Elucidating the molecular mechanisms underlying the endothelial-to-hematopoietic transition



Lucas Greder
Green Templeton College
University of Oxford

A Thesis submitted for the degree of Doctor of Philosophy

Michaelmas 2017

Candidate Number: 682210
Word Count: 43,581

Acknowledgements

I would like to thank all of the people who have helped me over the past 4 years. The list extends well beyond the brilliant scientists that I've had the pleasure to work with. So I apologize in advance for not thanking you directly if you think you deserve one. I kept the circle tight. Everyone who is in knows they are in.

My time in the WIMM was much more then just about what made it into this thesis—or what didn't make it in. At the beginning of my DPhil, I didn't believe in the "you have to make it through the cloud" mantra. No one needed to aimlessly wander only to find their way through after they've lost site of the things that really matter. That being said, I wondered. A lot. I am grateful that I did. I learned so much about Science. I learned so much about myself. I can't imagine having had any other project. I got to live on the technological edge, and occasionally fall off it. For this I have to thank my supervisor, Marella de Bruijn. You trusted me from the very beginning to chart my own path. You guided me even when I didn't know it, or didn't want to know it. You genuinely cared about me not only as a scientist, but also as a person. That is what made the wandering worth it. Thank you.

Vincent: The only thing I value you more than your scientific prowess is your friendship. So much of the work in this thesis is the product of our discussions and collaborations. You've been an incredible friend. Thank you. For the record, you're still my best French friend.

Chrissy: Without your guidance on all things embryo I would've been lost. Thank you for being such a great friend and lab mate.

Joe: My thesis wouldn't be the same without. Your contributions were brilliant. Thank you.

Emanuele and Dominic: You guys are in the circle. Thank you for the great discussions and help along the way.

Table of Contents

Acknowledgements **3**

Table of Contents **4**

List of abbreviations: **9**

Abstract **11**

Chapter 1: Introduction **12**

1.1 Hematopoietic stem cells **13**

HSC sit atop the adult hematopoietic hierarchy 13

De novo generation of HSC 15

1.2 Developmental hematopoiesis **16**

1.3 The origin of the hematopoietic lineage **18**

Hemangioblast 18

Hemogenic endothelium and EHT 18

Competent and specified HE and the point of hematopoietic commitment 19

HSC precursors 22

1.4 Other hematopoietic tissues and the case for AGM dHSCs **23**

Embryonic and definitive hematopoiesis arise independently 23

Extra-embryonic yolk sac 25

Intra-embryonic PAS/AGM 25

Other sites of hematopoiesis – Placenta and Umbilical 26

Fetal liver 27

1.5 Extrinsic regulators of developmental hematopoiesis **27**

BMP signaling 27

Hedgehog-Vegfa signaling 28

Notch signaling 29

TGFβ signaling 30

1.6 Intrinsic regulators and transcriptional hierarchy at onset of EHT **32**

Transcriptional regulation of Flk1+ mesoderm 32

Transcriptional regulation at onset of EHT in hemogenic endothelium 33

1.7 Transcriptional regulation at onset of EHT – Runx1 focused **34**

Requirement for Runx1 during development 36

TFs downstream of Runx1 38

1.8 Studying transcriptional networks (Gene regulatory interactions) **39**

1.7 Project aims **42**

Chapter 2: Materials & Methods 45
2.1 Cell culture 46
mESC Culture 46
mESC Differentiation 46
mESC Targeting 47
2.2 Mice 49
Animal husbandry 49
Ear notching and genotyping 49
Timed matings, embryo collection and dissections 51
Dissections and generation of cell suspensions 51
Flow cytometry and cell sorting 53
2.3 RNA-sequencing 53
RNA-sequencing library preparation and sequencing 53
RNA-sequencing analysis 54
2.4 Chromatin assays 54
ATAC-seq library preparation and sequencing 54
ATAC-seq data processing 55
Categorization of elements into subsets 56
Motif enrichment, GREAT and LOLA 57
Intersecting RNA-seq and ATAC-seq data 57
Construction of gene regulatory interaction maps 58
2.5 Molecular cloning 58
Generation of targeting vectors 58
gRNA design and cloning 59
In vitro gene knockouts – vector and edit detection 62
Lentivirus generation 65
2.6 Ex vivo gene knockouts, lenti infections, and re-aggregation cultures 65
2.7 Fluidigm 66
Microfluidic quantitative RT-PCR. 66
Fluidigm analysis 69
Statistical analysis 69
2.8 Functional assays: 69
Mouse colony forming cell (CFU-C) 69
OP9 maintenance and co-culture assay for hematopoietic potential 70
2.9 General assays 71
Confocal 71
Western Blotting 73

Luciferase....73

Chapter 3: Expression profiling of EHT....75

3.1 Expression profiling of EHT by RNA-seq....78

Sorting, data processing, and quality controls....78

Differential expression analysis....80

Differentially expressed TFs....83

Expression of TFs used for reprogramming studies....86

3.2 Expression profiles during early EHT: competence and specification stages....90

Endo to cHE....90

cHE to sHE....92

3.3 Expression profiles during determination phase of EHT....95

sHE to proHSC....95

proHSC to preI....99

PreI to preII....101

3.4 RNA-seq: enriched pathways – Regulation of actin cytoskeleton organization....103

Blocking actin reorganization blocks EHT....105

Runx1-associated changes in actin organization and cell shape....107

Runx1 may regulate actin reorganization through direct activation of Gelsolin107

3.5 Conclusion....112

Chapter 4: Chromatin profiling of EHT....114

4.1 Chromatin profiling of EHT....117

Sorting, sequencing, data processing and quality control....117

Cell type specific accessibility....119

Principal component analysis....121

4.2 Differentially accessible chromatin: Global and Pairwise....123

Global differential accessibility....124

Pairwise differential accessibility....127

4.3 Chromatin accessibility in cHE and sHE....131

Differential accessibility....132

Mechanisms underlying chromatin accessibility dynamics....134

4.4 Chromatin accessibility in sHE and proHSC....136

Differential accessibility....138

Mechanisms underlying chromatin accessibility dynamics....140

4.5 Conclusion and discussion....144

Chapter 5: Gene regulatory interactions during EHT 145
5.1 Strategy for integration of ATAC-seq and RNA-seq 146
5.2 Global integration of expression and chromatin profiles 149
Novel candidate regulator of preHSC maturation – *Stat4* 151
5.3 Gene regulatory interactions governing the cHE to sHE transition 153
Integration of expression and chromatin profiles 153
PI3K-AKT signaling during cHE to sHE transition (and throughout EHT) 156
Predicted TF network from integration of expression and chromatin profiles .. 159
Predicted TF network from in silico investigation of ATAC-seq elements 162
Overlay of RNA-seq/ATAC-seq integration with LOLA analysis 166
New predicted gene regulatory interaction: Runx1 and Bcl11a 166
5.4 Predicted gene regulatory interactions for sHE to proHSC transition ... 168
Integration of expression and chromatin profiles 170
Predicted TF network from integration of expression and chromatin profiles .. 171
Predicted TF network from in silico investigation of ATAC-seq elements 174
Overlay of RNA-seq/ATAC-seq integration with LOLA analysis 176
5.5 Conclusions and discussion 178
Chapter 6: Runx1-Ikzf1-Notch 182
6.1 Identification and characterization of the Ikzf1 -28 cis-element 183
-28 regulates *Ikzf1* expression during mESC differentiation 186
Ikzf1-28 regulates *Ikzf1* expression *in vivo* 188
6.2 Runx1 regulation of Ikzf1 via binding to -28 element 190
Runx1 regulation of Ikzf1 during *in vitro* mESC differentiations 190
Runx1 regulation of Ikzf1 during ex vivo EHT 192
Runx1 dependent activity of Ikzf1-28 cis-element 194
Summary 196
6.3 Dynamic regulation of Notch signaling in vivo 196
Runx1-Ikzf1-Notch hypothesis 197
Notch expression and chromatin profiles during EHT 199
6.4 sc-qRT-PCR for *Runx1*, *Ikzf1* and Notch target during in vivo EHT 201
6.5 Investigation of Runx1-Ikzf1-Notch regulatory interaction 203
Runx1-ERT2 fusion to study Runx1-dependent events 204
RUNX1-ERt2 controls Ikzf1 and Notch targets 206
Generation of Ikzf1-/- mESC line 207
Notch regulation is dependent on Runx1 and Ikzf1 209
6.6 Discussion 211

Runx1-Ikzf1-Notch regulatory cascade 211

Chapter 7: General discussion 213

7.1 Summary of results 214

7.2 Molecular mechanisms of EHT – putative regulators and GRIs 215

In vivo expression profiling 215

In vitro regulatory networks 216

Integration of expression and chromatin profiles identified new regulators and GRIs 216

CRISPR/Cas9-based pooled library screen to identify new regulators 218

Comparing AGM EHT to YS (or in vitro) EHT 220

Resolving differences between in vitro and in vivo Notch requirements 222

7.3 Caveats of GRIs generated from expression and chromatin profiling 222

7.4 Runx1 as direct regulator of cytoskeleton organization 224

7.5 Concluding remarks 225

References 227

Appendix 243

List of abbreviations:

