-49
Rnaset2a	5.05E-53	-0.446491042	0.064	0.61	1.02E-48
Rpl10	5.22E-53	1.015002152	0.992	1	1.06E-48
mt-Co3	6.31E-53	0.901513321	1	1	1.28E-48

H3c3	1.79E-51	-1.863216031	0.041	0.479	3.62E-47
H2bc18	6.13E-49	-0.799284625	0.042	0.486	1.24E-44
H3c6	2.44E-48	-1.171723343	0.127	0.664	4.94E-44
H3c11	2.78E-48	-0.510235284	0.016	0.39	5.62E-44
Ubb	1.05E-47	1.003715735	0.998	1	2.13E-43
H1f1	1.04E-46	-1.496783738	0.063	0.507	2.11E-42
Phgdh	6.12E-46	-1.005915729	0.044	0.466	1.24E-41
Nrgn	5.84E-45	-1.273815729	0.034	0.425	1.18E-40
Was	1.51E-44	-0.412505807	0.03	0.425	3.06E-40
Fcer1g	4.94E-44	-1.088152149	0.022	0.384	1.00E-39
H2bc6	8.86E-44	-0.485899555	0.025	0.397	1.79E-39
Plek	1.05E-41	-1.013094462	0.034	0.411	2.12E-37
H1f3	1.16E-41	-1.013197172	0.099	0.568	2.34E-37
Pla2g4a	1.64E-40	-0.448113288	0.025	0.377	3.33E-36
H3c7	1.20E-39	-0.351342144	0.02	0.356	2.44E-35
Phpt1	1.24E-39	-0.618899049	0.164	0.719	2.50E-35
Fbxo5	5.00E-37	-0.709323609	0.069	0.473	1.01E-32
Lgals1	8.31E-37	1.47336792	0.966	0.74	1.68E-32
Syce2	9.03E-37	-0.435308697	0.038	0.397	1.83E-32
Rps27	1.04E-36	0.637735349	0.998	1	2.11E-32
Pbk	1.11E-36	-0.518555659	0.049	0.425	2.25E-32
Rps19	1.94E-36	0.503219627	1	1	3.93E-32
Rplp1	7.44E-36	0.502327075	0.998	1	1.51E-31
H1f5	1.31E-35	-3.904037619	0.121	0.527	2.65E-31
Ptpn6	7.10E-35	-0.31219667	0.025	0.342	1.44E-30
H2bc14	1.55E-34	-0.282843089	0.008	0.267	3.14E-30
mt-Nd2	1.88E-34	0.787959009	0.997	1	3.80E-30
Plscr1	2.42E-34	-0.393489315	0.058	0.445	4.90E-30
Esco2	2.91E-34	-0.428560139	0.039	0.384	5.89E-30

Dok2	5.43E-34	-0.425686725	0.022	0.322	1.10E-29
Rps28	8.04E-34	0.483351694	0.997	1	1.63E-29
Cenpm	1.66E-33	-0.489369616	0.055	0.418	3.35E-29
Tk1	2.66E-33	-0.834832083	0.085	0.466	5.39E-29
Rrm2	7.92E-33	-0.93316988	0.11	0.527	1.60E-28
H1f4	8.05E-33	-1.840149121	0.479	0.87	1.63E-28
Psmc3ip	8.11E-33	-0.483289253	0.097	0.514	1.64E-28
P2rx1	9.05E-33	-0.256249981	0.005	0.24	1.83E-28
Hspa8	1.45E-32	0.674291798	0.998	1	2.93E-28
Tipin	1.89E-32	-0.820600365	0.246	0.733	3.83E-28
Pold2	2.25E-32	-0.564723935	0.163	0.651	4.55E-28
Rpl38	2.37E-32	0.45890246	1	1	4.79E-28
Ikzf1	3.82E-32	-0.258326912	0.002	0.219	7.73E-28
Tmem108	3.91E-32	-0.522370444	0.042	0.377	7.91E-28
H2ac8	4.00E-32	-1.288073437	0.125	0.534	8.09E-28
Ccna2	4.27E-32	-0.572500651	0.086	0.486	8.65E-28
Tuba1b	6.30E-32	-1.387818838	0.664	0.945	1.28E-27
Rgs18	1.98E-31	-0.280614042	0.006	0.24	4.00E-27
Gmnn	2.63E-31	-0.956867313	0.175	0.616	5.32E-27
Col4a2	4.73E-31	1.075525006	0.991	0.788	9.57E-27
3110082I17Rik	6.85E-31	-0.328979177	0.088	0.507	1.39E-26
Gm10076	8.36E-31	1.576375486	0.931	0.877	1.69E-26
Rack1	8.52E-31	0.423068497	0.997	1	1.73E-26
Eef1a1	1.05E-30	0.442428786	1	1	2.12E-26
H4c4	2.05E-30	-1.65650388	0.471	0.856	4.15E-26
Rps14	2.12E-30	0.401742407	1	1	4.28E-26
Dscc1	6.30E-30	-0.261340564	0.023	0.301	1.28E-25
Dhfr	1.05E-29	-0.457987397	0.083	0.466	2.13E-25
Mcm4	1.46E-29	-0.787477142	0.214	0.664	2.96E-25
Usp1	1.74E-29	-0.636341244	0.141	0.555	3.51E-25

Shcbp1	2.09E-29	-0.51220169	0.077	0.445	4.22E-25
Syngr1	2.52E-29	-0.315577811	0.014	0.26	5.11E-25
H3f3b	2.71E-29	-0.667890838	0.995	1	5.48E-25
Rad51	3.38E-29	-0.367006873	0.063	0.418	6.84E-25
Trip13	3.78E-29	-0.40774001	0.102	0.507	7.64E-25
Asf1b	5.20E-29	-0.546710807	0.092	0.473	1.05E-24
Clspn	5.23E-29	-0.629876511	0.095	0.479	1.06E-24
Ccne1	5.84E-29	-0.428209552	0.034	0.329	1.18E-24
Aurkb	8.47E-29	-0.646166466	0.086	0.452	1.71E-24
Wdr76	8.89E-29	-0.378917243	0.078	0.452	1.80E-24
Prim1	1.03E-28	-0.606395127	0.125	0.527	2.08E-24
Rfc5	1.86E-28	-0.500323032	0.11	0.507	3.78E-24
Peg3	2.29E-28	1.195692488	0.95	0.699	4.63E-24
Cenpw	2.38E-28	-0.664061704	0.202	0.651	4.82E-24
Ckap2l	2.44E-28	-0.46291245	0.066	0.411	4.93E-24
Mcm7	2.67E-28	-0.818969602	0.241	0.671	5.40E-24
Spc24	2.74E-28	-0.711428068	0.105	0.473	5.55E-24
Ptma	5.52E-28	-0.981918041	0.995	1	1.12E-23
Blm	6.23E-28	-0.31423883	0.078	0.452	1.26E-23
Spi1	9.39E-28	-0.30358293	0.019	0.267	1.90E-23
Knl1	1.71E-27	-0.404890604	0.059	0.39	3.47E-23
Brca1	1.94E-27	-0.366372635	0.064	0.397	3.93E-23
Ggct	2.17E-27	-0.270605505	0.045	0.356	4.40E-23
Kif11	2.31E-27	-0.636307053	0.069	0.404	4.68E-23
Sumo2	2.39E-27	-0.709822134	0.948	0.993	4.83E-23
Pf4	2.99E-27	-0.914855942	0.075	0.425	6.06E-23
Rpl37a	4.95E-27	0.407072172	1	1	1.00E-22
Cenpq	7.18E-27	-0.353967718	0.08	0.438	1.45E-22
Eme1	7.84E-27	-0.286205546	0.049	0.356	1.59E-22

Nuf2	1.14E-26	-0.258483231	0.055	0.377	2.31E-22
Hypk	1.20E-26	-0.656994126	0.407	0.856	2.42E-22
Kif4	1.32E-26	-0.380337236	0.058	0.377	2.68E-22
Car2	2.71E-26	-1.818826632	0.199	0.575	5.48E-22
Mcm5	2.88E-26	-0.666976297	0.189	0.616	5.83E-22
Umps	5.69E-26	-0.524856509	0.158	0.575	1.15E-21
Shmt1	6.24E-26	-0.321565636	0.07	0.411	1.26E-21
Gem	7.20E-26	-0.482614409	0.097	0.466	1.46E-21
Nrm	7.58E-26	-0.293084017	0.053	0.363	1.53E-21
Cenph	8.96E-26	-0.465047519	0.103	0.479	1.81E-21
Cldn5	8.99E-26	1.103974631	0.905	0.493	1.82E-21
Col18a1	1.00E-25	0.981403111	0.922	0.788	2.03E-21
Fignl1	1.35E-25	-0.304937429	0.081	0.438	2.73E-21
Pclaf	2.05E-25	-1.550516357	0.167	0.521	4.16E-21
Fen1	3.87E-25	-0.477447522	0.119	0.5	7.83E-21
U2af1	5.53E-25	-0.738521617	0.77	0.973	1.12E-20
Tpt1	7.84E-25	0.35958268	0.997	1	1.59E-20
Nop10	9.13E-25	-0.886282995	0.685	0.945	1.85E-20
Coro1a	1.02E-24	-0.321731854	0.025	0.267	2.06E-20
Ncapd3	1.12E-24	-0.350532884	0.135	0.548	2.26E-20
Mnd1	1.41E-24	-0.351534543	0.069	0.39	2.86E-20
Anp32b	1.46E-24	-0.99454247	0.873	0.979	2.95E-20
Bub1	1.77E-24	-0.362044678	0.063	0.37	3.59E-20
Cdc6	2.19E-24	-0.35722885	0.052	0.342	4.43E-20
Ranbp1	2.25E-24	-0.912468945	0.847	0.973	4.56E-20
Mns1	2.50E-24	-0.25846336	0.074	0.404	5.07E-20
Rpl9-ps6	3.20E-24	-0.53041674	0.329	0.822	6.47E-20
Isoc1	4.16E-24	-0.391339056	0.139	0.555	8.43E-20
Ncapg	4.23E-24	-0.261029093	0.05	0.342	8.57E-20

mt-Nd1	4.64E-24	0.658286519	0.995	1	9.39E-20
Hirip3	4.91E-24	-0.454819945	0.103	0.459	9.94E-20
Tuba1c	7.78E-24	-0.567090372	0.15	0.541	1.58E-19
Neil3	7.83E-24	-0.250842352	0.042	0.315	1.58E-19
Chaf1b	9.00E-24	-0.357796943	0.13	0.534	1.82E-19
Plk1	9.23E-24	-0.598928318	0.078	0.397	1.87E-19
Hapln1	9.44E-24	-0.607941863	0.139	0.521	1.91E-19
Dut	1.20E-23	-1.068147948	0.368	0.733	2.42E-19
Cenps	1.29E-23	-0.315786927	0.088	0.425	2.62E-19
Incenp	1.52E-23	-0.524287265	0.166	0.575	3.07E-19
Dtl	1.86E-23	-0.282540422	0.089	0.438	3.77E-19
Gnai2	2.00E-23	0.651557998	0.997	0.993	4.05E-19
Rpa2	2.15E-23	-0.404608466	0.161	0.582	4.35E-19
Psip1	2.16E-23	-0.779217214	0.588	0.911	4.37E-19
Dctpp1	2.16E-23	-0.822420351	0.366	0.788	4.37E-19
Cenpk	2.23E-23	-0.342722476	0.083	0.411	4.52E-19
Pola1	2.29E-23	-0.423636029	0.114	0.473	4.64E-19
Cbx5	3.99E-23	-0.727809794	0.648	0.938	8.09E-19
Sgo1	4.25E-23	-0.331845559	0.067	0.377	8.60E-19
Anp32e	4.67E-23	-0.696459465	0.326	0.733	9.45E-19
Mki67	6.55E-23	-0.844465032	0.114	0.466	1.33E-18
Med30	7.64E-23	-0.475670494	0.199	0.63	1.55E-18
Orc6	8.23E-23	-0.782854523	0.452	0.822	1.67E-18
Rpa3	8.93E-23	-0.448779581	0.166	0.568	1.81E-18
Rps18	9.71E-23	0.373906979	0.998	1	1.97E-18
Trmt112	1.28E-22	-0.613544915	0.565	0.932	2.59E-18
Ccne2	1.35E-22	-0.299246358	0.034	0.281	2.73E-18
Car8	1.45E-22	-	0.081	0.404	2.93E-18

		0.397684828			
Rbbp7	1.51E-22	-0.686513213	0.678	0.945	3.06E-18
Oaz1	1.78E-22	-0.505267904	0.941	1	3.60E-18
Prc1	1.85E-22	-0.906169782	0.13	0.486	3.75E-18
Cdk1	1.97E-22	-0.847788836	0.247	0.644	3.98E-18
Lig1	2.08E-22	-0.973883964	0.288	0.644	4.21E-18
Cenpf	2.82E-22	-0.676093494	0.121	0.486	5.70E-18
Smc4	3.14E-22	-1.01879286	0.418	0.801	6.36E-18
Rbbp4	3.20E-22	-0.675429367	0.684	0.959	6.49E-18
Pimreg	3.24E-22	-0.381644062	0.105	0.459	6.56E-18
Noc4l	4.01E-22	-0.327775498	0.178	0.61	8.13E-18
Cdh5	4.48E-22	0.801524954	0.989	0.842	9.07E-18
Chaf1a	4.74E-22	-0.330374612	0.117	0.479	9.60E-18
Cdca8	4.96E-22	-0.935688637	0.197	0.562	1.00E-17
Ska1	5.05E-22	-0.260509045	0.061	0.349	1.02E-17
Hist1h2ap	5.21E-22	-0.368715691	0.111	0.459	1.06E-17
Kif15	5.62E-22	-0.359962964	0.077	0.384	1.14E-17
Snrpd3	6.07E-22	-0.771128943	0.778	0.952	1.23E-17
Rps13	6.53E-22	0.417186923	0.997	1	1.32E-17
Vrk1	7.22E-22	-0.479797084	0.25	0.705	1.46E-17
Nusap1	7.70E-22	-0.844095358	0.105	0.438	1.56E-17
Eif5a	7.83E-22	-0.836347859	0.991	1	1.58E-17
Stat4	9.96E-22	-0.310631527	0.033	0.267	2.02E-17
Uqcrb	1.07E-21	-0.501056265	0.931	0.993	2.17E-17
Anp32a	1.15E-21	-0.657615318	0.767	0.966	2.33E-17
Cxxc5	1.32E-21	-0.537011405	0.235	0.664	2.67E-17
Hnrnpd	1.52E-21	-0.825200256	0.676	0.938	3.07E-17
Anln	1.61E-21	-0.319714659	0.088	0.411	3.25E-17

H2az1	1.65E-21	-1.592801132	0.883	0.979	3.33E-17
Agfg2	1.74E-21	-0.278123715	0.078	0.39	3.52E-17
Vars	2.22E-21	-0.426935695	0.246	0.692	4.50E-17
Mcm6	2.47E-21	-0.587587746	0.207	0.589	5.00E-17
Rps3a1	2.71E-21	-0.371818138	0.997	1	5.48E-17
Tubg1	3.00E-21	-0.471674962	0.235	0.651	6.07E-17
Spdl1	3.14E-21	-0.408454709	0.102	0.432	6.37E-17
Nme2	3.58E-21	-0.424060779	0.239	0.705	7.25E-17
Bub1b	3.95E-21	-0.292589261	0.066	0.356	8.00E-17
Hnrnpa2b1	4.22E-21	-0.748467447	0.955	0.993	8.54E-17
Dnmt1	4.37E-21	-0.766893024	0.311	0.699	8.85E-17
Kif22	4.92E-21	-0.355978437	0.086	0.397	9.97E-17
Hmgb1	5.18E-21	-1.046222976	0.936	0.986	1.05E-16
Arhgap11a	5.83E-21	-0.320108503	0.11	0.452	1.18E-16
Lsm3	7.75E-21	-0.635523421	0.383	0.795	1.57E-16
Fabp5	8.90E-21	-0.974164229	0.415	0.808	1.80E-16
Lyl1	9.07E-21	-0.52858427	0.205	0.603	1.84E-16
Myl12b	9.17E-21	-0.528640041	0.36	0.836	1.86E-16
Uhrf1	9.74E-21	-0.391033028	0.106	0.432	1.97E-16
B930036N10Rik	1.00E-20	-0.330576185	0.108	0.452	2.03E-16
Itgb3bp	1.29E-20	-0.270635656	0.152	0.555	2.61E-16
Lsm2	1.46E-20	-0.82829804	0.557	0.863	2.95E-16
Meaf6	2.60E-20	-0.264228157	0.18	0.61	5.27E-16
Atad2	2.98E-20	-0.7720411	0.235	0.596	6.02E-16
Atp5g1	3.25E-20	-0.607357324	0.959	0.993	6.58E-16
Col4a1	3.63E-20	0.83891395	1	0.842	7.34E-16
Rfc4	3.71E-20	-0.658873222	0.233	0.596	7.50E-16
Cdca7	5.13E-20	-0.395306805	0.135	0.479	1.04E-15

Ccnf	5.37E-20	-0.253283639	0.085	0.397	1.09E-15
Tmsb4x	6.10E-20	0.592995528	0.997	0.993	1.23E-15
Pcna	6.12E-20	-0.607365274	0.286	0.692	1.24E-15
Ung	6.37E-20	-0.259259936	0.028	0.24	1.29E-15
Atad5	8.87E-20	-0.295516736	0.121	0.466	1.80E-15
Birc5	9.73E-20	-1.177899539	0.207	0.541	1.97E-15
Tyms	1.00E-19	-0.86354186	0.347	0.699	2.03E-15
Hspd1	1.18E-19	-1.007380697	0.859	0.959	2.38E-15
Srm	1.31E-19	-0.736417101	0.415	0.774	2.65E-15
Uba52	1.69E-19	-0.510660651	0.833	0.986	3.43E-15
mt-Nd3	1.70E-19	0.680715182	0.92	0.959	3.45E-15
Rnaseh2b	1.75E-19	-0.306376357	0.171	0.575	3.54E-15
Nup43	1.85E-19	-0.28575156	0.127	0.479	3.74E-15
Dlgap5	2.20E-19	-0.260297495	0.081	0.384	4.45E-15
Ybx1	2.42E-19	-0.479564867	0.991	1	4.90E-15
Card19	2.61E-19	-0.397080055	0.185	0.582	5.29E-15
Top2a	3.05E-19	-1.574430448	0.233	0.562	6.17E-15
Cbx3	4.40E-19	-0.651173517	0.74	0.938	8.91E-15
Rab27b	4.47E-19	-0.27099964	0.083	0.377	9.05E-15
Sec11c	4.55E-19	-0.487816099	0.299	0.726	9.22E-15
Ncaph	4.66E-19	-0.282596947	0.089	0.39	9.43E-15
Ccnd3	4.77E-19	-0.699497445	0.757	0.952	9.65E-15
H3f3a	5.22E-19	-0.449306894	0.986	1	1.06E-14
Ing5	5.37E-19	-0.267151946	0.15	0.534	1.09E-14
Cycs	7.31E-19	-0.722766398	0.62	0.945	1.48E-14
Mgst3	7.63E-19	-0.622400809	0.343	0.74	1.54E-14
Alox5ap	8.57E-19	-0.994087921	0.315	0.664	1.74E-14
Sigmar1	9.07E-19	-0.307546682	0.192	0.603	1.84E-14

Mmrn2	9.19E-19	0.8830288	0.67	0.315	1.86E-14
Dnajc19	9.42E-19	-0.583485177	0.435	0.808	1.91E-14
Mndal	9.72E-19	-0.324353334	0.003	0.137	1.97E-14
Rrs1	1.24E-18	-0.370351843	0.213	0.623	2.51E-14
Tmem97	1.25E-18	-0.661798348	0.327	0.705	2.54E-14
Lmnb1	1.43E-18	-0.769070847	0.355	0.692	2.89E-14
Mis18bp1	1.47E-18	-0.338108029	0.102	0.411	2.98E-14
Hnrnpf	1.49E-18	-0.561661355	0.878	0.986	3.01E-14
Rpa1	1.51E-18	-0.324148634	0.147	0.507	3.06E-14
Samd1	1.68E-18	-0.255362217	0.13	0.479	3.41E-14
Dek	1.70E-18	-0.900954788	0.601	0.87	3.44E-14
Bloc1s1	1.91E-18	-0.277786692	0.182	0.582	3.87E-14
Mrpl12	1.99E-18	-0.600689344	0.782	0.952	4.03E-14
Ddx39	2.43E-18	-0.698563515	0.419	0.767	4.91E-14
Racgap1	2.46E-18	-0.335301072	0.108	0.411	4.99E-14
Map4k4	2.75E-18	0.754381395	0.898	0.822	5.56E-14
Exosc8	2.86E-18	-0.552487297	0.366	0.767	5.79E-14
Bzw1	2.94E-18	-0.46964413	0.884	0.986	5.95E-14
Cenpa	3.06E-18	-0.578972415	0.146	0.473	6.19E-14
Fam98b	3.63E-18	-0.367555325	0.31	0.76	7.34E-14
Maged2	3.83E-18	0.818210907	0.883	0.774	7.76E-14
Plk4	3.88E-18	-0.309366061	0.102	0.404	7.85E-14
Hpf1	3.97E-18	-0.718551969	0.607	0.877	8.03E-14
Kif23	4.53E-18	-0.280806241	0.089	0.384	9.17E-14
Rps6	4.92E-18	0.366089103	1	1	9.96E-14
Topbp1	5.07E-18	-0.377365585	0.233	0.644	1.03E-13
Hnrnpdl	5.58E-18	-0.550753023	0.524	0.911	1.13E-13
Msh6	5.96E-18	-0.269956668	0.146	0.5	1.21E-13

Tmsb10	7.61E-18	0.551994063	0.997	0.993	1.54E-13
Haspin	8.21E-18	-0.250497989	0.091	0.384	1.66E-13
Gins2	9.79E-18	-0.566126917	0.26	0.637	1.98E-13
Ppil1	1.01E-17	-0.517847711	0.322	0.712	2.05E-13
Nasp	1.08E-17	-0.897660681	0.534	0.801	2.18E-13
Rpl5	1.31E-17	0.403822233	0.997	1	2.66E-13
Chchd6	1.46E-17	-0.328948814	0.164	0.521	2.96E-13
Ptges3	1.74E-17	-0.552910516	0.826	0.986	3.52E-13
Eef1d	2.23E-17	-0.474483407	0.905	1	4.51E-13
Kif20b	2.33E-17	-0.38026775	0.153	0.493	4.72E-13
Nup37	2.36E-17	-0.281906042	0.194	0.582	4.77E-13
Kdr	2.44E-17	0.730628992	0.892	0.733	4.94E-13
Snrpf	3.09E-17	-0.652676697	0.808	0.959	6.27E-13
Tacc3	3.15E-17	-0.401972768	0.191	0.555	6.37E-13
Prmt1	3.43E-17	-0.590504586	0.74	0.925	6.94E-13
Ccnb1	3.72E-17	-0.383334448	0.113	0.411	7.54E-13
Cdca3	3.79E-17	-0.535744657	0.15	0.459	7.66E-13
Cmc2	3.79E-17	-0.533949981	0.277	0.637	7.68E-13
Uba2	4.18E-17	-0.436451612	0.377	0.801	8.46E-13
Tecr	4.35E-17	-0.557467375	0.822	0.959	8.81E-13
Sgf29	4.41E-17	-0.348066559	0.194	0.568	8.93E-13
Ifrd2	4.78E-17	-0.278213458	0.127	0.445	9.67E-13
H2ax	5.24E-17	-0.949543197	0.459	0.774	1.06E-12
Rps26	6.52E-17	-0.344122228	0.997	1	1.32E-12
Gpn1	7.12E-17	-0.297032414	0.2	0.596	1.44E-12
Sass6	7.24E-17	-0.270892884	0.116	0.418	1.46E-12
Ran	7.26E-17	-0.74312462	0.962	0.993	1.47E-12
Hpgd	7.37E-17	-0.447704632	0.081	0.342	1.49E-12

Cmss1	8.17E-17	-0.299416584	0.186	0.555	1.65E-12
Dnajc9	9.47E-17	-0.563823058	0.362	0.753	1.92E-12
Rpl12	1.02E-16	0.422074787	0.994	1	2.06E-12
Snrnp40	1.08E-16	-0.377194661	0.33	0.76	2.19E-12
Hspg2	1.23E-16	0.806228994	0.779	0.589	2.49E-12
Bckdk	1.43E-16	-0.299997714	0.213	0.616	2.90E-12
Hmmr	1.44E-16	-0.53664093	0.127	0.425	2.91E-12
Ost4	1.51E-16	-0.496931896	0.803	0.966	3.05E-12
Ndc80	1.58E-16	-0.327243976	0.124	0.432	3.20E-12
Runx1	1.58E-16	-0.407830292	0.094	0.363	3.20E-12
Hat1	1.62E-16	-0.470664893	0.269	0.637	3.27E-12
Exosc7	1.96E-16	-0.389462036	0.319	0.74	3.96E-12
Ezh2	1.96E-16	-0.555984866	0.355	0.726	3.96E-12
Timm10b	1.96E-16	-0.532687926	0.498	0.856	3.98E-12
Gap43	2.07E-16	0.721247835	0.903	0.589	4.18E-12
H2az2	2.10E-16	-0.693994619	0.795	0.938	4.26E-12
Dkc1	2.16E-16	-0.502304461	0.294	0.664	4.38E-12
Mrpl40	2.31E-16	-0.393744571	0.33	0.726	4.69E-12
Mettl1	2.45E-16	-0.304096293	0.169	0.521	4.96E-12
Dcps	2.46E-16	-0.348003946	0.343	0.795	4.99E-12
Atic	2.77E-16	-0.524791438	0.36	0.733	5.60E-12
Snu13	2.93E-16	-0.548314618	0.767	0.952	5.94E-12
Sparc	3.33E-16	0.584738278	0.998	0.979	6.75E-12
Marcks	3.37E-16	0.547761671	0.975	0.945	6.82E-12
Ndufb8	3.63E-16	-0.484299987	0.828	0.973	7.34E-12
Fkbp3	4.28E-16	-0.586753186	0.865	0.993	8.66E-12
Tcof1	4.30E-16	-0.441022921	0.29	0.671	8.69E-12
Anxa6	4.52E-16	0.694907854	0.873	0.705	9.14E-12
Dnmt3b	4.64E-16	-0.274908946	0.094	0.37	9.39E-12

Smc2	4.87E-16	-0.76812131	0.391	0.705	9.85E-12
Alyref	5.09E-16	-0.574575937	0.696	0.925	1.03E-11
Rhno1	6.54E-16	-0.258761385	0.18	0.555	1.32E-11
Cenpx	7.82E-16	-0.604615901	0.632	0.89	1.58E-11
Rrp15	8.21E-16	-0.411825388	0.335	0.726	1.66E-11
Pmf1	8.67E-16	-0.305446465	0.169	0.514	1.76E-11
Idh3a	9.07E-16	-0.350970909	0.266	0.671	1.84E-11
Mrpl18	9.45E-16	-0.607619473	0.657	0.918	1.91E-11
Ndufb10	1.02E-15	-0.466892659	0.717	0.945	2.06E-11
Aurka	1.15E-15	-0.312372172	0.086	0.342	2.33E-11
Tmem176b	1.17E-15	-0.734409224	0.621	0.904	2.36E-11
Mcm3	1.60E-15	-0.623186544	0.369	0.692	3.24E-11
Pold3	1.72E-15	-0.338632408	0.238	0.616	3.49E-11
Iqgap2	1.76E-15	-0.266173147	0.135	0.445	3.56E-11
Ddx39b	1.87E-15	-0.52360681	0.567	0.884	3.78E-11
Ncaph2	1.88E-15	-0.261688721	0.197	0.555	3.81E-11
Fxn	1.93E-15	-0.305180544	0.218	0.596	3.90E-11
Pnpo	1.95E-15	-0.256187517	0.156	0.493	3.95E-11
Exosc5	2.15E-15	-0.386293252	0.28	0.658	4.35E-11
Timm50	2.20E-15	-0.412157946	0.346	0.76	4.46E-11
Ldha	2.58E-15	-0.692417598	0.82	0.945	5.22E-11
Fn1	2.66E-15	0.919122699	0.851	0.767	5.39E-11
Rpl17	2.80E-15	0.286977397	1	1	5.66E-11
Tpx2	3.00E-15	-0.636086425	0.203	0.521	6.08E-11
Nxt1	3.18E-15	-0.297757154	0.257	0.651	6.44E-11
Tnni1	3.31E-15	-0.301512822	0.014	0.158	6.70E-11
Spc25	3.69E-15	-0.537506323	0.186	0.486	7.46E-11
Galk1	3.72E-15	-0.444062479	0.412	0.767	7.52E-11

Rrm1	3.87E-15	-0.448802736	0.211	0.541	7.83E-11
Ube2c	3.98E-15	-0.902248349	0.199	0.514	8.06E-11
Hells	4.33E-15	-0.510330528	0.26	0.589	8.77E-11
Rpl27a	5.80E-15	0.28135405	0.998	1	1.17E-10
Trim59	6.27E-15	-0.27883035	0.155	0.479	1.27E-10
Snrpb	6.57E-15	-0.447910648	0.887	0.986	1.33E-10
Acsl3	6.88E-15	-0.337919381	0.263	0.664	1.39E-10
Ccdc34	7.13E-15	-0.62688568	0.319	0.637	1.44E-10
Wdr75	7.59E-15	-0.334795695	0.302	0.726	1.54E-10
mt-Nd4	7.65E-15	0.403978254	0.997	1	1.55E-10
Lyrm9	7.67E-15	-0.25906353	0.15	0.473	1.55E-10
Snrpd1	8.12E-15	-0.596254419	0.844	0.973	1.64E-10
Lsm8	8.27E-15	-0.419366091	0.371	0.781	1.67E-10
Tcf4	9.50E-15	0.662675228	0.912	0.788	1.92E-10
Pa2g4	1.06E-14	-0.599818779	0.756	0.945	2.14E-10
Tmpo	1.11E-14	-0.538625128	0.424	0.781	2.24E-10
Icam2	1.14E-14	0.668483046	0.956	0.678	2.32E-10
Idi1	1.17E-14	-0.496944695	0.311	0.671	2.36E-10
Map1b	1.21E-14	0.822261395	0.739	0.466	2.46E-10
Rpl11	1.34E-14	0.274350554	0.998	1	2.71E-10
Ptpre	1.82E-14	-0.356548977	0.136	0.418	3.68E-10
Ndufb2	1.98E-14	-0.496013116	0.659	0.932	4.01E-10
Dtymk	2.01E-14	-0.559781717	0.488	0.774	4.08E-10
Nrp2	2.22E-14	0.651706265	0.858	0.712	4.50E-10
Rasip1	2.25E-14	0.584336976	0.923	0.808	4.55E-10
Sephs2	3.13E-14	-0.295545549	0.213	0.575	6.34E-10
Cdkn2d	3.33E-14	-0.340764081	0.175	0.493	6.75E-10
mt-Nd4l	3.46E-14	0.634779035	0.942	0.952	7.00E-10
Snrnp25	3.56E-14	-0.401230997	0.268	0.63	7.21E-10
Cox20	3.74E-14	-0.362627739	0.366	0.781	7.58E-10
Slbp	3.88E-14	-0.563525054	0.396	0.719	7.86E-10

Cep295	4.15E-14	-0.25763289	0.214	0.575	8.40E-10
Tfdp1	4.17E-14	-0.297131875	0.29	0.678	8.44E-10
Cenpe	4.48E-14	-0.384693907	0.16	0.459	9.08E-10
Hnrnph3	5.30E-14	-0.358706419	0.368	0.788	1.07E-09
Atp5j	5.47E-14	-0.394421892	0.944	0.979	1.11E-09
Eif3c	5.57E-14	-0.353644955	0.958	1	1.13E-09
Ppih	5.72E-14	-0.399192884	0.269	0.61	1.16E-09
Eef1b2	5.76E-14	0.291427676	0.998	1	1.17E-09
Cox17	6.21E-14	-0.491622044	0.593	0.897	1.26E-09
Lyar	6.41E-14	-0.582367503	0.365	0.705	1.30E-09
Uqcrq	6.84E-14	-0.424804392	0.941	0.993	1.38E-09
Cald1	7.77E-14	0.547802996	0.986	0.897	1.57E-09
Tubb4b	8.09E-14	-0.58278317	0.715	0.945	1.64E-09
Ccnd1	8.37E-14	1.098507777	0.67	0.479	1.69E-09
Aplp2	9.15E-14	0.668443314	0.842	0.808	1.85E-09
Nhp2	9.87E-14	-0.609857554	0.848	0.973	2.00E-09
Set	9.89E-14	-0.593949718	0.947	0.979	2.00E-09
Nsmce4a	1.06E-13	-0.400514372	0.473	0.849	2.15E-09
Gspt1	1.09E-13	-0.465610033	0.75	0.945	2.20E-09
Hspa4	1.09E-13	-0.503333554	0.822	0.966	2.20E-09
Cops3	1.12E-13	-0.432424144	0.491	0.856	2.28E-09
Dst	1.30E-13	0.676076801	0.728	0.473	2.64E-09
Cirbp	1.35E-13	-0.28708187	0.28	0.685	2.74E-09
Nop16	1.37E-13	-0.361589875	0.318	0.692	2.77E-09
Fkbp4	1.37E-13	-0.60033369	0.581	0.836	2.78E-09
H1f10	1.41E-13	-0.506405941	0.167	0.466	2.85E-09
Ifrd1	1.43E-13	-0.505887557	0.416	0.767	2.90E-09
Hmbs	1.44E-13	-0.365776249	0.365	0.753	2.91E-09
Gar1	1.51E-13	-0.416588367	0.357	0.692	3.05E-09
Haus1	1.55E-13	-0.275550229	0.263	0.658	3.14E-09

Magoh	1.66E-13	-0.476687498	0.579	0.884	3.37E-09
Nme1	1.67E-13	-0.527631015	0.93	0.993	3.38E-09
Ptrh2	1.88E-13	-0.369207338	0.343	0.733	3.81E-09
Cacybp	1.94E-13	-0.558629342	0.696	0.925	3.92E-09
Dtnbp1	1.94E-13	-0.395901456	0.358	0.74	3.93E-09
Srsf7	2.40E-13	-0.664270655	0.659	0.863	4.86E-09
Trnt1	2.58E-13	-0.270954698	0.3	0.719	5.22E-09
Fos	2.81E-13	-0.672772865	0.147	0.418	5.69E-09
Uchl3	3.05E-13	-0.394763201	0.326	0.678	6.18E-09
Wdr89	3.13E-13	-0.291536726	0.296	0.685	6.33E-09
Chchd2	3.24E-13	-0.298069106	0.992	1	6.56E-09
Sh3bp5	3.56E-13	0.675139512	0.776	0.582	7.21E-09
Parp1	3.72E-13	-0.422591038	0.454	0.808	7.52E-09
Rnaseh2c	3.83E-13	-0.423686299	0.43	0.795	7.76E-09
mt-Atp8	3.87E-13	0.722628793	0.737	0.678	7.83E-09
Cnbp	4.02E-13	-0.489226448	0.781	0.959	8.13E-09
Cgnl1	4.11E-13	0.665343183	0.678	0.404	8.33E-09
Timm8a1	4.63E-13	-0.424229982	0.285	0.61	9.36E-09
Lsm7	4.75E-13	-0.419529418	0.599	0.904	9.62E-09
Cks1b	4.84E-13	-0.515084733	0.468	0.788	9.79E-09
Snrpa1	5.62E-13	-0.436994667	0.565	0.863	1.14E-08
Cdc123	5.91E-13	-0.381553598	0.455	0.829	1.20E-08
Rps2	6.00E-13	-0.30873964	0.998	1	1.21E-08
Sf3b6	6.07E-13	-0.38521586	0.682	0.945	1.23E-08
Nrros	6.12E-13	-0.272670426	0.124	0.384	1.24E-08
Naa10	6.98E-13	-0.354708993	0.332	0.685	1.41E-08
Apex1	6.99E-13	-0.477891701	0.607	0.897	1.41E-08
Rnps1	7.13E-13	-0.448902688	0.631	0.918	1.44E-08

Eif1	7.25E-13	0.338855482	0.998	1	1.47E-08
Ruvbl2	7.33E-13	-0.299063523	0.286	0.671	1.48E-08
Nucks1	7.38E-13	-0.548241705	0.834	0.959	1.49E-08
Ltv1	7.52E-13	-0.250751519	0.254	0.637	1.52E-08
Mrps35	7.92E-13	-0.274725828	0.299	0.685	1.60E-08
Arl6ip6	8.63E-13	-0.274472835	0.218	0.555	1.75E-08
Ctcf	8.78E-13	-0.44031638	0.618	0.911	1.78E-08
Lsm4	9.07E-13	-0.522665916	0.736	0.959	1.84E-08
Ppid	1.06E-12	-0.471747594	0.615	0.89	2.14E-08
Eloc	1.12E-12	-0.415945351	0.732	0.979	2.27E-08
Ebna1bp2	1.21E-12	-0.417321664	0.554	0.877	2.46E-08
Arl4c	1.27E-12	-0.362035592	0.23	0.555	2.58E-08
Cox7c	1.42E-12	-0.311882486	0.997	1	2.87E-08
Hmgb2	1.42E-12	-1.598331588	0.806	0.89	2.88E-08
Ddx21	1.77E-12	-0.530450993	0.576	0.863	3.58E-08
Aprt	1.77E-12	0.566416878	0.966	0.938	3.59E-08
Ncl	1.90E-12	-0.673486729	0.989	0.993	3.84E-08
Exosc3	1.91E-12	-0.253364543	0.235	0.582	3.88E-08
Taf7	1.93E-12	-0.329366704	0.376	0.76	3.91E-08
Ppat	2.08E-12	-0.258001851	0.263	0.623	4.21E-08
Nelfe	2.08E-12	-0.283908543	0.316	0.699	4.21E-08
Fkbp2	2.19E-12	-0.372183583	0.358	0.726	4.44E-08
Septin7	2.73E-12	-0.359606747	0.925	0.979	5.53E-08
Polr2d	2.82E-12	-0.332041283	0.382	0.774	5.71E-08
Erg28	3.06E-12	-0.390738634	0.407	0.788	6.20E-08
Ccnb2	3.29E-12	-0.35230876	0.153	0.432	6.67E-08
Lsm6	3.37E-12	-0.568299033	0.645	0.856	6.83E-08
Mrps25	3.47E-12	-0.315842504	0.362	0.733	7.03E-08

Cotl1	3.76E-12	0.708220166	0.72	0.63	7.62E-08
Sarnp	3.83E-12	-0.35706995	0.728	0.952	7.75E-08
Tnfaip8	3.89E-12	-0.253844704	0.225	0.555	7.88E-08
Ppm1g	4.10E-12	-0.446238362	0.49	0.808	8.29E-08
Vkorc1	4.23E-12	-0.291853907	0.351	0.753	8.57E-08
Maf	4.80E-12	0.932112121	0.554	0.267	9.71E-08
Serpinb6a	4.94E-12	0.546164902	0.845	0.842	9.99E-08
Gmps	4.96E-12	-0.42258625	0.567	0.863	1.00E-07
Etfa	5.05E-12	-0.33002057	0.407	0.822	1.02E-07
Rfc3	6.08E-12	-0.464246237	0.441	0.753	1.23E-07
Rnaseh2a	6.39E-12	-0.326276168	0.423	0.788	1.29E-07
Anxa3	6.81E-12	0.585746849	0.817	0.63	1.38E-07
Ndufab1	6.83E-12	-0.389110974	0.872	0.986	1.38E-07
Isyna1	7.21E-12	-0.311261614	0.305	0.658	1.46E-07
Dpm3	7.32E-12	-0.439290331	0.563	0.863	1.48E-07
Anxa5	7.67E-12	0.518061639	0.897	0.842	1.55E-07
Atp5pb	7.70E-12	-0.323915888	0.947	0.993	1.56E-07
Htatsf1	8.09E-12	-0.271924869	0.282	0.637	1.64E-07
Arhgap31	8.30E-12	0.618705363	0.717	0.562	1.68E-07
Skap2	8.43E-12	-0.262609617	0.311	0.692	1.71E-07
Eif2b5	8.46E-12	-0.262115119	0.354	0.753	1.71E-07
Bcas2	9.46E-12	-0.441503009	0.634	0.904	1.92E-07
Paics	1.02E-11	-0.398208675	0.624	0.911	2.07E-07
S100a6	1.14E-11	0.488302293	0.831	0.692	2.31E-07
Emg1	1.20E-11	-0.392274065	0.545	0.884	2.43E-07
Mesd	1.27E-11	-0.275782134	0.427	0.87	2.58E-07
Stip1	1.29E-11	-0.421502393	0.507	0.829	2.61E-07
Cyc1	1.34E-11	-0.425018694	0.653	0.904	2.70E-07
Ndufb4	1.39E-11	-0.329074758	0.815	0.986	2.82E-07
Rasgrp2	1.45E-11	-0.52575946	0.363	0.699	2.93E-07
Mtdh	1.50E-11	-	0.883	0.966	3.04E-07

		0.450537813			
Coq7	1.63E-11	-0.280211538	0.297	0.651	3.29E-07
Nudc	1.69E-11	-0.463586327	0.745	0.918	3.42E-07
Srsf9	1.71E-11	-0.35997868	0.715	0.918	3.47E-07
Lmo4	1.74E-11	-0.393203241	0.286	0.637	3.53E-07
Rad23a	1.82E-11	-0.303282137	0.391	0.774	3.69E-07
Psma3	1.94E-11	-0.368447828	0.87	0.993	3.92E-07
Tax1bp3	2.21E-11	0.537870919	0.854	0.808	4.48E-07
Fh1	2.32E-11	-0.326198105	0.444	0.822	4.69E-07
Mmp2	2.36E-11	0.563956314	0.603	0.322	4.78E-07
Ctsl	2.39E-11	0.479439473	0.984	0.966	4.84E-07
Lrpap1	2.46E-11	-0.774969869	0.402	0.815	4.97E-07
Ndufa5	2.75E-11	-0.37157049	0.834	0.979	5.57E-07
Ndufs8	2.78E-11	-0.3797384	0.54	0.863	5.64E-07
Rangap1	3.04E-11	-0.311751163	0.341	0.712	6.16E-07
Flt1	3.15E-11	0.646797496	0.615	0.445	6.38E-07
Rfc2	3.20E-11	-0.393934219	0.377	0.726	6.49E-07
Chchd4	3.70E-11	-0.283846481	0.307	0.658	7.49E-07
Pfn1	3.82E-11	-0.283139626	0.998	1	7.72E-07
Pin1	3.82E-11	-0.376995053	0.631	0.938	7.74E-07
Hnrnpab	4.04E-11	-0.455549062	0.914	0.993	8.17E-07
Elf1	4.29E-11	-0.287091653	0.282	0.63	8.68E-07
Nop58	4.52E-11	-0.575764155	0.654	0.863	9.15E-07
mt-Cytb	4.57E-11	0.297667905	0.997	1	9.26E-07
D030007L05Rik	4.71E-11	-0.308425046	0.178	0.445	9.53E-07
Vapa	4.92E-11	-0.330516599	0.773	0.959	9.95E-07
Inka1	5.07E-11	-0.40985843	0.404	0.726	1.03E-06
Ftl1-ps1	5.25E-11	0.426181193	0.998	1	1.06E-06
Mrpl14	5.59E-11	-0.393858481	0.621	0.863	1.13E-06
Polr2k	5.66E-11	-0.349501341	0.432	0.788	1.15E-06
Hspa14	5.85E-11	-	0.405	0.767	1.19E-06