4-OHT	4 hydroxytamoxifen
AGM	aorta-gonad-mesonephros
ATAC-seq	assay for transposable-accessible chromatin sequencing
BL-CFC	blast-colony forming cell
BMP	bone morphogenetic protein
BM	bone marrow
cHE	competent hemogenic endohthelium
C	celsius
CBRG	computational biology research group
CFU-C	colony forming units culture
CFU-S	colony-forming units spleen
ChIP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation sequencing
CO2	carbon dioxide
CMF	calcium and magnesium free
CNE	conserved noncoding element
CPM	counts per million
CRISPR	clustered regularly interspaced short palindromic repeats
Cyto D	cytochalasinD
DAE	differentially accessible element
DB	dissection buffer
DEG	differentially expressed gene
dHSC	definitive hematopoietic stem cells
E	embryonic da
EC	endothelial cells
EHT	endothelial-to-hematopoietic transition
EMP	erythromyeloid progenitor
EMT	epithelial-to-mesencymal transition
FCS	fetal calf serum
FL	fetal liver
FMO	fluorescent minus one
gDNA	genomic DNA
GFP	green fluorescent protein
G-MEM	glasgow modified essential medium
gRNA	guide RNA
HE	hemogenic endothelium
HSPC	hematopoietic stem and progenitor cells
hPSC	human pluripotent stem cell
hsp68	heat shock promoter 68
IVC	individually ventilated cages
LT-HSC	long term HSC
mESC	mouse embryonic stem cells
min	minute
NC	newly closed
NO	newly open
O/N	over night
PAS	para-aortic splanchnopleura
PBS	phosphate buffered saline

PCR	polymerase chain reaction
PFA	paraformaldehyde
qRT-PCR	quantitative reverse-transcriptase PCR
RNA-seq	RNA sequencing
SD	standard deviation
sHE	specified hemogenic endothelium
Shh	sonic hedge hog
TF	Transcription factor
YS	yolk sac
UA	umbilical artery
UCSC	university of california santa clara
ul	microliter
UGR	uro-genital ridge
UVA	umbilical and vitellin artery
VE-Cad	ve-cadherin
WNT	
WTCHG	wellcome trust center for human genetics
WT	Wild type

Elucidating the molecular mechanisms underlying the endothelial-to-hematopoietic transition

Lucas Greder (Green Templeton College)
Doctor of Philosophy – Michaelmas 2017
University of Oxford

Abstract

Definitive hematopoietic stem and progenitor cells (HSPCs) are generated de novo from a subset of endothelium, the hemogenic endothelium (HE), via a process called the endothelial-to-hematopoietic transition (EHT). To facilitate the study of HE throughout this process, our lab generated a *Runx1* enhancer-reporter mouse model (23GFP) to dissect the cellular and molecular events that underlie EHT. Previous work in the lab highlighted that HE undergoes dynamic fate changes already early in development characterized by two distinct states, a competent state (cHE) and a specifying stage (sHE). In the context of EHT, this study rewrote the timeline of hematopoietic commitment during EHT and showed this is initiated in sHE before the proHSC and type I and type II preHSC stages. My aim was to identify and characterize the gene regulatory interactions (GRIs) underlying these cell fate transitions. We reasoned that integration of expression (RNA-seq) and chromatin profiles (ATAC-seq) would facilitate identification of these GRIs. After validating and characterizing the data sets, we focused on TFs where the timing and appearance of the motif coincided with expression of the TF and identified these as putative regulators of accessibility and subsequent changes in gene expression. Focusing on two key cell fate transitions: (1) cHE to sHE and (2) sHE to proHSC, we generated GRI maps. Analysis of these regulatory interactions identified putative causative regulators, including *Bcl11a*, *Mef2c*, *Ikzf1*, and *Ikzf2*. Looking more closely at *Ikzf1*, we identified a -28 element required for proper *Ikzf1* expression. Interestingly, -28 activity was Runx1-dependent. Further investigation identified Notch signaling as a downstream target of the novel the Runx1-Ikzf1 regulatory axis. Finally, interrogation of this regulatory axis using genetic perturbation revealed a complex interaction whereby *Runx1* and *Ikzf1* work both individually and synergistically to regulate Notch during *in vitro* EHT.

Chapter 1: Introduction

1.1 Hematopoietic stem cells

HSC sit atop the adult hematopoietic hierarchy

At the heart of the adult hematopoietic system lie definitive hematopoietic stem cells (HSC or dHSCs) that self renew and give rise to all lineages of the hematopoietic system. The adult hematopoietic hierarchy, with dHSCs at the top, consists of multipotent and lineage restricted blood progenitors (Figure 1.1). The progenitors replenish the different types of terminally differentiated blood cells needed to sustain an organism throughout life. Without HSCs the progenitor pools eventually exhaust, and the organism struggles to sustain requisite populations of effector cells. Thus, HSC-based therapies are essential for diseases requiring replenishment of the hematopoietic progenitor pools.

Unfortunately, our capacity to grow, expand and manipulate HSCs is limited. As a result, the potential for HSC-based therapies has yet to be truly realized. Studying how HSCs—and multipotent blood progenitors—are specified during development will help refine the conditions needed for growth and expansion of HSCs. Moreover, understanding HSC specification will aid the *de novo* generation of HSCs *in vitro* for regenerative medicine purposes. Despite recent advances that generated HSCs from human pluripotent stem cells (hPSCs) (Sugimura et al., 2017) and reprogrammed adult human endothelium (Lis et al., 2017), there remains an incomplete understanding of the precise cellular processes, signaling molecules, and the transcription factors (TFs) that underlie HSC emergence. A thorough understanding of the mechanisms that underlie blood stem and progenitor cell generation in the mammalian embryo will help address these gaps in our knowledge. In the following

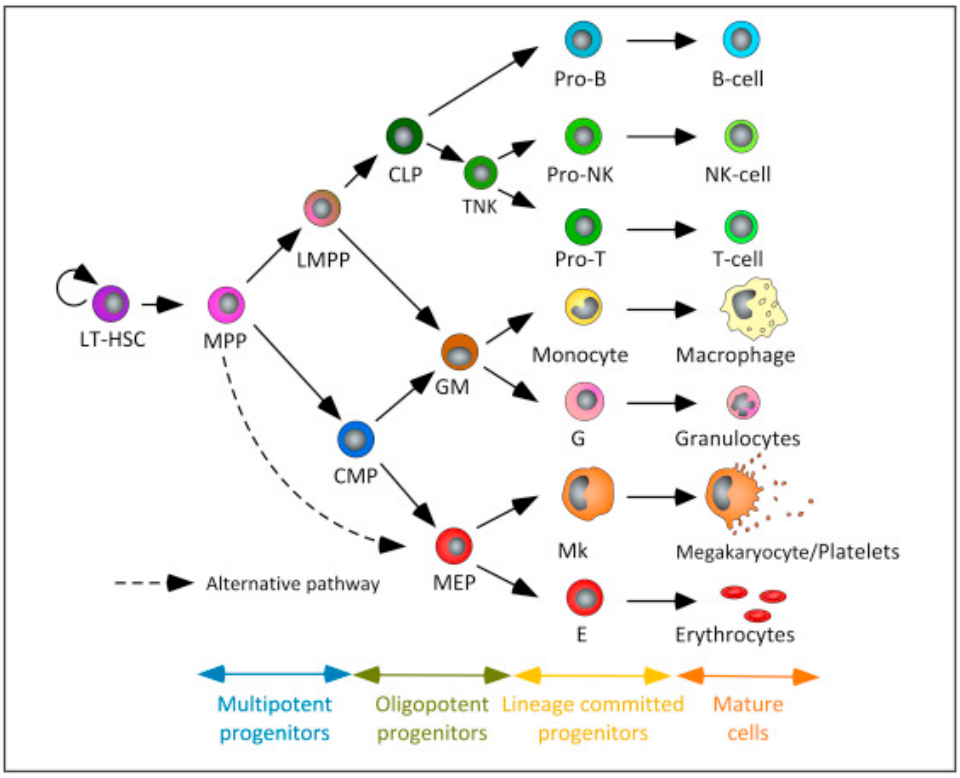


Figure 1.1: The hematopoietic hierarchy
Schematic of hierarchy for adult hematopoiesis as depicted by Godin et al., 2016.

sections, I address the scientific foundation upon which our knowledge of the field is currently based.

De novo generation of HSC

The two predominant methods for generating HSCs in vitro are (1) the direct reprogramming of cells into HSCs, and (2) the directed differentiation of pluripotent stem cells. Early studies using direct reprogramming used over-expression of *Oct4* to convert fibroblasts and mesenchymal stem cells to CD45+ blood progenitors and red blood cells, respectively (Szabo et al., 2010; Liu et al., 2015). The Rossi and Daley labs made considerable strides when they generated multipotent progenitors from committed blood cells (pre- and pro-B cells, respectively) by over-expressing a set of TFs (Riddell et al., 2014; Doulatov et al., 2012). Finally, a recent study by Rafii and colleagues used a cocktail of transcription factors to reprogram adult endothelial cells into HSCs (Lis et al., 2017). Using transient expression of Fosb, Gfi1, Runx1 and Sfpi1 (PU.1) and vascular-niche-derived angiocrine factors supplied by immortalized endothelial cells, the group generated definitive HSCs that no longer required expression of the exogenous factors.

Directed differentiation of embryonic stem cells (ESC) or induced pluripotent stem cells (iPSC) has been achieved for various blood lineages including platelets (Lu et al., 2011), NK cells (Woll et al., 2009), B cells (Carpenter et al., 2011) and T cells (Kennedy et al. 2012; Galic et al., 2006). However, no study to date has generated definitive HSC from ESC or iPSC alone. Recent work by Daley and colleagues combined directed differentiation with direct reprogramming to generate HSCs from hPSCs (Sugimura et al., 2017). Using cytokines, PSCs were differentiated into hemogenic endothelium followed by

induction of hematopoietic stem-cell-specifying TFs Sfpi1, Runx1, Lcor, Hoxa10, Hoxa9, Hoxa5, and Erg. This resulted in production of hematopoietic stem and progenitor cells.

Despite these recent gains, many questions and concerns remain unanswered. For example, both Lis et al. and Sugimura et al. use overexpression of TFs to generate HSC, some of which are potent regulators of transcription or have oncogenic properties. In addition, Sugimura et al. use *in vivo* specification of the hematopoietic precursors, highlighting the lack of knowledge surrounding the extrinsic and intrinsic signals needed to mature HE into HSCs. Thus, there is a need to study this process in the developing mouse embryo to help identify the key missing pieces.

1.2 Developmental hematopoiesis

Developmental hematopoiesis occurs in three waves with early events in the yolk sac and aorta-gonad-mesonephros (AGM) region (Figure 1.2). The first wave generates embryonic macrophages, megakarycotyes and primitive erythroid progenitors in the yolk sac. This wave initiates at the onset of gastrulation and continues through to E8.25 (Palis et al., 1999) after which the second, transient wave begins. In the second wave, the yolk sack produces erythromyeloid progenitors (EMPs) and megakaryocytes from HE (Palis et al., 1999; Tober et al., 2007). This wave coincides with the initiation of circulation at E8.25 (McGrath et al., 2003), which EMPs use to colonize the fetal liver (FL) (McGrath et al., 2011). This wave continues through E9.5 and also occurs in the AGM region. During this wave lymphoid progenitors emerge from HE in the YS and AGM. Finally, wave three begins with the first HSCs arising from HE/proHSCs/preHSCs in the AGM region at E9.5 and E10.5

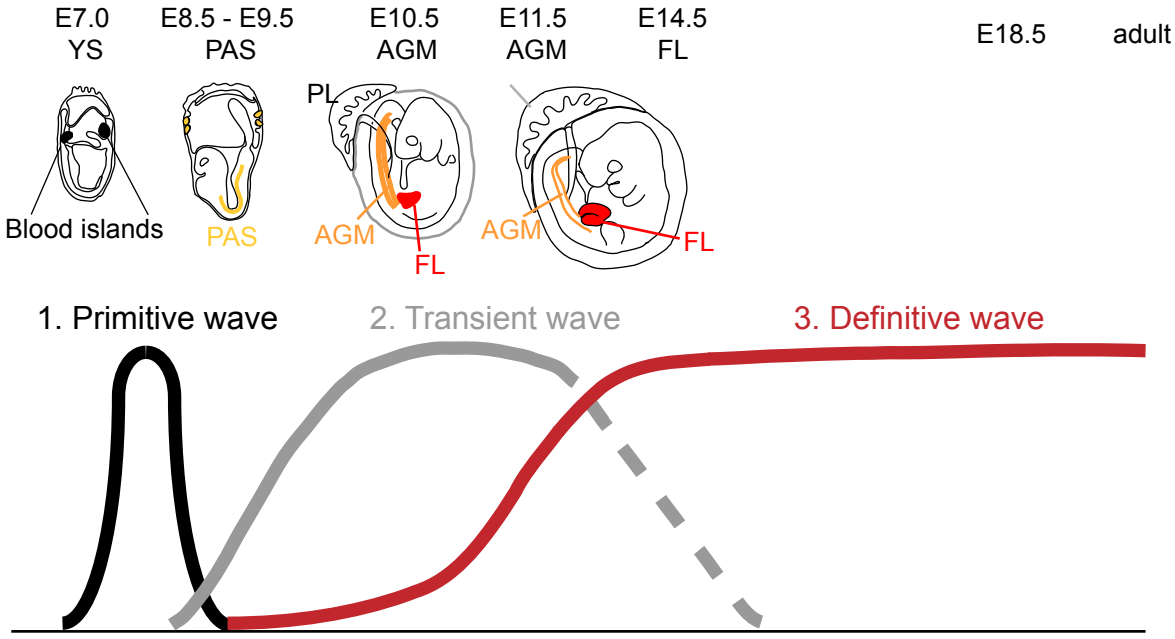


Figure 1.2: Developmental hematopoiesis
Schematic of the 3 waves of developmental hematopoiesis with the embryonic sites of developmenal hematopoiesis above the different waves.

(Rybtsov et al., 2011, 2014, Taoudi et al., 2008). In parallel, wave two continues to produce EMPs and lymphoid progenitors in the YS and AGM. Ultimately, the three waves culminate in the generation of HSCs that colonize and then expand in the FL before ultimately residing in the BM where they sustain blood progenitors throughout the life of the organism.

1.3 The origin of the hematopoietic lineage

Hemangioblast

Historically, the origin of the hematopoietic lineage was based on histological sections of yolk sac blood islands showing that early erythroid progenitors were surrounded by an outer endothelial lining (reviewed in Medvinski et al., 2011). Based on this proximity, it was hypothesized that bi-potent hemangioblasts gave rise to both endothelial and hematopoietic cells (Maximov 1909 and Sabin 1920). Although subsequent work exposed shortcomings of the original experiments, the existence of a hemangioblast was confirmed in *in vitro* embryonic stem cell (ESC) differentiation cultures where blast-colony forming cells (BL-CFC) exhibited bi-potential capacity (Lancrin et al., 2009; Kennedy et al., 2007; Choi et al., 1998). However, *in vivo* lineage tracing in the embryo has failed to identify a hemangioblast-like population. Many questions remain about their bi-potentiality and even their existence outside of an *in vitro* ES cell differentiation model.

Hemogenic endothelium and EHT

A proposed alternative to the hemangioblast hypothesis was that blood forms directly from a subset of endothelium called the hemogenic endothelium (HE)

via a process called the endothelial-to-hematopoietic transition (EHT) (Smith and Glomski, 1982). Early experiments in chick embryos indicated the existence of HE using dye-labeling experiments (Jaffredo et al., 1998). In addition, lineage-tracing studies in the mouse showed that endothelial cells (EC) labeled by a tamoxifen-inducible VE-Cadherin-CRE transgene contribute to the HSC lineage (Zovein et al., 2008). Further support for the HE hypothesis came from live imaging studies of ESC differentiation cultures (Eilken et al., 2009). After the formation of sheets of ECs, a subset of cells began to express CD41, a marker of hematopoietic potential (Ferkowicz et al., 2002, Mikkola et al., 2003), before going on to be CD41+CD45+ hematopoietic cells (Eilken et al., 2009). Additional live imaging studies in zebrafish (Bertrand et al., 2010; Kissa and Herbomel, 2010) and slices of mouse AGM (Boisset et al., 2010) confirmed the original observations. In an effort to reconcile the hemangioblast and EHT hypotheses, it was proposed that Flk1+ hemangioblasts first generate hemogenic endothelium and then produce hematopoietic cells (Lancrin et al., 2009). However, without additional studies confirming the in vivo presence of hemangioblasts, this hypothesis appears to be an in vitro phenomenon. In summary, it is now accepted that hematopoietic stem and progenitors (HSPCs) emerge from a specialized subset of endothelium called the HE through a process known as the endothelial-to-hematopoietic transition (EHT).

Competent and specified HE and the point of hematopoietic commitment

Despite EHT being imaged in real-time (Lancrin et al., 2009; Eilken et al., 2009; Boisset et al., 2010; Kissa and Herbomel 2010; Bertrand et al., 2010; Lam et al., 2010), questions remained regarding at which point cells initiated

their commitment to the hematopoietic lineage. Specifically, little was know about how this fate change was regulated at the molecular level. A great understanding of these molecular mechanisms will facilitate *de novo* generation of HSC *in vitro*. In order to study these cell fate changes, the field needed to be able to identify the cellular intermediates of EHT.

To facilitate the identification of cellular intermediates of EHT, our lab previously generated a Runx1 +23 enhancer-reporter mouse model (23GFP) (Bee et al., 2010; Nottingham et al., 2007). 23GFP expression preceded *Runx1* expression in E8.5 but otherwise recapitulated endogenous *Runx1* expression (Swiers et al., 2013). This allowed prospective isolation of hematopoietic-fated cells prior to the onset of Runx1 expression (Swiers et al. 2013). Importantly, endogenous Runx1 levels are unaffected, thereby allowing analysis of normal hematopoietic development.