		0.351937358			
Pcdh17	6.51E-11	0.767548993	0.71	0.589	1.32E-06
Ndufv2	6.63E-11	-0.382776938	0.723	0.945	1.34E-06
Srrm1	6.99E-11	-0.376213109	0.806	0.966	1.42E-06
Arpc4	7.07E-11	-0.380442327	0.667	0.945	1.43E-06
Esam	7.11E-11	0.532311915	0.837	0.747	1.44E-06
Pecam1	7.19E-11	0.404132833	0.958	0.842	1.46E-06
Gng2	7.30E-11	0.686822785	0.601	0.473	1.48E-06
Prkar2b	8.05E-11	-0.373178229	0.254	0.555	1.63E-06
Ftl1	8.16E-11	0.685140581	0.601	0.486	1.65E-06
Gm47283	8.51E-11	-0.377812275	0.46	0.801	1.72E-06
Actl6a	9.15E-11	-0.334308915	0.474	0.829	1.85E-06
Mrpl36	9.71E-11	-0.31149366	0.48	0.87	1.97E-06
Srsf2	1.12E-10	-0.446524158	0.554	0.822	2.26E-06
Eri1	1.14E-10	-0.274113582	0.294	0.623	2.31E-06
Tmem176a	1.18E-10	-0.461128373	0.543	0.877	2.39E-06
Mrpl42	1.28E-10	-0.437327539	0.545	0.795	2.58E-06
Ssb	1.41E-10	-0.347323345	0.887	0.979	2.86E-06
Cdk2ap1	1.47E-10	-0.281006095	0.366	0.747	2.98E-06
Rpl4	1.62E-10	0.296499723	0.994	1	3.28E-06
Ccdc88a	1.64E-10	0.616499655	0.764	0.774	3.32E-06
Cct6a	1.71E-10	-0.417302376	0.873	0.959	3.46E-06
Nrk	1.82E-10	-0.299188142	0.069	0.253	3.68E-06
Rpl29	1.84E-10	-0.32990489	0.815	0.979	3.73E-06
Cox7a2	1.87E-10	-0.339858274	0.92	0.993	3.79E-06
Amotl1	1.87E-10	0.579929395	0.573	0.342	3.79E-06
Supt16	1.93E-10	-0.493052304	0.703	0.89	3.92E-06
Brd7	2.22E-10	-0.425468699	0.731	0.932	4.49E-06
Ier2	2.24E-10	-0.49656146	0.718	0.925	4.53E-06
Hmgn5	2.25E-10	-0.443641223	0.41	0.733	4.55E-06
Txnl1	2.27E-10	-0.317437118	0.809	1	4.60E-06

Dlc1	2.40E-10	0.580341895	0.654	0.473	4.87E-06
Wdr18	2.46E-10	-0.329981373	0.488	0.849	4.98E-06
Ube2j2	2.47E-10	-0.288102676	0.429	0.822	5.01E-06
Nolc1	2.53E-10	-0.431331304	0.465	0.774	5.12E-06
Dpy30	2.67E-10	-0.358212915	0.512	0.842	5.41E-06
Tagln2	2.73E-10	-0.49062563	0.923	0.986	5.53E-06
Dlk1	2.74E-10	-0.411105462	0.07	0.253	5.55E-06
Snrpg	2.82E-10	-0.436283096	0.966	1	5.71E-06
Laptm5	2.92E-10	-0.302594218	0.272	0.589	5.92E-06
Ndufs7	2.94E-10	-0.373395087	0.714	0.938	5.94E-06
Gtf3c6	3.19E-10	-0.336767666	0.441	0.774	6.46E-06
Rhoj	3.23E-10	0.565501442	0.734	0.63	6.54E-06
Uqcrc2	3.27E-10	-0.356111929	0.712	0.952	6.61E-06
Nrp1	3.38E-10	0.682030276	0.685	0.541	6.84E-06
Phf5a	3.50E-10	-0.371898095	0.657	0.925	7.09E-06
Msh2	3.69E-10	-0.25178812	0.362	0.76	7.47E-06
Hnrnpk	3.86E-10	-0.321433784	0.942	0.993	7.81E-06
Ruvbl1	4.06E-10	-0.295696821	0.468	0.795	8.22E-06
Ybx3	4.07E-10	-0.42326374	0.77	0.952	8.24E-06
Ifitm3	4.32E-10	0.679057679	0.865	0.76	8.75E-06
Oat	4.41E-10	-0.256006167	0.249	0.555	8.94E-06
Kras	4.46E-10	-0.363359548	0.617	0.897	9.03E-06
Cd93	4.54E-10	0.625649944	0.754	0.685	9.19E-06
Mrpl10	5.59E-10	-0.26367993	0.36	0.719	1.13E-05
Selenoh	5.72E-10	-0.637235978	0.684	0.87	1.16E-05
Mrto4	5.81E-10	-0.409001644	0.537	0.801	1.18E-05
Pole3	7.00E-10	-0.261341059	0.371	0.733	1.42E-05
Fam174a	7.07E-10	-0.267814089	0.358	0.726	1.43E-05
Plin2	7.20E-10	0.54389015	0.761	0.74	1.46E-05
Ube2e3	7.46E-10	-0.37373936	0.668	0.938	1.51E-05
Ift57	7.60E-10	-	0.228	0.521	1.54E-05

		0.295283018			
Pttg1	8.06E-10	-0.421108647	0.152	0.377	1.63E-05
Mrps28	8.12E-10	-0.331897091	0.429	0.76	1.64E-05
Nap1l4	8.24E-10	-0.348544623	0.649	0.897	1.67E-05
Brd8	8.28E-10	-0.341238664	0.423	0.767	1.68E-05
Ckb	8.32E-10	-0.263857842	0.136	0.37	1.68E-05
Nrg1	8.38E-10	-0.275436396	0.081	0.267	1.70E-05
Thrap3	8.46E-10	-0.378947697	0.67	0.938	1.71E-05
Srrt	8.61E-10	-0.282401355	0.38	0.719	1.74E-05
Slc25a5	8.89E-10	-0.395815404	0.897	0.973	1.80E-05
Mrps15	8.92E-10	-0.379537359	0.546	0.842	1.81E-05
Rad51ap1	9.06E-10	-0.300696815	0.216	0.473	1.83E-05
Rcc2	9.11E-10	-0.369778267	0.427	0.74	1.84E-05
Cct7	9.20E-10	-0.334784995	0.878	0.966	1.86E-05
Rrp1	9.62E-10	-0.361926992	0.748	0.945	1.95E-05
Ppa1	1.01E-09	-0.378887387	0.344	0.658	2.05E-05
Bola1	1.19E-09	-0.251558509	0.349	0.712	2.40E-05
Nol7	1.20E-09	-0.359311386	0.798	0.952	2.42E-05
Prdx1	1.31E-09	-0.300578528	0.984	1	2.66E-05
Mtx1	1.34E-09	-0.286254239	0.338	0.678	2.72E-05
Cox7b	1.37E-09	-0.329876265	0.919	0.979	2.76E-05
Bola3	1.43E-09	-0.28863217	0.437	0.767	2.89E-05
Ilf2	1.43E-09	-0.319055779	0.585	0.884	2.89E-05
Adam19	1.47E-09	0.571912	0.499	0.274	2.97E-05
Tsfm	1.51E-09	-0.254329329	0.341	0.692	3.06E-05
Cuta	1.56E-09	-0.273190894	0.452	0.829	3.17E-05
Rbm3	1.58E-09	-0.391983057	0.87	0.952	3.19E-05
Ndufaf2	1.66E-09	-	0.399	0.726	3.36E-05

		0.320189437			
Ccng1	1.70E-09	0.53643084	0.823	0.781	3.45E-05
Mrps31	1.73E-09	-0.259648572	0.426	0.795	3.50E-05
Rnf187	1.97E-09	-0.354949964	0.704	0.938	3.99E-05
Arhgef12	2.03E-09	0.594379838	0.587	0.397	4.10E-05
Ctnnb1	2.06E-09	0.493494902	0.842	0.884	4.17E-05
Lgals9	2.08E-09	-0.295311346	0.421	0.781	4.21E-05
Rpl23a	2.09E-09	0.325590478	0.997	1	4.23E-05
S100a11	2.11E-09	0.386245713	0.915	0.87	4.28E-05
Mrpl22	2.38E-09	-0.305923732	0.588	0.87	4.81E-05
Leo1	2.39E-09	-0.264551656	0.376	0.747	4.84E-05
Mpdu1	2.64E-09	-0.321610907	0.485	0.836	5.35E-05
Eif4a3	2.71E-09	-0.338863932	0.779	0.966	5.49E-05
L1td1	2.81E-09	-0.270941379	0	0.055	5.68E-05
Smc6	2.85E-09	-0.39305031	0.557	0.829	5.76E-05
Park7	2.87E-09	-0.332936625	0.861	0.966	5.81E-05
Pum3	3.10E-09	-0.339685474	0.452	0.788	6.27E-05
Ralb	3.10E-09	0.427167297	0.903	0.884	6.28E-05
Rbmxl1	3.26E-09	-0.315259306	0.482	0.815	6.59E-05
Mrpl17	3.29E-09	-0.341214941	0.723	0.959	6.67E-05
Clpp	3.40E-09	-0.267809084	0.466	0.829	6.89E-05
Bub3	3.43E-09	-0.298879453	0.534	0.87	6.94E-05
Ktn1	3.47E-09	0.43385139	0.925	0.966	7.03E-05
Eif6	3.54E-09	-0.31697768	0.762	0.959	7.16E-05
Gadd45gip1	3.57E-09	-0.258380336	0.501	0.842	7.23E-05
Rsl1d1	3.79E-09	-0.397419621	0.698	0.925	7.67E-05
Hspe1	3.96E-09	-0.51800335	0.93	0.979	8.02E-05
St13	4.02E-09	0.399287275	0.961	0.979	8.14E-05
Rrbp1	4.09E-09	0.46757938	0.934	0.932	8.28E-05
Hspa9	4.13E-09	-0.331957184	0.51	0.808	8.36E-05
Gnl3	4.21E-09	-0.47180026	0.482	0.747	8.53E-05
Mettl9	4.40E-09	-	0.48	0.795	8.91E-05

		0.304093633			
Mbd3	4.50E-09	-0.327124269	0.651	0.938	9.11E-05
Rad21	4.57E-09	-0.273618655	0.369	0.705	9.25E-05
Tbc1d15	4.60E-09	-0.272259909	0.379	0.699	9.31E-05
Gtf2b	4.62E-09	-0.26304939	0.416	0.753	9.36E-05
Rap1a	4.66E-09	-0.3309788	0.643	0.904	9.44E-05
Mphosph10	4.68E-09	-0.294888083	0.446	0.774	9.46E-05
C1qbp	4.75E-09	-0.450892064	0.862	0.938	9.61E-05
Cdkn1a	4.97E-09	0.477755009	0.917	0.801	0.000100586
Mrps18c	5.24E-09	-0.301080717	0.612	0.897	0.000106168
Etfb	5.37E-09	-0.323437455	0.712	0.932	0.000108707
Ly6e	5.59E-09	-0.311697385	0.275	0.575	0.000113069
Id3	5.63E-09	0.605922106	0.721	0.527	0.000114006
Sash1	5.72E-09	0.531652658	0.634	0.507	0.000115877
Rab8a	5.93E-09	-0.318598209	0.62	0.925	0.000120075
Ndufv3	5.94E-09	-0.318200364	0.757	0.952	0.000120186
Mrpl41	6.43E-09	-0.262689495	0.424	0.774	0.000130071
Atp5g3	6.58E-09	-0.278556745	0.898	0.979	0.000133243
Tpst2	6.76E-09	-0.303517258	0.484	0.774	0.000136918
Nid1	6.81E-09	0.573715169	0.598	0.438	0.000137871
Plvap	7.08E-09	0.54905742	0.901	0.795	0.000143325
Smchd1	7.22E-09	-0.291651106	0.405	0.719	0.000146161
Psmb3	7.76E-09	-0.290173444	0.95	0.993	0.000157181
Abracl	8.22E-09	-0.361706091	0.753	0.973	0.000166413
Echs1	8.72E-09	-0.296856306	0.463	0.801	0.000176533
Adgrf5	8.91E-09	0.547641826	0.541	0.336	0.000180413
Nol11	9.19E-09	-0.310840478	0.51	0.836	0.00018595
Rex1bd	9.91E-09	-0.309679463	0.765	0.932	0.000200609
Fam181b	1.03E-08	-0.273216763	0.117	0.322	0.000208988
Casp3	1.07E-08	0.530672086	0.687	0.616	0.00021587

Rfc1	1.12E-08	-0.351483154	0.48	0.788	0.000226746
Slc39a4	1.29E-08	0.466201248	0.29	0.082	0.000262123
Glrx2	1.34E-08	-0.294442844	0.585	0.89	0.000271765
Flnb	1.34E-08	0.564465887	0.656	0.555	0.000271904
Psma4	1.42E-08	-0.303519032	0.897	0.979	0.00028699
Cks2	1.44E-08	-0.408245595	0.424	0.699	0.000292038
Tpm3	1.50E-08	-0.370595172	0.911	0.979	0.000302736
Plxnd1	1.54E-08	0.584453879	0.72	0.699	0.000311348
Thyn1	1.59E-08	-0.25844413	0.382	0.712	0.000322716
Mdh2	1.61E-08	-0.300415936	0.836	0.973	0.000325608
Basp1	1.63E-08	0.405957487	0.915	0.774	0.000329617
Elob	1.83E-08	-0.252335116	0.977	0.993	0.000370201
Dbi	1.91E-08	-0.345134091	0.886	0.979	0.000385926
Pcbp1	1.91E-08	-0.313616355	0.897	0.979	0.000386457
Higd1a	1.99E-08	-0.296637237	0.754	0.932	0.000403808
Psmb2	2.04E-08	-0.296691093	0.887	0.986	0.000412048
Limch1	2.07E-08	0.591701332	0.62	0.473	0.000419558
Tagln	2.14E-08	0.539331516	0.609	0.432	0.00043373
Lpp	2.23E-08	0.587421425	0.632	0.582	0.000451938
Mrps33	2.41E-08	-0.30153924	0.746	0.966	0.000487072
Dusp2	2.44E-08	-0.352319474	0.36	0.664	0.000494212
Pgp	2.60E-08	-0.338746378	0.579	0.87	0.000527136
Rps17	2.68E-08	0.251078502	1	1	0.000542684
Plec	2.77E-08	0.576657144	0.638	0.555	0.000560314
Nrep	2.81E-08	0.556903734	0.617	0.473	0.00056956
Pfdn6	2.86E-08	-0.304003763	0.479	0.753	0.000578354
Eif3f	2.94E-08	-0.335574095	0.98	1	0.000594609
Btg1	3.01E-08	0.43351567	0.842	0.747	0.000608898
H1f2	3.17E-08	-0.263906072	0.374	0.692	0.000640784
Epcam	3.23E-08	-0.348209843	0.006	0.075	0.000653095
C1d	3.24E-08	-0.322310773	0.681	0.911	0.000655045

Mrpl13	3.45E-08	-0.285391599	0.49	0.815	0.000699162
Hnrnpa3	3.57E-08	-0.548758252	0.889	0.938	0.000723021
Tomm5	3.63E-08	-0.369010754	0.723	0.959	0.000734216
Psmd12	3.64E-08	-0.295127239	0.876	0.979	0.000737197
Cetn3	3.93E-08	-0.328908268	0.671	0.911	0.000794608
Cox4i1	4.00E-08	0.269625156	0.998	1	0.000809452
Rpl13a	4.17E-08	0.42042847	0.845	0.877	0.000843584
Hmgn2	4.35E-08	-0.3702625	0.438	0.699	0.000880898
Vdac3	4.47E-08	-0.271137803	0.917	0.986	0.000904757
Stoml2	4.47E-08	-0.25828523	0.449	0.774	0.000904839
Anapc11	4.53E-08	-0.321505794	0.718	0.925	0.000917466
Gngt2	4.64E-08	0.429479893	0.844	0.767	0.000938802
Rpl7l1	5.02E-08	-0.256292416	0.543	0.877	0.001017085
Sf3b5	5.07E-08	-0.347150977	0.706	0.904	0.001025738
Mrpl15	5.35E-08	-0.278679472	0.715	0.938	0.001083377
Cyyr1	5.58E-08	0.471246814	0.441	0.247	0.001130522
Lims1	5.62E-08	-0.260705697	0.463	0.801	0.001137006
Tsn	5.80E-08	-0.276996377	0.782	0.945	0.001174422
Prelid3b	6.06E-08	-0.255361984	0.336	0.623	0.001227717
Afdn	6.13E-08	0.443918502	0.803	0.753	0.001240382
Ushbp1	6.16E-08	0.434420289	0.321	0.123	0.001247492
Gpx3	6.21E-08	0.479158468	0.756	0.664	0.001256672
Ddb1	6.48E-08	0.487146346	0.739	0.774	0.00131248
Mrpl20	6.63E-08	-0.360079192	0.757	0.932	0.001341839
Cisd1	6.79E-08	-0.339847897	0.512	0.801	0.001374579
Puf60	6.88E-08	-0.26763936	0.613	0.884	0.001392928
Arl1	7.36E-08	-0.285572371	0.596	0.938	0.001490818
Mrps16	7.59E-08	-0.314853272	0.684	0.911	0.001535469
Tanc1	7.69E-08	0.475294532	0.493	0.315	0.001557012
Cemip2	7.69E-08	0.517706052	0.54	0.404	0.001557183
Ak2	8.44E-08	-0.302292896	0.606	0.877	0.001708033

Hnrnpl	8.48E-08	-0.260475882	0.568	0.884	0.001717141
Ubc	8.49E-08	0.538683127	0.712	0.699	0.001719629
Snw1	9.58E-08	-0.302712967	0.651	0.89	0.001939213
Ndufaf8	9.75E-08	-0.263622915	0.532	0.877	0.001974525
Eprs	9.97E-08	-0.321458225	0.734	0.945	0.00201819
Spcs2	1.05E-07	-0.270004103	0.812	0.945	0.002128706
Nip7	1.06E-07	-0.296694022	0.554	0.836	0.002140241
Tle5	1.07E-07	-0.374361968	0.668	0.938	0.002162667
Psmb6	1.08E-07	-0.287357817	0.955	1	0.002177463
Hnrnpr	1.10E-07	-0.290193232	0.613	0.897	0.002219009
Rhoa	1.12E-07	-0.276661937	0.895	0.986	0.002270629
Apela	1.12E-07	-0.296066194	0.009	0.082	0.002273854
Cd9	1.16E-07	-0.400969112	0.46	0.726	0.002358236
Sf3b2	1.18E-07	-0.313597586	0.844	0.945	0.002380882
Emc2	1.19E-07	-0.268961186	0.567	0.87	0.002403819
Hnrnpu	1.23E-07	-0.363487149	0.911	0.973	0.002483088
Utp18	1.24E-07	-0.272876606	0.418	0.733	0.002512547
Idh2	1.28E-07	-0.282357144	0.521	0.836	0.002587932
Tubb2b	1.31E-07	0.540886408	0.551	0.411	0.002646807
Ubtf	1.35E-07	-0.252668108	0.479	0.774	0.002727176
Atp5o.1	1.38E-07	-0.27317045	0.93	0.993	0.00279592
Btg2	1.40E-07	0.534212196	0.628	0.466	0.002837245
Anxa2	1.41E-07	0.31453864	0.898	0.795	0.002853612
Immt	1.48E-07	-0.308966236	0.653	0.932	0.003000276
Psmd14	1.58E-07	-0.251678102	0.559	0.836	0.003201519
Nsmce2	1.58E-07	-0.251999384	0.455	0.774	0.003207695
Mrpl32	1.67E-07	-0.266653686	0.488	0.781	0.003382045
Tma7	1.68E-07	-0.272946649	0.936	1	0.00339858

Ola1	1.70E-07	-0.289597844	0.612	0.877	0.003443122
Arl4a	1.81E-07	-0.293311613	0.399	0.705	0.003655742
Clec1b	1.81E-07	-0.416686431	0.211	0.432	0.003659809
Coa3	1.83E-07	-0.263687238	0.834	0.986	0.003702
Mrpl28	1.87E-07	-0.277481823	0.546	0.849	0.00378189
Taf15	1.94E-07	-0.272518161	0.471	0.788	0.003933563
Mrpl30	1.98E-07	-0.265249795	0.751	0.945	0.004006429
Snrpd2	2.08E-07	-0.322603285	0.928	0.986	0.004212674
Smc3	2.14E-07	-0.321895787	0.653	0.904	0.00433882
Rasgrp3	2.19E-07	0.535975378	0.59	0.479	0.004428382
Malat1	2.42E-07	0.355565632	1	1	0.004892022
Cdv3	2.44E-07	-0.288975614	0.903	0.973	0.00494207
Sdhb	2.58E-07	-0.29060545	0.678	0.918	0.005223905
Sf3a3	2.61E-07	-0.30162805	0.543	0.842	0.005292093
Srsf10	2.65E-07	-0.305599132	0.584	0.849	0.005354821
Smu1	2.66E-07	-0.265573229	0.457	0.774	0.005382824
Polr2f	2.68E-07	-0.316974236	0.715	0.918	0.00543385
Fdps	2.75E-07	-0.355901672	0.485	0.733	0.005564836
Ywhaq	2.77E-07	-0.250890919	0.897	0.993	0.005616907
Dab2	2.79E-07	0.489866524	0.523	0.288	0.005643097
Smarca5	2.79E-07	-0.28416723	0.903	0.993	0.005643156
Sumo1	2.81E-07	-0.250850788	0.793	0.952	0.005697999
Eml4	2.94E-07	-0.361189732	0.638	0.897	0.005944841
Ets1	3.12E-07	0.411046888	0.833	0.692	0.006313903
Nfib	3.26E-07	0.531359346	0.62	0.521	0.006600851
Stub1	3.30E-07	-0.280044473	0.598	0.842	0.006672459
Smc1a	3.35E-07	-0.364490224	0.723	0.911	0.006779687
Cdc42se1	3.35E-07	-0.275959601	0.52	0.849	0.006780877
Vim	3.38E-07	0.278550317	0.998	1	0.006834169
Arl6ip1	3.42E-07	-0.359512414	0.795	0.959	0.006929159

Ssrp1	3.44E-07	-0.29652002	0.579	0.808	0.006967386
Cox7a2l	3.47E-07	0.348535533	0.947	0.973	0.007023445
Ddx18	3.57E-07	-0.284291579	0.574	0.877	0.007235002
Ifitm1	3.59E-07	-0.327219573	0.039	0.151	0.007259472
Cd47	3.63E-07	-0.401022119	0.706	0.87	0.007343541
Sqstm1	3.67E-07	0.379829588	0.884	0.877	0.007425544
Kxd1	3.78E-07	-0.301281071	0.563	0.829	0.00765916
Acp1	3.83E-07	-0.272442706	0.546	0.863	0.007757255
Dync1i2	4.04E-07	0.286515114	0.914	0.952	0.008172393
Nsmce1	4.17E-07	-0.264916376	0.548	0.842	0.008436291
Prcp	4.61E-07	0.532990341	0.653	0.603	0.009330053
Abce1	4.71E-07	-0.271666261	0.587	0.877	0.009529077
Mrpl57	4.72E-07	-0.303608113	0.698	0.966	0.009544744
Tomm40	4.73E-07	-0.275419638	0.559	0.842	0.009578126
Rheb	5.05E-07	-0.259549197	0.612	0.884	0.010223803
Pdia3	5.27E-07	-0.256430959	0.923	0.979	0.010662404
Fam136a	5.30E-07	-0.261278199	0.399	0.692	0.010731093
Ciao2b	5.34E-07	-0.268987726	0.588	0.897	0.010819612
Rhoc	5.37E-07	0.265479119	0.922	0.87	0.010868364
Nop56	5.60E-07	-0.293865807	0.546	0.822	0.011332815
Zfand5	5.70E-07	-0.277658306	0.692	0.904	0.011528759
Myct1	5.80E-07	0.409973785	0.757	0.603	0.011738795
Epn2	5.97E-07	0.482100786	0.603	0.521	0.012082614
Ece1	6.11E-07	0.440173852	0.604	0.466	0.012366456
Psme2	6.15E-07	-0.278206091	0.692	0.925	0.012456742
Bvht	6.32E-07	0.457032802	0.432	0.26	0.0128021
Gps2	7.32E-07	-0.254621	0.512	0.801	0.014817405
Ecscr	7.63E-07	0.358712155	0.97	0.952	0.01543913
Septin11	7.92E-07	0.43830943	0.79	0.808	0.016025695
Timm23	8.45E-07	-0.300813811	0.718	0.904	0.017107208
Utrn	8.45E-07	0.479662873	0.662	0.603	0.017109697
Arhgef28	8.72E-07	0.517446102	0.493	0.37	0.017647894

Ppdpf	8.88E-07	-0.256760227	0.518	0.788	0.01796595
Sinhcaf	8.95E-07	-0.26028094	0.407	0.699	0.01811203
Cdk4	9.12E-07	-0.250151403	0.906	0.979	0.01846918
Psmd13	9.30E-07	-0.267281245	0.69	0.911	0.018827107
Pno1	1.04E-06	-0.274091712	0.523	0.781	0.021123616
Adh5	1.07E-06	-0.255803313	0.698	0.959	0.021750628
Map1lc3b	1.17E-06	0.306045833	0.964	0.973	0.023648484
Psmd11	1.19E-06	-0.260421091	0.781	0.945	0.024158748
Cited2	1.33E-06	-0.257337063	0.432	0.705	0.027015784
Mast4	1.36E-06	0.512489922	0.685	0.658	0.027607617
Rflnb	1.42E-06	-0.346351511	0.108	0.253	0.028688893
Rad50	1.44E-06	-0.25579647	0.532	0.767	0.029197145
Cav1	1.48E-06	0.401638066	0.49	0.301	0.030060482
Phb2	1.53E-06	-0.313799074	0.817	0.966	0.030955374
Tpi1	1.54E-06	-0.276954804	0.552	0.822	0.031164192
Slirp	1.60E-06	-0.260819251	0.673	0.897	0.032443212
Pcm1	1.66E-06	-0.28945357	0.709	0.918	0.033690792
Kif2a	1.77E-06	-0.31729867	0.674	0.884	0.035752505
Crmp1	1.80E-06	0.428820745	0.41	0.253	0.03650819
Rbis	1.90E-06	-0.274948074	0.731	0.904	0.038496823
Cd109	1.92E-06	0.444728235	0.44	0.281	0.03891069
Nisch	2.09E-06	0.436358337	0.817	0.877	0.042398418
Hmcn1	2.11E-06	0.416479418	0.545	0.404	0.042696106
Cct3	2.17E-06	-0.297543314	0.848	0.952	0.043970969
Apoe	2.19E-06	0.725985439	0.523	0.404	0.044429822
Rbp1	2.25E-06	-0.450669467	0.429	0.671	0.045565648
Lsm5	2.31E-06	-0.266795447	0.585	0.849	0.046744223
Gatad1	2.37E-06	-0.255038782	0.579	0.849	0.048010338
Timm17a	2.41E-06	-0.258127483	0.698	0.932	0.048854581
Tspan18	2.43E-06	0.389569059	0.746	0.671	0.049111122

Supplementary Table 6.4- List of differentially expressed genes from sc-RNA-Seq HP clusters.

List of differentially expressed genes from All cytokines (+VEGF) HP (pct.1) cluster vs withdrawal VEGF (-VEGF) HP cluster (pct.2).

Gene	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
Rps10	5.33E-58	-2.055722649	0.926	1	1.08E-53
Rps29	2.08E-56	1.165821119	1	1	4.21E-52
Gm8797	3.13E-54	-1.639033084	0.443	1	6.34E-50
Rpl35	5.71E-53	1.423566074	1	1	1.16E-48
Rpl36al	1.91E-52	-1.510816346	0.846	1	3.86E-48
mt-Co2	1.07E-47	0.849069759	1	1	2.17E-43
mt-Atp6	6.77E-47	0.96045829	1	1	1.37E-42
mt-Co1	1.19E-43	0.967143406	1	1	2.41E-39
Rpl10	3.45E-43	0.965560193	1	1	6.98E-39
Rps3a1	2.37E-39	-0.607626788	1	1	4.79E-35
Gm10076	3.82E-39	1.935158158	0.987	0.935	7.73E-35
Nrgn	5.26E-37	-1.375977852	0.289	0.919	1.07E-32
Ubb	2.49E-36	1.082889131	1	1	5.04E-32
Tmem108	1.24E-35	-0.926611191	0.091	0.685	2.50E-31
Phpt1	1.91E-35	-0.783220299	0.141	0.758	3.86E-31
Hist1h2ap	4.59E-34	2.315813276	0.775	0.137	9.29E-30
Rpl9-ps6	9.83E-34	-0.877163999	0.285	0.871	1.99E-29
Cxxc5	1.52E-31	-0.878820788	0.218	0.806	3.09E-27
Ccnd3	3.18E-31	-1.071833776	0.782	0.992	6.43E-27
Rps13	1.60E-30	0.532008073	1	1	3.24E-26
Rps26	3.26E-30	-0.546934715	1	1	6.60E-26
Rpl38	3.36E-28	0.491723927	1	1	6.80E-24
mt-Co3	4.54E-28	0.639857974	1	1	9.19E-24
Rps27	5.35E-28	0.561384755	1	1	1.08E-23
Rplp1	8.94E-28	0.467372061	1	1	1.81E-23
Uqcrb	1.58E-26	-0.640918643	0.966	1	3.20E-22
Lgals1	2.96E-26	2.166219005	0.866	0.605	6.00E-22

Atp5g2	1.11E-25	-0.551332353	0.997	1	2.25E-21
Gem	3.29E-25	-0.94411136	0.138	0.645	6.67E-21
Rps28	6.85E-25	0.472913516	1	1	1.39E-20
H3f3a	1.05E-24	-0.597376817	1	1	2.13E-20
Mgst3	2.17E-24	-0.868106536	0.456	0.935	4.40E-20
Pf4	3.84E-24	-1.35319968	0.158	0.677	7.77E-20
H2ac8	4.09E-24	1.748031161	0.748	0.347	8.27E-20
Gap43	4.10E-24	1.855729961	0.668	0.194	8.30E-20
Col4a2	7.70E-24	1.783629718	0.819	0.556	1.56E-19
Rps19	1.39E-23	0.39885841	1	1	2.81E-19
Rnaset2a	4.22E-23	-0.470504503	0.084	0.508	8.53E-19
Alox5ap	5.73E-23	-1.075663772	0.453	0.919	1.16E-18
mt-Nd3	7.72E-23	0.805748733	0.98	0.976	1.56E-18
Peg3	1.17E-22	2.264811743	0.745	0.363	2.37E-18
Dusp2	1.78E-22	-0.983357029	0.326	0.806	3.60E-18
mt-Nd1	2.18E-22	0.657412682	1	1	4.40E-18
Rps11	3.40E-22	-0.366194267	1	1	6.89E-18
Hand2	4.55E-22	-0.70423723	0.034	0.387	9.20E-18
Uba52	8.27E-22	-0.625541903	0.919	1	1.67E-17
Gmfg	9.58E-22	-0.815827044	0.862	0.976	1.94E-17
Lyl1	1.03E-21	-0.751724182	0.336	0.815	2.09E-17
mt-Nd4l	2.72E-21	0.834168175	0.993	0.984	5.50E-17
Ssbp2	6.83E-21	-0.467875691	0.171	0.653	1.38E-16
Col4a1	7.98E-21	2.232014761	0.819	0.734	1.62E-16
Was	9.26E-21	-0.473980044	0.205	0.758	1.88E-16
Hypk	9.30E-21	-0.723840662	0.493	0.911	1.88E-16
Serpinb6b	1.19E-20	-0.593916825	0.128	0.581	2.42E-16
Map1b	1.20E-20	1.237472322	0.591	0.121	2.42E-16
Trmt112	1.20E-20	-0.647815651	0.678	0.984	2.43E-16
Clec1b	2.73E-20	-1.105495079	0.329	0.79	5.53E-16
Ndufb11	7.24E-20	-0.583021399	0.93	1	1.47E-15
Car2	7.60E-20	-	0.547	0.944	1.54E-15

		1.074771839			
Eno1	7.97E-20	0.881554867	0.674	0.323	1.61E-15
Rps4x	1.06E-19	-0.353171593	1	1	2.14E-15
Tmem176b	1.27E-19	-0.838408765	0.443	0.887	2.56E-15
Rps27a	1.54E-19	-0.28883444	1	1	3.12E-15
Il21r	1.89E-19	-0.422077568	0.037	0.371	3.83E-15
Rplp0	2.24E-19	-0.434783282	1	1	4.54E-15
Adgrf5	2.77E-19	0.862669567	0.497	0.032	5.61E-15
Chchd2	4.34E-19	-0.39957913	1	1	8.78E-15
Id3	7.57E-19	1.49628473	0.634	0.194	1.53E-14
Tpst2	8.54E-19	-0.738375842	0.5	0.855	1.73E-14
H1f5	1.12E-18	1.646311206	0.789	0.508	2.26E-14
Dok2	1.50E-18	-0.524789076	0.138	0.581	3.03E-14
Bloc1s1	1.58E-18	-0.51877856	0.225	0.677	3.20E-14
Rpl19	1.81E-18	-0.276185909	1	1	3.66E-14
Ppdpf	1.88E-18	-0.628423529	0.624	0.968	3.80E-14
Cald1	1.89E-18	1.55286425	0.832	0.798	3.84E-14
Tmsb10	3.20E-18	1.134225718	0.997	0.992	6.48E-14
Flna	3.34E-18	0.980682353	0.745	0.484	6.76E-14
Dhrs3	3.74E-18	-0.393653237	0.107	0.508	7.57E-14
Eif1	3.78E-18	0.41975953	1	1	7.65E-14
Gna15	5.46E-18	-0.388931087	0.107	0.508	1.11E-13
Ninj1	7.71E-18	-0.674274215	0.282	0.718	1.56E-13
Mfng	8.93E-18	-0.662655529	0.413	0.847	1.81E-13
Lgals9	1.10E-17	-0.732780434	0.339	0.806	2.23E-13
Hapln1	1.21E-17	-0.662986299	0.272	0.726	2.46E-13
Hspg2	1.40E-17	0.992721572	0.644	0.29	2.84E-13
Rasgrp2	1.42E-17	-0.860028273	0.557	0.871	2.88E-13
Cgnl1	1.95E-17	0.80449973	0.534	0.089	3.95E-13
Rpl30	2.42E-17	-0.308778168	1	1	4.90E-13
Gnai2	2.87E-17	0.687068495	0.997	1	5.81E-13
Tspan32	2.89E-17	-0.510289664	0.124	0.532	5.84E-13

Rps7	4.30E-17	-0.320991965	1	1	8.70E-13
Gucy1b1	4.38E-17	-0.416683869	0.01	0.258	8.86E-13
Fcer1g	5.01E-17	-0.995213521	0.258	0.694	1.01E-12
Oaz1	5.48E-17	-0.469926332	0.983	1	1.11E-12
Timm10b	5.57E-17	-0.654994627	0.55	0.919	1.13E-12
Tagln	5.63E-17	3.235207105	0.644	0.331	1.14E-12
Gria2	6.33E-17	-0.54469641	0.158	0.573	1.28E-12
Rbm38	6.65E-17	-0.663081679	0.312	0.718	1.35E-12
Tnni1	6.92E-17	-1.111602854	0.047	0.347	1.40E-12
Ccdc12	6.98E-17	-0.571820028	0.732	0.976	1.41E-12
Hsp90aa1	9.51E-17	0.74149084	0.99	0.992	1.93E-12
Ndufb8	9.66E-17	-0.554622946	0.899	0.992	1.96E-12
Sparc	1.17E-16	1.323294474	0.96	0.984	2.37E-12
Hpgd	1.34E-16	-1.115048862	0.191	0.605	2.72E-12
Rpl12	1.38E-16	0.434945711	1	1	2.79E-12
Limd2	1.55E-16	-0.607947358	0.93	0.992	3.14E-12
mt-Nd2	1.72E-16	0.536413578	1	1	3.48E-12
Ndufv3	1.94E-16	-0.571672062	0.849	0.992	3.92E-12
Emp3	3.15E-16	-0.692328523	0.638	0.96	6.38E-12
Col13a1	3.63E-16	-0.272264519	0.013	0.258	7.35E-12
Rpl18	4.01E-16	-0.30022614	1	1	8.13E-12
Syngr1	4.20E-16	-0.51677889	0.101	0.46	8.50E-12
Ost4	5.25E-16	-0.554042178	0.872	0.976	1.06E-11
Card19	5.90E-16	-0.469329477	0.208	0.621	1.19E-11
Cldn5	6.72E-16	1.196050316	0.621	0.218	1.36E-11
Erp27	7.99E-16	-0.352212069	0.01	0.242	1.62E-11
Anxa6	8.03E-16	1.08122639	0.708	0.435	1.62E-11
Myh10	8.46E-16	1.052669806	0.681	0.379	1.71E-11
Ctla2a	9.62E-16	-1.504843907	0.919	0.992	1.95E-11
Rps20	1.09E-15	-0.305991642	1	1	2.20E-11
Dst	1.17E-15	0.85753442	0.607	0.226	2.36E-11

Rspo3	1.17E-15	0.893426133	0.52	0.113	2.37E-11
Tmem176a	1.40E-15	-0.659220153	0.426	0.823	2.84E-11
Tcf4	1.98E-15	1.016628277	0.812	0.742	4.01E-11
Flnb	2.30E-15	0.863660712	0.584	0.25	4.65E-11
Cox17	2.31E-15	-0.564811893	0.711	0.944	4.68E-11
Acaa1a	2.39E-15	-0.455478631	0.326	0.742	4.84E-11
Eif3j2	2.54E-15	0.679651167	0.789	0.621	5.15E-11
Ptgs1	3.67E-15	-0.367720093	0.221	0.694	7.43E-11
Fau	3.87E-15	-0.293494236	1	1	7.84E-11
Mxra7	4.01E-15	0.612924897	0.45	0.065	8.11E-11
Rpl29	4.34E-15	-0.516477974	0.842	0.992	8.79E-11
D030007L05Rik	4.34E-15	-0.589511986	0.238	0.645	8.79E-11
Rpl21	4.61E-15	-0.28231128	1	1	9.33E-11
Bzw1	8.07E-15	-0.479170619	0.943	0.992	1.63E-10
C1d	8.15E-15	-0.567976036	0.839	0.992	1.65E-10
Apln	9.89E-15	0.704496864	0.369	0	2.00E-10
Adam19	1.12E-14	0.655341782	0.436	0.056	2.27E-10
Kxd1	1.15E-14	-0.6323867	0.54	0.895	2.33E-10
Eif3f	1.17E-14	-0.475137389	0.966	1	2.36E-10
S100a11	1.38E-14	1.095628713	0.772	0.661	2.80E-10
Plscr1	1.57E-14	-0.440629893	0.188	0.605	3.18E-10
Txnrd1	1.68E-14	0.647197654	0.899	0.879	3.41E-10
Gm16104	1.72E-14	-0.38952524	0.195	0.605	3.47E-10
Grb14	2.30E-14	-0.253022255	0.023	0.266	4.65E-10
Tecr	2.40E-14	-0.4583513	0.933	0.992	4.85E-10
Edn1	2.59E-14	1.108000006	0.426	0.04	5.24E-10
Nes	2.66E-14	1.134730595	0.574	0.274	5.38E-10
Lilr4b	2.70E-14	-0.284917169	0.013	0.234	5.46E-10
Tmbim6	2.70E-14	-0.541133853	0.789	0.976	5.46E-10
Edf1	2.80E-14	-0.41901433	0.936	1	5.68E-10
Cdkn1a	3.09E-14	0.892490639	0.889	0.742	6.26E-10
Stmn1	3.30E-14	0.748574007	0.943	0.984	6.68E-10
Stat4	4.85E-14	-0.303614818	0.107	0.468	9.82E-10

Tpt1	4.92E-14	0.261251868	1	1	9.96E-10
Vkorc1	5.23E-14	-0.460846225	0.319	0.718	1.06E-09
Top2a	5.76E-14	0.930474375	0.862	0.629	1.17E-09
Crmp1	5.77E-14	0.619773408	0.456	0.089	1.17E-09
Rack1	7.06E-14	0.316055514	1	1	1.43E-09
Basp1	7.09E-14	1.010123045	0.768	0.685	1.44E-09
Susd3	7.13E-14	-0.277625668	0.057	0.347	1.44E-09
Rpl9	7.65E-14	-0.31119982	1	1	1.55E-09
Dlc1	1.06E-13	0.779739561	0.507	0.153	2.15E-09
mt-Atp8	1.10E-13	0.751409381	0.879	0.774	2.24E-09
Tceal9	1.20E-13	1.029216768	0.856	0.879	2.43E-09
Vamp5	1.39E-13	-0.617566105	0.668	0.96	2.82E-09
Mmrn2	1.71E-13	0.632058627	0.473	0.105	3.47E-09
Amotl1	2.22E-13	0.553557722	0.446	0.081	4.50E-09
Rap1a	3.13E-13	-0.679742158	0.735	0.944	6.33E-09
Taf7	3.21E-13	-0.561823527	0.433	0.782	6.50E-09
Pla2g4a	3.70E-13	-0.322172204	0.188	0.581	7.49E-09
Plek	3.78E-13	-0.940065306	0.289	0.694	7.66E-09
Ktn1	3.94E-13	0.582783827	0.93	0.919	7.97E-09
Actn4	4.59E-13	0.81615664	0.785	0.677	9.29E-09
Tubb2b	4.65E-13	0.76685925	0.567	0.266	9.42E-09
Rbp1	4.73E-13	-1.039774256	0.413	0.782	9.57E-09
Aprt	4.77E-13	0.843142611	0.916	0.879	9.66E-09
Unc119	5.54E-13	-0.454123334	0.235	0.621	1.12E-08
Ndufb7	5.68E-13	-0.35965712	0.98	1	1.15E-08
Kif2a	5.89E-13	-0.64135206	0.775	0.976	1.19E-08
Palld	6.02E-13	1.107478098	0.624	0.411	1.22E-08
Casp3	6.79E-13	0.693988347	0.779	0.669	1.37E-08
Etfb	7.30E-13	-0.509682628	0.765	0.944	1.48E-08
S100a6	7.43E-13	0.996615153	0.768	0.54	1.50E-08
Ndufs8	7.51E-13	-0.456677574	0.654	0.944	1.52E-08
Atp5j	8.67E-13	-0.330271994	0.997	1	1.75E-08
Atp5k	8.95E-13	0.47709572	0.993	0.984	1.81E-08
Ftl1	9.11E-13	0.712857661	0.735	0.573	1.84E-08