Using this model, the lab performed the first clonal analysis of HE undergoing EHT in E8.5 to E10.5 mouse embryos (Swiers et al., 2013). Single cell functional and transcriptional analysis of HE revealed that gradual initiation of the hematopoietic program occurred earlier than expected and was accompanied by an unexpected early loss in endothelial potential (Figure 1.3). HE could be separated into two stages. E8.5 HE marked the compotent stage (cHE) where the cells were capable of responding to blood inducing signals. E9.5 marked the specifying stage (sHE) where the cells began to commit to the hematopoietic lineage. At E9.5, the initiation of commitment to the hematopoietic lineage coincided with expression of endogenous *Runx1* and occurred as the cells were still part of the endothelial wall (Swiers et al., 2013). Although not yet formally proven, it is believed that sHE go on to

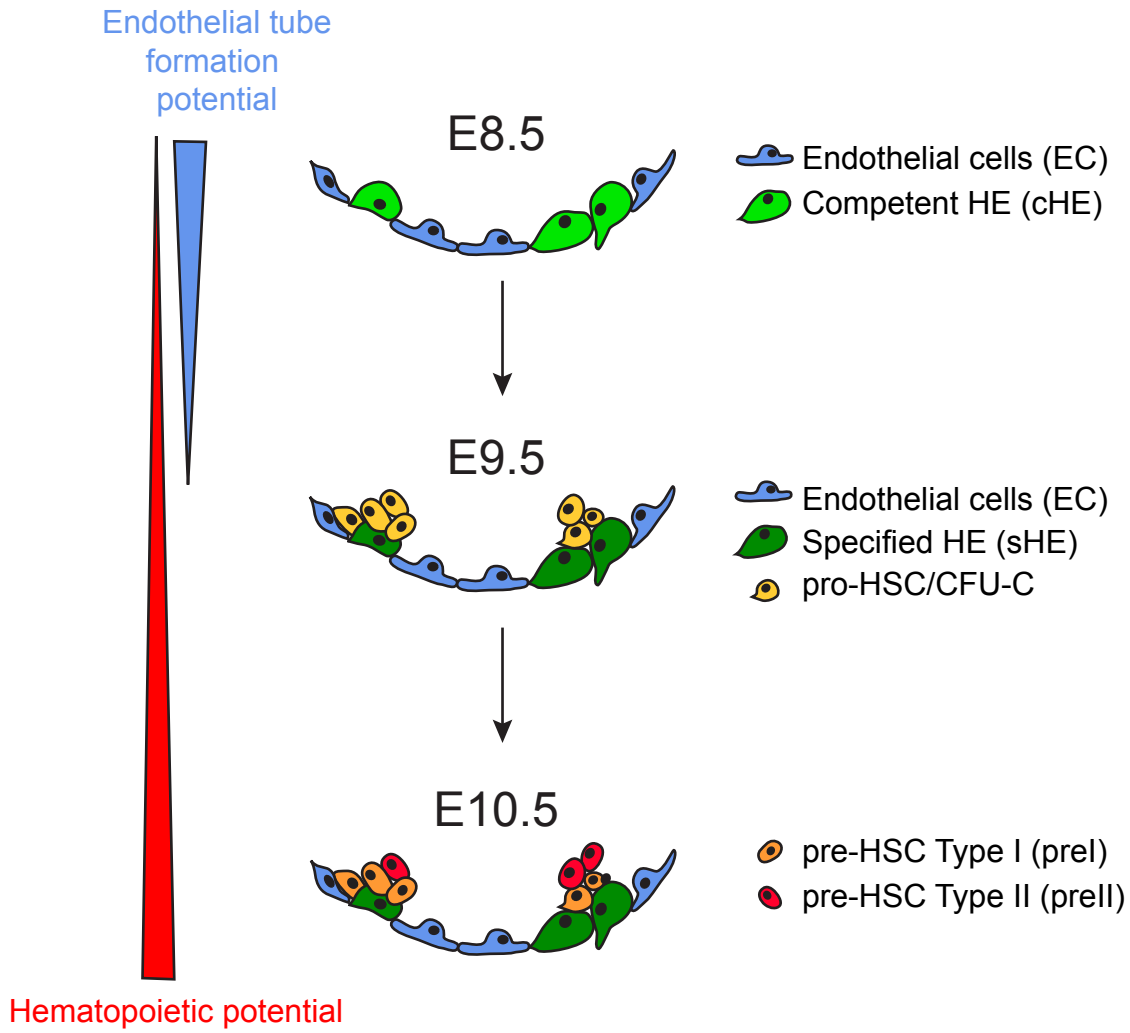


Figure 1.3: The celllular intermediates of the endothelial to hematopoietic transition. At E8.5, Endothelial cells (EC) and competent hemogenic endothelium (cHE) retain some endothelial tube formation potential. cHE has reduced enothelial capacity and is capable of responding to blood inducing signals. E9.5 includes EC, specified hemogenic endothelium (sHE), and pro-HSC/C-FU-C where the proHSC are fully committed to the hematopoietic lineage. E10.5 consists of pre-HSC Type I (preI) and pre-HSC Type II with few if any endothelial cells still undergoing a transition to endothelium with hematopoietic potential.

mature into the proHSC and preHSCs phenotypes discussed in the next section.

HSC precursors

An E11 embryo contains less than 3 total HSCs with one confirmed to reside in the AGM and one residing in the FL (Kumaravelu et al., 2002). An E12.5 AGM contains 3 HSCs where as the FL, after having been recently colonized, contains >50 HSCs (Kumaravelu et al., 2002). At E12.5, the number of HSPCs in circulation also increased. This rapid expansion is likely explained by the presence of immature HSC precursors present at earlier developmental time points (Rybtsov et al., 2016). Limitations with ex vivo culture conditions limited the study of such small cell number populations ex vivo and therefore the mechanisms of this expansion could not be studied.

Recent advancements in dissociation-reaggregation cultures enabled ex vivo maturation of hematopoietic precursor cells en route to HSC development (Sheridan et al., 2009). First insights from the system revealed that nearly all HSCs originate from a VEC+CD45+ preHSC (Taoudi et al., 2008). Subsequent work further characterized these preHSC subdividing HSC precursors into E9.5 proHSC, E10.5 preHSC Type I (preI), and E10.5 preHSC Type II (preII) (Rybtsov et al., 2011, 2014, Taoudi et al., 2008). Distinct cell surface phenotypes allow isolation and study of these populations including: proHSC (VE-Cad+CD41+CD43-CD45-), preI (VE-Cad+CD41+CD43+CD45-) and preII (VE-Cad+CD41+CD43+CD45+). Disassociation and re-aggregation followed by 4 days of culture was enough to mature preI and preII pre-HSCs into HSCs. However, 7 days in culture was required for pro-HSCs maturation into HSCs. A revised model of EHT includes cHE, sHE and HSC precursors

situated in the following hierarchy for definitive EHT: Endo → cHE → sHE → proHSC → preI → preII (Figure 1.4).

1.4 Other hematopoietic tissues and the case for AGM dHSCs

Sections 1.2 and 1.3 describe AGM hematopoiesis as the source of definitive HSC (dHSC) but other tissues are also involved in hematopoiesis. The existence of multiple sites has complicated how the field relates embryonic and adult hematopoietic hierarchies. Moreover, the adult hematopoietic hierarchy appears in reverse sequential order during embryogenesis (Medvinsky et al., 2011). Specifically, during development, the production of differentiated hematopoietic cells occurs prior to the development of dHSCs. There are currently two models that propose explanations for this reversal and account for the multiple sites of hematopoiesis.

Embryonic and definitive hematopoiesis arise independently

The first model suggests a common origin for the embryonic and adult hierarchies whereby the “ancestor” generates the first wave of hematopoiesis (Medvinsky et al., 2011). This “ancestor” then gradually develops “stemness” on its way to becoming a dHSC. The more widely accepted model posits that the embryo generates embryonic and adult (definitive) hematopoiesis independently (Medvinsky et al., 2011). The first wave, the embryonic wave, emerges first and provides primitive blood cells. This is followed by a wave of definitive hematopoiesis, which produces the dHSCs. Within the second model, the first progenitors and the dHSC develop independently as part of different hierarchies. The second model is supported by observations made in frog, zebrafish and chick (Ciau-Uitz et al., 2000; Dieterlen-Lievre, 1975). And in mouse, endothelium from the YS and AGM arise from different mesodermal