Pecam1	9.36E-13	0.865703548	0.812	0.847	1.89E-08
Cavin1	1.01E-12	0.759031583	0.634	0.363	2.04E-08
Vapa	1.05E-12	-0.487614577	0.779	0.976	2.13E-08
Eif3h	1.08E-12	-0.406179386	0.973	1	2.19E-08
Clec1a	1.12E-12	0.589547873	0.393	0.048	2.28E-08
Tpm1	1.13E-12	1.710660462	0.886	0.944	2.28E-08
Cd109	1.18E-12	0.513958553	0.372	0.032	2.39E-08
Tsc22d3	1.51E-12	-0.286586097	0.111	0.427	3.07E-08
Smagp	1.71E-12	-0.904729982	0.534	0.879	3.47E-08
Ankrd1	1.71E-12	2.089804102	0.383	0.048	3.47E-08
Mnd1	1.90E-12	-0.381848414	0.248	0.629	3.84E-08
Nop10	2.03E-12	-0.53992625	0.866	0.992	4.12E-08
Ier2	2.20E-12	-0.675452079	0.822	0.984	4.45E-08
Septin7	2.23E-12	-0.422585761	0.933	1	4.52E-08
Ptpre	2.29E-12	-0.521686624	0.299	0.685	4.64E-08
Ccdc88a	2.51E-12	0.639476698	0.809	0.718	5.07E-08
Gimap5	2.74E-12	-0.422536557	0.221	0.605	5.55E-08
H2az2	2.96E-12	-0.55508466	0.933	1	5.99E-08
Mta3	3.00E-12	-0.390120943	0.258	0.645	6.08E-08
H3c3	3.18E-12	0.838688689	0.594	0.29	6.44E-08
Acap1	3.37E-12	-0.290275712	0.114	0.435	6.82E-08
Cav1	3.96E-12	0.556301506	0.396	0.065	8.02E-08
Pmpcb	3.97E-12	-0.406421144	0.389	0.79	8.03E-08
Anp32a	4.14E-12	-0.49143798	0.886	0.992	8.38E-08
Homer3	4.44E-12	-0.473761719	0.607	0.887	8.99E-08
Smim26	4.54E-12	-0.402491324	0.322	0.71	9.20E-08
Adgrg3	4.73E-12	-0.269856286	0.144	0.492	9.58E-08
Nid1	4.80E-12	0.614773497	0.527	0.21	9.72E-08
Dpm3	5.38E-12	-0.46915347	0.725	0.944	1.09E-07
Perp	5.57E-12	-0.476939812	0.148	0.468	1.13E-07
Rps27rt	5.83E-12	-0.367321075	0.268	0.661	1.18E-07
Fn1	6.21E-12	0.870035415	0.836	0.718	1.26E-07

Cd93	6.56E-12	0.907886828	0.651	0.444	1.33E-07
Ebp	7.20E-12	-0.374314585	0.419	0.839	1.46E-07
Ssr4	8.02E-12	-0.439579544	0.903	0.992	1.62E-07
Sub1	8.62E-12	-0.413715384	0.96	1	1.75E-07
Ndufa13	9.60E-12	-0.348083236	0.99	1	1.94E-07
Nrg1	9.72E-12	-0.597302069	0.198	0.532	1.97E-07
Ndufb4	9.95E-12	-0.407038127	0.899	0.992	2.01E-07
Rab27b	1.27E-11	-0.298566462	0.138	0.476	2.57E-07
Cdh5	1.30E-11	0.862307804	0.836	0.839	2.63E-07
Sh3bp5	1.36E-11	0.874082	0.624	0.427	2.75E-07
Supt3	1.60E-11	-0.332726489	0.225	0.589	3.23E-07
Ahnak	1.65E-11	0.667597777	0.527	0.185	3.34E-07
Aplp2	1.79E-11	0.757802061	0.718	0.589	3.62E-07
Rpl36a-ps1	1.89E-11	-0.365944033	0.312	0.694	3.83E-07
Lasp1	1.96E-11	0.607820296	0.507	0.226	3.96E-07
Filip1l	2.38E-11	0.766273061	0.473	0.177	4.81E-07
Hnrnpa2b1	2.62E-11	-0.509138988	0.987	1	5.30E-07
Ddah1	2.63E-11	0.645615482	0.55	0.234	5.32E-07
Uqcrh	2.70E-11	-0.370197174	0.997	1	5.46E-07
Il11ra1	2.77E-11	-0.419125757	0.456	0.806	5.61E-07
Lpp	2.84E-11	0.821630596	0.604	0.395	5.75E-07
Skap2	2.88E-11	-0.3878592	0.369	0.766	5.83E-07
Vamp8	3.00E-11	-0.562701184	0.725	0.968	6.07E-07
Ccnd1	3.08E-11	1.099781749	0.658	0.452	6.23E-07
Epas1	3.25E-11	0.414872603	0.336	0.032	6.58E-07
Ndufb2	3.28E-11	-0.501724636	0.762	0.96	6.64E-07
Ipo11	3.45E-11	-0.557719944	0.735	0.944	6.99E-07
Ctsd	3.88E-11	-0.65481662	0.779	0.96	7.85E-07
Dpysl3	4.26E-11	0.865179434	0.621	0.427	8.62E-07
Wdr89	4.37E-11	-0.367472563	0.359	0.734	8.84E-07
S100a1	5.96E-11	-0.306074752	0.174	0.516	1.21E-06
Pdgfb	6.30E-11	0.803501046	0.55	0.298	1.28E-06

Arhgdib	6.49E-11	-0.530737182	0.869	0.984	1.31E-06
Rbis	7.01E-11	-0.419534578	0.805	0.976	1.42E-06
Mcam	7.18E-11	0.44288782	0.319	0.024	1.45E-06
Cuta	7.40E-11	-0.441184558	0.56	0.879	1.50E-06
Bex1	7.49E-11	1.150842198	0.765	0.718	1.52E-06
Tmem9	7.53E-11	-0.282763424	0.225	0.589	1.52E-06
Rps17	7.80E-11	0.305198907	1	1	1.58E-06
Mpp1	7.94E-11	-0.331845749	0.279	0.637	1.61E-06
Tle5	8.18E-11	-0.563157834	0.708	0.952	1.66E-06
Pcbp4	8.50E-11	0.464027667	0.48	0.177	1.72E-06
Dstn	8.84E-11	0.989282912	0.896	0.952	1.79E-06
Spcs1	8.93E-11	-0.431488218	0.732	0.935	1.81E-06
Sumo2	9.23E-11	-0.395224058	0.993	1	1.87E-06
St6galnac4	9.68E-11	-0.342822396	0.201	0.524	1.96E-06
Trmt10c	9.92E-11	-0.275478424	0.195	0.524	2.01E-06
Pcdh12	1.11E-10	0.356445818	0.336	0.04	2.25E-06
Anxa3	1.14E-10	0.846631736	0.638	0.508	2.30E-06
Nrros	1.18E-10	-0.338532225	0.211	0.556	2.40E-06
Itm2a	1.62E-10	-0.835851235	0.792	0.976	3.28E-06
Nudcd3	1.74E-10	0.525167406	0.862	0.871	3.52E-06
Psme2	2.00E-10	-0.460283732	0.711	0.919	4.04E-06
Jun	2.02E-10	-0.834780192	0.735	0.911	4.09E-06
Fermt3	2.31E-10	-0.271353939	0.191	0.548	4.67E-06
Bcl7c	2.32E-10	-0.378690259	0.47	0.831	4.70E-06
Car8	2.58E-10	-0.350114581	0.205	0.524	5.23E-06
Ctsl	2.65E-10	0.894106611	0.819	0.887	5.37E-06
Sarnp	2.73E-10	-0.429034915	0.836	0.976	5.52E-06
Ltc4s	2.83E-10	-0.719686326	0.56	0.871	5.72E-06
Isoc1	2.95E-10	-0.323696364	0.372	0.774	5.97E-06
Anxa2	3.25E-10	0.832306167	0.732	0.637	6.59E-06

Prc1	3.45E-10	0.470786926	0.718	0.363	6.99E-06
Rpl5	3.64E-10	0.305770607	0.997	1	7.37E-06
Tjp1	4.06E-10	0.656266076	0.57	0.315	8.22E-06
Tnfaip1	4.64E-10	0.710693973	0.601	0.403	9.39E-06
Ddx39b	4.66E-10	-0.489908264	0.782	0.952	9.43E-06
Prdx6	4.72E-10	0.555782058	0.923	0.984	9.55E-06
Ebf1	4.73E-10	-0.303518985	0.117	0.379	9.58E-06
Psmb1	4.76E-10	-0.308047018	0.98	1	9.64E-06
Icam2	4.79E-10	0.89900248	0.678	0.492	9.69E-06
Arl4a	4.94E-10	-0.433895021	0.319	0.669	1.00E-05
Tinagl1	4.97E-10	0.501779073	0.419	0.121	1.01E-05
Serhl	5.10E-10	-0.255317938	0.144	0.444	1.03E-05
Nisch	5.38E-10	0.494376769	0.842	0.839	1.09E-05
Cd55	5.63E-10	-0.309579569	0.114	0.387	1.14E-05
Runx1	5.67E-10	-0.323398741	0.312	0.726	1.15E-05
Nme2	5.73E-10	-0.350169712	0.312	0.653	1.16E-05
Itm2b	5.82E-10	-0.555583158	0.943	1	1.18E-05
S100a16	6.02E-10	0.535362318	0.346	0.048	1.22E-05
Dbn1	6.13E-10	0.63202429	0.466	0.194	1.24E-05
Rhoa	6.15E-10	-0.375561979	0.879	1	1.24E-05
Cox7a2	6.49E-10	-0.345444646	0.97	1	1.31E-05
Atf3	7.15E-10	0.428973827	0.379	0.073	1.45E-05
Rock2	7.64E-10	0.663006019	0.758	0.677	1.55E-05
B930036N10Rik	7.92E-10	-0.355890444	0.232	0.556	1.60E-05
Blvrb	7.98E-10	-0.408559984	0.383	0.742	1.62E-05
Atp5e	8.07E-10	-0.273746303	0.993	1	1.63E-05
Btf3	9.04E-10	-0.283686633	0.993	1	1.83E-05
Odc1	9.11E-10	0.734425002	0.809	0.782	1.84E-05
Rpl7a	9.42E-10	-0.259343046	1	1	1.91E-05
Brd7	1.00E-09	-0.410534102	0.832	0.976	2.03E-05
Dlk1	1.05E-09	-0.51766831	0.084	0.323	2.12E-05
Macf1	1.06E-09	0.616684683	0.812	0.774	2.15E-05

Ndufb10	1.07E-09	-0.433015367	0.846	0.96	2.16E-05
Clic1	1.10E-09	-0.365678233	0.983	1	2.24E-05
Rps2	1.21E-09	-0.271115396	1	1	2.45E-05
Mien1	1.21E-09	-0.381887302	0.554	0.879	2.46E-05
Pyurf	1.24E-09	-0.293190748	0.282	0.645	2.50E-05
Cep89	1.30E-09	-0.260598861	0.144	0.435	2.64E-05
Eif4ebp1	1.33E-09	0.500071881	0.899	0.96	2.70E-05
2200002D01Rik	1.36E-09	0.453948671	0.339	0.065	2.75E-05
Ndufs7	1.36E-09	-0.354388577	0.846	0.976	2.76E-05
Idi1	1.43E-09	-0.515258226	0.668	0.887	2.90E-05
Ddb1	1.53E-09	0.548671548	0.782	0.685	3.09E-05
Rhoc	1.55E-09	1.002871243	0.742	0.766	3.15E-05
Rnf187	1.57E-09	-0.413232352	0.775	0.968	3.17E-05
Rnaseh2c	1.58E-09	-0.402472535	0.738	0.96	3.20E-05
H1f4	1.69E-09	0.885713668	0.795	0.75	3.42E-05
Kdelr1	1.74E-09	-0.419884175	0.638	0.919	3.51E-05
Tbc1d15	1.77E-09	-0.378456966	0.487	0.815	3.58E-05
Cavin3	1.83E-09	0.982402612	0.638	0.492	3.71E-05
Arhgef12	1.89E-09	0.55897479	0.46	0.177	3.83E-05
Tead1	1.93E-09	0.408845994	0.332	0.056	3.90E-05
Cyren	2.02E-09	-0.319383484	0.309	0.645	4.09E-05
Ftl1-ps1	2.13E-09	0.357537345	1	1	4.31E-05
Arhgef15	2.25E-09	0.412666258	0.346	0.073	4.56E-05
Itga2b	2.62E-09	-0.36888257	0.265	0.621	5.30E-05
Sgf29	2.62E-09	-0.281786836	0.248	0.581	5.31E-05
Myl12b	2.74E-09	-0.358531749	0.386	0.726	5.54E-05
Gatd1	3.07E-09	-0.259667628	0.185	0.484	6.21E-05
G0s2	3.25E-09	-0.33037851	0.017	0.169	6.58E-05
H2ac18	3.36E-09	0.429687762	0.315	0.056	6.80E-05
Snrpf	3.41E-09	-0.391986612	0.94	0.984	6.90E-05
Smim20	3.45E-09	-0.255856434	0.282	0.629	6.98E-05

Arid5b	3.63E-09	0.612295144	0.483	0.218	7.35E-05
Gmps	3.93E-09	-0.357017808	0.758	0.96	7.96E-05
Ndufc2	4.28E-09	-0.3646847	0.893	1	8.66E-05
Mki67	4.48E-09	0.644099525	0.748	0.508	9.08E-05
D8Ertd738e	4.50E-09	-0.330688559	0.946	0.992	9.12E-05
Shroom2	4.51E-09	0.438906376	0.372	0.105	9.14E-05
Arhgap31	4.63E-09	0.49796436	0.604	0.419	9.37E-05
Naxe	4.63E-09	-0.363910001	0.51	0.806	9.38E-05
Klf6	4.68E-09	0.576393079	0.735	0.605	9.47E-05
Ccdc107	4.86E-09	-0.31052012	0.309	0.653	9.83E-05
St13	4.89E-09	0.439452554	0.956	0.968	9.89E-05
Ece1	4.95E-09	0.521728366	0.487	0.202	0.000100233
Erg28	5.05E-09	-0.365082133	0.587	0.895	0.000102156
Anp32b	5.05E-09	-0.389438006	0.983	1	0.000102207
Tax1bp3	5.16E-09	0.769813324	0.691	0.653	0.000104358
Tm2d2	5.29E-09	-0.374371078	0.617	0.944	0.00010717
Limch1	5.40E-09	0.544109139	0.493	0.21	0.00010932
Anxa5	5.80E-09	0.832785865	0.701	0.677	0.000117424
Rrbp1	5.91E-09	0.481903005	0.926	0.96	0.000119737
Sra1	6.25E-09	-0.395601516	0.544	0.863	0.000126465
Ednrb	6.55E-09	-0.287083029	0.148	0.419	0.000132595
Fam32a	6.60E-09	-0.361279423	0.772	0.952	0.000133647
Svil	6.62E-09	-0.357629375	0.393	0.742	0.000134085
Igfbp2	6.70E-09	0.573725641	0.409	0.153	0.000135566
Lsm4	6.83E-09	-0.36714167	0.916	0.992	0.000138332
Eef1b2	6.88E-09	0.25386553	1	1	0.000139254
Dab2	7.76E-09	0.424159638	0.305	0.048	0.000157176
Mmp2	8.23E-09	0.48679632	0.43	0.169	0.000166566
Rpl4	8.72E-09	0.273364155	1	1	0.000176449
Flt1	9.26E-09	0.506204912	0.507	0.242	0.000187361
BC004004	9.28E-09	-0.2789558	0.362	0.71	0.000187887
Mrpl33	9.53E-09	-0.352839651	0.829	0.976	0.000192827
B9d2	9.81E-09	-0.26366085	0.198	0.5	0.000198658
Nusap1	9.84E-09	0.457673505	0.685	0.331	0.000199241
Spc25	1.01E-08	0.533881262	0.681	0.427	0.000204918

Psmb5	1.04E-08	-0.313167518	0.946	0.992	0.000210692
Pigx	1.09E-08	-0.311975547	0.403	0.766	0.000220891
Emcn	1.10E-08	0.697539266	0.554	0.29	0.000222646
Ccn2	1.15E-08	0.56754182	0.396	0.121	0.00023261
Rtraf	1.24E-08	-0.339714672	0.94	1	0.000251586
Nectin2	1.28E-08	0.489641931	0.446	0.177	0.000258938
Psmb10	1.28E-08	-0.344015545	0.513	0.823	0.000259954
Tmsb4x	1.29E-08	0.439388359	1	1	0.000260456
Ssbp4	1.35E-08	-0.377557948	0.721	0.968	0.000273512
Rpl23a	1.36E-08	0.33261913	1	1	0.000275562
Hnrnpd	1.56E-08	-0.383024333	0.866	0.992	0.000314794
Fxyd6	1.62E-08	0.444665516	0.366	0.113	0.000327891
Ddx39	1.65E-08	-0.442077441	0.685	0.919	0.000334762
Cdc26	1.73E-08	-0.309585174	0.517	0.903	0.000350994
Senp7	1.83E-08	-0.328107663	0.295	0.597	0.000370469
Lama4	1.83E-08	0.442176223	0.393	0.137	0.00037143
Cited2	1.85E-08	-0.616059318	0.597	0.855	0.000373682
Serpinb1a	1.87E-08	-0.288603649	0.037	0.21	0.000378539
Sox11	1.90E-08	0.521498485	0.376	0.113	0.000384592
Sorbs2	1.90E-08	0.347599529	0.218	0	0.000384934
Upp1	1.93E-08	0.529195031	0.5	0.242	0.000390629
Clic4	2.03E-08	0.650596308	0.56	0.387	0.000410641
Pcbd2	2.10E-08	-0.310187328	0.399	0.734	0.000425499
Ciao2a	2.16E-08	-0.303280114	0.661	0.944	0.000437182
Gucy1a1	2.18E-08	-0.263729989	0.037	0.21	0.000441924
Timp1	2.23E-08	0.324862117	0.272	0.032	0.000452115
Crlf3	2.25E-08	-0.299674975	0.628	0.911	0.000455694
Afdn	2.26E-08	0.546574464	0.628	0.452	0.000457437
Ralb	2.31E-08	0.551337958	0.752	0.774	0.000467066
Tcim	2.39E-08	0.502239495	0.416	0.137	0.000483364
Esm1	2.43E-08	0.476665407	0.228	0.008	0.000492806
Phgdh	2.47E-08	-0.334368027	0.383	0.734	0.000500721

Rgs16	2.49E-08	0.315677965	0.232	0.008	0.000504213
Rhob	2.54E-08	0.745583254	0.695	0.645	0.000514617
Dok4	2.58E-08	-0.432202228	0.245	0.524	0.000523186
Sgo2a	2.72E-08	0.493707439	0.564	0.306	0.000550066
Ppp2r5a	2.74E-08	-0.352046916	0.416	0.742	0.00055386
Sgk1	2.78E-08	0.400549299	0.43	0.145	0.000561903
Psma2	2.81E-08	-0.300578526	0.913	0.992	0.000569271
Tuba1a	2.86E-08	0.905751988	0.792	0.855	0.00057853
Klk8	2.92E-08	-0.304864005	0.379	0.71	0.000591567
Atp5pb	2.95E-08	-0.309950094	0.973	1	0.000597944
Dtd2	2.96E-08	-0.26198412	0.326	0.677	0.000599332
Pik3ip1	3.04E-08	-0.30589964	0.164	0.419	0.000614385
Eif5	3.04E-08	0.380195669	0.96	0.984	0.000615883
Dctn6	3.05E-08	-0.387271634	0.577	0.847	0.000617679
Atp6v1f	3.06E-08	-0.330171822	0.859	0.992	0.000619814
Nrp1	3.06E-08	0.668243878	0.57	0.363	0.000620356
Gatad1	3.19E-08	-0.388959865	0.705	0.911	0.000644986
Bcar1	3.38E-08	0.38803923	0.409	0.145	0.00068388
Atp5b	3.41E-08	0.327537701	1	1	0.000691242
Acaa2	3.49E-08	-0.282754183	0.399	0.758	0.000706312
Ifitm2	3.52E-08	-0.35902073	0.96	1	0.000711965
Mrps35	3.53E-08	-0.291336894	0.399	0.726	0.0007142
Hnrnpdl	3.93E-08	-0.399306698	0.685	0.927	0.000795548
Rasgrp3	4.09E-08	0.577008995	0.527	0.298	0.000828925
Ndufa3	4.11E-08	-0.333998647	0.899	0.976	0.000832915
Psmg4	4.65E-08	-0.33048652	0.517	0.798	0.000942186
Mettl9	4.87E-08	-0.358195207	0.604	0.911	0.000985003
Tmem128	4.96E-08	-0.300642422	0.406	0.766	0.001004387
Elf1	4.97E-08	-0.314388871	0.376	0.702	0.001005264
Tma7	4.97E-08	-0.279754726	0.956	1	0.001007033
Sox17	5.00E-08	0.572232475	0.56	0.274	0.001011953
Smim10l1	5.21E-08	-0.283358026	0.396	0.71	0.001055414

Eif3k	5.44E-08	-0.320720947	0.96	1	0.001100427
Zfp703	5.59E-08	0.285084332	0.245	0.024	0.00113081
Sp3os	5.59E-08	-0.281626742	0.315	0.613	0.001131121
Klf2	5.74E-08	-0.989683856	0.57	0.839	0.001161617
Pdzd2	5.85E-08	0.267183759	0.218	0.008	0.001183304
Rrs1	5.93E-08	-0.262870261	0.295	0.621	0.0012003
Scamp2	6.42E-08	-0.293121852	0.443	0.798	0.001300422
Dnajc19	6.57E-08	-0.315971228	0.668	0.903	0.001330321
Mrpl32	6.61E-08	-0.317719163	0.601	0.887	0.001337495
Fam174a	6.73E-08	-0.313418292	0.342	0.645	0.001362865
Epn2	6.75E-08	0.455838353	0.45	0.218	0.001366856
Eif1b	6.85E-08	-0.327842751	0.732	0.927	0.00138611
Necap2	6.87E-08	-0.271350944	0.332	0.653	0.001390197
Nid2	7.13E-08	0.369710946	0.322	0.081	0.001443795
Lrpap1	7.16E-08	-0.288326078	0.423	0.766	0.001450304
S100a13	7.41E-08	-0.38257737	0.742	0.968	0.001499955
Eci1	7.68E-08	-0.333542247	0.419	0.718	0.001554183
Pdlim5	7.90E-08	0.629197436	0.487	0.282	0.001599958
Naa10	7.93E-08	-0.345955717	0.577	0.774	0.001605728
Pmp22	8.09E-08	-0.550472332	0.262	0.54	0.001637977
Hmgn1	8.17E-08	0.367069956	0.97	0.992	0.001654767
Tmem147	8.36E-08	-0.377901554	0.668	0.895	0.001692132
Cotl1	8.40E-08	0.571855943	0.775	0.718	0.001701273
Ppp1r15a	8.48E-08	-0.4085898	0.215	0.508	0.001717251
Ube2e1	9.01E-08	-0.319418691	0.493	0.79	0.001824472
Ndufa7	9.44E-08	-0.344787187	0.936	0.992	0.001910841
Itga1	9.79E-08	0.487427336	0.466	0.242	0.001982499
Dtnbp1	9.91E-08	-0.349714156	0.477	0.766	0.002005472
Atxn7l3b	1.00E-07	-0.376685813	0.789	0.952	0.002025008
Itga5	1.02E-07	0.341573071	0.336	0.097	0.002070332
Prdx3	1.04E-07	-	0.51	0.815	0.002100378

		0.311629203			
Tanc1	1.05E-07	0.366153852	0.379	0.137	0.002126839
Plekha3	1.10E-07	-0.275245505	0.309	0.629	0.002218693
Aurkb	1.15E-07	0.467743891	0.621	0.331	0.002318946
Arpc4	1.18E-07	-0.300374443	0.752	0.976	0.002383708
Ckap2	1.23E-07	0.64704188	0.644	0.484	0.002491817
Gkap1	1.24E-07	-0.37013967	0.523	0.855	0.002502151
Lonp2	1.24E-07	-0.283513296	0.416	0.774	0.002504704
Mtdh	1.26E-07	-0.397293714	0.926	1	0.00254252
Tnni2	1.26E-07	-0.362140718	0.044	0.21	0.002542946
Mrps28	1.27E-07	-0.343935513	0.517	0.758	0.002570135
Haus1	1.27E-07	-0.324619932	0.463	0.758	0.002575721
Mrpl36	1.33E-07	-0.328905384	0.631	0.887	0.002700211
Psmb6	1.34E-07	-0.276186665	0.977	1	0.00271093
Usp3	1.34E-07	-0.319284564	0.349	0.645	0.002718604
Igfbp7	1.42E-07	0.427369767	0.238	0.024	0.002864467
Txnip	1.44E-07	-0.415523036	0.547	0.815	0.002924166
Mrps24	1.64E-07	-0.263103555	0.936	0.992	0.003321937
Fmo1	1.71E-07	0.31483814	0.235	0.024	0.003466858
Tspan14	1.75E-07	-0.250662418	0.326	0.629	0.003534678
Myo1b	1.76E-07	0.437699564	0.456	0.226	0.003569219
Cnn2	1.82E-07	0.651978852	0.708	0.629	0.003693654
Mrps21	1.89E-07	-0.301164113	0.819	0.976	0.003819565
Ccng1	1.90E-07	0.507464038	0.745	0.734	0.003844684
Sfr1	1.93E-07	-0.320730695	0.94	1	0.003903409
Gjc1	1.94E-07	0.359242988	0.332	0.105	0.003931698
Rap1b	1.96E-07	-0.456160336	0.846	0.968	0.003964491
Pls3	2.02E-07	0.443741561	0.44	0.226	0.004083996
Vdac1	2.02E-07	-0.284179853	0.993	1	0.004096306
Sdhb	2.03E-07	-0.3511122	0.772	0.919	0.004110774
Lamtor2	2.14E-07	-0.328746783	0.826	0.968	0.004336029

Rell1	2.15E-07	0.418988917	0.393	0.161	0.004349512
Gabarapl2	2.23E-07	-0.404092085	0.809	0.944	0.004504313
Eif3c	2.23E-07	-0.259938768	0.993	1	0.004521613
H2aj	2.24E-07	-0.362263378	0.926	1	0.004533991
Plpp2	2.26E-07	0.342149699	0.322	0.097	0.004574913
Arl2	2.27E-07	-0.263576035	0.319	0.621	0.004601189
Ubtf	2.27E-07	-0.369910622	0.597	0.855	0.004603137
Tubb5	2.30E-07	0.396884483	0.993	1	0.004650765
Cd47	2.31E-07	-0.40572836	0.765	0.919	0.004669527
Mdk	2.42E-07	0.490325422	0.671	0.468	0.004907122
Tnc	2.48E-07	0.26252344	0.188	0	0.005024209
Cd9	2.54E-07	-0.453982966	0.399	0.653	0.005141591
Septin2	2.60E-07	0.386567404	0.372	0.137	0.005265649
Igfbp4	2.61E-07	-0.614239995	0.822	0.927	0.00528173
Robo4	2.64E-07	0.427176642	0.366	0.129	0.005344592
Nfe2l1	2.65E-07	0.446062479	0.775	0.774	0.00536186
Uqcrfs1	2.70E-07	-0.330214736	0.812	0.944	0.005458546
Atic	2.71E-07	-0.345697335	0.597	0.847	0.005482773
Myl9	2.80E-07	0.57222965	0.305	0.081	0.005663661
Ndufv2	2.87E-07	-0.333944275	0.852	0.976	0.005819197
Usp29	3.05E-07	0.284422275	0.255	0.04	0.006171311
Lamtor4	3.06E-07	-0.370381542	0.688	0.919	0.006193771
Ckb	3.16E-07	-0.898397703	0.336	0.589	0.006390469
Gpihbp1	3.28E-07	0.250139666	0.185	0	0.006644879
Cenpm	3.43E-07	-0.274572296	0.295	0.573	0.006945604
Psenen	3.46E-07	-0.331809615	0.725	0.935	0.006998714
Myh9	3.56E-07	0.627531792	0.725	0.718	0.007196607
Nrk	3.67E-07	-0.274596311	0.077	0.258	0.007422714
Mrps15	3.67E-07	-0.297648196	0.748	0.911	0.00743788
Vps28	3.71E-07	-0.325631138	0.644	0.903	0.007518868
Rbmxl1	3.74E-07	-0.318222526	0.654	0.895	0.007578873

Rrm2	3.85E-07	0.43746675	0.695	0.476	0.007803048
Kitl	3.93E-07	0.32571819	0.211	0.016	0.007946502
Maz	4.13E-07	-0.259579016	0.413	0.75	0.008351448
Efnb2	4.16E-07	0.359020546	0.352	0.113	0.008421157
Samd4	4.29E-07	0.264717093	0.221	0.024	0.008693488
Hopx	4.33E-07	-0.251227534	0.034	0.177	0.008760272
Eml4	4.50E-07	-0.361066835	0.792	0.911	0.00911428
Nrep	4.60E-07	0.385759474	0.477	0.242	0.009320738
Psmb2	4.77E-07	-0.294374119	0.923	0.992	0.009646589
Zwint	4.77E-07	0.430477248	0.819	0.855	0.009657285
Lsm7	4.91E-07	-0.320152738	0.792	0.952	0.009937994
Micos13	4.94E-07	-0.324041071	0.896	0.992	0.009995785
Acads	4.97E-07	-0.265499195	0.463	0.79	0.010069892
Commd7	5.02E-07	-0.314452838	0.567	0.863	0.010168044
Pclaf	5.03E-07	0.490968087	0.846	0.71	0.010188239
Kdr	5.10E-07	0.530438189	0.762	0.766	0.010319863
Sec13	5.10E-07	-0.339757584	0.564	0.823	0.010331043
Rcbtb2	5.11E-07	-0.341659244	0.604	0.879	0.010335586
Egfl7	5.15E-07	-0.80227797	0.909	0.992	0.010434754
Eef1d	5.28E-07	-0.276895193	0.956	1	0.010687672
Zfp422	5.29E-07	-0.297612874	0.406	0.726	0.010706637
Pdcd4	5.36E-07	-0.347198098	0.436	0.726	0.010846081
Trappc5	5.38E-07	-0.264738871	0.698	0.944	0.010896442
Tgfb1	5.40E-07	-0.352279663	0.701	0.919	0.010936125
Cbx3	5.46E-07	-0.334110148	0.913	0.984	0.011046493
Gadd45gip1	5.55E-07	-0.319075007	0.624	0.855	0.011234197
Ethe1	5.58E-07	0.560303071	0.443	0.258	0.011293641
Dctn3	5.60E-07	-0.305847112	0.772	0.944	0.011341038
Nexn	5.72E-07	0.443511819	0.255	0.048	0.011583342
Ccn1	5.82E-07	0.50687792	0.279	0.065	0.011781823
Ppp1r14a	5.87E-07	0.291315873	0.258	0.048	0.011887833

Lims1	6.17E-07	-0.414140225	0.55	0.831	0.012491683
Mrpl24	6.30E-07	-0.346539569	0.544	0.823	0.012752131
Ndufs5	6.53E-07	-0.289195261	0.943	1	0.013223954
Amotl2	6.72E-07	0.312673273	0.319	0.097	0.01359998
Ap2b1	6.73E-07	0.396528017	0.705	0.613	0.013616389
Emilin1	6.88E-07	-0.277139139	0.366	0.694	0.013935465
H2az1	6.96E-07	-0.472317563	0.997	1	0.014099156
Snrpa	7.04E-07	-0.277276824	0.654	0.911	0.014254224
Arhgef28	7.14E-07	0.396276268	0.409	0.185	0.014459901
Srsf9	7.25E-07	-0.296031968	0.812	0.992	0.014685121
Rgs10	7.42E-07	-0.424999855	0.151	0.379	0.015016197
Zmat5	7.51E-07	-0.285213122	0.487	0.774	0.015197686
Isyna1	7.64E-07	-0.319231742	0.396	0.694	0.015472385
Erh	7.72E-07	0.468947139	0.735	0.621	0.01563296
Serpine1	7.74E-07	0.435211057	0.205	0.016	0.015673274
Ift57	7.78E-07	-0.354982046	0.332	0.597	0.015747029
Plac1	7.84E-07	0.302414094	0.252	0.048	0.015864021
Ndufa2	7.90E-07	-0.294291697	0.963	1	0.015989755
Itgb3	7.94E-07	-0.298197391	0.215	0.492	0.016070953
Cd81	8.08E-07	-0.356143927	0.966	0.992	0.016352828
Sat1	8.34E-07	0.552334349	0.691	0.605	0.016887964
Scand1	8.64E-07	-0.263044561	0.96	1	0.017489
Cenpe	8.71E-07	0.444111827	0.691	0.427	0.017640034
Pxdn	9.03E-07	0.363471839	0.386	0.169	0.01828851
Ehd2	9.38E-07	0.320034243	0.346	0.121	0.01899527
Dock4	9.43E-07	0.385416347	0.359	0.145	0.019086081
Vav3	1.01E-06	-0.304929513	0.362	0.621	0.020420247
Ldha	1.03E-06	-0.363255325	0.96	0.992	0.0209048
Lsm6	1.05E-06	-0.388239164	0.852	0.96	0.021253409
Tnfaip2	1.05E-06	0.545916918	0.517	0.339	0.021317401
Cnrip1	1.06E-06	-0.553097004	0.413	0.653	0.021413549

Pttg1ip	1.07E-06	-0.270483368	0.346	0.645	0.021562101
Knop1	1.07E-06	0.393703921	0.933	0.895	0.02173382
Fkbp2	1.09E-06	-0.301925893	0.611	0.887	0.021966149
Tbca	1.09E-06	-0.29254041	0.903	1	0.022067127
Rnh1	1.10E-06	-0.308088613	0.54	0.806	0.022292805
Smim11	1.10E-06	-0.282640213	0.591	0.863	0.022327384
Fkbp8	1.14E-06	-0.352037902	0.795	0.976	0.023038032
Fdx1	1.15E-06	-0.310463112	0.446	0.75	0.023269508
Ushbp1	1.16E-06	0.251621387	0.242	0.04	0.023475608
Osgep	1.19E-06	-0.255340232	0.473	0.782	0.024188228
Nenf	1.21E-06	-0.282570604	0.641	0.911	0.024489253
Ifitm3	1.22E-06	0.527058237	0.732	0.605	0.024676459
Gpx8	1.22E-06	0.390932072	0.362	0.153	0.02470996
Rnf7	1.23E-06	-0.36259475	0.772	0.952	0.024865039
Ptpn14	1.29E-06	0.269206639	0.265	0.056	0.026178122
Crim1	1.30E-06	0.425879281	0.372	0.169	0.026411811
Mrps16	1.31E-06	-0.263436239	0.789	0.976	0.026610547
Snrpd1	1.32E-06	-0.323336259	0.956	0.992	0.026724355
Pdia3	1.38E-06	-0.304809997	0.953	1	0.02786579
Septin11	1.45E-06	0.482664021	0.775	0.774	0.029329084
Mrps30	1.50E-06	-0.264888343	0.446	0.742	0.030329847
Atp5g1	1.52E-06	-0.271775988	0.987	1	0.030839996
Smad1	1.53E-06	0.518457339	0.557	0.419	0.030950315
Mrps18c	1.57E-06	-0.290543409	0.735	0.935	0.031791829
H4c4	1.58E-06	0.634471883	0.755	0.766	0.031894562
Gm16286	1.59E-06	-0.311673143	0.685	0.895	0.03224556
Cr1l	1.61E-06	-0.279095752	0.443	0.718	0.032612088
Itgb1	1.64E-06	0.437946776	0.956	0.984	0.033204115
Specc1	1.64E-06	0.468117907	0.574	0.427	0.033265352
Dad1	1.72E-06	-0.357848102	0.903	0.984	0.034843709
Dync1h1	1.80E-06	0.470470444	0.681	0.621	0.036357218
Slc25a11	1.87E-06	-	0.705	0.935	0.037760362

		0.292090869			
Cdk2ap1	1.90E-06	-0.275269864	0.426	0.726	0.038510558
Supt4a	1.94E-06	-0.30343057	0.899	0.976	0.039320888
Fundc2	1.95E-06	-0.299923056	0.772	0.944	0.039498107
Iqgap1	1.98E-06	0.414047267	0.711	0.573	0.040086504
Sema6a	2.03E-06	0.262996017	0.282	0.073	0.041121214
Lcp1	2.04E-06	0.440920881	0.443	0.218	0.041206784
Fmc1	2.11E-06	-0.309599774	0.574	0.815	0.042669898
Cyp51	2.20E-06	-0.377231053	0.581	0.823	0.044559401
Fbxo6	2.26E-06	-0.264732707	0.315	0.581	0.045729985
Prr13	2.28E-06	-0.350658084	0.607	0.847	0.046096914
Cebpd	2.30E-06	-0.253440064	0.349	0.645	0.046484632
Snrpd3	2.32E-06	-0.299822639	0.926	1	0.046934091
Cmas	2.36E-06	-0.409684539	0.822	0.935	0.047799894

7 REFERENCES

ADELMAN, C. A., CHATTOPADHYAY, S. & BIEKER, J. J. 2002. The BMP/BMPR/Smad pathway directs expression of the erythroid-specific EKLF and GATA1 transcription factors during embryoid body differentiation in serum-free media. *Development,* 129**,** 539-49.

AKASHI, K., TRAVER, D., MIYAMOTO, T. & WEISSMAN, I. L. 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature,* 404**,** 193-7.

ALBRIGHT, S. C., WISEMAN, J. M., LANGE, R. A. & GARRARD, W. T. 1980. Subunit structures of different electrophoretic forms of nucleosomes. *J Biol Chem,* 255**,** 3673-84.

ALEXANDER, R. D., INNOCENTE, S. A., BARRASS, J. D. & BEGGS, J. D. 2010. Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell,* 40**,** 582-93.

ALFRED, R., GAREAU, T., KRAWETZ, R., RANCOURT, D. & KALLOS, M. S. 2010. Serum-free scaled up expansion and differentiation of murine embryonic stem cells to osteoblasts in suspension bioreactors. *Biotechnol Bioeng,* 106**,** 829-40.

ALLAN, J., HARBORNE, N., RAU, D. C. & GOULD, H. 1982. Participation of core histone "tails" in the stabilization of the chromatin solenoid. *J Cell Biol,* 93**,** 285-97.

ALLFREY, V. G., FAULKNER, R. & MIRSKY, A. E. 1964. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A,* 51**,** 786-94.

AMEMIYA, H. M., KUNDAJE, A. & BOYLE, A. P. 2019. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep,* 9**,** 9354.

ANDERSON, H., PATCH, T. C., REDDY, P. N., HAGEDORN, E. J., KIM, P. G., SOLTIS, K. A., CHEN, M. J., TAMPLIN, O. J., FRYE, M., MACLEAN, G. A., HUBNER, K., BAUER, D. E., KANKI, J. P., VOGIN, G., HUSTON, N. C., NGUYEN, M., FUJIWARA, Y., PAW, B. H., VESTWEBER, D., ZON, L. I., ORKIN, S. H., DALEY, G. Q. & SHAH, D. I. 2015. Hematopoietic stem cells develop in the absence of endothelial cadherin 5 expression. *Blood,* 126**,** 2811-20.

ANDERSSON, R., GEBHARD, C., MIGUEL-ESCALADA, I., HOOF, I., BORNHOLDT, J., BOYD, M., CHEN, Y., ZHAO, X., SCHMIDL, C., SUZUKI, T., NTINI, E., ARNER, E., VALEN, E., LI, K., SCHWARZFISCHER, L., GLATZ, D., RAITHEL, J., LILJE, B., RAPIN, N., BAGGER, F. O., JORGENSEN, M., ANDERSEN, P. R., BERTIN, N., RACKHAM, O., BURROUGHS, A. M., BAILLIE, J. K., ISHIZU, Y., SHIMIZU, Y., FURUHATA, E., MAEDA, S., NEGISHI, Y., MUNGALL, C. J., MEEHAN, T. F., LASSMANN, T., ITOH, M., KAWAJI, H., KONDO, N., KAWAI, J., LENNARTSSON, A., DAUB, C. O., HEUTINK, P., HUME, D. A., JENSEN, T. H., SUZUKI, H., HAYASHIZAKI, Y., MULLER, F., FORREST, A. R. R., CARNINCI, P., REHLI, M. & SANDELIN, A. 2014. An atlas of active enhancers across human cell types and tissues. *Nature,* 507**,** 455-461.

ANG, S. L., WIERDA, A., WONG, D., STEVENS, K. A., CASCIO, S., ROSSANT, J. & ZARET, K. S. 1993. The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development,* 119**,** 1301-15.

ARGELAGUET, R., CLARK, S. J., MOHAMMED, H., STAPEL, L. C., KRUEGER, C., KAPOURANI, C. A., IMAZ-ROSSHANDLER, I., LOHOFF, T., XIANG, Y., HANNA, C. W., SMALLWOOD, S., IBARRA-SORIA, X., BUETTNER, F., SANGUINETTI, G., XIE, W., KRUEGER, F., GOTTGENS, B., RUGG-GUNN, P. J., KELSEY, G., DEAN, W., NICHOLS, J., STEGLE, O., MARIONI, J. C. & REIK, W. 2019. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature,* 576**,** 487-491.

ARMESILLA, A. L., LORENZO, E., GOMEZ DEL ARCO, P., MARTINEZ-MARTINEZ, S., ALFRANCA, A. & REDONDO, J. M. 1999. Vascular endothelial growth factor activates nuclear factor of activated T cells in human endothelial cells: a role for tissue factor gene expression. *Mol Cell Biol,* 19**,** 2032-43.

ARNOLD, C. D., GERLACH, D., STELZER, C., BORYN, L. M., RATH, M. & STARK, A. 2013. Genome-wide

quantitative enhancer activity maps identified by STARR-seq. *Science,* 339**,** 1074-7.
AZZONI, E., FRONTERA, V., MCGRATH, K. E., HARMAN, J., CARRELHA, J., NERLOV, C., PALIS, J., JACOBSEN, S. E. W. & DE BRUIJN, M. F. 2018. Kit ligand has a critical role in mouse yolk sac and aorta-gonad-mesonephros hematopoiesis. *EMBO Rep,* 19.
BAKAYEV, V. V., BAKAYEVA, T. G. & VARSHAVSKY, A. J. 1977. Nucleosomes and subnucleosomes: heterogeneity and composition. *Cell,* 11**,** 619-29.
BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell,* 27**,** 299-308.
BANNISTER, A. J. & KOUZARIDES, T. 1996. The CBP co-activator is a histone acetyltransferase. *Nature,* 384**,** 641-3.
BANNISTER, A. J. & KOUZARIDES, T. 2011. Regulation of chromatin by histone modifications. *Cell Res,* 21**,** 381-95.
BARDOT, E. S. & HADJANTONAKIS, A. K. 2020. Mouse gastrulation: Coordination of tissue patterning, specification and diversification of cell fate. *Mech Dev,* 163**,** 103617.
BEE, T., ASHLEY, E. L., BICKLEY, S. R., JARRATT, A., LI, P. S., SLOANE-STANLEY, J., GOTTGENS, B. & DE BRUIJN, M. F. 2009. The mouse Runx1 +23 hematopoietic stem cell enhancer confers hematopoietic specificity to both Runx1 promoters. *Blood,* 113**,** 5121-4.
BEJERANO, G., PHEASANT, M., MAKUNIN, I., STEPHEN, S., KENT, W. J., MATTICK, J. S. & HAUSSLER, D. 2004. Ultraconserved elements in the human genome. *Science,* 304**,** 1321-5.
BELL, A. C., WEST, A. G. & FELSENFELD, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell,* 98**,** 387-96.
BERNSTEIN, B. E., KAMAL, M., LINDBLAD-TOH, K., BEKIRANOV, S., BAILEY, D. K., HUEBERT, D. J., MCMAHON, S., KARLSSON, E. K., KULBOKAS, E. J., 3RD, GINGERAS, T. R., SCHREIBER, S. L. & LANDER, E. S. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell,* 120**,** 169-81.
BERTRAND, J. Y., CHI, N. C., SANTOSO, B., TENG, S., STAINIER, D. Y. & TRAVER, D. 2010. Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature,* 464**,** 108-11.
BICKMORE, W. A. 2013. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet,* 14**,** 67-84.
BIGGIN, M. D. 2011. Animal transcription networks as highly connected, quantitative continua. *Dev Cell,* 21**,** 611-26.
BIRD, A. P. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res,* 8**,** 1499-504.
BISWAS, S. & RAO, C. M. 2018. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *Eur J Pharmacol,* 837**,** 8-24.
BLAKE, M. C., JAMBOU, R. C., SWICK, A. G., KAHN, J. W. & AZIZKHAN, J. C. 1990. Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. *Mol Cell Biol,* 10**,** 6632-41.
BLAYNEY, J., FRANCIS, H., CAMELLATO, B., MITCHELL, L., STOLPER, R., BOEKE, J., HIGGS, D. & KASSOUF, M. 2022. Super-enhancers require a combination of classical enhancers and novel facilitator elements to drive high levels of gene expression.
BLYSZCZUK, P. & WOBUS, A. M. 2006. In vitro differentiation of embryonic stem cells into the pancreatic lineage. *Methods Mol Biol,* 330**,** 373-85.
BOGGIANO, J. C., VANDERZALM, P. J. & FEHON, R. G. 2011. Tao-1 phosphorylates Hippo/MST kinases to regulate the Hippo-Salvador-Warts tumor suppressor pathway. *Dev Cell,* 21**,** 888-95.
BOISSET, J. C., VAN CAPPELLEN, W., ANDRIEU-SOLER, C., GALJART, N., DZIERZAK, E. & ROBIN, C. 2010. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature,* 464**,** 116-20.

BONIFER, C. 2000. Developmental regulation of eukaryotic gene loci: which cis-regulatory information is required? *Trends Genet,* 16**,** 310-5.
BONIFER, C. & BOWEN, D. T. 2010. Epigenetic mechanisms regulating normal and malignant haematopoiesis: new therapeutic targets for clinical medicine. *Expert Rev Mol Med,* 12**,** e6.
BONIFER, C. & COCKERILL, P. N. 2017. Chromatin priming of genes in development: Concepts, mechanisms and consequences. *Exp Hematol,* 49**,** 1-8.
BRADLEY, A., EVANS, M., KAUFMAN, M. H. & ROBERTSON, E. 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature,* 309**,** 255-6.
BRAHMA, S. & HENIKOFF, S. 2020. Epigenome Regulation by Dynamic Nucleosome Unwrapping. *Trends Biochem Sci,* 45**,** 13-26.
BRAY, S. J. 2016. Notch signalling in context. *Nat Rev Mol Cell Biol,* 17**,** 722-735.
BRENT, R. & PTASHNE, M. 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell,* 43**,** 729-36.
BRUVERIS, F. F., NG, E. S., LEITOGUINHO, A. R., MOTAZEDIAN, A., VLAHOS, K., SOURRIS, K., MAYBERRY, R., MCDONALD, P., AZZOLA, L., DAVIDSON, N. M., OSHLACK, A., STANLEY, E. G. & ELEFANTY, A. G. 2020. Human yolk sac-like haematopoiesis generates RUNX1-, GFI1- and/or GFI1B-dependent blood and SOX17-positive endothelium. *Development,* 147.
BRUVERIS, F. F., NG, E. S., STANLEY, E. G. & ELEFANTY, A. G. 2021. VEGF, FGF2, and BMP4 regulate transitions of mesoderm to endothelium and blood cells in a human model of yolk sac hematopoiesis. *Exp Hematol,* 103**,** 30-39 e2.
BUENROSTRO, J. D., WU, B., CHANG, H. Y. & GREENLEAF, W. J. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol,* 109**,** 21 29 1-21 29 9.
BUTKO, E., DISTEL, M., POUGET, C., WEIJTS, B., KOBAYASHI, I., NG, K., MOSIMANN, C., POULAIN, F. E., MCPHERSON, A., NI, C. W., STACHURA, D. L., DEL CID, N., ESPIN-PALAZON, R., LAWSON, N. D., DORSKY, R., CLEMENTS, W. K. & TRAVER, D. 2015. Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. *Development,* 142**,** 1050-61.
CALLEBAUT, I., COURVALIN, J. C. & MORNON, J. P. 1999. The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. *FEBS Lett,* 446**,** 189-93.
CANETE, A., CARMONA, R., ARIZA, L., SANCHEZ, M. J., ROJAS, A. & MUNOZ-CHAPULI, R. 2017. A population of hematopoietic stem cells derives from GATA4-expressing progenitors located in the placenta and lateral mesoderm of mice. *Haematologica,* 102**,** 647-655.
CARROLL, J. S., MEYER, C. A., SONG, J., LI, W., GEISTLINGER, T. R., EECKHOUTE, J., BRODSKY, A. S., KEETON, E. K., FERTUCK, K. C., HALL, G. F., WANG, Q., BEKIRANOV, S., SEMENTCHENKO, V., FOX, E. A., SILVER, P. A., GINGERAS, T. R., LIU, X. S. & BROWN, M. 2006. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet,* 38**,** 1289-97.
CARROLL, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell,* 134**,** 25-36.
CARTER, G. J. & VAN HOLDE, K. 1998. Self-association of linker histone H5 and of its globular domain: evidence for specific self-contacts. *Biochemistry,* 37**,** 12477-88.
CASIE CHETTY, S., ROST, M. S., ENRIQUEZ, J. R., SCHUMACHER, J. A., BALTRUNAITE, K., ROSSI, A., STAINIER, D. Y. & SUMANAS, S. 2017. Vegf signaling promotes vascular endothelial differentiation by modulating etv2 expression. *Dev Biol,* 424**,** 147-161.
CEINOS, R. M., TORRES-NUNEZ, E., CHAMORRO, R., NOVOA, B., FIGUERAS, A., RUANE, N. M. & ROTLLANT, J. 2013. Critical role of the matricellular protein SPARC in mediating erythroid progenitor cell development in zebrafish. *Cells Tissues Organs,* 197**,** 196-208.
CHAN, E. H., NOUSIAINEN, M., CHALAMALASETTY, R. B., SCHAFER, A., NIGG, E. A. & SILLJE, H. H. 2005. The Ste20-like kinase Mst2 activates the human large tumor suppressor kinase Lats1. *Oncogene,* 24**,** 2076-86.
CHANG, L. & KARIN, M. 2001. Mammalian MAP kinase signalling cascades. *Nature,* 410**,** 37-40.

CHAO, M. P., SEITA, J. & WEISSMAN, I. L. 2008. Establishment of a normal hematopoietic and leukemia stem cell hierarchy. *Cold Spring Harb Symp Quant Biol,* 73**,** 439-49.

CHEN, M. J., LI, Y., DE OBALDIA, M. E., YANG, Q., YZAGUIRRE, A. D., YAMADA-INAGAWA, T., VINK, C. S., BHANDOOLA, A., DZIERZAK, E. & SPECK, N. A. 2011. Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell,* 9**,** 541-52.

CHEN, M. J., YOKOMIZO, T., ZEIGLER, B. M., DZIERZAK, E. & SPECK, N. A. 2009. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature,* 457**,** 887-91.

CHEN, X., WANG, P., QIU, H., ZHU, Y., ZHANG, X., ZHANG, Y., DUAN, F., DING, S., GUO, J., HUANG, Y. & NA, J. 2022. Integrative epigenomic and transcriptomic analysis reveals the requirement of JUNB for hematopoietic fate induction. *Nat Commun,* 13**,** 3131.

CHENG, H., LIANG, P. H. & CHENG, T. 2013. Mouse hematopoietic stem cell transplantation. *Methods Mol Biol,* 976**,** 25-35.

CHOI, K., KENNEDY, M., KAZAROV, A., PAPADIMITRIOU, J. C. & KELLER, G. 1998. A common precursor for hematopoietic and endothelial cells. *Development,* 125**,** 725-32.

CHRISTENSEN, J. L., WRIGHT, D. E., WAGERS, A. J. & WEISSMAN, I. L. 2004. Circulation and chemotaxis of fetal hematopoietic stem cells. *PLoS Biol,* 2**,** E75.

CHUANG, L. S., IAN, H. I., KOH, T. W., NG, H. H., XU, G. & LI, B. F. 1997. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science,* 277**,** 1996-2000.

CHUNG, J. H., WHITELEY, M. & FELSENFELD, G. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell,* 74**,** 505-14.

CLAPIER, C. R., IWASA, J., CAIRNS, B. R. & PETERSON, C. L. 2017. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol,* 18**,** 407-422.

CLARKE, R. L., YZAGUIRRE, A. D., YASHIRO-OHTANI, Y., BONDUE, A., BLANPAIN, C., PEAR, W. S., SPECK, N. A. & KELLER, G. 2013. The expression of Sox17 identifies and regulates haemogenic endothelium. *Nat Cell Biol,* 15**,** 502-10.

COCKERILL, P. N. 2011. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J,* 278**,** 2182-210.

COLLEPARDO-GUEVARA, R. & SCHLICK, T. 2014. Chromatin fiber polymorphism triggered by variations of DNA linker lengths. *Proc Natl Acad Sci U S A,* 111**,** 8061-6.

COMOGLIO, F., PARK, H. J., SCHOENFELDER, S., BAROZZI, I., BODE, D., FRASER, P. & GREEN, A. R. 2018. Thrombopoietin signaling to chromatin elicits rapid and pervasive epigenome remodeling within poised chromatin architectures. *Genome Res,* 28**,** 295-309.

COMPERE, S. J. & PALMITER, R. D. 1981. DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells. *Cell,* 25**,** 233-40.

CORADA, M., ORSENIGO, F., MORINI, M. F., PITULESCU, M. E., BHAT, G., NYQVIST, D., BREVIARIO, F., CONTI, V., BRIOT, A., IRUELA-ARISPE, M. L., ADAMS, R. H. & DEJANA, E. 2013. Sox17 is indispensable for acquisition and maintenance of arterial identity. *Nat Commun,* 4**,** 2609.

CORE, L. J., WATERFALL, J. J. & LIS, J. T. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science,* 322**,** 1845-8.

COSTA, G., KOUSKOFF, V. & LACAUD, G. 2012. Origin of blood cells and HSC production in the embryo. *Trends Immunol,* 33**,** 215-23.

CREMER, T. & CREMER, M. 2010. Chromosome territories. *Cold Spring Harb Perspect Biol,* 2**,** a003889.

DANECEK, P., BONFIELD, J. K., LIDDLE, J., MARSHALL, J., OHAN, V., POLLARD, M. O., WHITWHAM, A., KEANE, T., MCCARTHY, S. A., DAVIES, R. M. & LI, H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience,* 10.

DARNELL, J. E., WALL, R. & TUSHINSKI, R. J. 1971. An adenylic acid-rich sequence in messenger RNA

of HeLa cells and its possible relationship to reiterated sites in DNA. *Proc Natl Acad Sci U S A,* 68**,** 1321-5.

DAVIDSON, I. F. & PETERS, J. M. 2021. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol,* 22**,** 445-464.

DE BRUIJN, M. F., MA, X., ROBIN, C., OTTERSBACH, K., SANCHEZ, M. J. & DZIERZAK, E. 2002. Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta. *Immunity,* 16**,** 673-83.

DE BRUIJN, M. F., SPECK, N. A., PEETERS, M. C. & DZIERZAK, E. 2000. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *EMBO J,* 19**,** 2465-74.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of transcription. *Genes Dev,* 25**,** 1010-22.

DECKER, T., PASCA DI MAGLIANO, M., MCMANUS, S., SUN, Q., BONIFER, C., TAGOH, H. & BUSSLINGER, M. 2009. Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis. *Immunity,* 30**,** 508-20.

DEL MONTE, G., GREGO-BESSA, J., GONZALEZ-RAJAL, A., BOLOS, V. & DE LA POMPA, J. L. 2007. Monitoring Notch1 activity in development: evidence for a feedback regulatory loop. *Dev Dyn,* 236**,** 2594-614.

DICKEL, D. E., ZHU, Y., NORD, A. S., WYLIE, J. N., AKIYAMA, J. A., AFZAL, V., PLAJZER-FRICK, I., KIRKPATRICK, A., GOTTGENS, B., BRUNEAU, B. G., VISEL, A. & PENNACCHIO, L. A. 2014. Function-based identification of mammalian enhancers using site-specific integration. *Nat Methods,* 11**,** 566-71.

DIETERLEN-LIEVRE, F. 1975. On the origin of haemopoietic stem cells in the avian embryo: an experimental approach. *J Embryol Exp Morphol,* 33**,** 607-19.

DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature,* 485**,** 376-80.

DODONOVA, S. O., ZHU, F., DIENEMANN, C., TAIPALE, J. & CRAMER, P. 2020. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature,* 580**,** 669-672.

DOETSCHMAN, T. C., EISTETTER, H., KATZ, M., SCHMIDT, W. & KEMLER, R. 1985. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol,* 87**,** 27-45.

DOMCKE, S., BARDET, A. F., ADRIAN GINNO, P., HARTL, D., BURGER, L. & SCHUBELER, D. 2015. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature,* 528**,** 575-9.

DOMINSKI, Z., YANG, X. C. & MARZLUFF, W. F. 2005. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell,* 123**,** 37-48.

DONCZEW, R. & HAHN, S. 2018. Mechanistic Differences in Transcription Initiation at TATA-Less and TATA-Containing Promoters. *Mol Cell Biol,* 38.

DOVEY, O. M., FOSTER, C. T. & COWLEY, S. M. 2010. Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proc Natl Acad Sci U S A,* 107**,** 8242-7.

DURAND, C., ROBIN, C., BOLLEROT, K., BARON, M. H., OTTERSBACH, K. & DZIERZAK, E. 2007. Embryonic stromal clones reveal developmental regulators of definitive hematopoietic stem cells. *Proc Natl Acad Sci U S A,* 104**,** 20838-43.

DURINCK, S., SPELLMAN, P. T., BIRNEY, E. & HUBER, W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc,* 4**,** 1184-91.

ECKMANN, C. R., RAMMELT, C. & WAHLE, E. 2011. Control of poly(A) tail length. *Wiley Interdiscip Rev RNA,* 2**,** 348-61.

EDGINTON-WHITE, B. & BONIFER, C. 2022. The transcriptional regulation of normal and malignant blood cell development. *FEBS J,* 289**,** 1240-1255.

EDGINTON-WHITE, B., MAYTUM, A., KELLAWAY, S. G., GOODE, D. K., KEANE, P., PAGNUCO, I., ASSI, S. A., AMES, L., CLARKE, M., COCKERILL, P. N., GOTTGENS, B., CAZIER, J. B. & BONIFER, C. 2023. A genome-wide relay of signalling-responsive enhancers drives hematopoietic specification. *Nat Commun,* 14**,** 267.

EDGINTON-WHITE, B., MAYTUM, A., KELLAWAY, S. G., GOODE, D. K., KEANE, P., PAGNUCO, I., ASSI, S. A., AMES, L., CLARKE, M., COCKERILL, P. N., GÖTTGENS, B., CAZIER, J. B. & BONIFER, C. 2022. A genome-wide relay of signalling-responsive enhancers drives haematopoietic specification.

EDMONDS, M., VAUGHAN, M. H., JR. & NAKAZATO, H. 1971. Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. *Proc Natl Acad Sci U S A,* 68**,** 1336-40.

EFERL, R., HASSELBLATT, P., RATH, M., POPPER, H., ZENZ, R., KOMNENOVIC, V., IDARRAGA, M. H., KENNER, L. & WAGNER, E. F. 2008. Development of pulmonary fibrosis through a pathway involving the transcription factor Fra-2/AP-1. *Proc Natl Acad Sci U S A,* 105**,** 10525-30.

EFERL, R., SIBILIA, M., HILBERG, F., FUCHSBICHLER, A., KUFFERATH, I., GUERTL, B., ZENZ, R., WAGNER, E. F. & ZATLOUKAL, K. 1999. Functions of c-Jun in liver and heart development. *J Cell Biol,* 145**,** 1049-61.

EILKEN, H. M., NISHIKAWA, S. & SCHROEDER, T. 2009. Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature,* 457**,** 896-900.

EMMEL, V. E. 1916. The cell clusters in the dorsal aorta of mammalian embryos. *American Journal of Anatomy,* 19**,** 401-421.

ERLER, J., ZHANG, R., PETRIDIS, L., CHENG, X., SMITH, J. C. & LANGOWSKI, J. 2014. The role of histone tails in the nucleosome: a computational study. *Biophys J,* 107**,** 2911-2922.

EVANS, C. J., HARTENSTEIN, V. & BANERJEE, U. 2003. Thicker than blood: conserved mechanisms in Drosophila and vertebrate hematopoiesis. *Dev Cell,* 5**,** 673-90.

EVANS, M. J. & KAUFMAN, M. H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature,* 292**,** 154-6.

EVANS, V., HATZOPOULOS, A., AIRD, W. C., RAYBURN, H. B., ROSENBERG, R. D. & KUIVENHOVEN, J. A. 2000. Targeting the Hprt locus in mice reveals differential regulation of Tie2 gene expression in the endothelium. *Physiol Genomics,* 2**,** 67-75.

EWELS, P., MAGNUSSON, M., LUNDIN, S. & KALLER, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics,* 32**,** 3047-8.

FADLULLAH, M. Z. H., NEO, W. H., LIE, A. L. M., THAMBYRAJAH, R., PATEL, R., MEVEL, R., AKSOY, I., DO KHOA, N., SAVATIER, P., FONTENILLE, L., BAKER, S. M., RATTRAY, M., KOUSKOFF, V. & LACAUD, G. 2022. Murine AGM single-cell profiling identifies a continuum of hemogenic endothelium differentiation marked by ACE. *Blood,* 139**,** 343-356.

FALOON, P., ARENTSON, E., KAZAROV, A., DENG, C. X., PORCHER, C., ORKIN, S. & CHOI, K. 2000. Basic fibroblast growth factor positively regulates hematopoietic development. *Development,* 127**,** 1931-41.

FIELD, A. & ADELMAN, K. 2020. Evaluating Enhancer Function and Transcription. *Annu Rev Biochem,* 89**,** 213-234.

FRANKEL, A. D. & KIM, P. S. 1991. Modular structure of transcription factors: implications for gene regulation. *Cell,* 65**,** 717-9.

FRANKLIN, R. E. & GOSLING, R. G. 1953. Molecular configuration in sodium thymonucleate. *Nature,* 171**,** 740-1.

FU, M., WANG, C., REUTENS, A. T., WANG, J., ANGELETTI, R. H., SICONOLFI-BAEZ, L., OGRYZKO, V., AVANTAGGIATI, M. L. & PESTELL, R. G. 2000. p300 and p300/cAMP-response element-binding protein-associated factor acetylate the androgen receptor at sites governing hormone-dependent transactivation. *J Biol Chem,* 275**,** 20853-60.

FULTON, D. L., SUNDARARAJAN, S., BADIS, G., HUGHES, T. R., WASSERMAN, W. W., ROACH, J. C. &

SLADEK, R. 2009. TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol,* 10**,** R29.

GEERTZ, M., SHORE, D. & MAERKL, S. J. 2012. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc Natl Acad Sci U S A,* 109**,** 16540-5.

GEIMER LE LAY, A. S., ORAVECZ, A., MASTIO, J., JUNG, C., MARCHAL, P., EBEL, C., DEMBELE, D., JOST, B., LE GRAS, S., THIBAULT, C., BORGGREFE, T., KASTNER, P. & CHAN, S. 2014. The tumor suppressor Ikaros shapes the repertoire of notch target genes in T cells. *Sci Signal,* 7**,** ra28.

GEYER, P. K. & CORCES, V. G. 1992. DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes Dev,* 6**,** 1865-73.

GHISLETTI, S., BAROZZI, I., MIETTON, F., POLLETTI, S., DE SANTA, F., VENTURINI, E., GREGORY, L., LONIE, L., CHEW, A., WEI, C. L., RAGOUSSIS, J. & NATOLI, G. 2010. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity,* 32**,** 317-28.

GIANNOS, P., PROKOPIDIS, K., RALEIGH, S. M., KELAIDITI, E. & HILL, M. 2022. Altered mitochondrial microenvironment at the spotlight of musculoskeletal aging and Alzheimer's disease. *Sci Rep,* 12**,** 11290.

GIFFORD, C. A., ZILLER, M. J., GU, H., TRAPNELL, C., DONAGHEY, J., TSANKOV, A., SHALEK, A. K., KELLEY, D. R., SHISHKIN, A. A., ISSNER, R., ZHANG, X., COYNE, M., FOSTEL, J. L., HOLMES, L., MELDRIM, J., GUTTMAN, M., EPSTEIN, C., PARK, H., KOHLBACHER, O., RINN, J., GNIRKE, A., LANDER, E. S., BERNSTEIN, B. E. & MEISSNER, A. 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell,* 153**,** 1149-63.

GILLETTE, T. G. & HILL, J. A. 2015. Readers, writers, and erasers: chromatin as the whiteboard of heart disease. *Circ Res,* 116**,** 1245-53.

GILMOUR, J., ASSI, S. A., NOAILLES, L., LICHTINGER, M., OBIER, N. & BONIFER, C. 2018. The Co-operation of RUNX1 with LDB1, CDK9 and BRD4 Drives Transcription Factor Complex Relocation During Haematopoietic Specification. *Sci Rep,* 8**,** 10410.

GILMOUR, J., O'CONNOR, L., MIDDLETON, C. P., KEANE, P., GILLEMANS, N., CAZIER, J. B., PHILIPSEN, S. & BONIFER, C. 2019. Robust hematopoietic specification requires the ubiquitous Sp1 and Sp3 transcription factors. *Epigenetics Chromatin,* 12**,** 33.

GISSEL, C., DOSS, M. X., HIPPLER-ALTENBURG, R., HESCHELER, J. & SACHINIDIS, A. 2006. Generation and characterization of cardiomyocytes under serum-free conditions. *Methods Mol Biol,* 330**,** 191-219.

GLANTSCHNIG, H., RODAN, G. A. & RESZKA, A. A. 2002. Mapping of MST1 kinase sites of phosphorylation. Activation and autophosphorylation. *J Biol Chem,* 277**,** 42987-96.

GOODE, D. K., OBIER, N., VIJAYABASKAR, M. S., LIE, A. L. M., LILLY, A. J., HANNAH, R., LICHTINGER, M., BATTA, K., FLORKOWSKA, M., PATEL, R., CHALLINOR, M., WALLACE, K., GILMOUR, J., ASSI, S. A., CAUCHY, P., HOOGENKAMP, M., WESTHEAD, D. R., LACAUD, G., KOUSKOFF, V., GOTTGENS, B. & BONIFER, C. 2016. Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Dev Cell,* 36**,** 572-87.

GOTTSCHLING, D. E. 2004. Summary: epigenetics--from phenomenon to field. *Cold Spring Harb Symp Quant Biol,* 69**,** 507-19.

GREGORETTI, I. V., LEE, Y. M. & GOODSON, H. V. 2004. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *J Mol Biol,* 338**,** 17-31.

GRIBNAU, J., DIDERICH, K., PRUZINA, S., CALZOLARI, R. & FRASER, P. 2000. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell,* 5**,** 377-86.

GROSCHEL, S., SANDERS, M. A., HOOGENBOEZEM, R., DE WIT, E., BOUWMAN, B. A. M., ERPELINCK, C., VAN DER VELDEN, V. H. J., HAVERMANS, M., AVELLINO, R., VAN LOM, K., ROMBOUTS, E. J., VAN DUIN, M., DOHNER, K., BEVERLOO, H. B., BRADNER, J. E., DOHNER, H., LOWENBERG, B., VALK, P. J. M., BINDELS, E. M. J., DE LAAT, W. & DELWEL, R. 2014. A single oncogenic

enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell,* 157**,** 369-381.

GRUNBERG, S. & HAHN, S. 2013. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci,* 38**,** 603-11.

GUALDI, R., BOSSARD, P., ZHENG, M., HAMADA, Y., COLEMAN, J. R. & ZARET, K. S. 1996. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev,* 10**,** 1670-82.

GUENTHER, M. G. & YOUNG, R. A. 2010. Transcription. Repressive transcription. *Science,* 329**,** 150-1.

GUIU, J., SHIMIZU, R., D'ALTRI, T., FRASER, S. T., HATAKEYAMA, J., BRESNICK, E. H., KAGEYAMA, R., DZIERZAK, E., YAMAMOTO, M., ESPINOSA, L. & BIGAS, A. 2013. Hes repressors are essential regulators of hematopoietic stem cell development downstream of Notch signaling. *J Exp Med,* 210**,** 71-84.

GYURKOVICS, H., GAUSZ, J., KUMMER, J. & KARCH, F. 1990. A new homeotic mutation in the Drosophila bithorax complex removes a boundary separating two domains of regulation. *EMBO J,* 9**,** 2579-85.

HAAR, J. L. & ACKERMAN, G. A. 1971. A phase and electron microscopic study of vasculogenesis and erythropoiesis in the yolk sac of the mouse. *Anat Rec,* 170**,** 199-223.

HAMELIN, V., LETOURNEUX, C., ROMEO, P. H., PORTEU, F. & GAUDRY, M. 2006. Thrombopoietin regulates IEX-1 gene expression through ERK-induced AML1 phosphorylation. *Blood,* 107**,** 3106-13.

HANSCOMBE, O., WHYATT, D., FRASER, P., YANNOUTSOS, N., GREAVES, D., DILLON, N. & GROSVELD, F. 1991. Importance of globin gene order for correct developmental expression. *Genes Dev,* 5**,** 1387-94.

HAO, Y., HAO, S., ANDERSEN-NISSEN, E., MAUCK, W. M., 3RD, ZHENG, S., BUTLER, A., LEE, M. J., WILK, A. J., DARBY, C., ZAGER, M., HOFFMAN, P., STOECKIUS, M., PAPALEXI, E., MIMITOU, E. P., JAIN, J., SRIVASTAVA, A., STUART, T., FLEMING, L. M., YEUNG, B., ROGERS, A. J., MCELRATH, J. M., BLISH, C. A., GOTTARDO, R., SMIBERT, P. & SATIJA, R. 2021. Integrated analysis of multimodal single-cell data. *Cell,* 184**,** 3573-3587 e29.

HAYASHI, H. & KUME, T. 2008. Foxc transcription factors directly regulate Dll4 and Hey2 expression by interacting with the VEGF-Notch signaling pathways in endothelial cells. *PLoS One,* 3**,** e2401.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

HEKIMOGLU, B. & RINGROSE, L. 2009. Non-coding RNAs in polycomb/trithorax regulation. *RNA Biol,* 6**,** 129-37.

HENRIQUES, T., SCRUGGS, B. S., INOUYE, M. O., MUSE, G. W., WILLIAMS, L. H., BURKHOLDER, A. B., LAVENDER, C. A., FARGO, D. C. & ADELMAN, K. 2018. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev,* 32**,** 26-41.

HEWITT, K. J., KATSUMURA, K. R., MATSON, D. R., DEVADAS, P., TANIMURA, N., HEBERT, A. S., COON, J. J., KIM, J. S., DEWEY, C. N., KELES, S., HAO, S., PAULSON, R. F. & BRESNICK, E. H. 2017. GATA Factor-Regulated Samd14 Enhancer Confers Red Blood Cell Regeneration and Survival in Severe Anemia. *Dev Cell,* 42**,** 213-225 e4.

HILL, C. S. & TREISMAN, R. 1995. Transcriptional regulation by extracellular signals: mechanisms and specificity. *Cell,* 80**,** 199-211.

HIRAI, H., SAMOKHVALOV, I. M., FUJIMOTO, T., NISHIKAWA, S., IMANISHI, J. & NISHIKAWA, S. 2005. Involvement of Runx1 in the down-regulation of fetal liver kinase-1 expression during transition of endothelial cells to hematopoietic cells. *Blood,* 106**,** 1948-55.

HIRANO, T. & MITCHISON, T. J. 1994. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell,* 79**,** 449-58.

HNISZ, D., ABRAHAM, B. J., LEE, T. I., LAU, A., SAINT-ANDRE, V., SIGOVA, A. A., HOKE, H. A. & YOUNG, R. A. 2013. Super-enhancers in the control of cell identity and disease. *Cell,* 155**,** 934-47.

HO, L. & CRABTREE, G. R. 2010. Chromatin remodelling during development. *Nature,* 463**,** 474-84.

HODAWADEKAR, S. C. & MARMORSTEIN, R. 2007. Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene,* 26**,** 5528-40.

HOEFFEL, G., WANG, Y., GRETER, M., SEE, P., TEO, P., MALLERET, B., LEBOEUF, M., LOW, D., OLLER, G., ALMEIDA, F., CHOY, S. H., GRISOTTO, M., RENIA, L., CONWAY, S. J., STANLEY, E. R., CHAN, J. K., NG, L. G., SAMOKHVALOV, I. M., MERAD, M. & GINHOUX, F. 2012. Adult Langerhans cells derive predominantly from embryonic fetal liver monocytes with a minor contribution of yolk sac-derived macrophages. *J Exp Med,* 209**,** 1167-81.

HOGGA, I. & KARCH, F. 2002. Transcription through the iab-7 cis-regulatory domain of the bithorax complex interferes with maintenance of Polycomb-mediated silencing. *Development,* 129**,** 4915-22.

HOLLIDAY, R. & PUGH, J. E. 1975. DNA modification mechanisms and gene activity during development. *Science,* 187**,** 226-32.

HOOGENKAMP, M., LICHTINGER, M., KRYSINSKA, H., LANCRIN, C., CLARKE, D., WILLIAMSON, A., MAZZARELLA, L., INGRAM, R., JORGENSEN, H., FISHER, A., TENEN, D. G., KOUSKOFF, V., LACAUD, G. & BONIFER, C. 2009. Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. *Blood,* 114**,** 299-309.

HOTCHKISS, R. D. 1948. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem,* 175**,** 315-32.

HOWELL, E. D., YZAGUIRRE, A. D., GAO, P., LIS, R., HE, B., LAKADAMYALI, M., RAFII, S., TAN, K. & SPECK, N. A. 2021. Efficient hemogenic endothelial cell specification by RUNX1 is dependent on baseline chromatin accessibility of RUNX1-regulated TGFbeta target genes. *Genes Dev,* 35**,** 1475-1489.

HUANG, H. M., LI, J. C., HSIEH, Y. C., YANG-YEN, H. F. & YEN, J. J. 1999. Optimal proliferation of a hematopoietic progenitor cell line requires either costimulation with stem cell factor or increase of receptor expression that can be replaced by overexpression of Bcl-2. *Blood,* 93**,** 2569-77.

HUME, M. A., BARRERA, L. A., GISSELBRECHT, S. S. & BULYK, M. L. 2015. UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res,* 43**,** D117-22.

IMAI, Y., KUROKAWA, M., YAMAGUCHI, Y., IZUTSU, K., NITTA, E., MITANI, K., SATAKE, M., NODA, T., ITO, Y. & HIRAI, H. 2004. The corepressor mSin3A regulates phosphorylation-induced activation, intranuclear location, and stability of AML1. *Mol Cell Biol,* 24**,** 1033-43.

ISHITOBI, H., WAKAMATSU, A., LIU, F., AZAMI, T., HAMADA, M., MATSUMOTO, K., KATAOKA, H., KOBAYASHI, M., CHOI, K., NISHIKAWA, S., TAKAHASHI, S. & EMA, M. 2011. Molecular basis for Flk1 expression in hemato-cardiovascular progenitors in the mouse. *Development,* 138**,** 5357-68.

IVANOVS, A., RYBTSOV, S., WELCH, L., ANDERSON, R. A., TURNER, M. L. & MEDVINSKY, A. 2011. Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *J Exp Med,* 208**,** 2417-27.

IZPISUA-BELMONTE, J. C., FALKENSTEIN, H., DOLLE, P., RENUCCI, A. & DUBOULE, D. 1991. Murine genes related to the Drosophila AbdB homeotic genes are sequentially expressed during development of the posterior part of the body. *EMBO J,* 10**,** 2279-89.

JACK, J., DORSETT, D., DELOTTO, Y. & LIU, S. 1991. Expression of the cut locus in the Drosophila wing

margin is required for cell type specification and is regulated by a distant enhancer. *Development,* 113**,** 735-47.
JACOB, F. & MONOD, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol,* 3**,** 318-56.
JAFFREDO, T., GAUTIER, R., EICHMANN, A. & DIETERLEN-LIEVRE, F. 1998. Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development,* 125**,** 4575-83.
JANSSON, L. & LARSSON, J. 2012. Normal hematopoietic stem cell function in mice with enforced expression of the Hippo signaling effector YAP1. *PLoS One,* 7**,** e32013.
JIA, J., YE, T., CUI, P., HUA, Q., ZENG, H. & ZHAO, D. 2016. AP-1 transcription factor mediates VEGF-induced endothelial cell migration and proliferation. *Microvasc Res,* 105**,** 103-8.
JOHNSON, P. F. & MCKNIGHT, S. L. 1989. Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem,* 58**,** 799-839.
JOLMA, A., YAN, J., WHITINGTON, T., TOIVONEN, J., NITTA, K. R., RASTAS, P., MORGUNOVA, E., ENGE, M., TAIPALE, M., WEI, G., PALIN, K., VAQUERIZAS, J. M., VINCENTELLI, R., LUSCOMBE, N. M., HUGHES, T. R., LEMAIRE, P., UKKONEN, E., KIVIOJA, T. & TAIPALE, J. 2013. DNA-binding specificities of human transcription factors. *Cell,* 152**,** 327-39.
JOLMA, A., YIN, Y., NITTA, K. R., DAVE, K., POPOV, A., TAIPALE, M., ENGE, M., KIVIOJA, T., MORGUNOVA, E. & TAIPALE, J. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature,* 527**,** 384-8.
JONES, P. L., VEENSTRA, G. J., WADE, P. A., VERMAAK, D., KASS, S. U., LANDSBERGER, N., STROUBOULIS, J. & WOLFFE, A. P. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet,* 19**,** 187-91.
JONKERS, I., KWAK, H. & LIS, J. T. 2014. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife,* 3**,** e02407.
JORDAN, H. E. 1916. Evidence of hemogenic capacity of endothelium. *Anatomical Record,* 10**,** 417-420.
KATSUMURA, K. R., BRESNICK, E. H. & GROUP, G. F. M. 2017. The GATA factor revolution in hematology. *Blood,* 129**,** 2092-2102.
KELLER, G. 2005. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev,* 19**,** 1129-55.
KELLER, G. M. 1995. In vitro differentiation of embryonic stem cells. *Curr Opin Cell Biol,* 7**,** 862-9.
KELLUM, R. & SCHEDL, P. 1991. A position-effect assay for boundaries of higher order chromosomal domains. *Cell,* 64**,** 941-50.
KELLUM, R. & SCHEDL, P. 1992. A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol,* 12**,** 2424-31.
KIM, J. J. & KINGSTON, R. E. 2022. Context-specific Polycomb mechanisms in development. *Nat Rev Genet,* 23**,** 680-695.
KIM, K., KIM, I. K., YANG, J. M., LEE, E., KOH, B. I., SONG, S., PARK, J., LEE, S., CHOI, C., KIM, J. W., KUBOTA, Y., KOH, G. Y. & KIM, I. 2016. SoxF Transcription Factors Are Positive Feedback Regulators of VEGF Signaling. *Circ Res,* 119**,** 839-52.
KIM, M., VASILJEVA, L., RANDO, O. J., ZHELKOVSKY, A., MOORE, C. & BURATOWSKI, S. 2006. Distinct pathways for snoRNA and mRNA termination. *Mol Cell,* 24**,** 723-734.
KIM, T. K., HEMBERG, M., GRAY, J. M., COSTA, A. M., BEAR, D. M., WU, J., HARMIN, D. A., LAPTEWICZ, M., BARBARA-HALEY, K., KUERSTEN, S., MARKENSCOFF-PAPADIMITRIOU, E., KUHL, D., BITO, H., WORLEY, P. F., KREIMAN, G. & GREENBERG, M. E. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature,* 465**,** 182-7.
KINDER, S. J., TSANG, T. E., QUINLAN, G. A., HADJANTONAKIS, A. K., NAGY, A. & TAM, P. P. 1999. The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development,* 126**,** 4691-701.

KIRMIZITAS, A., MEIKLEJOHN, S., CIAU-UITZ, A., STEPHENSON, R. & PATIENT, R. 2017. Dissecting BMP signaling input into the gene regulatory networks driving specification of the blood stem cell lineage. *Proc Natl Acad Sci U S A,* 114**,** 5814-5821.

KISSA, K. & HERBOMEL, P. 2010. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature,* 464**,** 112-5.

KLEIN, G., CONZELMANN, S., BECK, S., TIMPL, R. & MULLER, C. A. 1995. Perlecan in human bone marrow: a growth-factor-presenting, but anti-adhesive, extracellular matrix component for hematopoietic cells. *Matrix Biol,* 14**,** 457-65.

KLOET, S. L., MAKOWSKI, M. M., BAYMAZ, H. I., VAN VOORTHUIJSEN, L., KAREMAKER, I. D., SANTANACH, A., JANSEN, P., DI CROCE, L. & VERMEULEN, M. 2016. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol,* 23**,** 682-690.

KONDO, M., WEISSMAN, I. L. & AKASHI, K. 1997. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell,* 91**,** 661-72.

KONTARAKI, J., CHEN, H. H., RIGGS, A. & BONIFER, C. 2000. Chromatin fine structure profiles for a developmentally regulated gene: reorganization of the lysozyme locus before trans-activator binding and gene expression. *Genes Dev,* 14**,** 2106-22.

KOONTZ, L. M., LIU-CHITTENDEN, Y., YIN, F., ZHENG, Y., YU, J., HUANG, B., CHEN, Q., WU, S. & PAN, D. 2013. The Hippo effector Yorkie controls normal tissue growth by antagonizing scalloped-mediated default repression. *Dev Cell,* 25**,** 388-401.

KORNBERG, R. D. & LORCH, Y. 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell,* 98**,** 285-94.

KUEHNER, J. N., PEARSON, E. L. & MOORE, C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol,* 12**,** 283-94.

KWAK, H., FUDA, N. J., CORE, L. J. & LIS, J. T. 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science,* 339**,** 950-3.

LAMBERT, S. A., JOLMA, A., CAMPITELLI, L. F., DAS, P. K., YIN, Y., ALBU, M., CHEN, X., TAIPALE, J., HUGHES, T. R. & WEIRAUCH, M. T. 2018. The Human Transcription Factors. *Cell,* 175**,** 598-599.

LANCRIN, C., MAZAN, M., STEFANSKA, M., PATEL, R., LICHTINGER, M., COSTA, G., VARGEL, O., WILSON, N. K., MOROY, T., BONIFER, C., GOTTGENS, B., KOUSKOFF, V. & LACAUD, G. 2012. GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood,* 120**,** 314-22.

LANCRIN, C., SROCZYNSKA, P., STEPHENSON, C., ALLEN, T., KOUSKOFF, V. & LACAUD, G. 2009. The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature,* 457**,** 892-5.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LARA-ASTIASO, D., WEINER, A., LORENZO-VIVAS, E., ZARETSKY, I., JAITIN, D. A., DAVID, E., KEREN-SHAUL, H., MILDNER, A., WINTER, D., JUNG, S., FRIEDMAN, N. & AMIT, I. 2014. Immunogenetics. Chromatin state dynamics during blood formation. *Science,* 345**,** 943-9.

LEDDIN, M., PERROD, C., HOOGENKAMP, M., GHANI, S., ASSI, S., HEINZ, S., WILSON, N. K., FOLLOWS, G., SCHONHEIT, J., VOCKENTANZ, L., MOSAMMAM, A. M., CHEN, W., TENEN, D. G., WESTHEAD, D. R., GOTTGENS, B., BONIFER, C. & ROSENBAUER, F. 2011. Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. *Blood,* 117**,** 2827-38.

LEE, H. G., KAHN, T. G., SIMCOX, A., SCHWARTZ, Y. B. & PIRROTTA, V. 2015. Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome Res,* 25**,** 1170-81.

LEE, J. H., VOO, K. S. & SKALNIK, D. G. 2001. Identification and characterization of the DNA binding domain of CpG-binding protein. *J Biol Chem,* 276**,** 44669-76.

LEE, S. Y., YOON, J., LEE, M. H., JUNG, S. K., KIM, D. J., BODE, A. M., KIM, J. & DONG, Z. 2012. The

role of heterodimeric AP-1 protein comprised of JunD and c-Fos proteins in hematopoiesis. *J Biol Chem,* 287**,** 31342-8.