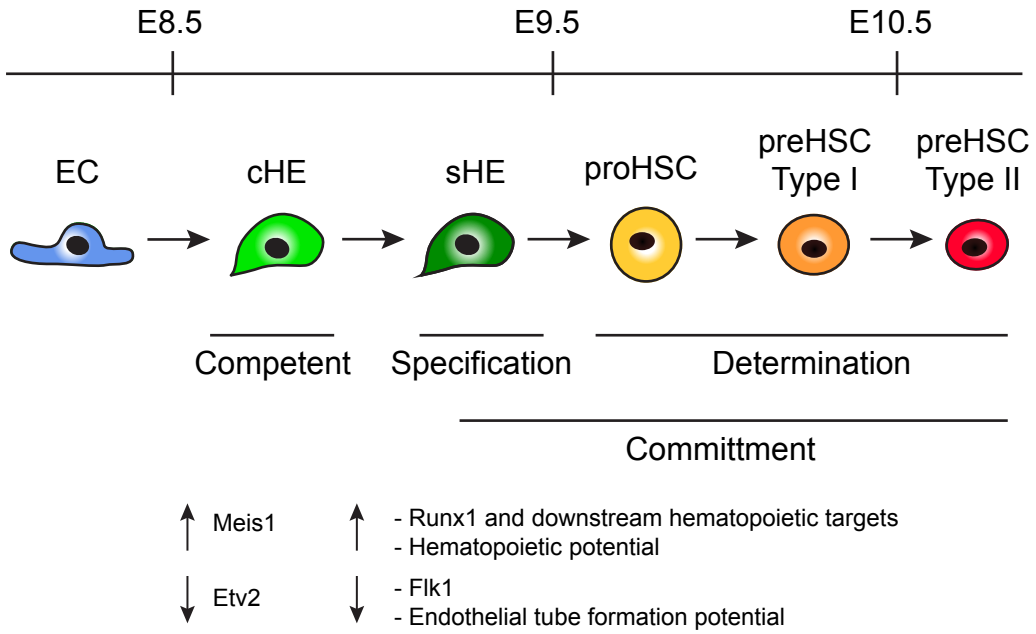


Figure 1.4: The celllular intermediates of the endothelial to hematopoietic transition. Schematic showing the six different cell types in the revised model of EHT. Competent stage consists of 23GFP+ cHE (but Runx1-) capapble of responding to blood inducing signals and marked by increase in *Meis1* and decrease in *Etv2*. Specificaiton includes 23GFP+ Runx1+ sHE where Runx1 is instituting a hematopoietic program and repressing an endothelial program. Commitment phase includes all cell types where the hematopoeitic program is locked in. Determination phase includes all pro- and preHSCs as they mature towards progenitor or HSCs.

progenitors suggesting the YS-derived embryonic wave and AGM-derived definitive wave arise independently (Kinder et al., 1999; Lawson et al., 1991).

Extra-embryonic yolk sac

Beginning at E7.0 of mouse development in the yolk sac, hematopoietic differentiation produces the first wave of progenitors made up of macrophages, megakaryocytes, and primitive erythoid progenitors (Silver and Palis, 1997; Palis et al., 1999). Additional work showed YS potential included myeloid progenitors (CFU-Cs) but lacked CFU-S at E9.5 (colony-forming-units-splenic) and dHSC at E11.5 (Moore and Metfalf, 1970; Medvinsky et al., 1993; Medvinksky and Dzierzak, 1996; Muller et al., 1994). Also originating in the YS is the second, transient wave, which consists mainly of EMPs (Tober et al., 2007). However, recent work showed a immune-restricted lymphoid-primed progenitor emerging from the Yolk Sac during this second wave (Boiers et al., 2013). Prior to the onset of circulation at E8.25, there is no circulation to distribute progenitors to the embryo proper (Lux et al., 2008). Thus, the absence of CFU-Cs present within the pre-circulation supports the idea that EMP of the second wave derived from the YS.

Intra-embryonic PAS/AGM

After E9, hematopoietic potential was detected in the intra-embryonic para-aortic splanchnopleura (PAS) region (Moore and Metcalf, 1970) with CFU-S arising at E9.5 (Medvinsky et al., 1999). To ensure an intra-embryonic origin, YS and PAS separated prior to the onset of circulation revealed immature blood progenitors only in intra-embryonic PAS (Cumano et al., 1996). As the PAS region matures into the AGM around E10, this region was found to give rise to dHSCs capable of long-term repopulation potential (LT-HSC)

(Medvinsky and Dzierzak, 1996; Mueller et al., 1994). Further dissection of the AGM region into the dorsal aorta (DA) and the uro-genital ridges (UGR) revealed the DA harbored the dHSC (de Bruijn et al., 2000). Additional work separating these two components revealed that each was responsible for producing specific inductive signals (Soulihol et al., 2016a).

Other sites of hematopoiesis – Placenta and Umbilical

The placenta was first known to harbor dHSCs almost 50 years ago (Dancis et al., 1977; Dancis et al., 1968). However, debate still lingers as to whether dHSC come from placenta *de novo*. Mouse placenta at E9.0 contained hematopoietic cells and expressed Runx1 (Gekas et al., 2005). dHSC appeared in the placenta during the same developmental time window as the AGM, E10.5-E11.5 (Gekas et al., 2005). During explant culture the placenta was not capable of expanding the number of dHSC where other sites with *de novo* dHSC capability were able to expand dHSC numbers (Robin et al., 2006). Although more work is needed to confirm these findings, the current working hypothesis was that the placenta was not a site of *de novo* dHSC formation. Similar to the placenta, the mouse umibilical cord and vessles were hematopoietically active and contained Runx1-postive clusters (Caprioli et al., 1998; Liakhovitskaia et al., 2009; North et al., 1999). dHSC were present at E10.5 but further work was needed to show *de novo* dHSC potential (de Bruijn et al., 2000b). In summary, while both sites undergo a Runx1-dependent endothelial-to-hematopoietic transition as was evidenced by the appearance of hematopoietic clusters, it was not clear whether placenta or umbilical harbor *de novo* dHSC potential.

Fetal liver

At E11.5, AGM, YS and placenta each contain approximately one dHSC, with a total of three HSCs in the whole conceptus (Kumaravelu et al., 2002). However, at E12.5, greater than 60 dHSCs are present in the embryo with only three found in the AGM region and the rest residing in the FL (Kumaravelu et al., 2002). Expansion of HSCs occurs through maturation of preHSCs that begins as early as E9.5 in the AGM (Rybotsov et al., 2016). Further expansion and maturation occur in the FL likely through induction of proliferation by angiopoietin-like factors (Medvinsky et al., 2011; Zhang et al., 2006). Ultimately, dHSCs migrate to and reside in the bone marrow (BM) in a quiescent, inactive state (Wilson et al., 2008).

1.5 Extrinsic regulators of developmental hematopoiesis

BMP signaling

In mice, bone morphogenetic protein (BMP) signaling together with WNT ligand was essential for patterning the ventroposterior mesoderm, a key step in mesoderm specification (Miura et al., 2006). Prior to EHT, BMP antagonists Chordin and Noggin prohibit the fusion of the paired dorsal aortae—a prerequisite to HSC emergence in quail embryos (Reese et al., 2004). In zebrafish, immediately preceding the onset of *Runx1* expression, BMP antagonists are removed by the pro-BMP signal Tolloid (Wilkinson et al., 2009). In mice, one study established that shear stress from blood flow induced PKA-CREB activation in ECs triggered Bmp4 release to neighboring HE cells (Kim et al., 2015). Activation of Bmp4 via flow induced shear stress matches timing requirements for Bmp signaling during HSC emergence.

Bmp4 signaling then triggers HSC specification from HE in the DA (Wilkinson et al., 2009).

In mice, a Bmp4 gradient drives HSPC emergence and impacts maintenance of the HSC pool (Durand et al., 2007; Goldman et al., 2009; Wilkinson et al., 2009). Two recent reports highlight the complex nature of Bmp4 signaling after HSC emergence. One study found increased levels of BMP antagonists in preHSC suggesting that inhibition of BMP is required for preHSC maturation (Soulihoul et al., 2016a). Using a BMP-response element to report BMP activity, the other study found that a subset of preHSC are BMP-activated and a subset are non-BMP-activated (Crisan et al., 2016). At the RNA level, non-BMP-activated preHSC were more like HSPCs and BMP-activated were more like HE. Despite the recently discovered complexity, it is accepted that Bmp4 signaling is required for HSC emergence but not for arterial specification (Wilkinson et al., 2009).

Hedgehog-Vegfa signaling

In zebrafish, the absence of sonic hedge hog (Shh) results in defects in HSC formation (Gering and Patient, 2005). Among other properties, Shh signaling mediates the midline convergence of pre-hematopoietic Kdr (Flk1) expressing mesoderm (Gering and Patient, 2005; Williams et al., 2010; Wilkinson et al., 2012). Moreover, Shh activation of Vegfa regulates angioblast migration (Cleaver et al., 1998). Therefore, prior to HSC specification, Shh-vegfa regulates arteriovenous specification by regulating a precursor's ability to converge at the midline (Clements and Traver, 2013). In some instances where precursors do reach the midline in the absence of Shh, the cells still lack hematopoietic potential, suggesting that Shh regulates arteriovenous and

HSC specification (Gering and Patient, 2005; Williams et al., 2010; Wilkinson et al., 2012).

The Shh-vegfa signaling axis directly regulates Notch activation of HSC specification (Lawson et al., 2002, Clements and Traver, 2013), suggesting that it sits at the apex of definitive hematopoiesis. Vegfa exerts its influence through binding to *Kdr* (*Flk1*) in EC and regulating expression of key downstream targets (Lawson et al., 2002), which include notch receptors (*Notch3*) and genes involved in arterial specification, like *Efnb2a* (Lengerke et al., 2008, Lawson et al., 2002). This further supports the idea that Shh-vegfa is required for arteriovenous and HSC specification.