LEE, T. I. & YOUNG, R. A. 2013. Transcriptional regulation and its misregulation in disease. *Cell,* 152**,** 1237-51.

LEE-THEDIECK, C., SCHERTL, P. & KLEIN, G. 2022. The extracellular matrix of hematopoietic stem cell niches. *Adv Drug Deliv Rev,* 181**,** 114069.

LEONHARDT, H., PAGE, A. W., WEIER, H. U. & BESTOR, T. H. 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell,* 71**,** 865-73.

LERNER, J., GOMEZ-GARCIA, P. A., MCCARTHY, R. L., LIU, Z., LAKADAMYALI, M. & ZARET, K. S. 2020. Two-Parameter Mobility Assessments Discriminate Diverse Regulatory Factor Behaviors in Chromatin. *Mol Cell,* 79**,** 677-688 e6.

LI, E., BESTOR, T. H. & JAENISCH, R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69**,** 915-26.

LI, P., LAHVIC, J. L., BINDER, V., PUGACH, E. K., RILEY, E. B., TAMPLIN, O. J., PANIGRAHY, D., BOWMAN, T. V., BARRETT, F. G., HEFFNER, G. C., MCKINNEY-FREEMAN, S., SCHLAEGER, T. M., DALEY, G. Q., ZELDIN, D. C. & ZON, L. I. 2015. Epoxyeicosatrienoic acids enhance embryonic haematopoiesis and adult marrow engraftment. *Nature,* 523**,** 468-71.

LICHTINGER, M., INGRAM, R., HANNAH, R., MULLER, D., CLARKE, D., ASSI, S. A., LIE, A. L. M., NOAILLES, L., VIJAYABASKAR, M. S., WU, M., TENEN, D. G., WESTHEAD, D. R., KOUSKOFF, V., LACAUD, G., GOTTGENS, B. & BONIFER, C. 2012. RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J,* 31**,** 4318-33.

LIE, A. L. M., MARINOPOULOU, E., LILLY, A. J., CHALLINOR, M., PATEL, R., LANCRIN, C., KOUSKOFF, V. & LACAUD, G. 2018. Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. *Development,* 145.

LIEBERMAN-AIDEN, E., VAN BERKUM, N. L., WILLIAMS, L., IMAKAEV, M., RAGOCZY, T., TELLING, A., AMIT, I., LAJOIE, B. R., SABO, P. J., DORSCHNER, M. O., SANDSTROM, R., BERNSTEIN, B., BENDER, M. A., GROUDINE, M., GNIRKE, A., STAMATOYANNOPOULOS, J., MIRNY, L. A., LANDER, E. S. & DEKKER, J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science,* 326**,** 289-93.

LIU, F., LI, D., YU, Y. Y., KANG, I., CHA, M. J., KIM, J. Y., PARK, C., WATSON, D. K., WANG, T. & CHOI, K. 2015. Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. *EMBO Rep,* 16**,** 654-69.

LIU, J. K., DIPERSIO, C. M. & ZARET, K. S. 1991. Extracellular signals that regulate liver transcription factors during hepatic differentiation in vitro. *Mol Cell Biol,* 11**,** 773-84.

LIZAMA, C. O., HAWKINS, J. S., SCHMITT, C. E., BOS, F. L., ZAPE, J. P., CAUTIVO, K. M., BORGES PINTO, H., RHYNER, A. M., YU, H., DONOHOE, M. E., WYTHE, J. D. & ZOVEIN, A. C. 2015. Repression of arterial genes in hemogenic endothelium is sufficient for haematopoietic fate acquisition. *Nat Commun,* 6**,** 7739.

LOLLIES, A., HARTMANN, S., SCHNEIDER, M., BRACHT, T., WEISS, A. L., ARNOLDS, J., KLEIN-HITPASS, L., SITEK, B., HANSMANN, M. L., KUPPERS, R. & WENIGER, M. A. 2018. An oncogenic axis of STAT-mediated BATF3 upregulation causing MYC activity in classical Hodgkin lymphoma and anaplastic large cell lymphoma. *Leukemia,* 32**,** 92-101.

LOMELI, H. & CASTILLO-CASTELLANOS, F. 2020. Notch signaling and the emergence of hematopoietic stem cells. *Dev Dyn,* 249**,** 1302-1317.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* 389**,** 251-60.

LUNDIN, V., SUGDEN, W. W., THEODORE, L. N., SOUSA, P. M., HAN, A., CHOU, S., WRIGHTON, P. J., COX, A. G., INGBER, D. E., GOESSLING, W., DALEY, G. Q. & NORTH, T. E. 2020. YAP Regulates Hematopoietic Stem Cell Formation in Response to the Biomechanical Forces of Blood Flow. *Dev Cell,* 52**,** 446-460 e5.

MAGIN, T. M., MCWHIR, J. & MELTON, D. W. 1992. A new mouse embryonic stem cell line with good germ line contribution and gene targeting frequency. *Nucleic Acids Res,* 20**,** 3795-6.

MAJETI, R., PARK, C. Y. & WEISSMAN, I. L. 2007. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell,* 1**,** 635-45.

MANDEL, C. R., KANEKO, S., ZHANG, H., GEBAUER, D., VETHANTHAM, V., MANLEY, J. L. & TONG, L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature,* 444**,** 953-6.

MANN, R. S. & CARROLL, S. B. 2002. Molecular mechanisms of selector gene function and evolution. *Curr Opin Genet Dev,* 12**,** 592-600.

MARSDEN, M. P. & LAEMMLI, U. K. 1979. Metaphase chromosome structure: evidence for a radial loop model. *Cell,* 17**,** 849-58.

MARSHALL, C. J., KINNON, C. & THRASHER, A. J. 2000. Polarized expression of bone morphogenetic protein-4 in the human aorta-gonad-mesonephros region. *Blood,* 96**,** 1591-3.

MARTIN, G. R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A,* 78**,** 7634-8.

MASTON, G. A., EVANS, S. K. & GREEN, M. R. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet,* 7**,** 29-59.

MATHELIER, A., FORNES, O., ARENILLAS, D. J., CHEN, C. Y., DENAY, G., LEE, J., SHI, W., SHYR, C., TAN, G., WORSLEY-HUNT, R., ZHANG, A. W., PARCY, F., LENHARD, B., SANDELIN, A. & WASSERMAN, W. W. 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res,* 44**,** D110-5.

MATYS, V., KEL-MARGOULIS, O. V., FRICKE, E., LIEBICH, I., LAND, S., BARRE-DIRRIE, A., REUTER, I., CHEKMENEV, D., KRULL, M., HORNISCHER, K., VOSS, N., STEGMAIER, P., LEWICKI-POTAPOV, B., SAXEL, H., KEL, A. E. & WINGENDER, E. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res,* 34**,** D108-10.

MAZZARELLA, L., JORGENSEN, H. F., SOZA-RIED, J., TERRY, A. V., PEARSON, S., LACAUD, G., KOUSKOFF, V., MERKENSCHLAGER, M. & FISHER, A. G. 2011. Embryonic stem cell-derived hemangioblasts remain epigenetically plastic and require PRC1 to prevent neural gene expression. *Blood,* 117**,** 83-7.

MCGRATH, K. E., FRAME, J. M. & PALIS, J. 2015. Early hematopoiesis and macrophage development. *Semin Immunol,* 27**,** 379-87.

MEDVINSKY, A. L., SAMOYLINA, N. L., MULLER, A. M. & DZIERZAK, E. A. 1993. An early pre-liver intraembryonic source of CFU-S in the developing mouse. *Nature,* 364**,** 64-7.

MELCHER, M., SCHMID, M., AAGAARD, L., SELENKO, P., LAIBLE, G. & JENUWEIN, T. 2000. Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression. *Mol Cell Biol,* 20**,** 3728-41.

MENG, Z., MOROISHI, T. & GUAN, K. L. 2016. Mechanisms of Hippo pathway regulation. *Genes Dev,* 30**,** 1-17.

MERCER, E. M., LIN, Y. C., BENNER, C., JHUNJHUNWALA, S., DUTKOWSKI, J., FLORES, M., SIGVARDSSON, M., IDEKER, T., GLASS, C. K. & MURRE, C. 2011. Multilineage priming of enhancer repertoires precedes commitment to the B and myeloid cell lineages in hematopoietic progenitors. *Immunity,* 35**,** 413-25.

MICALLEF, S. J., LI, X., ELEFANTY, A. G. & STANLEY, E. G. 2007. Pancreas differentiation of mouse ES cells. *Curr Protoc Stem Cell Biol,* Chapter 1**,** Unit 1G 2.

MICHAEL, A. K., GRAND, R. S., ISBEL, L., CAVADINI, S., KOZICKA, Z., KEMPF, G., BUNKER, R. D., SCHENK, A. D., GRAFF-MEYER, A., PATHARE, G. R., WEISS, J., MATSUMOTO, S., BURGER, L., SCHUBELER, D. & THOMA, N. H. 2020. Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science,* 368**,** 1460-1465.

MIFSUD, B., MARTINCORENA, I., DARBO, E., SUGAR, R., SCHOENFELDER, S., FRASER, P. & LUSCOMBE, N. M. 2017. GOTHiC, a probabilistic model to resolve complex biases and to

identify real interactions in Hi-C data. *PLoS One,* 12**,** e0174744.
MIKKOLA, H. K., FUJIWARA, Y., SCHLAEGER, T. M., TRAVER, D. & ORKIN, S. H. 2003. Expression of CD41 marks the initiation of definitive hematopoiesis in the mouse embryo. *Blood,* 101**,** 508-16.
MINAMI, T., DONOVAN, D. J., TSAI, J. C., ROSENBERG, R. D. & AIRD, W. C. 2002. Differential regulation of the von Willebrand factor and Flt-1 promoters in the endothelium of hypoxanthine phosphoribosyltransferase-targeted mice. *Blood,* 100**,** 4019-25.
MIRSHEKAR-SYAHKAL, B., HAAK, E., KIMBER, G. M., VAN LEUSDEN, K., HARVEY, K., O'ROURKE, J., LABORDA, J., BAUER, S. R., DE BRUIJN, M. F., FERGUSON-SMITH, A. C., DZIERZAK, E. & OTTERSBACH, K. 2013. Dlk1 is a negative regulator of emerging hematopoietic stem and progenitor cells. *Haematologica,* 98**,** 163-71.
MIZZEN, C. A., YANG, X. J., KOKUBO, T., BROWNELL, J. E., BANNISTER, A. J., OWEN-HUGHES, T., WORKMAN, J., WANG, L., BERGER, S. L., KOUZARIDES, T., NAKATANI, Y. & ALLIS, C. D. 1996. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell,* 87**,** 1261-70.
MORGUNOVA, E. & TAIPALE, J. 2017. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol,* 47**,** 1-8.
MULLER, A. M., MEDVINSKY, A., STROUBOULIS, J., GROSVELD, F. & DZIERZAK, E. 1994. Development of hematopoietic stem cell activity in the mouse embryo. *Immunity,* 1**,** 291-301.
MURTHA, M., TOKCAER-KESKIN, Z., TANG, Z., STRINO, F., CHEN, X., WANG, Y., XI, X., BASILICO, C., BROWN, S., BONNEAU, R., KLUGER, Y. & DAILEY, L. 2014. FIREWACh: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat Methods,* 11**,** 559-65.
NAFRIA, M., BONIFER, C., STANLEY, E. G., NG, E. S. & ELEFANTY, A. G. 2020. Protocol for the Generation of Definitive Hematopoietic Progenitors from Human Pluripotent Stem Cells. *STAR Protoc,* 1**,** 100130.
NAIK, S. H., PERIE, L., SWART, E., GERLACH, C., VAN ROOIJ, N., DE BOER, R. J. & SCHUMACHER, T. N. 2013. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature,* 496**,** 229-32.
NAKANO, T., KODAMA, H. & HONJO, T. 1994. Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science,* 265**,** 1098-101.
NASMYTH, K. 2001. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet,* 35**,** 673-745.
NECHAEV, S., FARGO, D. C., DOS SANTOS, G., LIU, L., GAO, Y. & ADELMAN, K. 2010. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science,* 327**,** 335-8.
NEPH, S., STERGACHIS, A. B., REYNOLDS, A., SANDSTROM, R., BORENSTEIN, E. & STAMATOYANNOPOULOS, J. A. 2012. Circuitry and dynamics of human transcription factor regulatory networks. *Cell,* 150**,** 1274-86.
NG, E. S., DAVIS, R., STANLEY, E. G. & ELEFANTY, A. G. 2008. A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. *Nature Protocols,* 3**,** 768-776.
NIKI, H., IMAMURA, R., KITAOKA, M., YAMANAKA, K., OGURA, T. & HIRAGA, S. 1992. E.coli MukB protein involved in chromosome partition forms a homodimer with a rod-and-hinge structure having DNA binding and ATP/GTP binding activities. *EMBO J,* 11**,** 5101-9.
NISHIKAWA, S. I., NISHIKAWA, S., HIRASHIMA, M., MATSUYOSHI, N. & KODAMA, H. 1998. Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. *Development,* 125**,** 1747-57.
NIWA, H., BURDON, T., CHAMBERS, I. & SMITH, A. 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev,* 12**,** 2048-60.
NOLL, M. & KORNBERG, R. D. 1977. Action of micrococcal nuclease on chromatin and the location

of histone H1. *J Mol Biol,* 109**,** 393-404.
NORA, E. P., LAJOIE, B. R., SCHULZ, E. G., GIORGETTI, L., OKAMOTO, I., SERVANT, N., PIOLOT, T., VAN BERKUM, N. L., MEISIG, J., SEDAT, J., GRIBNAU, J., BARILLOT, E., BLUTHGEN, N., DEKKER, J. & HEARD, E. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature,* 485**,** 381-5.
NORTH, T., GU, T. L., STACY, T., WANG, Q., HOWARD, L., BINDER, M., MARIN-PADILLA, M. & SPECK, N. A. 1999. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development,* 126**,** 2563-75.
NOTTA, F., DOULATOV, S., LAURENTI, E., POEPPL, A., JURISICA, I. & DICK, J. E. 2011. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science,* 333**,** 218-21.
NOTTINGHAM, W. T., JARRATT, A., BURGESS, M., SPECK, C. L., CHENG, J. F., PRABHAKAR, S., RUBIN, E. M., LI, P. S., SLOANE-STANLEY, J., KONG, A. S. J. & DE BRUIJN, M. F. 2007. Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood,* 110**,** 4188-97.
NOVO, C. L., JAVIERRE, B. M., CAIRNS, J., SEGONDS-PICHON, A., WINGETT, S. W., FREIRE-PRITCHETT, P., FURLAN-MAGARIL, M., SCHOENFELDER, S., FRASER, P. & RUGG-GUNN, P. J. 2018. Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. *Cell Rep,* 22**,** 2615-2627.
O'KANE, C. J. & GEHRING, W. J. 1987. Detection in situ of genomic regulatory elements in Drosophila. *Proc Natl Acad Sci U S A,* 84**,** 9123-7.
OATLEY, M., BOLUKBASI, O. V., SVENSSON, V., SHVARTSMAN, M., GANTER, K., ZIRNGIBL, K., PAVLOVICH, P. V., MILCHEVSKAYA, V., FOTEVA, V., NATARAJAN, K. N., BAYING, B., BENES, V., PATIL, K. R., TEICHMANN, S. A. & LANCRIN, C. 2020. Single-cell transcriptomics identifies CD44 as a marker and regulator of endothelial to haematopoietic transition. *Nat Commun,* 11**,** 586.
OBIER, N., CAUCHY, P., ASSI, S. A., GILMOUR, J., LIE, A. L. M., LICHTINGER, M., HOOGENKAMP, M., NOAILLES, L., COCKERILL, P. N., LACAUD, G., KOUSKOFF, V. & BONIFER, C. 2016. Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate. *Development,* 143**,** 4324-4340.
OTTERSBACH, K. 2019. Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochem Soc Trans,* 47**,** 591-601.
OWENS, D. D. G., ANSELMI, G., OUDELAAR, A. M., DOWNES, D. J., CAVALLO, A., HARMAN, J. R., SCHWESSINGER, R., BUCAKCI, A., GREDER, L., DE ORNELLAS, S., JEZIORSKA, D., TELENIUS, J., HUGHES, J. R. & DE BRUIJN, M. 2022. Dynamic Runx1 chromatin boundaries affect gene expression in hematopoietic development. *Nat Commun,* 13**,** 773.
OZDEMIR, I. & GAMBETTA, M. C. 2019. The Role of Insulation in Patterning Gene Expression. *Genes (Basel),* 10.
PADRON-BARTHE, L., TEMINO, S., VILLA DEL CAMPO, C., CARRAMOLINO, L., ISERN, J. & TORRES, M. 2014. Clonal analysis identifies hemogenic endothelium as the source of the blood-endothelial common lineage in the mouse embryo. *Blood,* 124**,** 2523-32.
PALIS, J., ROBERTSON, S., KENNEDY, M., WALL, C. & KELLER, G. 1999. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development,* 126**,** 5073-84.
PARTHUN, M. R. 2007. Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene,* 26**,** 5319-28.
PATEL, A. B., GREBER, B. J. & NOGALES, E. 2020. Recent insights into the structure of TFIID, its assembly, and its binding to core promoter. *Curr Opin Struct Biol,* 61**,** 17-24.
PATRINOS, G. P., DE KROM, M., DE BOER, E., LANGEVELD, A., IMAM, A. M., STROUBOULIS, J., DE LAAT, W. & GROSVELD, F. G. 2004. Multiple interactions between regulatory regions are

required to stabilize an active chromatin hub. *Genes Dev,* 18**,** 1495-509.
PEARSON, S., CUVERTINO, S., FLEURY, M., LACAUD, G. & KOUSKOFF, V. 2015. In vivo repopulating activity emerges at the onset of hematopoietic specification during embryonic stem cell differentiation. *Stem Cell Reports,* 4**,** 431-44.
PEARSON, S., SROCZYNSKA, P., LACAUD, G. & KOUSKOFF, V. 2008. The stepwise specification of embryonic stem cells to hematopoietic fate is driven by sequential exposure to Bmp4, activin A, bFGF and VEGF. *Development,* 135**,** 1525-35.
PEIFER, M. & BENDER, W. 1986. The anterobithorax and bithorax mutations of the bithorax complex. *EMBO J,* 5**,** 2293-303.
PELLIN, D., LOPERFIDO, M., BARICORDI, C., WOLOCK, S. L., MONTEPELOSO, A., WEINBERG, O. K., BIFFI, A., KLEIN, A. M. & BIASCO, L. 2019. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat Commun,* 10**,** 2395.
PEREIRA, C. F., CHANG, B., QIU, J., NIU, X., PAPATSENKO, D., HENDRY, C. E., CLARK, N. R., NOMURA-KITABAYASHI, A., KOVACIC, J. C., MA'AYAN, A., SCHANIEL, C., LEMISCHKA, I. R. & MOORE, K. 2013. Induction of a hemogenic program in mouse fibroblasts. *Cell Stem Cell,* 13**,** 205-18.
PICK, M., AZZOLA, L., OSBORNE, E., STANLEY, E. G. & ELEFANTY, A. G. 2013. Generation of Megakaryocytic Progenitors from Human Embryonic Stem Cells in a Feeder- and Serum-Free Medium. *Plos One,* 8.
PICKAR-OLIVER, A. & GERSBACH, C. A. 2019. The next generation of CRISPR-Cas technologies and applications. *Nat Rev Mol Cell Biol,* 20**,** 490-507.
PIMKIN, M., KOSSENKOV, A. V., MISHRA, T., MORRISSEY, C. S., WU, W., KELLER, C. A., BLOBEL, G. A., LEE, D., BEER, M. A., HARDISON, R. C. & WEISS, M. J. 2014. Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. *Genome Res,* 24**,** 1932-44.
POPE, S. D. & MEDZHITOV, R. 2018. Emerging Principles of Gene Expression Programs and Their Regulation. *Mol Cell,* 71**,** 389-397.
PORCHER, C., SWAT, W., ROCKWELL, K., FUJIWARA, Y., ALT, F. W. & ORKIN, S. H. 1996. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell,* 86**,** 47-57.
PRADHAN, M., ESTEVE, P. O., CHIN, H. G., SAMARANAYKE, M., KIM, G. D. & PRADHAN, S. 2008. CXXC domain of human DNMT1 is essential for enzymatic activity. *Biochemistry,* 47**,** 10000-9.
PRASKOVA, M., KHOKLATCHEV, A., ORTIZ-VEGA, S. & AVRUCH, J. 2004. Regulation of the MST1 kinase by autophosphorylation, by the growth inhibitory proteins, RASSF1 and NORE1, and by Ras. *Biochem J,* 381**,** 453-62.
PROUDFOOT, N. J. & BROWNLEE, G. G. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature,* 263**,** 211-4.
PUGH, B. F. & TJIAN, R. 1991. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev,* 5**,** 1935-45.
QI, Q., CHENG, L., TANG, X., HE, Y., LI, Y., YEE, T., SHRESTHA, D., FENG, R., XU, P., ZHOU, X., PRUETT-MILLER, S., HARDISON, R. C., WEISS, M. J. & CHENG, Y. 2021. Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis. *Blood,* 137**,** 1327-1339.
QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.
RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-2308.
RAO, S. S., HUNTLEY, M. H., DURAND, N. C., STAMENOVA, E. K., BOCHKOV, I. D., ROBINSON, J. T., SANBORN, A. L., MACHOL, I., OMER, A. D., LANDER, E. S. & AIDEN, E. L. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell,* 159**,** 1665-80.

REMENYI, A., LINS, K., NISSEN, L. J., REINBOLD, R., SCHOLER, H. R. & WILMANNS, M. 2003. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev,* 17**,** 2048-59.

REYES, A. & HUBER, W. 2018. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res,* 46**,** 582-592.

REYNOLDS, G. A., BASU, S. K., OSBORNE, T. F., CHIN, D. J., GIL, G., BROWN, M. S., GOLDSTEIN, J. L. & LUSKEY, K. L. 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5' untranslated regions. *Cell,* 38**,** 275-85.

RHEE, H. S. & PUGH, B. F. 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature,* 483**,** 295-301.

RICHARD, C., DREVON, C., CANTO, P. Y., VILLAIN, G., BOLLEROT, K., LEMPEREUR, A., TEILLET, M. A., VINCENT, C., ROSSELLO CASTILLO, C., TORRES, M., PIWARZYK, E., SPECK, N. A., SOUYRI, M. & JAFFREDO, T. 2013. Endothelio-mesenchymal interaction controls runx1 expression and modulates the notch pathway to initiate aortic hematopoiesis. *Dev Cell,* 24**,** 600-11.

RICHARD, P. & MANLEY, J. L. 2009. Transcription termination by nuclear RNA polymerases. *Genes Dev,* 23**,** 1247-69.

RICHMOND, T. J., FINCH, J. T., RUSHTON, B., RHODES, D. & KLUG, A. 1984. Structure of the nucleosome core particle at 7 A resolution. *Nature,* 311**,** 532-7.

RIEGER, M. A. & SCHROEDER, T. 2012. Hematopoiesis. *Cold Spring Harb Perspect Biol,* 4.

RIGGS, A. D. 1990. DNA methylation and late replication probably aid cell memory, and type I DNA reeling could aid chromosome folding and enhancer function. *Philos Trans R Soc Lond B Biol Sci,* 326**,** 285-97.

RINGROSE, L. & PARO, R. 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet,* 38**,** 413-43.

RIVERA, L. B., BRADSHAW, A. D. & BREKKEN, R. A. 2011. The regulatory function of SPARC in vascular biology. *Cell Mol Life Sci,* 68**,** 3165-73.

ROBB, L., ELWOOD, N. J., ELEFANTY, A. G., KONTGEN, F., LI, R., BARNETT, L. D. & BEGLEY, C. G. 1996. The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. *EMBO J,* 15**,** 4123-9.

ROBERT-MORENO, A., ESPINOSA, L., DE LA POMPA, J. L. & BIGAS, A. 2005. RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells. *Development,* 132**,** 1117-26.

ROBIN, C., OTTERSBACH, K., DURAND, C., PEETERS, M., VANES, L., TYBULEWICZ, V. & DZIERZAK, E. 2006. An unexpected role for IL-3 in the embryonic development of hematopoietic stem cells. *Dev Cell,* 11**,** 171-80.

ROSAMOND, J., ENDLICH, B., TELANDER, K. M. & LINN, S. 1979. Mechanisms of action of the type-I restriction endonuclease, ecoB, and the recBC DNase from Escherichia coli. *Cold Spring Harb Symp Quant Biol,* 43 Pt 2**,** 1049-57.

ROUNTREE, M. R., BACHMAN, K. E. & BAYLIN, S. B. 2000. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet,* 25**,** 269-77.

ROUTH, A., SANDIN, S. & RHODES, D. 2008. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci U S A,* 105**,** 8872-7.

RUTHENBURG, A. J., ALLIS, C. D. & WYSOCKA, J. 2007. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell,* 25**,** 15-30.

RYBTSOV, S., BATSIVARI, A., BILOTKACH, K., PARUZINA, D., SENSERRICH, J., NERUSHEV, O. & MEDVINSKY, A. 2014. Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43(-) embryonic precursor. *Stem Cell Reports,* 3**,** 489-501.

SABIN, F. R. 1920. Studies on the origin of blood-vessels and of red blood-corpuscles as seen in the living blastoderm of chicks during the second day of incubation. *Contributions to Embryology,* 9**,** 215-U56.

SAKA, Y., SUTANI, T., YAMASHITA, Y., SAITOH, S., TAKEUCHI, M., NAKASEKO, Y. & YANAGIDA, M. 1994. Fission yeast cut3 and cut14, members of a ubiquitous protein family, are required for chromosome condensation and segregation in mitosis. *EMBO J,* 13**,** 4938-52.

SALDANHA, A. J. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics,* 20**,** 3246-8.

SAPONARO, M., KANTIDAKIS, T., MITTER, R., KELLY, G. P., HERON, M., WILLIAMS, H., SODING, J., STEWART, A. & SVEJSTRUP, J. Q. 2014. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell,* 157**,** 1037-49.

SARDINA, J. L., COLLOMBET, S., TIAN, T. V., GOMEZ, A., DI STEFANO, B., BERENGUER, C., BRUMBAUGH, J., STADHOUDERS, R., SEGURA-MORALES, C., GUT, M., GUT, I. G., HEATH, S., ARANDA, S., DI CROCE, L., HOCHEDLINGER, K., THIEFFRY, D. & GRAF, T. 2018. Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. *Cell Stem Cell,* 23**,** 905-906.

SCHLEUSSNER, N., MERKEL, O., COSTANZA, M., LIANG, H. C., HUMMEL, F., ROMAGNANI, C., DUREK, P., ANAGNOSTOPOULOS, I., HUMMEL, M., JOHRENS, K., NIEDOBITEK, A., GRIFFIN, P. R., PIVA, R., SCZAKIEL, H. L., WOESSMANN, W., DAMM-WELK, C., HINZE, C., STOIBER, D., GILLISSEN, B., TURNER, S. D., KAERGEL, E., VON HOFF, L., GRAU, M., LENZ, G., DORKEN, B., SCHEIDEREIT, C., KENNER, L., JANZ, M. & MATHAS, S. 2018. The AP-1-BATF and -BATF3 module is essential for growth, survival and TH17/ILC3 skewing of anaplastic large cell lymphoma. *Leukemia,* 32**,** 1994-2007.

SCHMIERER, B. & HILL, C. S. 2007. TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility. *Nat Rev Mol Cell Biol,* 8**,** 970-82.

SCHMITT, S., PRESTEL, M. & PARO, R. 2005. Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev,* 19**,** 697-708.

SCHNEIDER, R., BANNISTER, A. J., MYERS, F. A., THORNE, A. W., CRANE-ROBINSON, C. & KOUZARIDES, T. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol,* 6**,** 73-7.

SCHUBERT, H. L., WITTMEYER, J., KASTEN, M. M., HINATA, K., RAWLING, D. C., HEROUX, A., CAIRNS, B. R. & HILL, C. P. 2013. Structure of an actin-related subcomplex of the SWI/SNF chromatin remodeler. *Proc Natl Acad Sci U S A,* 110**,** 3345-50.

SCHULZ, C., GOMEZ PERDIGUERO, E., CHORRO, L., SZABO-ROGERS, H., CAGNARD, N., KIERDORF, K., PRINZ, M., WU, B., JACOBSEN, S. E., POLLARD, J. W., FRAMPTON, J., LIU, K. J. & GEISSMANN, F. 2012. A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science,* 336**,** 86-90.

SCHWARTZ, Y. B. & PIRROTTA, V. 2013. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat Rev Genet,* 14**,** 853-64.

SCHWERING, I., BRAUNINGER, A., DISTLER, V., JESDINSKY, J., DIEHL, V., HANSMANN, M. L., RAJEWSKY, K. & KUPPERS, R. 2003. Profiling of Hodgkin's lymphoma cell line L1236 and germinal center B cells: identification of Hodgkin's lymphoma-specific genes. *Mol Med,* 9**,** 85-95.

SEITAN, V. C., FAURE, A. J., ZHAN, Y., MCCORD, R. P., LAJOIE, B. R., ING-SIMMONS, E., LENHARD, B., GIORGETTI, L., HEARD, E., FISHER, A. G., FLICEK, P., DEKKER, J. & MERKENSCHLAGER, M. 2013. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res,* 23**,** 2066-77.

SEKIYA, T. & ZARET, K. S. 2007. Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo. *Mol Cell,* 28**,** 291-303.

SEN, R. & BALTIMORE, D. 1986. Inducibility of kappa immunoglobulin enhancer-binding protein Nf-kappa B by a posttranslational mechanism. *Cell,* 47**,** 921-8.

SEO, S., FUJITA, H., NAKANO, A., KANG, M., DUARTE, A. & KUME, T. 2006. The forkhead transcription

factors, Foxc1 and Foxc2, are required for arterial specification and lymphatic sprouting during vascular development. *Dev Biol,* 294**,** 458-70.

SERRANO, A. G., GANDILLET, A., PEARSON, S., LACAUD, G. & KOUSKOFF, V. 2010. Contrasting effects of Sox17- and Sox18-sustained expression at the onset of blood specification. *Blood,* 115**,** 3895-8.

SHAH, S., TAKEI, Y., ZHOU, W., LUBECK, E., YUN, J., ENG, C. L., KOULENA, N., CRONIN, C., KARP, C., LIAW, E. J., AMIN, M. & CAI, L. 2018. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell,* 174**,** 363-376 e16.

SHALABY, F., ROSSANT, J., YAMAGUCHI, T. P., GERTSENSTEIN, M., WU, X. F., BREITMAN, M. L. & SCHUH, A. C. 1995. Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature,* 376**,** 62-6.

SHI, J. & VAKOC, C. R. 2014. The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol Cell,* 54**,** 728-36.

SHI, Y., DI GIAMMARTINO, D. C., TAYLOR, D., SARKESHIK, A., RICE, W. J., YATES, J. R., 3RD, FRANK, J. & MANLEY, J. L. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell,* 33**,** 365-76.

SHI, Y., LAN, F., MATSON, C., MULLIGAN, P., WHETSTINE, J. R., COLE, P. A., CASERO, R. A. & SHI, Y. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell,* 119**,** 941-53.

SHLYUEVA, D., STAMPFEL, G. & STARK, A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet,* 15**,** 272-86.

SIMEONOV, D. R., GOWEN, B. G., BOONTANRART, M., ROTH, T. L., GAGNON, J. D., MUMBACH, M. R., SATPATHY, A. T., LEE, Y., BRAY, N. L., CHAN, A. Y., LITUIEV, D. S., NGUYEN, M. L., GATE, R. E., SUBRAMANIAM, M., LI, Z., WOO, J. M., MITROS, T., RAY, G. J., CURIE, G. L., NADDAF, N., CHU, J. S., MA, H., BOYER, E., VAN GOOL, F., HUANG, H., LIU, R., TOBIN, V. R., SCHUMANN, K., DALY, M. J., FARH, K. K., ANSEL, K. M., YE, C. J., GREENLEAF, W. J., ANDERSON, M. S., BLUESTONE, J. A., CHANG, H. Y., CORN, J. E. & MARSON, A. 2017. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature,* 549**,** 111-115.

SIMPSON, R. T. 1978. Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry,* 17**,** 5524-31.

SING, A., PANNELL, D., KARAISKAKIS, A., STURGEON, K., DJABALI, M., ELLIS, J., LIPSHITZ, H. D. & CORDES, S. P. 2009. A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell,* 138**,** 885-97.

SMIT, W. L., DE BOER, R. J., MEIJER, B. J., SPAAN, C. N., VAN ROEST, M., KOELINK, P. J., KOSTER, J., DEKKER, E., ABBINK, T. E. M., VAN DER KNAAP, M. S., VAN DEN BRINK, G. R., MUNCAN, V. & HEIJMANS, J. 2021. Translation initiation factor eIF2Bepsilon promotes Wnt-mediated clonogenicity and global translation in intestinal epithelial cells. *Stem Cell Res,* 55**,** 102499.

SMITH, A. G. 2001. Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol,* 17**,** 435-62.

SMITH, A. G., HEATH, J. K., DONALDSON, D. D., WONG, G. G., MOREAU, J., STAHL, M. & ROGERS, D. 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature,* 336**,** 688-90.

SMITH, Z. D. & MEISSNER, A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet,* 14**,** 204-20.

SNEPPEN, K. & RINGROSE, L. 2019. Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent. *Nat Commun,* 10**,** 2133.

SONG, F., CHEN, P., SUN, D., WANG, M., DONG, L., LIANG, D., XU, R. M., ZHU, P. & LI, G. 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science,* 344**,** 376-80.

SONG, J., TEPLOVA, M., ISHIBE-MURAKAMI, S. & PATEL, D. J. 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science,* 335**,** 709-12.

SOSHNIKOVA, N. & DUBOULE, D. 2009. Epigenetic temporal control of mouse Hox genes in vivo. *Science,* 324**,** 1320-3.

SPITZ, F. & FURLONG, E. E. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet,* 13**,** 613-26.

SROCZYNSKA, P., LANCRIN, C., KOUSKOFF, V. & LACAUD, G. 2009. The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. *Blood,* 114**,** 5279-89.

STEFANSKA, M., BATTA, K., PATEL, R., FLORKOWSKA, M., KOUSKOFF, V. & LACAUD, G. 2017. Primitive erythrocytes are generated from hemogenic endothelial cells. *Sci Rep,* 7**,** 6401.

STEWART, C. L., KASPAR, P., BRUNET, L. J., BHATT, H., GADI, I., KONTGEN, F. & ABBONDANZO, S. J. 1992. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature,* 359**,** 76-9.

STOCK, J. K., GIADROSSI, S., CASANOVA, M., BROOKES, E., VIDAL, M., KOSEKI, H., BROCKDORFF, N., FISHER, A. G. & POMBO, A. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol,* 9**,** 1428-35.

STROUBOULIS, J., DILLON, N. & GROSVELD, F. 1992. Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev,* 6**,** 1857-64.

STRUNNIKOV, A. V., LARIONOV, V. L. & KOSHLAND, D. 1993. SMC1: an essential yeast gene encoding a putative head-rod-tail protein is required for nuclear division and defines a new ubiquitous protein family. *J Cell Biol,* 123**,** 1635-48.

SWIERS, G., RODE, C., AZZONI, E. & DE BRUIJN, M. F. 2013. A short history of hemogenic endothelium. *Blood Cells Mol Dis,* 51**,** 206-12.

SYKES, S. M. & SCADDEN, D. T. 2013. Modeling human hematopoietic stem cell biology in the mouse. *Semin Hematol,* 50**,** 92-100.

SZABO, Q., BANTIGNIES, F. & CAVALLI, G. 2019. Principles of genome folding into topologically associating domains. *Sci Adv,* 5**,** eaaw1668.

TABERLAY, P. C., ACHINGER-KAWECKA, J., LUN, A. T., BUSKE, F. A., SABIR, K., GOULD, C. M., ZOTENKO, E., BERT, S. A., GILES, K. A., BAUER, D. C., SMYTH, G. K., STIRZAKER, C., O'DONOGHUE, S. I. & CLARK, S. J. 2016. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res,* 26**,** 719-31.

TAGOH, H., MELNIK, S., LEFEVRE, P., CHONG, S., RIGGS, A. D. & BONIFER, C. 2004. Dynamic reorganization of chromatin structure and selective DNA demethylation prior to stable enhancer complex formation during differentiation of primary hematopoietic cells in vitro. *Blood,* 103**,** 2950-5.

TANAKA, T., KUROKAWA, M., UEKI, K., TANAKA, K., IMAI, Y., MITANI, K., OKAZAKI, K., SAGATA, N., YAZAKI, Y., SHIBATA, Y., KADOWAKI, T. & HIRAI, H. 1996. The extracellular signal-regulated kinase pathway phosphorylates AML1, an acute myeloid leukemia gene product, and potentially regulates its transactivation ability. *Mol Cell Biol,* 16**,** 3967-79.

THAMBYRAJAH, R., MAZAN, M., PATEL, R., MOIGNARD, V., STEFANSKA, M., MARINOPOULOU, E., LI, Y., LANCRIN, C., CLAPES, T., MOROY, T., ROBIN, C., MILLER, C., COWLEY, S., GOTTGENS, B., KOUSKOFF, V. & LACAUD, G. 2016. GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat Cell Biol,* 18**,** 21-32.

THOMA, F., KOLLER, T. & KLUG, A. 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol,* 83**,** 403-27.

THOMAS, M. C. & CHIANG, C. M. 2006. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol,* 41**,** 105-78.

THOMSON, J. A., ITSKOVITZ-ELDOR, J., SHAPIRO, S. S., WAKNITZ, M. A., SWIERGIEL, J. J., MARSHALL,

V. S. & JONES, J. M. 1998. Embryonic stem cell lines derived from human blastocysts. *Science,* 282**,** 1145-7.

TIAN, B. & MANLEY, J. L. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol,* 18**,** 18-30.

TIAN, X., MORRIS, J. K., LINEHAN, J. L. & KAUFMAN, D. S. 2004. Cytokine requirements differ for stroma and embryoid body-mediated hematopoiesis from human embryonic stem cells. *Exp Hematol,* 32**,** 1000-9.

TIRODE, F., BUSSO, D., COIN, F. & EGLY, J. M. 1999. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell,* 3**,** 87-95.

TOLHUIS, B., PALSTRA, R. J., SPLINTER, E., GROSVELD, F. & DE LAAT, W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell,* 10**,** 1453-65.

TRAPNELL, C., CACCHIARELLI, D., GRIMSBY, J., POKHAREL, P., LI, S., MORSE, M., LENNON, N. J., LIVAK, K. J., MIKKELSEN, T. S. & RINN, J. L. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol,* 32**,** 381-386.

TSAI, F. Y., KELLER, G., KUO, F. C., WEISS, M., CHEN, J., ROSENBLATT, M., ALT, F. W. & ORKIN, S. H. 1994. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature,* 371**,** 221-6.

TSANKOV, A. M., GU, H., AKOPIAN, V., ZILLER, M. J., DONAGHEY, J., AMIT, I., GNIRKE, A. & MEISSNER, A. 2015. Transcription factor binding dynamics during human ES cell differentiation. *Nature,* 518**,** 344-9.

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., SWANSON, S., O, J. T., ZON, L. I., THOMSON, J. A., BERNSTEIN, I. D. & SLUKVIN, II 2018. NOTCH signaling specifies arterial-type definitive hemogenic endothelium from human pluripotent stem cells. *Nat Commun,* 9**,** 1828.

UHLMANN, F. 2016. SMC complexes: from DNA to chromosomes. *Nat Rev Mol Cell Biol,* 17**,** 399-412.

VAN INGEN, H., VAN SCHAIK, F. M., WIENK, H., BALLERING, J., REHMANN, H., DECHESNE, A. C., KRUIJZER, J. A., LISKAMP, R. M., TIMMERS, H. T. & BOELENS, R. 2008. Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure,* 16**,** 1245-56.

VAQUERIZAS, J. M., KUMMERFELD, S. K., TEICHMANN, S. A. & LUSCOMBE, N. M. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet,* 10**,** 252-63.

VERMA, I. M., STEVENSON, J. K., SCHWARZ, E. M., VAN ANTWERP, D. & MIYAMOTO, S. 1995. Rel/NF-kappa B/I kappa B family: intimate tales of association and dissociation. *Genes Dev,* 9**,** 2723-35.

VERMEULEN, M., MULDER, K. W., DENISSOV, S., PIJNAPPEL, W. W., VAN SCHAIK, F. M., VARIER, R. A., BALTISSEN, M. P., STUNNENBERG, H. G., MANN, M. & TIMMERS, H. T. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell,* 131**,** 58-69.

VIJAYABASKAR, M. S., GOODE, D. K., OBIER, N., LICHTINGER, M., EMMETT, A. M. L., ABIDIN, F. N. Z., SHAR, N., HANNAH, R., ASSI, S. A., LIE, A. L. M., GOTTGENS, B., LACAUD, G., KOUSKOFF, V., BONIFER, C. & WESTHEAD, D. R. 2019. Identification of gene specific cis-regulatory elements during differentiation of mouse embryonic stem cells: An integrative approach using high-throughput datasets. *PLoS Comput Biol,* 15**,** e1007337.

VOIGT, P., TEE, W. W. & REINBERG, D. 2013. A double take on bivalent promoters. *Genes Dev,* 27**,** 1318-38.

VOSS, T. C., SCHILTZ, R. L., SUNG, M. H., YEN, P. M., STAMATOYANNOPOULOS, J. A., BIDDIE, S. C., JOHNSON, T. A., MIRANDA, T. B., JOHN, S. & HAGER, G. L. 2011. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell,* 146**,** 544-54.

WALTER, K., BONIFER, C. & TAGOH, H. 2008. Stem cell-specific epigenetic priming and B cell-specific transcriptional activation at the mouse Cd19 locus. *Blood,* 112**,** 1673-82.

WAMSTAD, J. A., ALEXANDER, J. M., TRUTY, R. M., SHRIKUMAR, A., LI, F., EILERTSON, K. E., DING, H., WYLIE, J. N., PICO, A. R., CAPRA, J. A., ERWIN, G., KATTMAN, S. J., KELLER, G. M., SRIVASTAVA, D., LEVINE, S. S., POLLARD, K. S., HOLLOWAY, A. K., BOYER, L. A. & BRUNEAU, B. G. 2012. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell,* 151**,** 206-20.

WANG, A., YUE, F., LI, Y., XIE, R., HARPER, T., PATEL, N. A., MUTH, K., PALMER, J., QIU, Y., WANG, J., LAM, D. K., RAUM, J. C., STOFFERS, D. A., REN, B. & SANDER, M. 2015. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell,* 16**,** 386-99.

WANG, R., ZHENG, D., YEHIA, G. & TIAN, B. 2018. A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Res,* 28**,** 1427-1441.

WANG, X., FREIRE VALLS, A., SCHERMANN, G., SHEN, Y., MOYA, I. M., CASTRO, L., URBAN, S., SOLECKI, G. M., WINKLER, F., RIEDEMANN, L., JAIN, R. K., MAZZONE, M., SCHMIDT, T., FISCHER, T., HALDER, G. & RUIZ DE ALMODOVAR, C. 2017. YAP/TAZ Orchestrate VEGF Signaling during Developmental Angiogenesis. *Dev Cell,* 42**,** 462-478 e7.

WAREING, S., ELIADES, A., LACAUD, G. & KOUSKOFF, V. 2012. ETV2 expression marks blood and endothelium precursors, including hemogenic endothelium, at the onset of blood development. *Dev Dyn,* 241**,** 1454-64.

WASSERMAN, W. W. & SANDELIN, A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet,* 5**,** 276-87.

WATSON, J. D. & CRICK, F. H. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature,* 171**,** 737-8.

WEI, Y., MA, D., GAO, Y., ZHANG, C., WANG, L. & LIU, F. 2014. Ncor2 is required for hematopoietic stem cell emergence by inhibiting Fos signaling in zebrafish. *Blood,* 124**,** 1578-85.

WEIRAUCH, M. T., YANG, A., ALBU, M., COTE, A. G., MONTENEGRO-MONTERO, A., DREWE, P., NAJAFABADI, H. S., LAMBERT, S. A., MANN, I., COOK, K., ZHENG, H., GOITY, A., VAN BAKEL, H., LOZANO, J. C., GALLI, M., LEWSEY, M. G., HUANG, E., MUKHERJEE, T., CHEN, X., REECE-HOYES, J. S., GOVINDARAJAN, S., SHAULSKY, G., WALHOUT, A. J. M., BOUGET, F. Y., RATSCH, G., LARRONDO, L. F., ECKER, J. R. & HUGHES, T. R. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell,* 158**,** 1431-1443.

WHYTE, W. A., ORLANDO, D. A., HNISZ, D., ABRAHAM, B. J., LIN, C. Y., KAGEY, M. H., RAHL, P. B., LEE, T. I. & YOUNG, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell,* 153**,** 307-19.

WIDMANN, C., GIBSON, S., JARPE, M. B. & JOHNSON, G. L. 1999. Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol Rev,* 79**,** 143-80.

WILEY, S. R., KRAUS, R. J. & MERTZ, J. E. 1992. Functional binding of the "TATA" box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. *Proc Natl Acad Sci U S A,* 89**,** 5814-8.

WILKINSON, A. C., GOODE, D. K., CHENG, Y. H., DICKEL, D. E., FOSTER, S., SENDALL, T., TIJSSEN, M. R., SANCHEZ, M. J., PENNACCHIO, L. A., KIRKPATRICK, A. M. & GOTTGENS, B. 2013. Single site-specific integration targeting coupled with embryonic stem cell differentiation provides a high-throughput alternative to in vivo enhancer analyses. *Biol Open,* 2**,** 1229-38.

WILKINSON, R. N., POUGET, C., GERING, M., RUSSELL, A. J., DAVIES, S. G., KIMELMAN, D. & PATIENT, R. 2009. Hedgehog and Bmp polarize hematopoietic stem cell emergence in the zebrafish

dorsal aorta. *Dev Cell,* 16**,** 909-16.
WILLCOCKSON, M. A., TAYLOR, S. J., GHOSH, S., HEALTON, S. E., WHEAT, J. C., WILSON, T. J., STEIDL, U. & SKOULTCHI, A. I. 2019. Runx1 promotes murine erythroid progenitor proliferation and inhibits differentiation by preventing Pu.1 downregulation. *Proc Natl Acad Sci U S A,* 116**,** 17841-17847.
WILLIAMS, R. L., HILTON, D. J., PEASE, S., WILLSON, T. A., STEWART, C. L., GEARING, D. P., WAGNER, E. F., METCALF, D., NICOLA, N. A. & GOUGH, N. M. 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature,* 336**,** 684-7.
WILSON, B. G. & ROBERTS, C. W. 2011. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer,* 11**,** 481-92.
WILSON, N. K., SCHOENFELDER, S., HANNAH, R., SANCHEZ CASTILLO, M., SCHUTTE, J., LADOPOULOS, V., MITCHELMORE, J., GOODE, D. K., CALERO-NIETO, F. J., MOIGNARD, V., WILKINSON, A. C., JIMENEZ-MADRID, I., KINSTON, S., SPIVAKOV, M., FRASER, P. & GOTTGENS, B. 2016. Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model. *Blood,* 127**,** e12-23.
WINGETT, S., EWELS, P., FURLAN-MAGARIL, M., NAGANO, T., SCHOENFELDER, S., FRASER, P. & ANDREWS, S. 2015. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res,* 4**,** 1310.
WINKLER, M. E., MAURITZ, C., GROOS, S., KISPERT, A., MENKE, S., HOFFMANN, A., GRUH, I., SCHWANKE, K., HAVERICH, A. & MARTIN, U. 2008. Serum-free differentiation of murine embryonic stem cells into alveolar type II epithelial cells. *Cloning and Stem Cells,* 10**,** 49-64.
WOO, C. J., KHARCHENKO, P. V., DAHERON, L., PARK, P. J. & KINGSTON, R. E. 2010. A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell,* 140**,** 99-110.
WOOD, C. & TONEGAWA, S. 1983. Diversity and joining segments of mouse immunoglobulin heavy chain genes are closely linked and in the same orientation: implications for the joining mechanism. *Proc Natl Acad Sci U S A,* 80**,** 3030-4.
WOODCOCK, C. L. & DIMITROV, S. 2001. Higher-order structure of chromatin and chromosomes. *Curr Opin Genet Dev,* 11**,** 130-5.
WU, X. & ZHANG, Y. 2017. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet,* 18**,** 517-534.
XU, J., POPE, S. D., JAZIREHI, A. R., ATTEMA, J. L., PAPATHANASIOU, P., WATTS, J. A., ZARET, K. S., WEISSMAN, I. L. & SMALE, S. T. 2007. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc Natl Acad Sci U S A,* 104**,** 12377-82.
XU, M. J., MATSUOKA, S., YANG, F. C., EBIHARA, Y., MANABE, A., TANAKA, R., EGUCHI, M., ASANO, S., NAKAHATA, T. & TSUJI, K. 2001. Evidence for the presence of murine primitive megakaryocytopoiesis in the early yolk sac. *Blood,* 97**,** 2016-22.
YANG, X. J. & SETO, E. 2007. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene,* 26**,** 5310-8.
YANG, X. J. & SETO, E. 2008. The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat Rev Mol Cell Biol,* 9**,** 206-18.
YIN, F., YU, J., ZHENG, Y., CHEN, Q., ZHANG, N. & PAN, D. 2013. Spatial organization of Hippo signaling at the plasma membrane mediated by the tumor suppressor Merlin/NF2. *Cell,* 154**,** 1342-55.
YIN, Y., MORGUNOVA, E., JOLMA, A., KAASINEN, E., SAHU, B., KHUND-SAYEED, S., DAS, P. K., KIVIOJA, T., DAVE, K., ZHONG, F., NITTA, K. R., TAIPALE, M., POPOV, A., GINNO, P. A., DOMCKE, S., YAN, J., SCHUBELER, D., VINSON, C. & TAIPALE, J. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science,* 356.
YUDKOVSKY, N., RANISH, J. A. & HAHN, S. 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature,* 408**,** 225-9.

YUE, F., CHENG, Y., BRESCHI, A., VIERSTRA, J., WU, W., RYBA, T., SANDSTROM, R., MA, Z., DAVIS, C., POPE, B. D., SHEN, Y., PERVOUCHINE, D. D., DJEBALI, S., THURMAN, R. E., KAUL, R., RYNES, E., KIRILUSHA, A., MARINOV, G. K., WILLIAMS, B. A., TROUT, D., AMRHEIN, H., FISHER-AYLOR, K., ANTOSHECHKIN, I., DESALVO, G., SEE, L. H., FASTUCA, M., DRENKOW, J., ZALESKI, C., DOBIN, A., PRIETO, P., LAGARDE, J., BUSSOTTI, G., TANZER, A., DENAS, O., LI, K., BENDER, M. A., ZHANG, M., BYRON, R., GROUDINE, M. T., MCCLEARY, D., PHAM, L., YE, Z., KUAN, S., EDSALL, L., WU, Y. C., RASMUSSEN, M. D., BANSAL, M. S., KELLIS, M., KELLER, C. A., MORRISSEY, C. S., MISHRA, T., JAIN, D., DOGAN, N., HARRIS, R. S., CAYTING, P., KAWLI, T., BOYLE, A. P., EUSKIRCHEN, G., KUNDAJE, A., LIN, S., LIN, Y., JANSEN, C., MALLADI, V. S., CLINE, M. S., ERICKSON, D. T., KIRKUP, V. M., LEARNED, K., SLOAN, C. A., ROSENBLOOM, K. R., LACERDA DE SOUSA, B., BEAL, K., PIGNATELLI, M., FLICEK, P., LIAN, J., KAHVECI, T., LEE, D., KENT, W. J., RAMALHO SANTOS, M., HERRERO, J., NOTREDAME, C., JOHNSON, A., VONG, S., LEE, K., BATES, D., NERI, F., DIEGEL, M., CANFIELD, T., SABO, P. J., WILKEN, M. S., REH, T. A., GISTE, E., SHAFER, A., KUTYAVIN, T., HAUGEN, E., DUNN, D., REYNOLDS, A. P., NEPH, S., HUMBERT, R., HANSEN, R. S., DE BRUIJN, M., et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature,* 515**,** 355-64.

YUN, M., WU, J., WORKMAN, J. L. & LI, B. 2011. Readers of histone modifications. *Cell Res,* 21**,** 564-78.

ZAIDAN, N., NITSCHE, L., DIAMANTI, E., HANNAH, R., FIDANZA, A., WILSON, N. K., FORRESTER, L. M., GOTTGENS, B. & OTTERSBACH, K. 2022. Endothelial-specific Gata3 expression is required for hematopoietic stem cell generation. *Stem Cell Reports,* 17**,** 1788-1798.

ZAIDAN, N. & OTTERSBACH, K. 2018. The multi-faceted role of Gata3 in developmental haematopoiesis. *Open Biol,* 8.

ZAMBIDIS, E. T., OBERLIN, E., TAVIAN, M. & PEAULT, B. 2006. Blood-forming endothelium in human ontogeny: lessons from in utero development and embryonic stem cell culture. *Trends Cardiovasc Med,* 16**,** 95-101.

ZARET, K. S. & CARROLL, J. S. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev,* 25**,** 2227-41.

ZHANG, C. C. & LODISH, H. F. 2008. Cytokines regulating hematopoietic stem cell function. *Curr Opin Hematol,* 15**,** 307-11.

ZHANG, J., KLOS, M., WILSON, G. F., HERMAN, A. M., LIAN, X., RAVAL, K. K., BARRON, M. R., HOU, L., SOERENS, A. G., YU, J., PALECEK, S. P., LYONS, G. E., THOMSON, J. A., HERRON, T. J., JALIFE, J. & KAMP, T. J. 2012. Extracellular matrix promotes highly efficient cardiac differentiation of human pluripotent stem cells: the matrix sandwich method. *Circ Res,* 111**,** 1125-36.

ZHANG, W. & LIU, H. T. 2002. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res,* 12**,** 9-18.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol,* 9**,** R137.

ZHANG, Z., FU, J. & GILMOUR, D. S. 2005. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev,* 19**,** 1572-80.

ZHAO, B., YE, X., YU, J., LI, L., LI, W., LI, S., YU, J., LIN, J. D., WANG, C. Y., CHINNAIYAN, A. M., LAI, Z. C. & GUAN, K. L. 2008. TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev,* 22**,** 1962-71.

ZHENG, Y., WANG, W., LIU, B., DENG, H., USTER, E. & PAN, D. 2015. Identification of Happyhour/MAP4K as Alternative Hpo/Mst-like Kinases in the Hippo Kinase Cascade. *Dev Cell,* 34**,** 642-55.

ZHOU, X. & MA, H. 2008. Evolutionary history of histone demethylase families: distinct evolutionary

patterns suggest functional divergence. *BMC Evol Biol,* 8**,** 294.

ZHU, Q., GAO, P., TOBER, J., BENNETT, L., CHEN, C., UZUN, Y., LI, Y., HOWELL, E. D., MUMAU, M., YU, W., HE, B., SPECK, N. A. & TAN, K. 2020. Developmental trajectory of prehematopoietic stem cell formation from endothelium. *Blood,* 136**,** 845-856.

ZIEGLER, B. L., VALTIERI, M., PORADA, G. A., DE MARIA, R., MULLER, R., MASELLA, B., GABBIANELLI, M., CASELLA, I., PELOSI, E., BOCK, T., ZANJANI, E. D. & PESCHLE, C. 1999. KDR receptor: a key marker defining hematopoietic stem cells. *Science,* 285**,** 1553-8.

ZLATANOVA, J., LEUBA, S. H. & VAN HOLDE, K. 1999. Chromatin structure revisited. *Crit Rev Eukaryot Gene Expr,* 9**,** 245-55.

ZOVEIN, A. C., HOFMANN, J. J., LYNCH, M., FRENCH, W. J., TURLO, K. A., YANG, Y., BECKER, M. S., ZANETTA, L., DEJANA, E., GASSON, J. C., TALLQUIST, M. D. & IRUELA-ARISPE, M. L. 2008. Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell,* 3**,** 625-36.

ZUIN, J., DIXON, J. R., VAN DER REIJDEN, M. I., YE, Z., KOLOVOS, P., BROUWER, R. W., VAN DE CORPUT, M. P., VAN DE WERKEN, H. J., KNOCH, T. A., VAN, I. W. F., GROSVELD, F. G., REN, B. & WENDT, K. S. 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A,* 111**,** 996-1001.

ZUIN, J., ROTH, G., ZHAN, Y., CRAMARD, J., REDOLFI, J., PISKADLO, E., MACH, P., KRYZHANOVSKA, M., TIHANYI, G., KOHLER, H., EDER, M., LEEMANS, C., VAN STEENSEL, B., MEISTER, P., SMALLWOOD, S. & GIORGETTI, L. 2022. Nonlinear control of transcription through enhancer-promoter interactions. *Nature,* 604**,** 571-577.

nature communications

Article

https://doi.org/10.1038/s41467-023-35910-9

A genome-wide relay of signalling-responsive enhancers drives hematopoietic specification

Received: 11 March 2022

Accepted: 6 January 2023

Published online: 17 January 2023

Check for updates

B. Edginton-White[1,4] ✉, A. Maytum[1,4], S. G. Kellaway[1], D. K. Goode[2], P. Keane[1], I. Pagnuco[3,1], S. A. Assi[1], L. Ames[1], M. Clarke[1], P. N. Cockerill[1], B. Göttgens[2], J. B. Cazier[1,3] & C. Bonifer[1] ✉

Developmental control of gene expression critically depends on distal cis-regulatory elements including enhancers which interact with promoters to activate gene expression. To date no global experiments have been conducted that identify their cell type and cell stage-specific activity within one developmental pathway and in a chromatin context. Here, we describe a high-throughput method that identifies thousands of differentially active cis-elements able to stimulate a minimal promoter at five stages of hematopoietic progenitor development from embryonic stem (ES) cells, which can be adapted to any ES cell derived cell type. We show that blood cell-specific gene expression is controlled by the concerted action of thousands of differentiation stage-specific sets of cis-elements which respond to cytokine signals terminating at signalling responsive transcription factors. Our work provides an important resource for studies of hematopoietic specification and highlights the mechanisms of how and where extrinsic signals program a cell type-specific chromatin landscape driving hematopoietic differentiation.

The blueprint for the developmental regulation of gene expression is encoded in our genome in the form of cis-regulatory elements that exist as nuclease hypersensitive sites in chromatin. These elements are scattered over large distances and integrate multiple intrinsic and extrinsic signals regulating the activity of transcription factors (TFs) that bind to such elements[1,2]. TFs and TF encoding genes together with their targets form gene regulatory networks (GRNs) that define the identity of a cell. TFs together with chromatin remodellers/modifiers form multimolecular complexes that assemble on cis-regulatory elements and interact with each other within intranuclear space to activate gene expression[2]. To answer the question of how one GRN transits into another in development, it is essential (i) to identify and characterize the full complement of cell type and cell stage-specific transcription regulatory elements, (ii) to identify the TFs binding to them and (iii) to understand how they respond to external cues.

For several decades, reporter gene assays have been used to define cis-regulatory elements as enhancers or promoters, with enhancers being able to increase transcription from promoters independent of their orientation[3,4]. However, assessing enhancer activity within a chromatin context is more difficult. Studies inserting individual enhancer-promoter combinations at different genomic locations revealed that the local chromatin environment strongly influences gene expression. Such effects can typically only be overcome if the full complement of cis-regulatory elements is present on a transgene, making the analysis of individual elements difficult as their deletion make transgenes again

[1]Institute of Cancer and Genomic Sciences, School of Medicine and Dentistry, University of Birmingham, B152TT Birmingham, UK. [2]Department of Haematology, Wellcome and Medical Research Council Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge CB2 0AW, UK. [3]Centre for Computational Biology, Institute of Cancer and Genomic Sciences, University of Birmingham, B152TT Birmingham, UK. [4]These authors contributed equally: B. Edginton-White, A. Maytum. ✉e-mail: B.Edginton-White@bham.ac.uk; c.bonifer@bham.ac.uk

susceptible to position effects as reviewed in[5]. The deletion of individual elements within a gene locus can uncover enhancer function, but often misses developmental stage-specific elements, at least in part due to functional redundancy with neighbouring elements. Therefore, multiple surrogate markers have been identified that correlate with a high activity of the gene linked to the respective cis-regulatory element, including DNaseI hypersensitivity, TF binding, histone acetylation/mono-methylation and enhancer transcription[3,6–11]. None of these features, alone or in combination was fully predictive of enhancer activity with TF binding being the best predictor[12,13]. Consequently, functional assays remain essential to ascertain whether any given element can stimulate transcription in a chromatin context.

During embryonic development, definitive blood cells including hematopoietic stem cells develop from mesoderm derived endothelial cells within the dorsal aorta[14,15]. The specification of hematopoietic cells in the embryo and hematopoietic cell differentiation in the adult have served as an important model to reveal general principles of the control of gene expression in mammalian development[16]. The roles of the most important TFs together with signals such as cytokines controlling different developmental stages are known, and most intermediate cell types have been identified. Moreover, in vitro differentiation of human and mouse embryonic stem cells (ESCs) recapitulates embryonic hematopoietic development, thus facilitating deep molecular analysis into the developmentally-controlled transition of gene regulatory networks (GRNs)[17].

We previously reported a multi-omics analysis revealing dynamic GRNs that are specific for each of the major stages of blood cell specification from ESCs. We identified the locations and dynamic activities of sets of cis-regulatory elements associated with developmental gene regulation and based on this data uncovered important pathways, such as Hippo signalling that are required for blood cell formation[18–20]. However, major questions are still open. Whilst we could correlate chromatin alterations with dynamic gene expression[21], our data did not provide functional evidence for which cis-regulatory elements have enhancer activity, how they are controlled at different developmental stages, how extrinsic signals control their activity and importantly, which TFs mediate signalling responsiveness.

In the work presented here, we describe the development of a high-throughput method identifying thousands of enhancer and promoter elements specifically active at defined stages of blood cell specification in a chromatin environment and correlate their activity with gene expression in the same cells. We are using mouse ESCs to be able to integrate our results with our previously global multi-omics collected data, but the method can be expanded to any cell type that can be differentiated from ESCs and can easily be adapted to human ESCs. We show that the same elements exist as active chromatin in vivo in the appropriate mouse cell types. Finally, we identify cytokine responsive enhancer elements, and for one cytokine, VEGF, characterize the TFs mediating its activity within the GRN driving blood cell development. Our work provides a tool to significantly advance our understanding of developmental gene expression control in the hematopoietic system and beyond.

Results

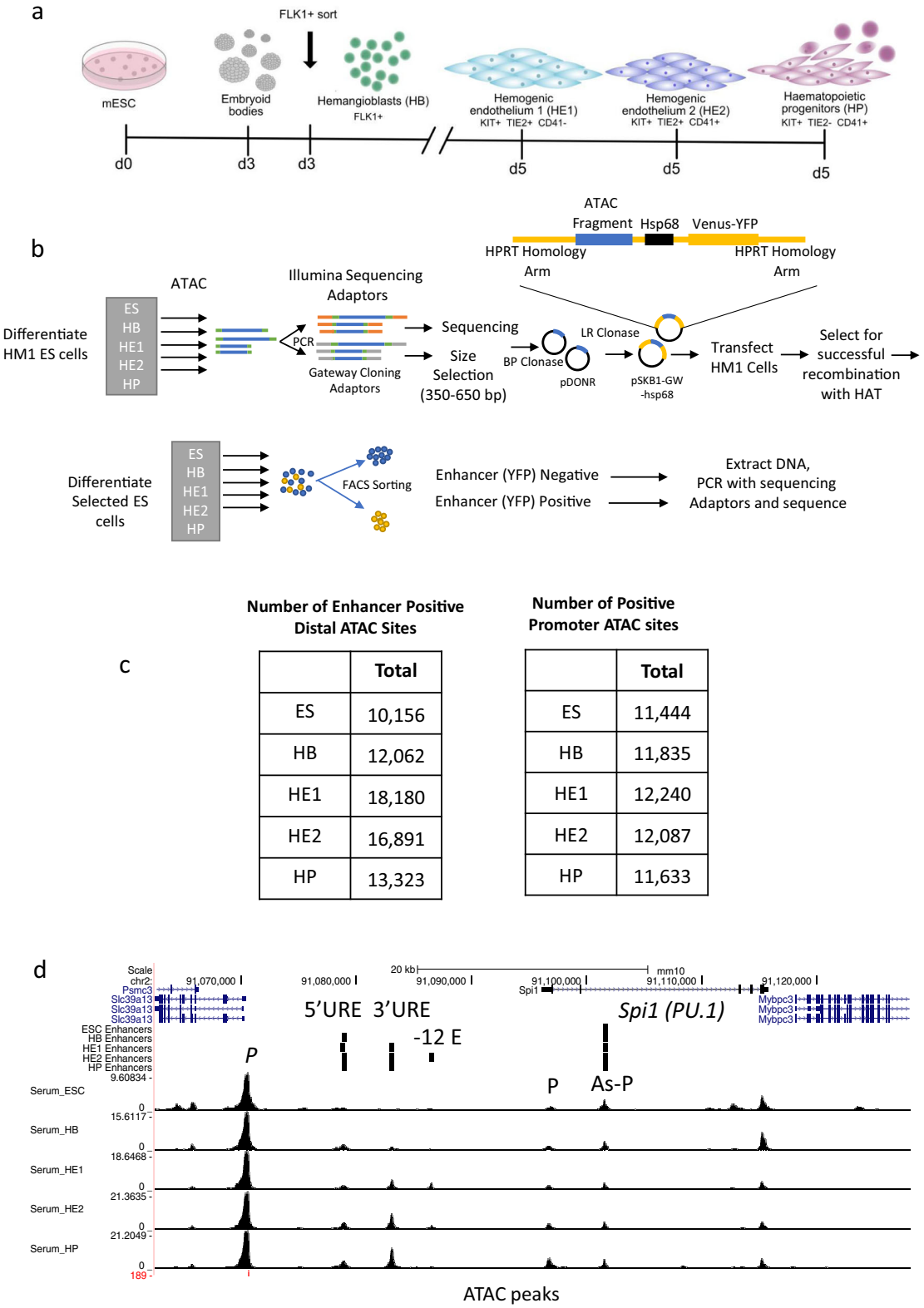
Establishing a high-throughput method for functional enhancer testing in a chromatin environment

We identified functional enhancer elements in the chromatin of mouse embryonic stem cells and their differentiated progeny representing different stages of hematopoietic specification (Fig. 1a)[18,19]. The first stage analysed here consists of FLK1-expressing hemangioblast-like cells[22] which have the ability to differentiate into cardiac, endothelial and hematopoietic cells and which are purified from embryoid bodies (EBs) after day 3 of culture. These cells are then placed into blast culture and form a mixture of (i) hemogenic endothelium 1 (HE1) which expresses a low level of RUNX1, (ii) hemogenic endothelium 2 (HE2) cells which up-regulate RUNX1 and CD41 but are still adherent and (iii) blast-like hematopoietic progenitor cells (HP) that underwent the endothelial-hematopoietic transition (EHT) and float off into the culture medium. Using the differential expression of specific surface markers (KIT, TIE-2 and CD41) each cell type can be purified to near homogeneity.

Figure 1 and Supplementary Fig. 1 show the enhancer identification pipeline which is based on the system developed by Wilkinson et al. 2013[23,24] and was adapted for a high-throughput genome-wide screen. Essentially, we differentiate cells, sort them into different developmental stages as shown in Fig. 1a, purify ATAC-Seq fragments for each differentiation stage, clone them into a targeting vector to generate a fragment library which is then integrated into a defined target site in the *HPRT* locus carrying a minimal promoter to drive a reporter gene (Venus-YFP) (Fig. 1b). By employing this targeting system, we ensure that only one fragment and reporter construct is active in each cell and that it is in an accessible chromatin environment throughout differentiation. We then differentiate cells and purify cells from each stage of development for reporter activity measurements (Supplementary Fig. 1a, b). Cell populations expressing high, medium and low YFP levels are isolated together with YFP negative cells (Supplementary Fig. 1b). Fragment inserts are sequenced after amplification using barcoded primers recognising the ATAC linkers. Each unique aligned read in the sequencing data is representative of a fragment that was cloned into the reporter construct and is assigned activity based on which FACS (YFP) population it was sorted into. Sequences then undergo a rigorous filtering against different criteria as detailed in Supplementary Fig. 1c (discussed in more detail in Supplementary Notes). The original plasmid libraries cover between 90% and 98% of all ATAC-Seq peaks (Supplementary Fig. 1d). Our screen was conducted in two replicates identifying several hundred-thousand fragments with transcription-stimulatory activity (Supplementary Fig. 1f). 22–31% of fragments were located within annotated distal elements (Supplementary Fig. 1d), covering more than 70,000 enhancer-positive ATAC sites across all stages (Fig. 1c left panel, Supplementary Fig. 1e, Supplementary Dataset 1). The remaining fragments were promoter sequences which were classified as being within 1.5 kb of an annotated transcription start site (Fig. 1c, right panel, Supplementary Fig. 1e). An example for enhancer annotation is shown in Fig. 1d, depicting the well-characterized Spi*1* (PU.1) locus, which captures all previously identified enhancer elements together with their known stage-specific activity[25] with other examples shown in Supplementary Notes. Most ATAC-fragments displaying stimulatory activity in our reporter assay (see scheme in Supplementary Fig. 2a, Supplementary Fig. 1e) overlapped with fragments within the same ATAC site that did not score, indicating that the vast majority of captured open chromatin regions can regulate transcription. Moreover, between 30% and 50% of all distal ATAC sites and around 80% of all promoter sites in the ATAC library contained a fragment scoring in our assay (Fig. 2a; Supplementary Fig. 2b). The median number of positive fragments per distal ATAC site was between 3 and 6 (Supplementary Fig. 1g).

We next integrated our enhancer data with previously published chromatin immunoprecipitation (ChIP)-Seq data characterizing histone modifications at cis-regulatory elements in the same experimental system[18]. About 30% to 60% of all enhancer sites overlap with H3K27Ac regions (Fig. 2b, Supplementary Fig. 2c). We find a significant overlap of our positive but not our negative/unknown enhancer ATAC sites with the VISTA enhancer database which describes 1061 functionally identified enhancers[26] (Fig. 2c, Supplementary Fig. 2d). The size distribution of positive and negative (non-scoring) fragments was the same (Supplementary Fig. 2e) indicating an absence of size selection for active fragments. Most enhancer positive ATAC fragments

https://doi.org/10.1038/s41467-023-35910-9

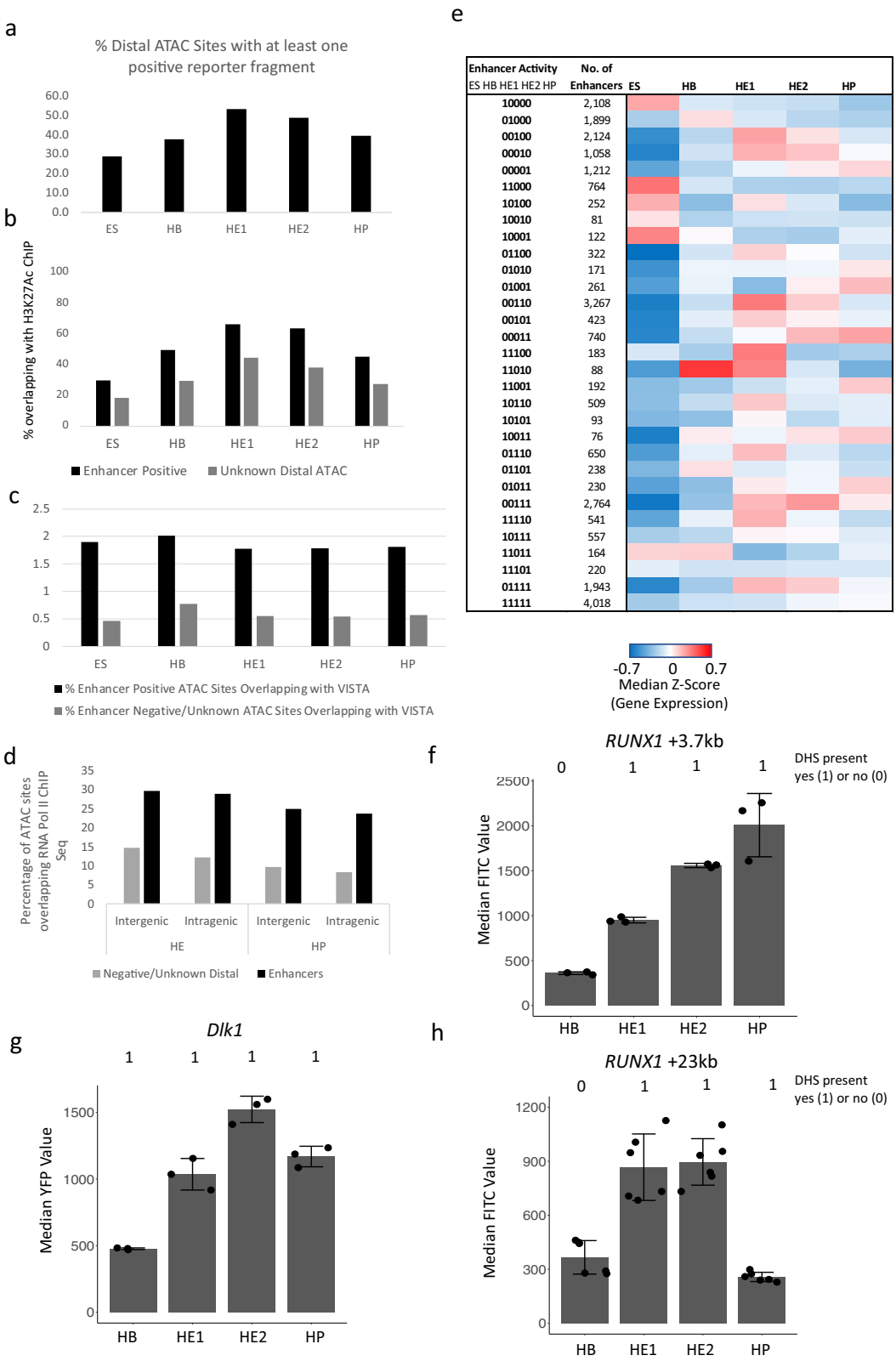


	Total
ES	10,156
HB	12,062
HE1	18,180
HE2	16,891
HP	13,323

	Total
ES	11,444
HB	11,835
HE1	12,240
HE2	12,087
HP	11,633

Fig. 1 | Establishing a high-throughput method for functional enhancer testing in a chromatin environment. a Depiction of the ES cell differentiation system and the cell types analysed. **b** Overview of the screening procedure using ATAC-Seq fragments from 5 different cell stages. **c** Total number of distal and promoter ATAC-sites scoring positive in two replicates of the assay. **d** UCSC Browser screenshot showing an example of well characterized enhancers (5'URE, 3'URE, −12 kb enhancer, promoter (P) and antisense promoter (As-P) from the SPI*1* (PU.1) locus[25] marked by ATAC-Seq peaks which were recovered in our screen. Distal cis regulatory elements active at the different stages scoring positive in our assay (enhancers) are indicated by vertical bars.

https://doi.org/10.1038/s41467-023-35910-9



overlap with open chromatin sites found in purified hemogenic endothelium and endothelial cells from day 9.5 and day 13.5 mouse embryos[27,28] indicating that they are active in vivo (Supplementary Fig. 2f). Finally, the comparison with previously collected TF ChIP data[18,20,29–31] shows that between 17% (in HB) and 76% (in HP cells) of enhancer fragments are bound by ubiquitous and differentiation stage-specific TFs, depending on the number of available ChIP experiments for each stage.

It was reported that active enhancers are bound by RNA-Polymerase II (Pol II) and are transcribed[8,32]. To examine the correlation between the ability of distal elements identified in our study to drive RNA production during development we examined whether they

Fig. 2 | Characterization of enhancer features and association of cell stage-specific enhancer activity with cell stage-specific gene expression. a Distal ATAC sites containing at least one fragment scoring positive in our assay. Source data are provided as a Source Data file. **b** Percentage of functionally identified distal elements scoring positive in our screen overlapping with histone H3 lysine 27 (H3K27Ac) peaks as compared to inactive elements or where activity is not known. ChIP data from ref. [18]. Source data are provided as a Source Data file. **c** Overlap of fragments scoring positive in our assay with known enhancers from the VISTA database[26]. Source data are provided as a Source Data file. **d** Percentage intergenic and intragenic fragments scoring positive in our assay overlapping with RNA polymerase II binding sites in the hemogenic endothelium and HP cells. Pol II data from ref. [30]. Source data are provided as a Source Data file. **e** Presence or absence of enhancer activity at the different developmental stages expressed as binary code (0 = scoring negative; 1 = scoring positive) and sorted by the activity pattern across differentiation. The heatmap depicts the activity of the genes associated with these elements. Enhancer - promoter association was determined by using the union of HiC data from ES and HPC7 (HP) cells[29] together with co-regulation data determined by ref. [21] in a total of 21,671 elements. All promoters not covered by these data (5599) were associated by being the nearest to the enhancer element. **f–h** Activity profiles of individual enhancers identified in our screen in single ES cell clones during differentiation. The presence or absence of an ATAC-peak is indicated by the binary code used in (**e**). **f** *RUNX1* + 3.7 enhancer element (chr16:92822182-92822601). **g** *Dlk1* enhancer element (Chr12: 109437601-109438085), **h** *RUNX1* + 23 kb enhancer element[49]. Data are presented as mean values +/− standard deviation (SD). Dots showing individual values for *Dlk1* and *Runx1* + 3.7 kb $n = 3$ biologically independent experiments and for *Runx1* + 23 kb $n = 6$ biologically independent experiments. For sequence details see Supplementary Notes. Source data are provided as a Source Data file.

were capable of binding Pol II using a previously published data-set from ESC generated HE and HP cells[30]. Figure 2d shows that up to 30% of identified enhancers are associated with Pol II binding.

Sequential stages of hematopoietic specification are defined by distinct enhancer sets

The genomic information for tissue-specific gene expression manifests itself in the activity pattern of distal regulatory elements[18,33]. An important feature of our method is therefore the identification of developmental stage-specific enhancer activity (Supplementary data-set 2). Between 10% and 20% of all distal and between 15% and 50% of all promoter sites showed stage specific activity in our assay (Supplementary Fig. 3a). To assign genes to regulatory elements, we used publicly available HiC and co-regulation data[21,29,34] and examined how gene expression correlated with enhancer activity by generating a binarized matrix cataloguing enhancers as active (1) and inactive (0) at each of the five differentiation stages (Fig. 2e, Supplementary data-sets 2, 3). This analysis which depicts gene expression at the different stages in a heatmap shows that (i) enhancer activity during development is largely continuous and (ii) stage-specific gene expression is strongly associated with stage specific enhancer activity. Examples for stage-specific active enhancers can be found in Fig. 2f–h and Supplementary Fig. 2g.

We then determined for each differentiation stage, which TF motif combinations were associated with enhancer activity (Fig. 3 and Supplementary Fig. 3). We first examined the motif content of distal ATAC-Seq sites at specific differentiation stages by defining distal elements specific for each stage[18]. In line with previously published ChIP data[18,20,29], cell stage-specific cis-element patterns are associated with stage-specific TF motifs (Fig. 3a), with those for hematopoietic TFs such as RUNX1 or PU.1 being enriched in HP cells and those for the HIPPO signalling mediator TEAD and SOX factors enriched in the HE. This pattern was also seen with stage-specifically active enhancer but not with promoter fragments (Supplementary Fig. 3b). In contrast, ubiquitously active enhancer fragments displayed a similar motif signature at all developmental stages, reinforcing the notion that tissue-specificity is encoded in distal elements (Supplementary Fig. 3c, left panel). The motif for the Zn ++ finger factor CTCF was enriched in active distal enhancer but not promoter fragments. It was recently shown in differentiating erythroid cells that dynamically bound CTCF cooperates with lineage-specific TFs bound to distal elements to interact with promoters[35]. Our data are consistent with this finding.

We next asked whether stage-specific enhancer activity was correlated with a pattern of cooperating TFs. To this end, for each developmental stage we performed a motif co-localization analysis which examines whether specific binding motif pairs located within 50 bp of each other were enriched in stage-specific enhancer fragments as compared to all open chromatin sites (analysis scheme depicted in Supplementary Fig. 3d). Motifs with a 100% overlap were removed. Stage-specific enhancer activity correlated with enriched colocalizations of motifs for developmental-stage specific TFs (Fig. 3b–f). In line with their generally open chromatin structure[36], ESC-specific enhancers showed a great variety of paired motifs such as those for the pluripotency factors NANOG/SOX2/OCT4. Interestingly, AP-1 motifs showed a very high co-occurrence both with itself and other motifs, potentially linking enhancer activity to signalling processes (Fig. 3b). AP-1 homo-typic motif associations were also found at HB-specific enhancer fragments but co-localization with pluripotency factor motifs was lost (Fig. 3c). The HE1 stage showed a strong enrichment in RBPJ, AP-1 and SMAD motif pairs, with AP-1 motifs colocalizing with high frequency with most other factors (Fig. 3d). The number of HE2-specific enhancers was low as it is a transitory stage. Here, PU.1 motifs show increased co-localization with SMAD and OCT motifs (Fig. 3e). This observation is well-supported by studies that TGFβ and BMP-induced SMADs often follow lineage-specific master transcription factors to cell stage-specific enhancers[37,38]. As expected from previous ChIP studies[39], we find a co-localization of motifs for hematopoietic TFs such as C/EBP, RUNX, PU.1 and GATA at the HP-stage. AP-1 motifs again co-localized with a variety of other motifs, including C/EBP, RUNX1 and PU.1 (Fig. 3f). Although generally enriched in enhancer fragments at all stages (Supplementary Fig. 3c), CTCF motifs were not significantly paired with any other motif.

Identification of cytokine-responsive enhancer elements

Cell differentiation involves extracellular signals which alter growth and differentiation states by changing gene expression. Signalling molecules such as cytokine receptors, integrins and kinase molecules are well characterized, but less is known of how different signals are integrated at the level of the genome. In spite of a number of efforts looking at specific genomic regions such as[40,41], we have limited global information about which cis-regulatory elements can respond to signals, which TF combinations are involved and how they cooperate to ensure that the genome responds to outside signals in a coordinated and balanced fashion. Our global cis-element collection allows us to answer these questions.

To this end, we employed a serum-free in vitro differentiation system[42] that is based on the sequential addition of growth factors such as BMP4, VEGF and hematopoietic cytokines. We tested how individual cytokines affected the differentiation profile and the open chromatin landscape of HE1, HE2 and HP cells sorted as described in Fig. 4a. The comparison of the chromatin signature of cells differentiated in serum and under serum-free conditions (Supplementary Fig. 4a), showed that around 80–90% of all promoters active in cells from serum-free culture overlapped in both conditions, demonstrating the reproducibility of our differentiation system. However, whilst the overlap for HB and HE1 cells was high (> 70%) for HP cells we noticed changes in the bulk open chromatin landscape which only affected the distal elements. This result indicates that although the cellular identity seemed to be largely preserved in sorted cells

https://doi.org/10.1038/s41467-023-35910-9

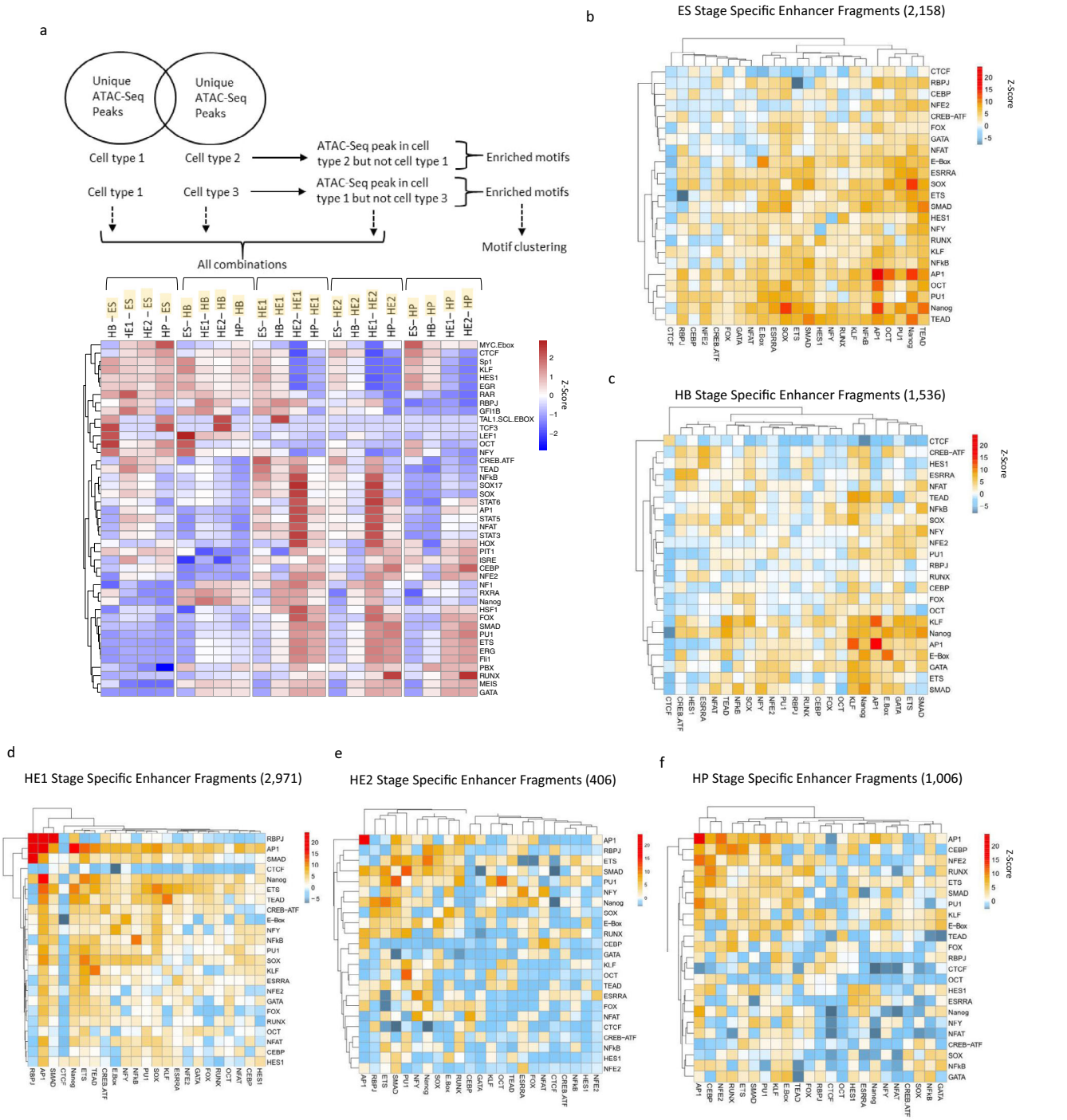


Fig. 3 | Stage-specific enhancer elements show a specific TF motif and motif co-localization pattern. a Identification of specific TF binding motifs in cell type specific open chromatin regions (distal elements only) as described in ref. [18]. The upper panel shows the filtering strategy. The lower panel shows how specific binding motifs are enriched in ATAC-Sites that are specific for each cell type (highlighted in yellow) as compared to another cell type (indicated by brackets). **b**–**f** Motif co-localization analysis. Heatmaps depicting Z-Score enrichments for pairs of TF binding motifs within 50 bp for five cell differentiation stages as indicated on top of each heat-map, together with the number of fragments analyzed. Binding motifs for the indicated TFs are listed on the right and the bottom of the heat-map. Position-weight matrices used for this analysis can be found in Supplementary Notes Table 1. The overall strategy is outlined in Supplementary Fig. 3d. Note that we do not expect the diagonal to be a uniform line. In the heatmap the diagonal represents the tendency for a motif to co-localize with itself which only yields a high score with groups of multiple of the same motif within 50 bp of each other, and if these groups occur more frequently compared to the background peak set. This feature will be different for each motif.

expressing the right combination of surface markers, the difference in the signalling environment exerted a strong effect on the chromatin landscape.