Notch signaling

Notch signaling is required for establishing arterial identity (Kumano et al., 2003; Duarte et al., 2004; Krebs et al., 2000; Krebs et al., 2004; Lawson et al., 2001; Lawson et al., 2002), specification of HE (Clements et al. 2011; Kim et al., 2014; Kobayashi et al., 2014), and has an essential role in HSC emergence (Kumano et al., 2003; Robert-Moreno et al., 2005; Robert-Moreno et al. 2008). In contrast, Notch signaling is dispensable for maintenance of adult HSCs (Malliard et al., 2008). Notch ligand (Jag1, Jag2, Dll1, Dll4) binding to a Notch receptor (Notch1-4) results in release of the Notch intracellular domain that works with RBP-JK to activate transcriptional targets (*Hey1*, *Hey2*, *Hes1*, among others) (Andersson et al. 2011).

Prior to HSC specification, Notch signaling is required for arteriovenous specification (Lawson et al., 2001). Mice lacking *Notch1* or other Notch components (e.g., RBP-JK, Hey2) lose arterial-specific marker expression of Ephrin B2 (*Efnb2*) (Krebs et al., 2000; Zhong et al., 2001; Fischer et al.,

2004). In addition, Notch signals mediated by Dll4 drive growth of the arterial endothelium from which hematopoietic precursors arise (Leslie et al., 2007). Jagged1 expression at low levels drives endothelial cell fate towards the hematopoietic lineage at the expense of the endothelial program (Ayllon et al., 2015, Gama-Norton et al., 2015). This effect is mediated, in part, by a Notch1-Hes1 signaling axis (Lee et al., 2013). Overall, Notch is required for the specification of HE (Burns et al., 2005).

Mutants for Notch components including *Jag1*, *Notch1*, or *RBP-JK* are embryonically lethal due to severe hematopoietic defects with increased numbers of HE cells, suggesting that *Notch1-Jag1-RBP-JK* regulates HSC emergence from HE (Kumano et al., 2003; Robert-Moreno et al., 2005; Robert-Moreno et al. 2008). A recent study suggests that Notch is actually down-regulated during HSC emergence (Butko et al., 2016). Other groups found that Notch is down-regulated in hematopoietic clusters but remains expressed in HE (Guiu et al., 2013; Lizama et al., 2015). Another recent study highlighted the precise point in which HSC emergence is no longer dependent on Notch signaling (Soulihol et al., 2016b). Enforced Notch activity abrogated maturation of preI into preII preHSCs showing that proper down-regulation is needed for preI to preII maturation. Taken together, tight regulation of Notch is necessary throughout EHT.

TGFβ signaling

Like Notch, TGFβ appears to have dual roles during EHT. In zebrafish, TGFβ signaling regulated the Notch ligand Jagged1a in endothelial cells prior to EHT (Monteiro et al., 2016). Knockdown of Tbgfβ1 and TgfβR2 reduced Runx1 and cMyb expression in the DA, showing that TGFβ is upstream of

Notch which is upstream of Runx1. After formation of HE, specification of HSPCs is dependent on TGFβ3 (Monteiro et al., 2016). In mESC differentiations, overexpression of TGFβ signaling inhibits EHT and inhibition enhances generation of hematopoietic progenitors (Vargel et al., 2016; Wang et al., 2012). This is likely due to activation of the TF Sox17, which in turn promotes an arterial fate over a hematopoietic fate through repression of *Runx1* (Lizama et al., 2015; Bos et al., 2015). From a transcription regulation point of view, activated TGFβ phosphorylates a group of SMAD proteins that activate targets that are essential for HSC emergence (Zhang et al., 2014).

Of note, TGFβ signaling has been shown to induce EMT (Lamouille et al., 2014; Van Meeteren et al., 2012). The mechanism of activation may involve activation of SMAD proteins. Another means of activation may be through induction of actin cytoskeletal reorganization via *Cdc42* and *RhoA* (Edlund et al., 2002). Similarities between EMT and EHT (Vargel et al., 2016) support the idea that TGFβ signaling may induce EHT, in part, through modulation of the actin cytoskeleton. In support of this hypothesis, sustained induction of TGFβ causes formation of stress fibers (Edlund et al., 2002). Cells undergoing shape change, especially from adherent to non-adherent, undergo actin cytoskeletal reorganization that includes a shift from stress fiber to cortical actin organization (Haynes et al., 2011). Therefore, TGFβ signaling at the onset of EHT may encourage stress fiber formation whereas its inhibition may allow actin cytoskeletal reorganization to cortical actin structures.

1.6 Intrinsic regulators and transcriptional hierarchy at onset of EHT

Transcriptional regulation of Flk1+ mesoderm

As stated above, *in vivo* proof of a hemangioblast-like cell remains elusive. However, there is proof that both endothelium and blood arise from cells expressing the VEGF receptor, Flk1 (Kabrun et al., 1997; Shalaby et al., 1995; Lugus et al., 2009). Flk1 is expressed at E7.5 in mesoderm and later in yolk sac (YS) blood islands and the embryonic vasculature (Drake and Fleming, 2000; Dumont et al., 1995; Yamaguchi et al., 1993). In the absence of Flk1, embryos die between E8.5 and E9.5 due to defects in endothelial and blood development (Shalaby et al., 1997). The Flk1+ mesoderm is believed to be one of the only markers of mesoderm slated for "hemato-endothelial fates" (Nishikawa et al., 1998b). Therefore, our current understanding places Flk1 at the apex of the developmental hematopoiesis hierarchy.

Downstream of Flk1, a candidate regulator of hemogenic mesoderm is Etv2 (a member of Ets-family transcription factors) (Sumanas et al., 2005). Without Etv2, embryos die due to blood vessel defects and concomitant induction of the cardiac mesoderm (Lee et al., 2008). Exogenous Etv2 can induce Flk1 and rescue blood cell formation when normal induction is inhibited via BMP, Wnt or Notch siginaling (Lee et al., 2008; Moignard et al., 2013). Etv2 is expressed in YS blood islands but is gradually down regulated and is no longer expressed in E10.5 endothelium (Eliades et al., 2016). Recently, Etv2 was shown to interact with Gata2 to regulate hematopoietic targets (e.g., Pu.1) (Shi et al., 2014). In $Etv2^{-/-}$ Flk1+ mesoderm several key hemato-endothelial genes are down-regulated, including *Scl* and *Fli1*. Taken together, Etv2 has a critical role in establishing the blood cell transcriptional network.

One of Etv2's direct targets is another ETS TF, Fli1 (Abedin et al., 2014). In zebrafish and xenopus, loss of Fli1 results in reduction in the number of hemangioblasts (Liu et al., 2008). Fli1 is part of the heptad of TFs regulating hematopoietic progenitors (Wilson et al., 2010a). Transcriptional regulation in hematopoietic cells occurs through a -15kb element (Beck et al., 2013) and +12kb element (Pimanda et al., 2007). Fli1 was shown to be downstream of Bmp signaling and upstream of Gata2 and Scl (Liu et al., 2008).

Transcriptional regulation at onset of EHT in hemogenic endothelium

Scl, or Tal1, is essential for the specification of HE from endothelial precursors but not for the specification of hematopoietic cells from HE. Scl is a basic helix-loop-helix TF expressed in the early stages of EHT in mesoderm. Scl-null embryos die at around E9.5 with a lack of hematopoietic cells but with other mesodermal lineages being unaffected (Robb et al., 1995; Shivdasani et al., 1995). A lack of Scl does not impact expression of Etv2 (Sumanas et al. 2008), but Scl is not expressed in $Etv2^{-/-}$ or $Flk1^{-/-}$ embryos (Kataoka et al., 2013, Wareing et al., 2012; Ema et al., 2003). However, Scl expression in $Etv2^{-/-}$ can rescue blood cell development (Kataoka et al., 2013, Wareing et al., 2012). Scl has been implicated as a direct target as Etv2 binds directly at the Scl locus (Kataoka et al., 2013). Scl has a specific temporal window during development. Using VEC-Cre and Tie2-Cre models, deletion with VEC-Cre and not Tie2-Cre ablated hematopoietic development, suggesting that any requirement for Scl is done by the time Tie2 is expressed in HE (Schlaeger et al., 2005; Chan et al, 2007; Souroullas et al., 2009). Taken together, Scl is no longer required for EHT after HE specification.

Gata2, a member the GATA family of zinc finger TFs, is essential for the generation of HSCs during EHT and for HSC survival after emergence (de Pater et al., 2013). *Gata2* is first expressed at E7.5 in endothelial cells in the paired dorsal aorta and later in the endothelial cells lining the dorsal aorta, but also in the YS (Robert-Moreno et al. 2005; Kobayashi-Osaki et al., 2005; Minegishi et al., 2003; Minegishi et al., 1999). Gata2$^{-/-}$ embryos die at E10 just prior to the appearance of HSCs (Kaimakis et al., 2016; Ng et al., 1994; Tsai et al., 1994; Robert-Moreno et al., 2005). Interestingly, *Gata2* is not required for the generation of all HSPCs but all HSCs express *Gata2* (Kaimakis et al., 2016). It is part of a heptad of transcription factors that act on the regulatory elements of hematopoietic genes (Wilson et al., 2010a). In addition, it has been specifically found to act in a triad with Fli1 and Scl1 (Pimanda et al., 2007). Interestingly, despite this cooperation with other TFs, including Runx1, conditional deletion of Gata2 shows that the developmental requirements for GATA2 and RUNX1 differ during the generation of HSPCs with both being required for generation of HSCs but only Gata2 is required thereafter for survival (de Prater et al., 2013).