To elucidate the role of different cytokines in chromatin programming, we differentiated cells in the presence and absence of BMP4, VEGF, IL-6 and IL-3, respectively, sorted the different cell types and examined which open chromatin regions changed at least two-fold in response to cytokine removal (Fig. 4a). In total, more than 10,000 unique open chromatin regions were up or down regulated (Supplementary Fig. 4b) which included both distal elements and promoters.

https://doi.org/10.1038/s41467-023-35910-9

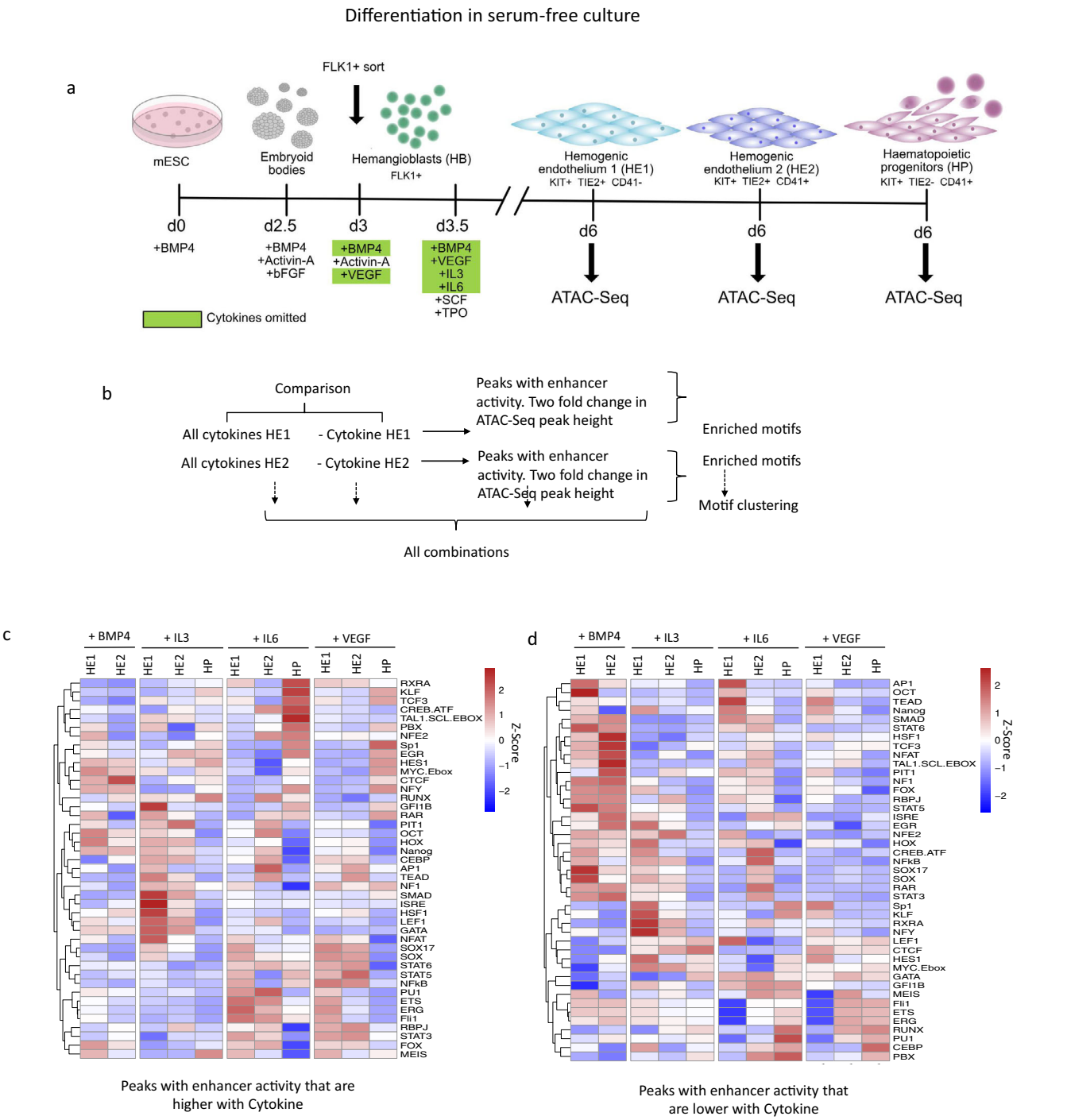
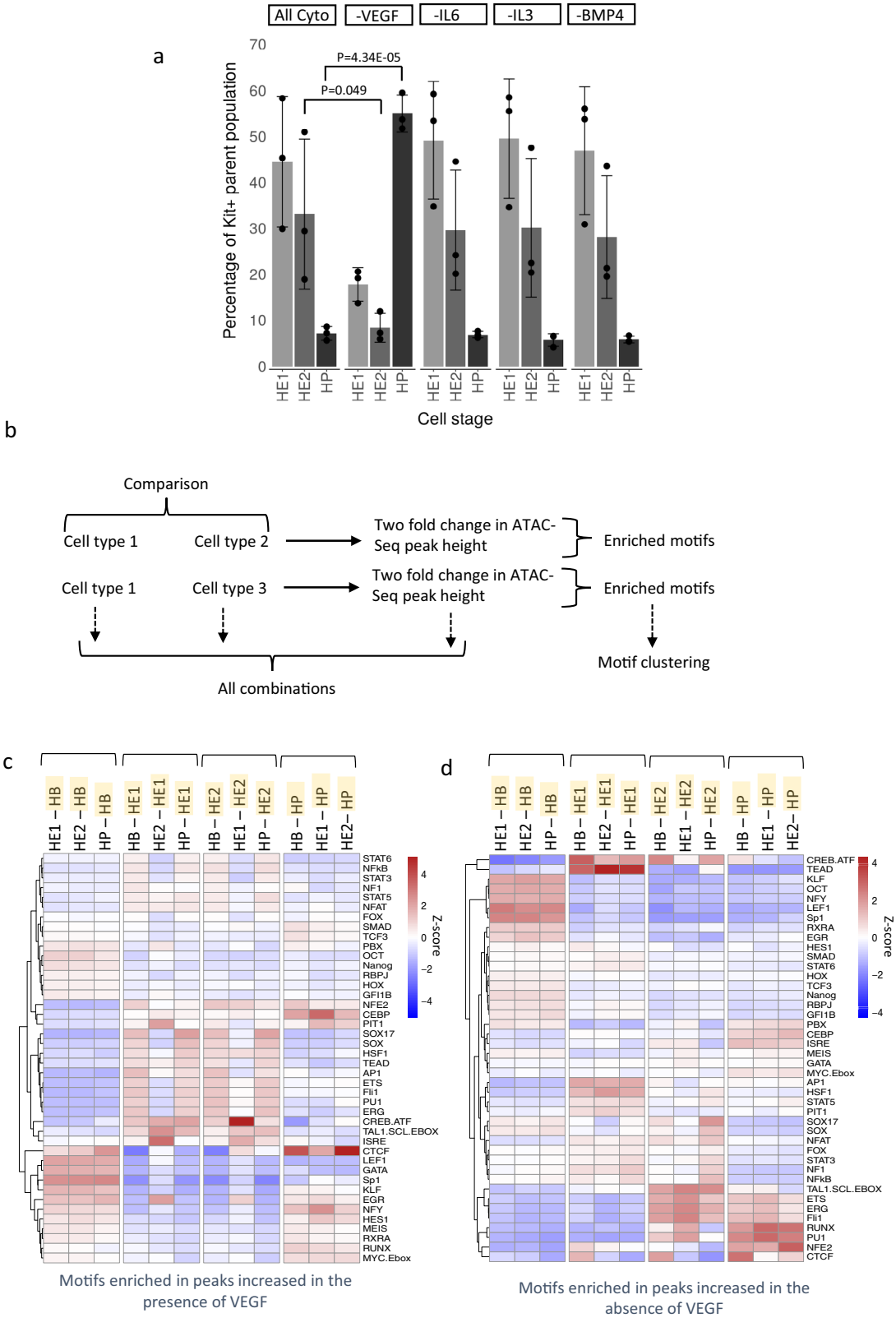


Fig. 4 | Identification of cytokine -responsive enhancer elements. a Overview of serum-free in vitro blood progenitor differentiation modified from[42]. Specific cytokines that were left out at the beginning of blast culture are highlighted in green. **b**–**d** Identification of cytokine responsive distal elements scoring positive in our assay in the presence or absence of the indicated cytokines. **b** Overview of motif enrichment analysis strategy, **c**, **d** TF binding motif enrichment in distal ATAC peaks containing fragments scoring positive in our assay that are increased (**c**) or decreased (**d**) at least 2-fold in the presence of the indicated cytokines. Position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Notes.

We next examined, which TF binding motifs were enriched in cytokine-responsive ATAC-Seq peaks with enhancer activity, i.e., peaks that were at least 2-fold higher or lower in the presence of cytokines (Fig. 4b–d). Relative enrichment scores were calculated for each binding motif which were transformed to a Z-Score, hierarchically clustered and plotted as a heatmap (see methods). The alteration of cytokine conditions had a profound influence on chromatin programming (Fig. 4c, d). The absence of BMP4 from the blast culture onwards was incompatible with HP formation and open chromatin regions with enhancer activity containing SMAD, HOX, RAR and NOTCH motif signatures were lost in HE (Fig. 4d). This finding is in keeping with these factors being required to form the HE[43,44]. The presence of VEGF led to a loss of peaks with a hematopoietic motif signature in HE2/HP, such as RUNX1, FLI1, GATA and PU.1 motifs. A similar motif enrichment pattern was found when all open chromatin regions were analysed (Supplementary Fig. 4c–e).

https://doi.org/10.1038/s41467-023-35910-9



To assess the effect of cytokine withdrawal on differentiation, we examined the effect of cytokine omission on the frequency of HE1, HE2 and HP cells (Fig. 5a). The presence or absence of BMP4, IL-6 and IL-3 did not influence the proportion of generated HE1, HE2 and HP cells as compared to the all-cytokine condition (Fig. 5a). In contrast, VEGF strongly suppressed hematopoietic progenitor formation.

The balance between hematopoietic and endothelial development is regulated by VEGF-responsive cis-regulatory elements

VEGF had the greatest influence on the formation of hematopoietic progenitor cells (Fig. 5a). We therefore examined its role in regulating enhancer activity in more detail. VEGF omission resulted in a significant decrease in the proportion of HE2 cells gained ($P = 0.049$) and a significant increase in the proportion of HP cells formed

Fig. 5 | The presence of VEGF suppresses multipotent progenitor (HP) development. a Proportion of HE1/HE2/HP cells within the differentiation culture in the absence of the indicated cytokines as measured by FACS. Data are presented as mean values +/− SD. Dots showing individual values for $n = 3$ biologically independent experiments. *P*-Values were calculated using two-sided Student's *t*-test. Source data are provided as a Source Data file. **b**–**d** Un-supervised motif clustering analysis examining enrichment of the TF binding motifs as described in Fig.3a, analyzing in all ATAC peaks in two culture conditions. **b** Motif enrichment strategy. **c** Motifs enriched in peaks of specific cell types (highlighted in yellow) which were at least 2-fold increased in peak height in the All Cytokine Condition, **d** Motifs enriched in peaks from specific cell types (highlighted in yellow) which were at least 2-fold increased in peak height in the VEGF omission condition. TF binding motifs are depicted at the right of the panel. Relative motif enrichment scores Z-score were calculated by columns. Position-weight matrices used for this analysis can be found in Supplementary Table 1 in Supplementary Notes.

($P = 4.34E-05$) at day 3 blast culture. VEGF omission cultures contained a smaller number of cells but an increased proportion of HP cells (Supplementary Fig. 5a) with VEGF withdrawal at around 12–14 h being optimal for the generation of these two cell types (Supplementary Fig. 5b). Moreover, VEGF removal at different time points of blast culture did not influence the proportion of HE1 cells but led to an inverse correlation of HE2 and HP cell numbers (Supplementary Fig. 5b), indicating that this cytokine impacted on the EHT. These results are consistent with previous observations showing that the receptor for VEGF, FLK1, is essential for the formation of blood islands from hemogenic endothelium cells[45] but once hematopoietic cells are formed, cells become dependent on hematopoietic cytokines. However, the molecular basis of this finding, i.e. which genomic events are responsible for this phenomenon has so far been unclear.

Our ATAC-Seq analysis found 7814 chromatin regions carrying enhancer elements that responded to VEGF (Supplementary Dataset 4). To identify VEGF responsive TFs, we repeated the supervised motif clustering analysis that highlighted cell type specific motif enrichments in the presence and absence of VEGF (Fig. 5b–d). HP cells in VEGF cultures maintain an enrichment of motifs for HES1 which is a mediator of NOTCH signalling[46] (Fig. 5c). In contrast, the omission of VEGF activates enhancers with a hematopoietic motif signature with RUNX1 and PU.1 motifs (Fig. 5d). Moreover, ATAC peaks in -VEGF HE1 cells were enriched in TEAD motifs together with binding motifs for factors linked to inflammatory signalling (AP-1, NFkB and CREB/ATF) which have been shown to be important for stem cell development[47]. A similar motif signature in -VEGF cultures was also seen when ATAC sites were directly compared in a pair-wise fashion (Supplementary Fig. 5c–e) with ATAC peaks derived from HE1 cells from +VEGF cultures enriched for TEAD (29.61% of targets, $P = 1e-193$), AP-1 (16.23% targets, $P = 1e-317$) and SOX (41.39% targets, $P = 1e-696$).

VEGF blocks the upregulation of *RUNX1* at the chromatin and gene expression level

The data shown above demonstrate that VEGF interferes with the EHT and blood progenitor formation. Both processes are crucially dependent on the TF RUNX1[19] which activates hematopoietic genes and represses endothelial genes in cooperation with the TF GFI1[48]. To examine why the EHT is deficient, we examined the chromatin structure of *Runx1* in the presence and absence of VEGF (Fig. 6a). In the presence of VEGF multiple distal DHSs of *Runx1* fail to form. 5 of these elements score in our enhancer assay, including the previously characterized + 23 kb enhancer (2)[49] together with an enhancer at + 3.7 kb (3) (Fig. 2f, h). Of note, several, of these elements, such as the + 3.7 kb enhancer are bound by TEAD and AP-1 as well (Supplementary Notes, Fig. 1a).

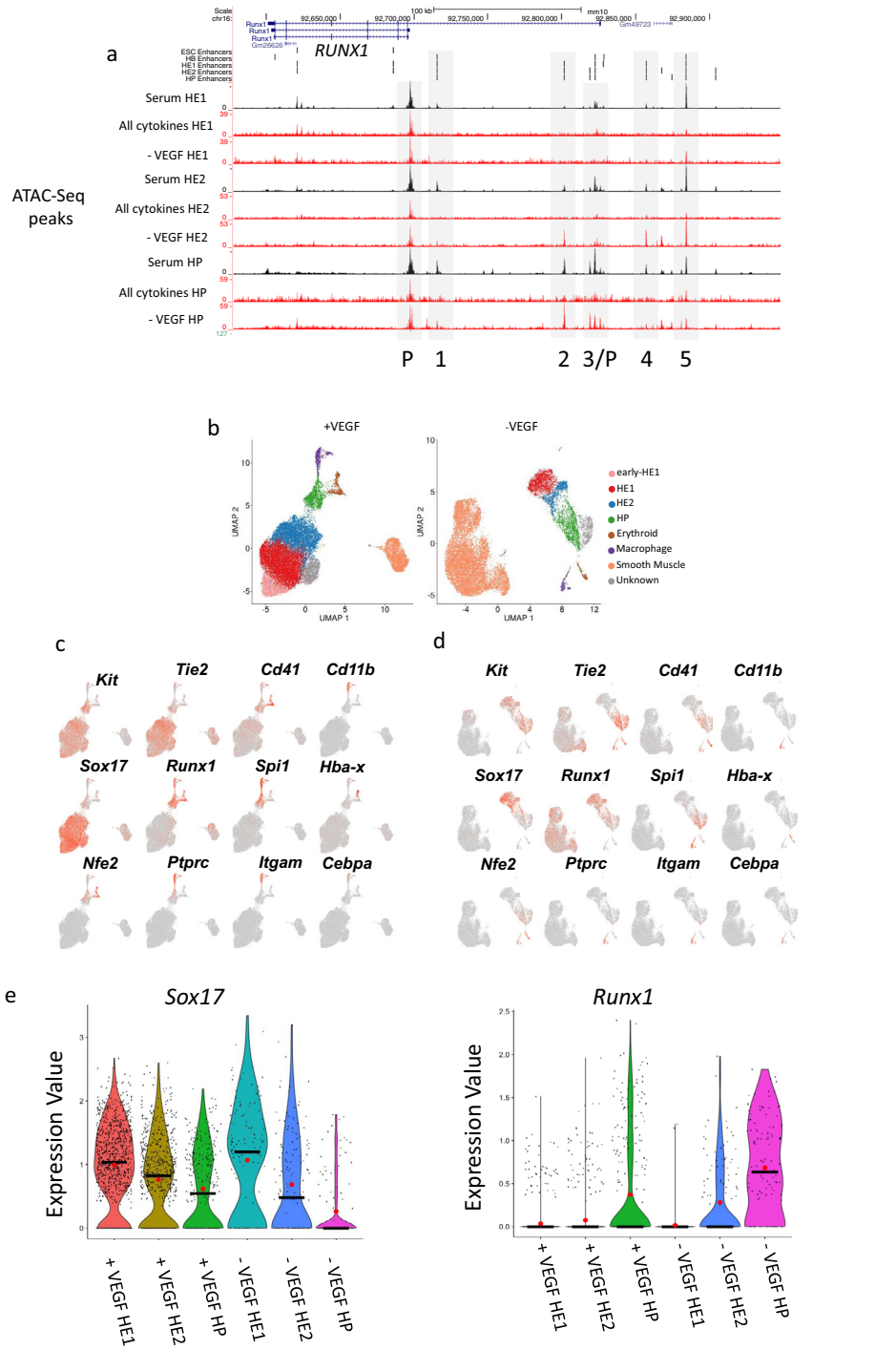
The results described so far suggested that the reduced ability of VEGF cultures to undergo the EHT was a result of a failure of *Runx1* enhancer activation and a failure of its transcriptional upregulation in HE2. Although we profiled chromatin in FACS purified cells, these alterations could still be caused by shifts in cell composition. We therefore studied VEGF-mediated changes in gene expression in HE1/2 cells at the single cell level (Fig. 6b–e, Supplementary Fig. 6a–e). Without VEGF the overall numbers of HE1/HE2/HP cells were reduced (Fig. 6b) but the population showed an increased proportion of HP cells, together with an increase in the proportion of smooth muscle cells (Fig. 6b). However, the cellular identity and the overall differentiation trajectory were not altered (Supplementary Fig. 6d). This shift was in concordance with an incomplete down-regulation of endothelial genes such as *Sox17* in HE1 (Avg Log2FC 1.07, P Val Adj < 1.38E-303) and in HE2 (Avg Log2FC 0.49, P Val Adj 2.18E-174) and Tie2 in HE1 (Avg log2FC 0.17, Adj P Val 1.05E-108) and in HE2 (Avg Log2FC 0.09, P Adj Val 3.55E-12) as seen by sc-RNA-seq differential gene expression analysis (Fig. 6c–e). sc-RNA-seq also revealed a lack of upregulation of *Runx1* in HE2 and HP cells cultured with VEGF (Fig. 6e, right panel). The balance between endothelial and hematopoietic gene expression is controlled by SOX17 which represses *Runx1*[50,51]. After the EHT, RUNX1 together with GFI1 binds to *Sox17* and *Notch1* enhancers (Supplementary notes, Figs. 2, 3)[18,48,52,53] forming a feed forward loop driving EHT progression. *Sox17* and *Runx1* are thus expressed in a mutually exclusive fashion with the former being high before the EHT[51] and then being downregulated and the latter being upregulated from a low level during the EHT[52]. This balance shifted after VEGF was omitted (Fig. 6d, e). We therefore conclude that VEGF signalling interferes with the activation of *Runx1* enhancers in the HE, driving gene expression required for the EHT.

To investigate whether VEGF-responsive enhancers were connected to VEGF-regulated genes, we paired each element that required VEGF activation in HE1, HE2 and HP with its respective gene as shown in Fig. 2e. We then integrated this data with their expression as measured in our single-cell experiments (Supplementary Fig. 6e, Supplementary dataset 4). This analysis uncovered that at high-stringency analysis, between 10% and 19% of VEGF responsive enhancers are linked to VEGF responsive up- or down-regulated genes, thus directly linking alteration in chromatin with changes in gene expression. The analysis of gene ontology (GO) terms of VEGF dependent genes highlighted angiogenesis as the top pathway with the VEGF receptor FLK1 (*Kdr)* being a prominent example ($P = 2.88E-13$, Supplementary dataset 4). VEGF-responsive down-regulated genes linked to down-regulated enhancer elements include hematopoietic regulator genes such as *Runx1, Klf2, Jun* and *Elf1*, validating our general approach (Supplementary Dataset 4).

The VEGF response involves the TEAD – AP-1 axis

We next examined which specific TFs were connected to endothelium-specific gene expression and VEGF responsiveness. Distal ATAC-Seq peaks specific for HE1 cells derived from cultures containing VEGF were enriched for SOX, E-BOX, AP-1 and TEAD TF motifs, indicating a distinct endothelial TF motif signature (Supplementary Fig. 5c). Similarly, distal ATAC-Seq peaks specific for HE2 cells derived from cultures containing VEGF were also enriched for the endothelial TF motifs SOX (36.91% targets $P = 1e-858$), TEAD (48.27% targets, $P = 1e-366$) and AP-1 (21.45% targets, $P = 1e-543$). We therefore examined several VEGF-responsive and endothelial specific enhancer sequences in more detail, which included the *Sparc, Pxn* and *Hspg2* enhancers (Fig. 7a–c, Supplementary Notes, Figs. 4–6). Inspection of the binding motifs of these elements uncovered that they contained SOX, AP-1, TEAD and the NOTCH-signalling responsive factor RBPJ, which is in concordance with the global HE signature (Fig. 5c), but also RUNX1 motifs.

https://doi.org/10.1038/s41467-023-35910-9



	+ VEGF HE1	+ VEGF HE2	+ VEGF HP	- VEGF HE1	- VEGF HE2	- VEGF HP
Sox17	2.26	1.66	1.27	3.23	1.82	0.52

Mean Normalised Expression Value

	+ VEGF HE1	+ VEGF HE2	+ VEGF HP	- VEGF HE1	- VEGF HE2	- VEGF HP
Runx1	0.051	0.125	0.838	0.032	0.492	1.300

Mean Normalised Expression Value

Fig. 6 | VEGF withdrawal shifts the proportion of endothelial/smooth muscle and hemogenic endothelium/blood progenitor cells by blocking *Runx1* up-regulation. a UCSC genome browser screenshot depicting open chromatin regions and the locations of distal elements scoring positive in our assay (vertical bars) in the *Runx1* locus under the indicated culture conditions (Serum culture, Serum-free with all cytokines and without VEGF (-VEGF)) and at the different developmental stages. Enhancers (numbered 1–5) and promoter elements (P) are indicated by blue boxes. Enhancer 2 is located at + 23 kb, enhancer 3 is at + 3.7 kb. **b** Single cell gene expression analysis with purified HE1 and HE2 populations in the presence (+ VEGF) and absence of VEGF (-VEGF). UMAP clustering analysis was performed after all data were pooled. **c**, **d** Gene expression levels of the indicated genes with (**c**) and without VEGF (**d**) are projected on the clusters. **e** Violin plots of expression values from 6 clusters for $n = 2266$ cells (black line representing the median, red dot the mean and black dots the spread of the data) and mean normalized (Seurat log normalized) gene expression levels of *Sox17* and *Runx1* in the indicated cell types in the presence and absence of VEGF calculated from single cell expression data.

https://doi.org/10.1038/s41467-023-35910-9

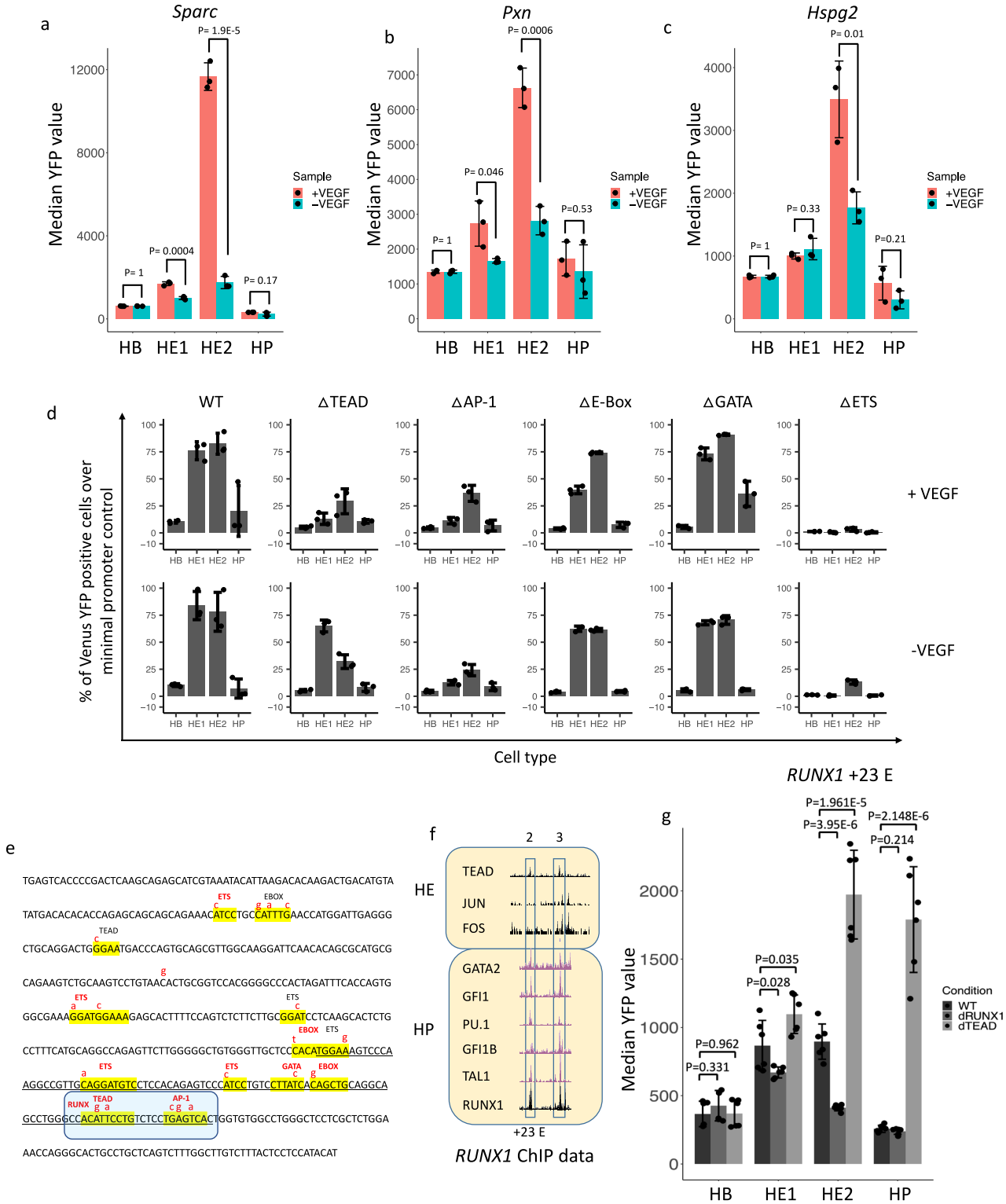


Fig. 7 | HE-specific and cytokine responsive gene expression is mediated by the interplay of TEAD, AP-1 and RUNX1. a–c Activity of the *Sparc, Pxn* and *Hspg2* enhancers (for details see Supplementary Notes) in isolation in the presence and absence of VEGF. Data are presented as mean values +/− SD. Dots showing individual values for $n = 3$ biologically independent experiments. *P*-values calculated using a two-sided Student's *t*-test. Source data are provided as a Source Data file. **d** HE-specific expression of *Galnt1* (data from ref. [18]), reporter gene activity driven by wild type (WT) and mutated *Galnt1* enhancer elements in the presence and absence of the indicated cytokines. Data are presented as mean values +/− SD. Dots showing individual values for $n = 3$ biologically independent experiments. **e** Sequence of the *Galnt1* enhancer with TF binding motif mutations indicated in red. Strong consensus sequences are highlighted in red. Source data are provided as a Source Data file. **f** Chip data showing the binding of the indicated transcription factors to *RUNX1* enhancers in HE and HP cells. **g** Effect of mutations of the TEAD and RUNX1 binding sites on the individual activity of the *RUNX1* + 23 kb enhancer in serum culture. Data are presented as mean values +/− SD. Dots showing individual values for $n = 6$ biologically independent experiments. *P*-values calculated using a two-sided Student's *t*-test. Source data are provided as a Source Data file.

We previously identified HIPPO signalling as being crucial for HE development[18] and others have shown that shear stress activating the TEAD partner YAP via RHO-GTP induces the formation of HSCs from the HE via activation of *Runx1*[54]. Our ChIP data show that AP-1 and TEAD bind to such motifs in the HE[20] but once HP cells have formed, TEAD binding is lost[18]. We previously reported that TEAD and AP-1 often show a specific spacing between binding sites where binding is interdependent[20]. The expression of a dominant negative (dn)FOS peptide blocking all AP-1 DNA binding reduced the binding of TEAD at such sites, suggesting that HIPPO and MAP kinase signalling interface at these elements. In HP cells, RUNX1 has been shown to repress the endothelial program[48,53], we were therefore interested in understanding the role of RUNX1 in regulating enhancer activity. To examine the role of these TFs in more detail, we identified enhancer elements which were (i) active specifically in the HE, (ii) were bound by several of the above-described factors and (iii) were VEGF responsive (based on Supplementary dataset 4). We created ES cell lines carrying such enhancer sequences and examined their activity in serum-free cultures with and without VEGF. The *Galnt1* enhancer fitted these criteria (Fig. 7d, e, Supplementary Fig. 7a–c). It contains a composite AP-1/TEAD motifs (highlighted in blue) which overlaps with a RUNX1 binding site, together with a number of other motifs (Fig. 7e) allowing us to study the interplay between these factors in response to VEGF. Our ChIP data confirmed their binding (Supplementary Fig. 7d). dnFOS induction reduced *Galnt1* mRNA expression and TEAD binding specifically in the HE (Supplementary Fig. 7a, b), indicating that AP-1 activates the element and is required for the binding of TEAD factors within this composite module. We therefore decided to study the role of these factors in more detail by mutating the different binding sites.

To examine the response of the *Galnt1* enhancer to VEGF at four differentiation stages (HB – HP), we first measured the activity of the intact enhancer element (Fig. 7d) as compared to the promoter control by using flow cytometry to assay the number of YFP positive cells. The activity profile of the enhancer mirrored the gene expression profile in the presence of all cytokines (Fig. 7d). We also measured median YFP florescence confirming the results (Supplementary Fig. 7e). The analysis of mutant enhancer elements revealed that the binding motifs for TEAD, ETS and AP-1 were the most important for endothelial-specific median reporter activity compared to the wild-type sequence activity in the HE whereas the mutation of others had no impact (Fig. 7d, Supplementary Fig. 7e). The mutation of the TEAD and the overlapping TEAD/RUNX1 binding sites led to an increase in the percentage of YFP + cells in HE1 cells, suggesting that here TEAD and RUNX1 are in balance to restrict enhancer activity. Analysis of the TFs binding to the *Galnt1* enhancer revealed that in HP cells it is bound by RUNX1 and GFI1 (Supplementary Fig. 7d)[18] at a site overlapping the TEAD motif, suggesting a factor exchange and the establishment of a repressive RUNX1 complex. Taken together, these data demonstrate that VEGF modulates enhancer activity by the balanced interplay of RUNX1, AP-1 and TEAD TFs.

We next used the information obtained above to gain insight into the role of HE-and HP-specific TFs with regards to *RUNX1* regulation. To this end, we constructed ES lines harbouring mutations of the TEAD and RUNX1 binding sites in the + 23 kb enhancer (for ChIP data see Fig. 7e) to test how mutation affected developmental regulation in serum culture (for sequence details see Supplementary Notes, Fig. 7a). The + 23 kb element is activated in the HE and is then down-regulated (Fig. 7f). Consistent with what is seen with the *Galnt1* enhancer, mutation of the TEAD-site in the + 23 kb element led to an upregulation of the element in the HE, consistent with the repressive activity of this factor in the HE whilst mutating the autoregulatory RUNX1 site caused a reduction in enhancer activity.

Finally, we examined how VEGF impacted on other signalling factors involved in regulating HP development. The formation of the arterial HE as a source of HP cells is dependent on NOTCH1 signalling[46,55]. VEGF and SOX17 activate *Notch1* and lack of its down-regulation blocks the EHT[50,51,56,57]. *Notch1* and *Sox17* are down-regulated after the EHT and after upregulation in HP cells, RUNX1 and GFI1 repress *Sox17*[18,48,52,53] (Supplementary Notes, Fig. 2), thus forming a feed forward loop driving EHT progression. We examined whether any of the components of this pathway was modulated by VEGF. We found that the *Dlk1* gene encoding a repressor of NOTCH1 activity[58] was strongly upregulated in the absence of VEGF, both at the chromatin, enhancer activity and the gene expression level (Fig. 8a–c, Supplementary Notes, Fig. 4) as seen by sc-RNA-seq differential gene expression analysis in HE1 thus contributing to the regulation of the switch from an endothelial to a hematopoietic program. *Dlk1* is associated with two enhancers, one of which (enhancer 1) is VEGF-responsive and carries a TEAD / GATA / RUNX1 / E-Box signature (Supplementary Notes Fig. 8a, b).

Taken together, our data show that VEGF signalling in the hemogenic endothelium impacts on a gene regulatory network (Fig. 8d–f) that controls NOTCH1 signalling responsive enhancer elements which bind TEAD and AP-1 and are enriched in NOTCH signature (RBPJ/HES) and SOX motifs. The signalling-responsive activation pattern of these enhancers represents the molecular engine driving the EHT.

Discussion

Our study introduces a versatile method that identifies developmental stage-specific, functional enhancers for any cell type that can be differentiated from ES cells. Importantly, our method recovers many enhancers that have been previously characterised in the context of whole loci, for example those in the Spi*1* or the *RUNX1* locus[25,49]. The enhancer collection described here thus comprises an important resource for gene targeting approaches and for studies of how signals modulate enhancer activity. The method can also easily be adapted to human cells.

All identified cis-elements tested in our assay are derived from DNA-sequences existing as open chromatin and are enriched for motifs of differentiation stage-specific TFs known to be expressed at this stage. Integration with ChIP-Seq data from multiple sources confirms that these factors bind these motifs. The proportion of non-scoring fragments outside of a chromatin region scoring positive was surprisingly small, demonstrating that most open chromatin regions surrounding genes can impact on transcription. Inactive fragments most likely represent enhancer sub-fragments providing information about enhancer sub-structure. Distal elements identified in this study only partially overlap with chromatin features that have been associated with enhancer sequences, such as H3K27Ac. However, enhancers are nucleosome-free and are only flanked by modified histones. Nucleosomes cover 200 bp of DNA, each increasing the sequence space in which a functional element could be located. We therefore believe that assaying the function and sequence composition of open chromatin regions as described here will provide a much more precise tool to home in on those sequences with true transcription-regulatory activity. We see an association of distal elements with sites of RNA-Polymerase binding, confirming that enhancer transcription is widespread, but not ubiquitous.

We find that thousands of regulatory elements are connected to extracellular signals and contain binding motifs for signalling responsive TFs. In response to VEGF, thousands of chromatin regions are opened or closed, directly impacting on the expression of their associated genes. Our single cell experiments show that VEGF truly impedes differentiation, with *Runx1* failing to be upregulated, *Notch1* and *Sox17* not being downregulated, the EHT being delayed, and fewer HP cells being formed, thus directly linking signalling dependent enhancer activity to cell fate decisions. Based on this data, we suggest that the activity of most enhancer elements is fine-tuned by signalling which is therefore truly instructive. Our data comparing alternate

https://doi.org/10.1038/s41467-023-35910-9

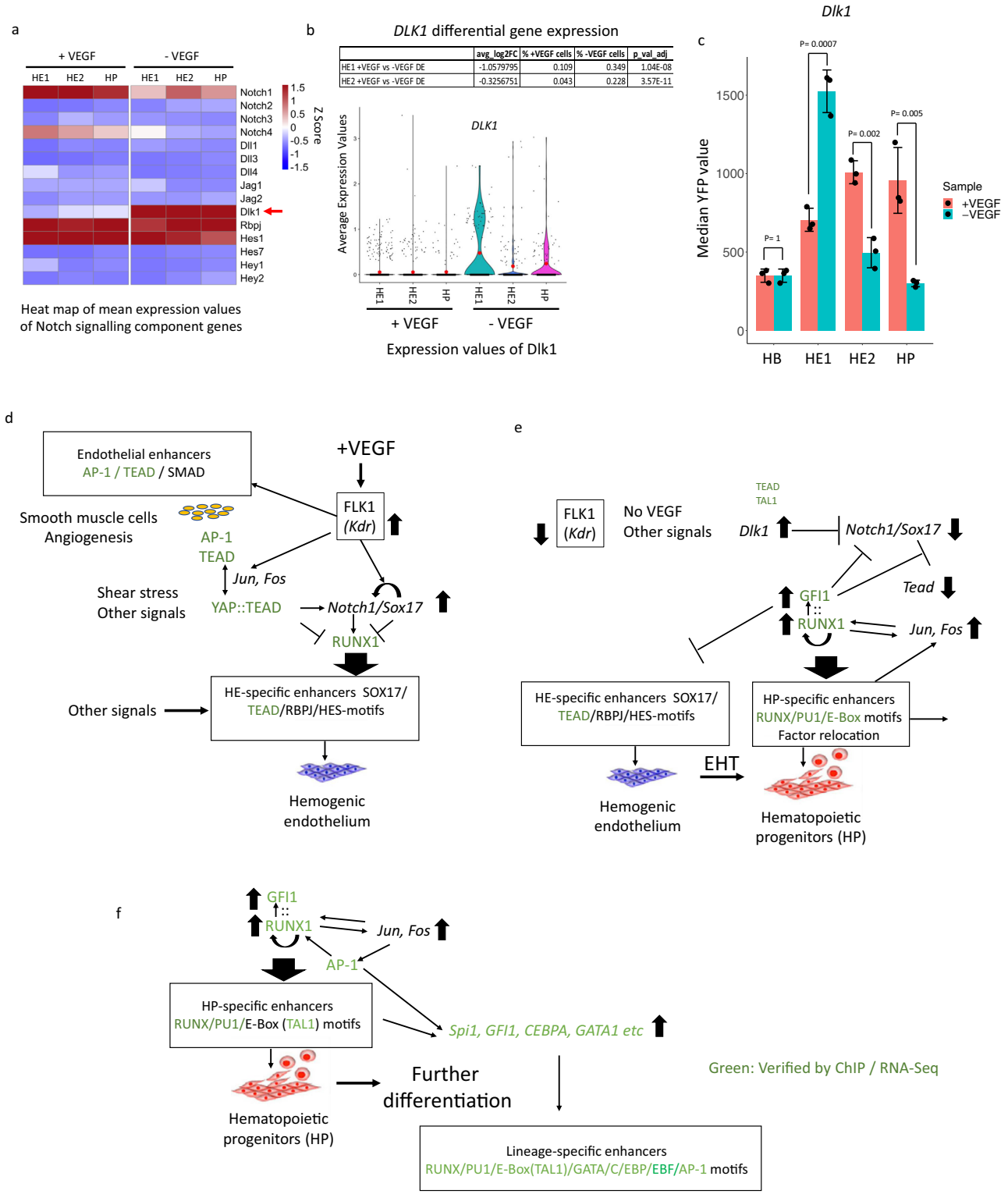


Fig. 8 | VEGF regulates the balance between endothelial and hemogenic development by controlling NOTCH1 activity. a Heatmap showing the expression of NOTCH signaling genes with (+VEGF) and without (−VEGF) based on single cell expression data. **b** Differential gene expression analysis of *Dlk1* depicted as Violin plots as described in Fig. 7 from cells cultured with and without VEGF across 6 clusters for $n = 2266$ cells (black line representing the median and red dot the mean). Differential gene expression taken as +/−0.25 average log2 fold change and adjusted *P*-value < 0.01 (Bonferroni corrected). **c** The *Dlk1* enhancer 1 analyzed in isolation (Chr12: 109,437,601-109438085) is VEGF responsive. Data are presented as mean values +/−SD. Dots showing individual values for $n = 3$ biologically independent experiments. *P*-values calculated using a two-sided Student's *t*-test. Source data are provided as a Source Data file. **d**–**f** Core gene regulatory and signaling network regulating blood stem cell emergence. For further details see text.

differentiation conditions show that signalling impact is highly culture dependent, which will make it imperative to seek differentiation conditions mimicking those found in vivo. In this context it is relevant that AP-1 motifs are enriched in all stage-specific enhancers and co-localize with multiple tissue-specifically expressed TF motifs, with the AP-1:TEAD / RUNX1 site in the *Galnt1* enhancer being an example. AP-1 can interact with chromatin remodellers to assist in the binding of other TFs[59], and multiple studies showed that it is involved in cell fate decisions in response to signals. For example, AP-1 binding is essential for signalling dependent chromatin priming and enhancer activation during T cell differentiation[60]. Blocking its activity at the HP stage during ES cell differentiation leads to a complete abolition of myelopoiesis[20]. The AP-1 family therefore represents a ubiquitous axis integrating the genomic response to specific external signals driving differentiation with pre-existing internal gene expression programs.