1.7 Transcriptional regulation at onset of EHT – Runx1 focused

Runx1 (also known as Cbfa2 (“core binding factor alpha”) or AML) is part of the Runx transcription factor family (Runx1, Runx2, Runx3). Each is involved in regulation of key lineage-specific genes during development. Runx1 has seven alternatively spliced exons (Bee et al., 2009a). Runx1 was first identified by its ability to bind DNA in conjunction with its DNA binding partner Cbfb (“core binding factor beta”) (Speck et al., 1987). Further studies

identified a requirement for Runx1 and Cbfb in the generation of dHSCs during development (Okuda et al., 1996; Wang et al., 1996).

Runx1 is expressed in all hematopoietic sites in the embryo (North et al., 1999; North et al., 2002). Its expression marks commitment to the hematopoietic lineage (Swiers et al., 2013). Although Runx1 has a role in adult hematopoiesis, only its role during developmental hematopoiesis is discussed here. Its critical importance during developmental hematopoiesis means both its upstream and downstream regulation has been examined in great detail. Transcriptional regulation of Runx1 occurs via two alternative promoters, the distal P1 promoter and the proximal P2 promoter (Bee et al., 2009a). Promoter usage varies during developmental hematopoiesis (Bee et at al., 2009a). Runx1 expression in mature hematopoietic progenitors in both YS and AGM requires both promoters compared to primitive blood cells that use only the P2 (Bee et al., 2009a). Starting after FL colonization and continuing through to adulthood, Runx1 is predominantly regulated via the P1 (Bee et al., 2009a). DNAseI-seq and ChIP-seq from HPC7 and 416b hematopoietic progenitor lines identified putative regulatory elements capable of regulating *Runx1* expression (Schutte et al., 2016; Viestra et al., 2014). Ongoing work in the lab demonstrated stage-specific expression conferred by these different elements (Schutte et al., 2016; Swiers et al., 2013; Bee et al., 2010; Nottingham et al., 2007). The +23 enhancer marks all dHSCs, all clonogenic blood progenitors and cellular intermediates of EHT fated to become hematopoietic cells (Swiers et al., 2013; Nottingham et al., 2007).

Requirement for Runx1 during development

Specifically in the AGM, Runx1 marks HE, hematopoietic clusters and HSCs (North et al., 1999; North et al., 2002; Swiers et al. 2013) but also marks subaortic smooth muscle cells (Swiers et al., 2013a). Runx1-null embryos exhibited a complete block in definitive hematopoiesis and embryonic lethality by E12.5 (Cai et al., 2000; North et al., 1999; Okuda et al., 1996; Wang et al., 1996; Yokomizo et al., 2001). Embryos lacked CFU-C and LT-HSC potential (Okuda et al., 1996; Wang et al., 1996a) and were thought incapable of hematopoietic differentiation, generating only endothelial cells in culture (Mukouyama et al., 2000; Yokomizo et al., 2001). However, a recent study reported embryos producing preHSC type I but failing to transition to CD45+ preHSC type II stage (Liakhovitskaia et al., 2014).

Deletion of *Runx1* at specific time points using an inducible Cre-floxed transgenic mouse system showed *Runx1* is absolutely required for normal EHT (Chen et al., 2009). Moreover, deletion prior to E10.5 most greatly affected EMP formation and function (Tober et al., 2013). HSCs rely on *Runx1* expression up until E11.5, after which HSCs become *Runx1* independent (Tober et al., 2013; Chen et al., 2009).

To complicate its absolute requirement, studies have shown a gene dosage effect of *Runx1*. *Runx1* haploinsufficiency affects the timing and location of HSC emergence by speeding up the appearance of HSC in the AGM (Cai et al., 2000). *Runx1*$^{+/-}$ E10.5 AGMs were found to have HSCs but E11.5 AGMs failed to repopulate recipients (Cai et al., 2000), which is the opposite of what happens in the wild type. Overall, *Runx1* haploinsufficiency reduces definitive

hematopoietic progenitor numbers and overall cluster formation (Cai et al., 2000; Mukoouyama et al., 2000; North et al., 1999; Wang et al., 1996).

Mechanistically, Runx1 works as a transcription activator and repressor. Runx1 binds regulatory elements and up regulates expression of direct targets including but not limited to Sfpi1 (Huang et al., 2008). Mutant forms of Runx1 lacking the C-terminal transactivation domain were unable to rescue hematopoiesis in Runx1-null ES cells (Nishimura et al., 2004; Okuda et al., 2000), or p-Sp from Runx1-null embryos (Goyama et al., 2004). Moreover, ectopic expression of Runx1 in E8.5 newly specified endothelial progenitors imparts full hemogenic potential (Eliades et al., 2016). In contrast, expression of endothelial markers Flk1 and VE-cadherin are down-regulated in cells after Runx1 expression (Hirai et al., 2003). Runx1 may mediate site-specific recruitment of Polycomb repressive complex 1 as a means of silencing the endothelial program (Yu et al., 2012). On a genome-wide scale, Runx1 also facilitates chromatin unfolding and global alterations of transcription factor binding patterns (Hoogencamp et al., 2009; Lichtinger et al., 2012).

During EHT, HE cells lose their endothelial identity and become free moving hematopoietic cells. In vitro, this occurs when cells change from adherent, flat cells with front-rear polarity into round cells with apical-basal polarity consistent with hematopoietic progenitors (Lancrin et al., 2012). Runx1 facilitates the change in cell shape, at least in part, by activating its target genes Gfi1 and Gfi1b that, even in the absence of Runx1, trigger the down-regulation of endothelial markers and facilitate the formation of round cells (Lancrin et al. 2012). In support of this work, Runx1 has been shown to regulate a cell adhesion program (Lie-A-Ling et al., 2014).

TFs downstream of Runx1

Given that Runx1 is essential for EHT, it is critical to understand its downstream effectors. Two of those effectors, Gfi1 and Gfi1b, are direct targets of Runx1 (Lancrin et al., 2012). Gfi1 marks HE that generates HSCs (Thambyrajah et al., 2015). Gfi1 and Gfi1b mark CD41+ hematopoietic cells in the YS and AGM (Lancrin et al., 2012). In the absence of Runx1, Gfi1 and Gfi1b trigger down-regulation of endothelial genes and the formation of round cells. Gfi1 (Gfi1 and Gfi1b) proteins are required for HSCs and hematopoietic progenitor formation in the AGM (Thambyrajah et al., 2015). $Gfi1^{-/-}$::$Gfi1b^{-/-}$ embryos maintain endothelial gene expression despite the presence of Runx1 and fail to detach from the endothelial layer (Lancin et al., 2012). Transcriptional regulation of Gfi1b is controlled by Runx1 and Gata2 binding to the +16kb and +17kb elements (Moignard et al., 2013; Wilson et al. 2010a; Wilson et al., 2010b). Transcription regulation of Gfi1 is controlled mainly by the -33kb element, which is bound by Scl, Gata2, Meis1, Runx1, Pu.1, and Erg (Wilson et al. 2010a).

An additional downstream target of Runx1 is Sfpi1 (Pu.1), which is an ETS transcription factor that is expressed at the onset of EHT. In Runx1 deficient ES-derived hemangioblasts, cells expressed Scl1, Flk1 and Fli1 but did not express Sfpi1 (Hoogencamp et al., 2009). Runx1 regulates Sfpi1 through 3 RUNX motifs in an enhancer 14kb upstream of the TSS (Huang et al., 2008). Sfpi1 is first expressed in HE and increases expression as cells become more hematopoietic (Swiers et al., 2013). In its absence, embryos die at birth from a defect in T and B cell production, in addition to defects in myeloid cells (Scott

et al., 1994). Of note, Sfpi1 positively regulates hematopoietic commitment of HE cells (Hoogenkamp et al., 2007).

Although there are many other targets to discuss, the last gene discussed her is Myb whose expression is restricted to hematopoietic cells during EHT (Wilkinson et al., 2014; Mukouyama et al., 1999; Thompson et al., 1998). Moreover, it is required for the generation of mature hematopoietic progenitors whereby in its absence embryos experience severe anemia after E13.5 (Mucenski et al., 1991; Mukouyama et al., 1999). Previous identification as a target of Runx1 requires additional work to show decisive regulation by Runx1.

1.8 Studying transcriptional networks (Gene regulatory interactions)

Gene regulatory interactions (GRIs) or gene regulatory networks (GRNs) are the set of regulatory genes and their interactions that determine transcriptional networks and ultimately cell fate (Davidson and Peter, 2015). This includes regulatory genes (i.e., TFs), genes involved in intercellular signaling, and effector genes that drive downstream functions. The ability to link genes to upstream transcriptional regulators allows identification of GRIs. GRIs organize the spatial expression of transcriptional regulatory genes that drive developmental fate.