In addition to basic insights into enhancer function, our study provided additional mechanistic insights into how the core gene regulatory and signalling network regulating hematopoietic specification is connected to the genome (Fig. 8d–f). Via its receptor, FLK1, VEGF orchestrates differential enhancer and promoter activity which regulates the balance between the NOTCH1/SOX17 axis operating at specific enhancers establishing HE identity and RUNX1 driving the EHT and hematopoietic development. VEGF signals to AP-1[61,62], and we previously showed that this TF regulates the balance between hemogenic and vascular smooth muscle, i.e the endothelial fate[20]. AP-1 also interfaces with HIPPO signalling via TEAD TFs. Switching off HIPPO signalling via YAP activation induces *Runx1* expression in response to shear stress[54] and in endothelial cells promotes angiogenesis and endothelial gene expression[63]. During the EHT, RUNX1 is upregulated, autoregulates itself by binding to its own enhancers, represses the endothelial program by interacting with GFI1/LSD1 and relocates hematopoietic transcription factors such as TAL1/SCL and FLI-1 to hematopoietic genes[52] leading to further differentiation (Fig. 8e). Taken together with these findings, our ChIP, single-cell gene expression and reporter gene studies place the interplay between RUNX1, SOX17, NOTCH, HIPPO-signalling and AP-1 mediating VEGF responsive TFs at specific enhancer elements at the heart of the balance between endothelial or hematopoietic fate. Although most of the network components and their different roles are known from many perturbation and knock-out experiments as reviewed in[44], it was unclear how and where signals and transcription factors impact on the genome and how they are connected within cell stage specific gene regulatory networks. In this study, we have now identified the factors involved in their regulation and the genomic elements upon which they act, providing a rich resource for studies of identifying the signals required for the activation of the correct gene expression program required for efficient blood cell production.

Methods

HM-1 ES cell culture

The HM-1 targeting ES cell line originally described by Magin et al (1992)[64] was cultured on gelatinised tissue culture plates in DMEM-ES media (DMEM (Merck, D5796) with 15% FSC, 1 mM sodium pyruvate (Merck, S8636), 1 x Penicillin/Streptomycin (Merck, P4333), 1 x L-Glutamine (Merck, G7513), 1 x Non-Essential amino acids (Merck, M7145), 1000 U/ml ESGRO®LIF (Merck, ESG1107), 0.15 mM MTG and 25 mM Hepes buffer (Merck, H0887)) at 37 °C and 5% CO_2. Cell colonies were dissociated with TrypLE™ Express (ThermoFisher, 12605010) at 48 h intervals and re-plated at 1.2×10^4 per cm^2.

In vitro hematopoietic differentiation in Serum

Hematopoietic In vitro differentiation (I.V.D) in serum was essentially performed as previously described in Obier et al.[20]. ES cells colonies were dissociated to form a single cell suspension using TrypLE™ Express (ThermoFisher, 12605010) and trypsin activity was stopped by addition of DMEM-ES media at 1:1 ratio. Cells were then resuspended at 2.5×10^4/ml in I.V.D media (IMDM (Merck, I3390), 15% FCS, 1 x Penicillin/Streptomycin (Merck, P4333), 1 x L-Glutamine (Merck, G7513), 0.15 mM MTG, 50 μg/ml Ascorbic acid and 180 μg/ml Human transferrin (Merck, T8158)) and plated onto non-adherent dishes (ThermoFisher, 501 V). The cells were the incubated at 37 °C, 5% CO_2 for 3 days until floating embryoid bodies (EBs) were formed.

Embryoid bodies were harvested by transferring EB containing media to 50 ml centrifuge tubes and allowing the EBs to settle by gravity. The EBs were washed with PBS, allowed to settle and then were dissociated with TrypLE™ Express (ThermoFisher, 12605010) ensuring a single cell suspension was achieved while avoiding damage to the cells. The dissociated cells were then sorted for the FLK1 surface marker (forming the population referred to as HB), by incubation with a FLK1 biotin-coupled antibody (1:200) (eBioscience, 13-5821-82), followed by mixing cells with MACS anti-biotin beads (Miltenyi Biotec, 130-090,485) and separation using a MACS LS column (Miltenyi Biotec, 130-042-401). The sorted cells were then used for flow cytometry and sequencing experiments as well as further differentiation in blast culture into HE1, HE2 and HP.

Blast culture was performed by resuspending the FLK1+ cells and plating on a gelintin coated tissue culture flask at 1.6×10^4 cells per cm^2 in Blast media (IMDM (Merck, I3390), 10% FCS, 1 x Penicillin/Streptomycin (Merck, P4333), 1 x L-Glutamine (Merck, G7513), 0.45 mM MTG, 25 μg/ml Ascorbic acid, 180 μg/ml Human transferrin (Merck, T8158), 20% D4T Conditioned Media, 5 ng/ml VEGF(PeproTech, 450-32), 10 ng/ml IL-6 (PeproTech, 216-16)) and incubating for 2.5 days at 37 °C, 5% CO_2.

FACS sorting of the HE1, HE2 and HP populations was achieved after incubation with antibodies KIT-APC (1:100)(BD Pharmingen, 553356), TIE2-PE (1:200)(eBioscience, 12-5987-82) and CD41-PECY7 (1:100)(eBioscience, 25-0411-82) and then sorted into HE1 (KIT+, TIE2+, CD41-), HE2 (KIT+, TIE2+, CD41+) and HP (KIT+, TIE2-, CD41+) populations.

Serum free in vitro hematopoietic differentiation

Serum Free I.V.D Culture derived Embryoid bodies (EBs) were generated from HM-1 mouse embryonic stem cells by plating at 5.0×10^5 cells/ml in StemPro™–34 SFM media (ThermoFisher, 10639011) into non-adhesive dishes (ThermoFisher, 501 V). BMP4 (PeproTech, 315-27) was added at a concentration of 5 ng/ml. Cultures were left to incubate at 37 °C and 5% CO_2 for 60 h before bFGF (PeproTech, 450-33) and Activin A (PeproTech, 120-14E) were added at a concentration of 5 ng/ml each. The cells were incubated for 16 h at 37 °C, 5% CO2 and then sorted for FLK1 + cells as described in Obier et al.[20]. FLK1 + cells were plated in StemPro™–34 SFM media on 0.1% gelatine coated plates or tissue culture flasks at 2.25×10^4 cells per cm^2. BMP4, Activin-A and bFGF were added to a concentration of 5 ng/ml for 16 h. Media was then removed, the blast culture was washed with PBS (Merck, D8662) and fresh StemPro™–34 SFM media was added containing BMP4 (5 ng/ml), VEGF (5 ng/ml)(PeproTech, 450-32), TPO (5 ng/ml))(PeproTech, 315-14), SCF (100 ng/ml))(PeproTech, 250-03), IL6 (10 ng/ml)) (PeproTech, 216-16) and IL3 (1 ng/ml))(PeproTech, 213-13). For cytokine withdrawal experiments, one of BMP4, VEGF, IL6, or IL3 was not added at this stage. Blast cultures were left to incubate at 37 °C and 5% CO_2 for 72 h before cells were harvested for cell sorting and FACS analysis. Inhibition of trypsin was achieved using Trypsin Inhibitor (ThermoFisher, 17075029) at a 1:1 ratio.

Enhancer reporter library cloning

A genome-wide enhancer reporter assay was designed based on an enhancer reporter system designed by Wilkinson et al. (2013)[23]. Briefly the enhancer reporter functions by inserting a fragment of interest upstream of a HSP68 minimal promoter and Venus-YFP reporter by Gateway® cloning. The reporter construct is then transfected into the

HM-1 ES cell line which has a non-functional HPRT locus. Using HPRT homology arms the report cassette becomes integrated into the HPRT locus by homologous recombination, also repairing the locus, enabling selection of clones with successful recombination. This enhancer reporter system was modified and optimised for genome-wide screening. To obtain genome-wide enhancer fragments for cloning into the reporter we isolated tn5 tagmented open-chromatin fragments based on the ATAC-Seq protocol[33] from cells of the five differentiation stages. Cells were obtained (2.5×10^5) from each differentiation stage (ES, HB, HE1, HE2, HP) as detailed in the serum IVD method. Cells were pelleted in 5 aliquots of 5×10^4 cells and following the ATAC-Seq protocol each cell pellet had a transposition mix added (2x TD Buffer 25 μl, TN5 Transposase 2.5 μl (Illumina, 20034197), PBS 16.5 μl, 1% Digitonin 0.5 μl (ThermoFisher, BN2006), 10 % Tween-20 0.5 μl (Merck, P9416), H_2O 5 μl) and cells were gently resuspended by pipetting. The transposition reaction was then incubated in a shaker at 700 RPM, 37 °C for 30 min. The 5 reactions were then combined and DNA purified using the Qiagen MinElute® Reaction Clean-up kit (28206) and eluted in 26.5 μl H_2O. One fifth of the purified reaction from each stage was used to produce the baseline ATAC-Seq libraries (see ATAC-Seq method) and the remainder had linkers added incorporating AttB Gateway® cloning sites to enable insertion into a Gateway® donor vector (pDONR 221, Thermo Fisher). The linkers were designed based on the tn5 transposase sequence and added by PCR in 2 steps. Initially 5 cycles of PCR were performed (Transposed DNA 20 μl, H_2O 20 μl, 25 μM AttB tn5 Fwd primer 5 μl, 25 μM AttB tn5 Rev primer 5 μl, NEBNext Master Mix 50 μl (NEB, M0541)) with the following conditions, 72 °C for 5 min, 98 °C for 30 s, 5 cycles of 98 °C 10 s, 63 °C 30 s and 72 °C 1 min and a final hold at 4 °C.

To optimize the required cycle number, the material from the initial PCR reaction was split and a 5 μl was used for a separate qPCR reaction (0.125 μM AttB tn5 Fwd Primer, 0.125 μM AttB tn5 Rev Primer, 1 x SYBR Green I (Merck, S9430), NEBNext Master Mix (NEB, M0541)) with the following conditions: 72 °C for 5 min, 98 °C for 30 s, 35 cycles of 98 °C 10 s, 63 °C 30 s and 72 °C 1 min using a StepOnePlus™ Real-Time PCR System (Thermo Fisher). Using the raw data from the cyan channel the number of cycles for 25% amplification was calculated. The number of cycles obtained from the qPCR side reaction was then used to amplify the remainder of the material as follows, 98 °C for 30 s, cycles of 98 °C 10 s, 63 °C 30 s and 72 °C 1 min with a final hold at 4 °C. The resulting material was size selected, to optimise for material originating from enhancers, for fragments between 350 and 650 bp by electrophoresis on a 1.2% agarose gel with ethidium bromide and gel extraction was performed using the Qiagen MinElute® Gel Extraction Kit (Qiagen, 28604) following manufactures protocol.

To generate attL-flanked entry clones compatible with cloning into the reporter vector the size selected material was cloned into the pDONR™221 (Thermo Fisher) by BP Gateway® reaction (100 fmol pDONR221, 100 fmol PCR material, 2 μl BP Clonase II, made up to 20 μl with TE buffer (pH 8.0) the reaction was incubated for 18 h at 25 °C and ended by treatment with 0.2 μg proteinase K and incubation at 37 °C for 10 min. NEB 10-beta Electrocompetent cells (NEB, C3019H) were transformed with the plasmid by electroporation with a Bio-Rad GenePulser using program EC1 (2.0 kV, 200 Ohm, 25 μF), after addition of outgrowth media and incubation at 37 °C, 250 rpm for 1 h the bacteria were spread onto 150 mm agar plates containing 50 μg/ml kanamycin. Following overnight incubation at 37 °C, 10 ml of LB was added to each plate and the bacterial lawn scraped off using a cell scraper. Plasmid DNA was extracted from the collected bacteria by Maxi-prep using the NucleoBond Xtra Midi EF kit (Macherey-Nagel, 740420.50) following manufacturer's instructions. The resulting product was eluted in 500 μl H_2O.

The chromatin fragments were then transferred into the enhancer reporter cassette in the pSKB-GW-Hsp68-Venus vector[23] by Gateway® LR reaction (150 ng Entry clone (pDONR), 150 ng pSKB, 2 μl LR Clonase II, made up to 10 μl with TE buffer (pH 8.0)) the reaction was incubated for 1 h at 25 °C and ended by treatment with 0.2 μg proteinase K and incubation at 37 °C for 10 min. As with the BP reaction product, NEB 10-beta Electrocompetent cells were transformed with the material and after overnight incubation on agar plates containing 200 μg/ml ampicillin the bacteria were collected and the plasmid library extracted by Maxi-Prep. To obtain the highest possible diversity of fragments the cloning was carried out 10 times for material from each differentiation stage with the pDONR 221 library material being combined before moving to the LR clonase reaction.

Enhancer reporter library transfection

The HM-1 ES cell line was cultured on gelatinised culture plates in ES-Media (see IVD method) at 37 °C, 10% CO_2 and split every 48 h to maintain healthy cells. Prior to transfection the cells were treated for 1 week with 1 × 6-TG (6-Thioguanine) (Merck, A4660) to remove any cells with a functional HPRT locus. The plasmid libraries were mixed and digested with PmeI (NEB, R0560) to linearise the plasmids prior to transfection. To obtain optimum homologous recombination efficiency CRISPR guides were designed flanking the region and cloned into the PX458 Cas9 and sgRNA expression vector[65] (pSpCas9(BB)−2A-GFP (PX458) was a gift from Feng Zhang (Addgene plasmid # 48138; http://n2t.net/addgene:48138; RRID:Addgene_48138).

For each experiment, 500×10^6 cells were transfected using a Nucleofector®−4D (Lonza) with the P3 Primary Cell X kit (Lonza, V4XP-3024) with 5×10^6 per cuvette 10 ug pSKB enhancer reporter plasmid and 10 ug of PX458-HPRT CRISPR plasmid using program CG-104. Following electroporation, the cells were plated on gelatinised plates in ES-media with the addition of 50 μM SCR7 pyrazine (Merck, SML1546) to inhibit Non-homologous end joining, promoting homologous recombination following cutting by Cas9. After 12 h the media was changed and 1 x HAT (Hypoxanthine-Aminopterin-Thymidine (Merck, H0262)) was added to select for clones with successful recombination. The cells were then maintained in media containing HAT for one week and HT (hypoxanthine and thymidine (Merck, H0137)) for a further 2 passages to prevent the cells from dying after the withdrawal of HAT, whereby all cells were kept and replated after splitting.

The selected cells were then put into IVD cultures and differentiated into HB, HE1, HE2, and HP and sorted by FACS (See Serum IVD Method). In addition to the previously mentioned gating for obtaining the cell populations the cells were also sorted into negative, low, medium and high Venus-YFP populations representing the activity of the enhancer fragment driven minimal promoter. The YFP-Venus negative population was gated based on a cell line containing a reporter construct with only the minimal promoter and reporter with no enhancer fragment.

Enhancer reporter library preparation

Genomic DNA was extracted from the sorted cell populations using the Qiagen DNA Micro kit following manufacturer's instructions. To amplify the fragments contained in the reporter cassettes and add sequencing adaptors a PCR using indexed primers compatible with Illumina sequencing and targeted against the tn5 transposase sequence ((0.5 μM Nextera PCR Primer i5, 0.5 μM Nextera PCR Primer i7, 25 μl NEBNext Master Mix). Dependent on the amount of DNA obtained from the sorted cells between 25 and 30 PCR cycles were performed using the following conditions 98 °C for 30 s, cycles of 98 °C 10 s, 63 °C 30 s and 72 °C 1 min with a final hold at 4 °C. Libraries were quantified using a Bioanalyzer 2100 (Agilent) with a High

Sensitivity DNA Chip and pooled at equal concentration for sequencing. The final pool concentration and size was then confirmed by Kappa Library Quantification Kit (Roche, KK4824) and Bioanalyzer. The pooled libraries were sequenced on a Next-Seq 550 as a paired end run with a 150 cycle High Output kit by the Genomics Birmingham Sequencing Facility.

Trouble-shooting note. During the development of the method, we discovered by sequencing that the different cloning and targeting vectors reproducibly preserved the original complexity of the ATAC-Seq library. The number of individual sequences equalled the sequence complexity contained in the original ATAC-Seq experiment. We recommend performing such controls for every experiment. However, as shown in the manuscript, the screening procedure can give different levels of enhancer coverage. It turned out that the quality and viability of the transfected cells and their differentiation capacity, i.e., the ability of each individual clone carrying one construct to form the different cell types, was crucial to obtain a maximal coverage of the initial complexity of the library. Therefore, care needs to be taken to maintain the cells in optimal growth condition and use a large number of cells as recommended above.

Generation of cell lines carrying individual reporter constructs

A putative enhancer for the *Galnt1* gene containing multiple transcription factor binding sites was identified from DNaseI hypersensitive site data[18]. To test the functionality of the enhancer throughout differentiation the same enhancer reporter system[23] as used for the genome-wide screen was employed. The genomic sequence for the enhancer was obtained and flanking attB1 and attB2 sites were added to enable Gateway® cloning (Supplementary Notes Table 3). This fragment was then synthesised as an Invitrogen GeneArt Strings DNA Fragment (ThermoFisher). The resulting fragment was cloned by the one step Gateway protocol whereby in a single reaction the fragment was inserted into pDONR 221 to generate an attL-flanked entry clone by BP clonase II (ThermoFisher, 11789020) and from this vector into the pSKB GW-Hsp68-Venus reporter vector[23] by LR clonase II (ThermoFisher, 11791020). Briefly this reaction consisted of 100 ng of Enhancer DNA, 75 ng pDONR 221, 75 ng pSKB GW-Hsp68-Venus, TE pH 8.0 to 6 μl total volume, 1.5 μl LR clonase and 0.5 μl BP clonase. The reaction was incubated at 25 °C for 3 h and then had 0.2 μg proteinase K added and a further incubation for 10 min at 37 °C to stop the reaction. Competent DH5α bacteria (NEB, C2987H) were transformed with 1 μl of Gateway® reaction by incubation on ice for 30 min and heat shock at 42 °C for 45 s. The bacteria were then incubated for 1 h at 37 °C with SOC media (ThrmoFisher, 15544034) and plated onto agar plates containing 100 μg/ml ampicillin (Merck, A9518). After overnight incubation at 37 °C colonies were picked for mini-prep cultures and mini-preps were performed using the Qiaprep® Miniprep kit (Qiagen, 27106 × 4). The insert in the resulting plasmid was sanger sequenced (Source Biosciences, stock sequencing primer 'EGFP_Nrev') to check the sequence. Following confirmation of the insert sequence the same bacterial colony was used to grow cultures for Maxi-prep which was performed using the EndoFree® Plasmid Maxi Kit (Qiagen, 12362).

HM-1 cells were cultured and transformed in the same way as with the Enhancer screen method. Following 6-TG treatment 5×10^6 HM-1 cells were transfected using a Nucleofector®–4D (Lonza) with the P3 Primary Cell X kit (Lonza, V4XP-3024). The cells were then selected for those with successful integration of the reporter cassette by treatment for 5 days with 1 x HAT containing media. At 9 to 12 days after colonies appeared clones were picked and re-plated on a gelatinised 96 well plate. The clones were then grown in media containing 1 x HT for 2 passages and then used for IVD and flow cytometry as detailed.

To study the impact of different transcription factor binding on the *Galnt1* enhancer, sequences were produced where various transcription factor binding motifs were mutated. These mutant versions of the enhancer were cloned in the same way as with the original *Galnt1* enhancer sequence (Supplementary Notes Table 3).

To validate enhancer elements identified in our enhancer, enhancers associated to the *Sparc*, *Pxn, Eif2b3, Dlk1* and *Hspg2* genes were chosen. These enhancers were also identified as being VEGF signalling responsive taken as a 2-fold change in ATAC-seq peak height with and without VEGF in the culture. Primers (Supplementary Notes Table 4) were designed to amplify out the 400 bp enhancer element and then the enhancer element was cloned HM-1 mESCs as described above.

To validate the + 23 kb and + 3.7 kb *RUNX1* enhancers, the 400 bp enhancer sequences were synthesised by GeneArt String synthesis (Thermo Fisher). For each *RUNX1* enhancer with individual TF binding motifs mutated GeneArt String synthesis (Thermo Fisher) was used (Supplementary Notes Table 3). The sequences were then cloned into HM-1 mESCs as described above.

ATAC-sequencing. ATAC-Seq was performed as described in Buenrostro et al.[66]. Briefly 5000-50,000 HB, HE1, HE2 and HP cells were sorted by FACS and transposed in 1x tagment DNA buffer, Tn5 transposase (Illumina, 20034197) and 0.01% Digitonin (ThermoFisher, BN2006) for 30 min incubated at 37 °C with agitation. DNA was purified using a MinElute Reaction Cleanup Kit (Qiagen) and DNA was amplified by PCR using Nextera primers[66].

Single cell sorting and single cell RNA-Seq. 7.0×10^6 cells from the All cytokines condition blast culture and 2.8×10^6 cells from the VEFG withdrawal blast culture were taken for cell sorting and sc-RNA-Seq library preparation. Both samples were stained with CD41-PECY7 (1:100)(eBioscience, 25-0411-82), KIT-APC (1:100)(BD Pharmingen, 553356) and TIE2-PE (1:200)(eBioscience, 12-5987-81) and then sorted into HE1 (KIT + , TIE2 + , CD41-) and HE2 (KIT + , TIE2 + , CD41 +) populations. From the All cytokines sample, 758000 HE1 and 380000 HE2 cells were sorted. From the VEGF withdrawal sample 71500 HE1 and 33300 HE2 cells were purified by cell sorting. Cells were resuspended in 80 μL at a concentration of 1000–1200 cells/μL for evaluation of cell viability. The viability of the All cytokines populations as measured by Trypan Blue (Merck, T8154) staining was found to be 73% for HE1 and 92% for HE2. The viability of the VEGF withdrawal populations was found to be 73% and 75%. For each sample, 10,000 single cells were loaded on a Chromium Single Cell Instrument (10x Genomics) and processed.

Data analysis methods

Enhancer sequencing data analysis. Paired end sequencing reads were trimmed using Trim Galore (https://github.com/FelixKrueger/TrimGalore)[67] with the parameters --nextera --length 70 –paired to remove remaining adaptor sequences and poor-,quality sequences. The reads were then aligned to the mm10 genome using bowtie2[68] using parameters --very-sensitive --fr --no-discordant -X 600 --no-mixed, a maximum fragment length (-X) of 600 bp was set based on the largest size selected fragments during the initial cloning. Aligned reads were filtered for only those which were properly paired and with a mapq score of over 40 using Samtools and output as a bedpe file. Because we treat each unique read as a direct readout of enhancer activity, this filtering is an essential step to remove multimapping reads which would produce false positives. The bedpe file was further converted to a bed file of enhancer fragments by taking the first co-ordinate of read 1 and the final co-ordinate of read 2 for each pair. Duplicate fragments were removed using the unique function leaving a bed file of unique aligned fragments representing those PCR amplified from the FACS sorted cells.

Because we use the fragments generating a positive FACS signal as a direct readout of enhancer activity stringent filtering is essential.

To remove potential PCR artefacts, a library was produced using the same PCR conditions as the other libraries but using DNA from untransfected cells. Since there should be no complimentary sequences to the ATAC primers in wild type DNA any fragments produced in this background library represent PCR artefacts. Using the background library, fragments were further filtered using the intersect function on Bedtools[69] and any having a perfect overlap with the background library were removed. The fragments were then filtered using Bedtools intersect for those which were found in open chromatin, defined by DNaseI-Seq, in Goode et al. (2016)[18] and further by a corresponding ATAC peak in the correct differentiation stage in HM-1 cells. This enabled us to define which stage a reporter fragment would be in open chromatin and have potential to act as an enhancer. Finally, the enhancer fragments were annotated using the annotatePeaks function of Homer for those within 1.5 kb of a TSS (promoter fragments) and distal fragments representing genuine enhancers. This analysis produced a final list of enhancer fragments which was used for all further fragment analyses. We defined full enhancers by overlapping distal ATAC sites with enhancer fragments and called any open chromatin region with an overlapping positive enhancer fragment as enhancer positive. The remaining ATAC sites we define as negative/unknown because negative fragments may not overlap the required TF binding sites for enhancer activity.

ATAC-seq data analysis. Single-end reads from ATAC-seq experiments were processed with Trimmomatic (version 0.39) and aligned to the mm10 mouse genome using Bowtie2 2.4.4[68] using the options -very-sensitive-local. Open chromatin regions (peaks) were identified using MACS2 2.2.7.1[70] using the options -B -trackline -nomodel. The peak sets were then filtered against the mm10 blacklist[71]. Peaks were then annotated as either promoter-proximal if within 1.5 kb of a transcription start site and as a distal element if not. To conduct differential chromatin accessibility analysis, a peak union was first constructed by merging peaks from two comparisons. Tag-density in a 400 bp window as centred on the peak summits was derived from bedGraph files from MACS2 peak calling using the annotatePeaks.pl function in Homer 4.11[72] (with the options -size 400 -bedGraph. Tag-densities were then normalised as counts-per-million (CPM) in R v3.6.1 and further $\log_2$-transformed as $\log_2(\mathrm{CPM}+1)$. In cases were replicate samples were available, the average normalized tag count was used for all downstream comparisons. A peak which was two-fold increase or decrease different to the control was taken as being differentially accessible. A de-novo motif analysis was performed on sets of gained and lost peaks using the find -MotifsGenome.pl function in Homer using the options -size 200 -noknown. Tag density plots were constructed by retrieving the tag-density in a 2 kb window centred on the peak summits with the annotatePeaks.pl function in Homer with the options -size 2000 -hist 10 -ghist -bedGraph. These were then plotted as a heat map using Java TreeView v1.1.6[73].

Gene annotation using HiC data. Promoter capture HiC data from ESCs cells were downloaded from Novo CL et al. (accession numbers: GSM2753058, runs: SRR5972842, SRR5972842, SRR5972842, SRR5972842, SRR5972842)[34] and from HPC7 from Comoglio et al.[74] (accession numbers: GSM2702943 & GSM2702944, runs SRR5826938, SRR5826939). The CHi-C paired-end sequencing reads were aligned to the mouse genome mm10 build using HiCUP pipeline[75]. Initially, the raw sequencing reads were separated and then mapped against the reference genome. The aligned reads were then filtered for experimental artefacts and duplicate reads, and then re-paired. Statistically significant interactions were called using GOTHiC package[76] and HOMER software[72]. This protocol uses a cumulative binomial test to detect interactions between distal genomic loci that have significantly more reads than expected by chance, by using a background model of random interactions. This analysis assigns each interaction with a p-value, which represents its significance. The union of all CHiC interactions from both ESC and HPC7 cells were used to annotate positive enhancer to their related promoter.

Motif co-localization analysis. Genomic co-ordinates for each transcription factor (TF) binding motif were retrieved from the sets of stage specific positive enhancer fragments and from within all distal ATAC-Seq peaks using the annotatePeaks.pl function in Homer and exported as a BED file using the -mbed option. Motif co-occurrence was then measured for each stage (ES, HB, HE1, HE2 and HP) by counting the number of times a pair of TF motifs were found within 50 bp of each other in the set of specific positive enhancer fragments.

To assess the significance of this co-occurrence, we carried out a re-sampling analysis whereby a number of ATAC-Seq sites equal to the number of stage specific enhancer fragments was randomly sampled from the set of all distal ATAC sites found in that stage. The number of motif pairs was then counted in this random set. This procedure was repeated 1000 times and resulted in a distribution of motif pair counts for each pair of TF binding motifs. A z-score was then calculated for each motif pair using as:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

Where x is the number of motif pairs found in the specific positive enhancer fragments, μ is the average number of motif pairs in 1000 random samples, and σ is the standard deviation for those motif pair counts. A positive z-score in this case suggests that the number of motif pairs found in the positive enhancer fragments is greater than could be expected by chance. The resulting z-score matrix was then hierarchically clustered using complete linkage of the Euclidean distance in R and displayed as a heatmap.

Relative motif enrichment analysis from ATAC-Seq data. To identify transcription factor binding motifs which were enriched in a set of peaks relative to another, we calculated a relative enrichment motif score, S_{ij} for each motif i in each peak set j as:

$$S_{ij} = \frac{\frac{n_{ij}}{m_j}}{\sum_j \frac{n_{ij}}{\sum_j m_j}}, \tag{2}$$

where n_{ij} is the number of instances of motif i in peak set j and m_j is the total number of sites in peak set j. This score was calculated for each TF motif in each of the peak sets and a matrix of enrichment scores was produced which was then hierarchically clustered using complete linkage of the Euclidean distance in R and displayed as a heat map with results scaled by either row or column.

Single cell RNA-Seq analysis. Fastq files from scRNA-Seq experiments were aligned to the mouse genome (mm10) using the count function in CellRanger v6.0.1 from 10x Genomics and using gene models from Ensembl (release 102) as the reference transcriptome. The resulting Unique Molecular Identifier (UMI) count matrices were processed using the Seurat package v4.0.5[77] in R v4.1.2. Cell quality control was carried out for each of the four samples sequenced individually in order to remove cells which had few or a higher than expected number of detected genes or that had a high proportion of UMIs aligned to mitochondrial transcripts or quality control parameters for each sample. Genes detected in less than 3 cells were removed from the analysis.

The filtered data objects for each cell type were then combined according to their VEGF status and normalized using the Log-Normalize method. In order to calculate the cell cycle stage for each of the cells, the in-built S-phase and G2M-phase marker gene lists

from Seurat were first converted from human gene symbols to their corresponding mouse orthologs using the biomaRt package v2.50.0[78] in R. Cell cycle stage was then inferred using the CellCycleScoring function in Seurat. The possible effect of cell cycle stage on downstream analysis was then removed from the dataset by linear regression using the ScaleData function. Clustering of cells was performed using the first 20 principal components and visualized using the Uniform Manifold Approximation and Projection (UMAP) method. Cell marker genes for each of the clusters identified (Supplementary Dataset 5) were calculated using the FindAllMarkers function. A gene was considered a marker gene if it was expressed with a log2 fold-difference of 0.5 between the cluster being considered and all other cells as well as being detected in at least 50% of cells in that cluster. A gene was considered statistically significant if it had a Bonferroni-adjusted p-value < 0.05. Cell type was then inferred for each cluster identified by manual inspection of the marker gene lists and comparing these to the expression known surface markers genes (Tie2, Cd41) for HE1, HE2, and HP.

Cell trajectory (pseudotime) analysis was carried out using Monocle v3.1.0[79]. Data from Seurat was exported to Monocle using the seuratwrappers package in R (https://github.com/satijalab/seurat-wrappers). Trajectories were then inferred using the learn graph function and ordered along pseudotime by selecting a root node which corresponded to the earliest cell-type (HE1).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that support this study are available from the corresponding authors upon reasonable request. The genome-wide data generated in this study have been deposited in the Gene Expression Omnibus (GEO) database under accession code GSE198775. The genome-wide data generated in this study have also been provided as a UCSC Genome Browser Track-Hub (https://genome-trackhub.bham.ac.uk/data/EnhancerHub/hub.txt). The publicly available genome-wide data analysed in this study were deposited in GEO under accession code GSE69101, GSE143460, GSE126496 and GSE79320. Source data are provided with this paper.

References

1. Cockerill, P. N. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J.* **278**, 2182–2210 (2011).
2. Edginton-White, B. & Bonifer, C. The transcriptional regulation of normal and malignant blood cell development. *FEBS J.* **289**, 1240–1255 (2022).
3. Field, A. & Adelman, K. Evaluating enhancer function and transcription. *Annu. Rev. Biochem.* **89**, 213–234 (2020).
4. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
5. Bonifer, C. Developmental regulation of eukaryotic gene loci: Which cis-regulatory information is required? *Trends Genet.* **16**, 310–315 (2000).
6. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
7. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
8. Kim, T. K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
9. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
10. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
11. Hou, T. Y. & Kraus, W. L. Spirits in the material world: Enhancer RNAs in transcriptional regulation. *Trends Biochem. Sci.* **46**, 138–153 (2021).
12. Dogan, N. et al. Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin* **8**, 16 (2015).
13. Gasperini, M. et al. A Genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 377–390.e19 (2019).
14. Medvinsky, A. L., Samoylina, N. L., Muller, A. M. & Dzierzak, E. A. An early pre-liver intraembryonic source of CFU-S in the developing mouse. *Nature* **364**, 64–67 (1993).
15. de Bruijn, M. F., Speck, N. A., Peeters, M. C. & Dzierzak, E. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *EMBO J.* **19**, 2465–2474 (2000).
16. Orkin, S. H. & Zon, L. I. Hematopoiesis: An evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
17. Ditadi, A., Sturgeon, C. M. & Keller, G. A view of human haematopoietic development from the Petri dish. *Nat. Rev. Mol. Cell Biol.* **18**, 56–67 (2017).
18. Goode, D. K. et al. Dynamic gene regulatory networks drive hematopoietic specification and differentiation. *Dev. Cell* **36**, 572–587 (2016).
19. Lancrin, C. et al. The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature* **457**, 892–895 (2009).
20. Obier, N. et al. Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate. *Development* **143**, 4324–4340 (2016).
21. Vijayabaskar, M. S. et al. Identification of gene specific cis-regulatory elements during differentiation of mouse embryonic stem cells: An integrative approach using high-throughput datasets. *PLoS Comput Biol.* **15**, e1007337 (2019).
22. Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. & Keller, G. Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432**, 625–630 (2004).
23. Wilkinson, A. C. et al. Single site-specific integration targeting coupled with embryonic stem cell differentiation provides a high-throughput alternative to in vivo enhancer analyses. *Biol. Open* **2**, 1229–1238 (2013).
24. Dickel, D. E. et al. Function-based identification of mammalian enhancers using site-specific integration. *Nat. Methods* **11**, 566–571 (2014).
25. Leddin, M. et al. Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. *Blood* **117**, 2827–2838 (2011).
26. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA enhancer Browser-a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
27. Zhu, Q. et al. Developmental trajectory of prehematopoietic stem cell formation from endothelium. *Blood* **136**, 845–856 (2020).
28. Howell, E. D. et al. Efficient hemogenic endothelial cell specification by RUNX1 is dependent on baseline chromatin accessibility of RUNX1-regulated TGFbeta target genes. *Genes Dev.* **35**, 1475–1489 (2021).
29. Wilson, N. K. et al. Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model. *Blood* **127**, e12–e23 (2016).

30. Gilmour, J. et al. The Co-operation of RUNX1 with LDB1, CDK9 and BRD4 drives transcription factor complex relocation during haematopoietic specification. *Sci. Rep.* **8**, 10410 (2018).
31. Kellaway, S. G. et al. Different mutant RUNX1 oncoproteins program alternate haematopoietic differentiation trajectories. *Life Sci. Alliance* **4**, e202000864 (2021).
32. Lam, M. T., Li, W., Rosenfeld, M. G. & Glass, C. K. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.* **39**, 170–182 (2014).
33. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
34. Novo, C. L. et al. Long-range enhancer interactions are prevalent in mouse embryonic stem cells and are reorganized upon pluripotent state transition. *Cell Rep.* **22**, 2615–2627 (2018).
35. Qi, Q. et al. Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis. *Blood* **137**, 1327–1339 (2021).
36. Azuara, V. et al. Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8**, 532–538 (2006).
37. Mullen, A. C. et al. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**, 565–576 (2011).
38. Trompouki, E. et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**, 577–589 (2011).
39. Wilson, N. K. et al. Combinatorial transcriptional control in blood stem/progenitor cells: Genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532–544 (2010).
40. Simeonov, D. R. et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115 (2017).
41. Hewitt, K. J. et al. GATA Factor-Regulated Samd14 enhancer confers red blood cell regeneration and survival in severe anemia. *Dev. Cell* **42**, 213–225.e4 (2017).
42. Pearson, S., Sroczynska, P., Lacaud, G. & Kouskoff, V. The stepwise specification of embryonic stem cells to hematopoietic fate is driven by sequential exposure to Bmp4, activin A, bFGF and VEGF. *Development* **135**, 1525–1535 (2008).
43. Robert-Moreno, A., Espinosa, L., de la Pompa, J. L. & Bigas, A. RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells. *Development* **132**, 1117–1126 (2005).
44. Ottersbach, K. Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochem Soc. Trans.* **47**, 591–601 (2019).
45. Shalaby, F. et al. Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62–66 (1995).
46. Lomeli, H. & Castillo-Castellanos, F. Notch signaling and the emergence of hematopoietic stem cells. *Dev. Dyn.* **249**, 1302–1317 (2020).
47. Li, Y. et al. Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell production. *Genes Dev.* **28**, 2597–2612 (2014).
48. Lancrin, C. et al. GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood* **120**, 314–322 (2012).
49. Nottingham, W. T. et al. Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood* **110**, 4188–4197 (2007).
50. Lizama, C. O. et al. Repression of arterial genes in hemogenic endothelium is sufficient for haematopoietic fate acquisition. *Nat. Commun.* **6**, 7739 (2015).
51. Clarke, R. L. et al. The expression of Sox17 identifies and regulates haemogenic endothelium. *Nat. Cell Biol.* **15**, 502–510 (2013).
52. Lichtinger, M. et al. RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J.* **31**, 4318–4333 (2012).
53. Thambyrajah, R. et al. GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat. Cell Biol.* **18**, 21–32 (2016).
54. Lundin, V. et al. YAP regulates hematopoietic stem cell formation in response to the biomechanical forces of blood flow. *Dev. Cell* **52**, 446–460.e5 (2020).
55. Uenishi, G. I. et al. NOTCH signaling specifies arterial-type definitive hemogenic endothelium from human pluripotent stem cells. *Nat. Commun.* **9**, 1828 (2018).
56. Lin, F. J., Tsai, M. J. & Tsai, S. Y. Artery and vein formation: A tug of war between different forces. *EMBO Rep.* **8**, 920–924 (2007).
57. Richard, C. et al. Endothelio-mesenchymal interaction controls runx1 expression and modulates the notch pathway to initiate aortic hematopoiesis. *Dev. Cell* **24**, 600–611 (2013).
58. Mirshekar-Syahkal, B. et al. Dlk1 is a negative regulator of emerging hematopoietic stem and progenitor cells. *Haematologica* **98**, 163–171 (2013).
59. Voss, T. C. et al. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* **146**, 544–554 (2011).
60. Bevington, S. L. et al. IL-2/IL-7-inducible factors pioneer the path to T cell differentiation in advance of lineage-defining factors. *EMBO J.* **39**, e105220 (2020).
61. Armesilla, A. L. et al. Vascular endothelial growth factor activates nuclear factor of activated T cells in human endothelial cells: a role for tissue factor gene expression. *Mol. Cell Biol.* **19**, 2032–2043 (1999).
62. Jia, J. et al. AP-1 transcription factor mediates VEGF-induced endothelial cell migration and proliferation. *Microvasc. Res.* **105**, 103–108 (2016).
63. Wang, X. et al. YAP/TAZ orchestrate VEGF signaling during developmental angiogenesis. *Dev. Cell* **42**, 462–478.e7 (2017).
64. Magin, T. M., McWhir, J. & Melton, D. W. A new mouse embryonic stem cell line with good germ line contribution and gene targeting frequency. *Nucleic Acids Res.* **20**, 3795–3796 (1992).
65. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
66. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
67. Krueger, F., James, F., Ewels, P., Afyounian, E. & Schuster-Boeckler, B. FelixKrueger/TrimGalore: v0.6.7. https://doi.org/10.5281/zenodo.5127898 (2021).
68. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
69. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
70. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
71. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: Identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).
72. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
73. Saldanha, A. J. Java Treeview-extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
74. Comoglio, F. et al. Thrombopoietin signaling to chromatin elicits rapid and pervasive epigenome remodeling within poised chromatin architectures. *Genome Res.* **28**, 295–309 (2018).

75. Wingett, S. et al. HiCUP: Pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310 (2015).
76. Mifsud, B. et al. GOTHiC, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. *PLoS One* **12**, e0174744 (2017).
77. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
78. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
79. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

Acknowledgements

This research was funded by a project grant from the Biotechnology and Biological Sciences Research Council (BBSRC) to C.B. and J.B. (BB/R014809/1), a BBSRC MiDTP studentship to C.B. for A.M., a BBSRC MIBTP studentship to A.M., a BBSRC LoLa grant to C.B and B.G. (BB/I001220/2), as well as grants from Blood Cancer UK (15001) and the Medical Research Council (MR/S021469/1) to CB and P.N.C. We thank Genomics Birmingham for expert sequencing services, Nunzia Nirchio for technical support and the Birmingham Technology Hub for cell sorting facilities.

Author contributions

B.E.-W., A.M., S.K., L.A., D.G., and M.C. performed experiments, generated and analysed data, P.K., S.A.A., and I.P. analysed data, J.B. supervised data analysis and together with P.N.C. analysed data and helped writing the manuscript, B.G. contributed resources and helped writing the manuscript, C.B. conceived and directed the study and C.B. and B.E.-W. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-35910-9.

Correspondence and requests for materials should be addressed to B. Edginton-White or C. Bonifer.

Peer review information *Nature Communications* thanks Emery Bresnick, Leonard Zon, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available