Understanding GRIs requires understanding the underlying components. For example, to understand how the GRI controlled by Runx1 drives cell fate decisions we need to understand four components. First, this process begins with an understanding of the expression of the regulatory apparatus itself—how *Runx1* is regulated (Davidson and Peter, 2015). Transcriptional regulation of RUNX1 has been (Schutte et al., 2016; Swiers et al., 2013; Bee

et al., 2010; Nottingham et al., 2007) and continues to be studied by our lab. DNAseI-seq and ChIP-seq from HPC7 and 416b hematopoietic progenitor lines (Viestra et al., 2014; Schutte et al., 2016) identified greater than 13 putative Runx1 cis-elements. Ongoing work in the lab demonstrated the stage-specific expression conferred by several of these elements. The expression profiles generated and analyzed in Chapter 3 reiterate *Runx1*'s expression pattern during EHT. As a whole these pieces provide a detailed understanding of how transcriptional regulation of Runx1 leads to its precise expression pattern during EHT, confirming an understanding of the regulatory apparatus itself.

Second, we need to understand the DNA sequence that processes regulatory inputs (*i.e.,* cellular context in which *Runx1* is expressed) (Davidson and Peter, 2015). The RUNX family of TFs recognize the TGTGGT consensus motif (Bravo et al., 2001). TF binding is ameliorated by altering the motif with TG(T → G)(G → C)GT (pyrimidine → purine) and (purine → pyrimidine), giving a TGCAGT sequence. With this sequence, amino acids R174 and R177 of Runx1 protein facilitate the Runx1-DNA interaction needed for Runx1 binding (Tahirvo et al., 2001; Bravo et al., 2001). Moreover, without R174 or R177, the Runx1-DNA interaction is compromised (Tahirvo et al., 2001). Binding to DNA is only a part of the processing. What happens after DNA binding is what dictates the biologically relevant outcome. For example, Runx1-binding has been shown to recruit polycomb complexes (Yu et al., 2010) or transcriptional elongation complexes (Jiang et al., 2005). Each interaction has the potential to affect expression of local genes. Exactly how

Runx1 binding is processed by the DNA sequence remains poorly understood.

Third, we need to understand the different contexts in which the regulatory system operates (Davidson and Peter, 2015). The chromatin and TF binding techniques needed to study Runx1 in the developing embryo require starting cell numbers that far exceed what is feasible from embryonic material. However, systematic dissection of Runx1 binding and subsequent effect on chromatin dynamics and gene expression has been extensively studied in a mESC differentiation model system (Goode et al., 2016; Lichtinger et al., 2012). This gives us one context in which the Runx1 regulatory system operates. A second context is during adult hematopoiesis. Although there is not an absolute requirement for Runx1 in adult hematopoiesis, it does play an active role in regulating a transcriptional program in addition to regulating effector proteins (Hamey et al., 2017; Schutte et al., 2016; Cai et al., 2014; Wilson et al., 2010), among other biological processes. This data has enabled the generation of GRIs governing hematopoietic progenitors. Using expression and chromatin profiles from in vivo EHT will facilitate the generation of gene regulatory interaction critical to dHSC emergence.

Finally, we need validation of GRIs using experimentally acquired knowledge (Davidson and Peter, 2015). In the case where genetic perturbation is absent, computational gene network inference methods based on expression data can predict functional relationships (Hamey et al., 2017). Unfortunately, data collected on whole populations limits the power of the inference. In contrast, data from single cells can be used to discover simple regulatory relationships (Moignard et al., 2013) or create complex networks (Moignard et al., 2015). In

an ideal setting, experimental knowledge would combine genetic perturbation with single cell sequencing. One such approach that has recently been adopted is Perturb-seq, which allows perturbations to be evaluated for impact on gene expression in single cells (Dixit et al., 2016). Using CRISPR pooled screening combined with RNA-seq, researchers were able to identify specific perturbations performed in single cells while simultaneously measuring the effect of the perturbation on gene expression in those same single cells. This facilitated identification of gene targets affected by the individual gene modifications. In an elegant study, Schutte and colleagues used both genetic and computation approaches to validate the GRI governing hematopoietic progenitors (Schutte et al., 2016). Using a similar combination of genetic and computational approaches will enable generation of validated GRIs governing developmental hematopoiesis.

1.7 Project aims

Overall, the aim of my project was to study the molecular mechanisms underlying the endothelial-to-hematopoietic transition by combining expression and chromatin profiles of the cellular intermediates. To facilitate the identification of cellular intermediates of EHT, our lab previously generated a Runx1 +23 enhancer-reporter mouse model (23GFP) (Bee et al. 2010, Nottingham et al. 2007). 23GFP expression preceded *Runx1* expression in E8.5 but otherwise recapitulated endogenous *Runx1* expression (Swiers et al. 2013). This allowed prospective isolation of hematopoietic-fated cells prior to the onset of Runx1 (Swiers et al. 2013). Using this model, our lab can now isolate all of the cellular intermediates undergoing EHT.

My project focused on profiling these cellular intermediates with the aim to identify and characterize the GRIs underlying the cell fate transitions within EHT. We reasoned that integration of expression (RNA-seq) and chromatin profiles (ATAC-seq) would facilitate identification of these GRIs. Specifically, we focused on TFs where the timing and appearance of the motif coincided with expression of the TF and identified these as putative regulators of accessibility and subsequent changes in gene expression. To identify differentially expressed TFs, I analyzed RNA-seq data that was collected prior to my time in the lab (**Chapter 3**). In parallel, pathway enrichment analysis of differentially expressed genes revealed novel pathways important for regulation of EHT. Turning to chromatin profiles, I generated and analyzed ATAC-seq data from the same phenotypic populations as the RNA-seq (**Chapter 4**). Identification of enriched motifs in the differentially accessible elements identified putative regulators of accessibility. Next, for integration of the expression and chromatin data sets, I focused on two key transitions: (1) the cHE to sHE transition and (2) the sHE to proHSC transition (**Chapter 5**). Here, I integrated expression and chromatin profiles to generate GRI maps with the aim of identifying novel regulators for both transitions. Identification of known regulatory interactions validated our approach. Discovery of new predicted regulatory interactions, including Runx1 and the *Ikzf1*-28 element, highlighted the power of our data to provide new insights on EHT. Finally, in **Chapter 6** characterization of the Runx1-*Ikzf1*-28 interaction revealed Runx1-dependent activation of *Ikzf1* during EHT. Further investigation showed the Runx1-Ikzf1 regulatory interaction was necessary for proper regulation of Notch signaling. Overall, our data represents the first genome-wide profiling of

cells at the onset of EHT. Identification of validated regulators as well as the identification of a new regulatory interaction highlights the value of our data to the field of developmental hematopoiesis.

Chapter 2: Materials & Methods

2.1 Cell culture

mESC Culture

Mouse embryonic stem cells (mESC) (E14, strain 129/Ola) were cultured on plasticware treated with 0.1% gelatin (Sigma) or containing a layer of mitotically inactivated mouse embryonic fibroblasts (MEF) plated 24h prior to culture. Growth media consisted of (G-MEM BHK-21) (Thermo Scientific) supplemented with 10% fetal calf serum (FCS) (Gibco), 100mM NEAA (Gibco), 2mM L-glutamine (Gibco), 100mM Sodium Pyruvate (Gibco), 100um Beta-mercaptoethanol (Gibco), and leukemia inhibitor factor (prepared in house and used a concentration of ~ 1000U/ml). Cells were grown in a 37°C humidified incubator with 5% CO_2. Cells were passaged using 0.05% Trypsin (Gibco) every ~2-3 days or when cells reached 80% confluence. Each passage seeded ~10,000 cells / cm^2.

mESC Differentiation

The differentiation protocol was based on Pearson et al., 2015 with minor modifications. Briefly, one day prior to start of differentiation, mESC growth media was replaced with growth media minus LIF. On day 0, to generate embryoid bodies (EBs), mESC were harvested, counted and 750,000 cells were re-suspended in 20ml of mESC differentiation medium consisting of StemPro-34 (Thermo Scientific) supplemented with StemPro-Nutrient Supplement, 8 mg/ml Ascorbic Acid (Sigma), 100U/ml Transferrin (Roche), 2mm L-glutamine (Gibco), and 5 ng/ml BMP-4 (R&D Systems) and cultured in non-tissue culture treated 10cm dishes (Corning). On day 3, mESC differentiation medium was supplemented with 5 ng/ml bFGF (R&D Systems) and 5 ng/ml Activin A (Peprotech). On day 4, EBs were harvested and

disassociated in 0.05% Trypsin (Gibco) before Flk1+ cells were isolated by flow cytometry on an Becton Dickinson (BD) FACS Aria Fusion. 600,000 Flk1+ cells were sorted into 100% FCS (Gibco), centrifuged, and plated in fresh differentiation medium containing 5 ng/ml BMP-4 (R&D systems), 5 ng/ml Activin A (Peprotech), 10 ng/ml SCF (Peprotech), and 25 ng/ml VEGF (Peprotech) into a well of a 6-well plate coated with 0.1% gelatin. On day 5, differentiation medium was replaced with fresh medium containing 10 ng/ml SCF (Peprotech) and 25 ng/ml VEGF (Peprotech).

mESC Targeting

One day prior to transfection, 150,000 mESC were plated into a well of a 6-well plated coated with 0.1% gelatin. 2h prior to transfection, mESC growth medium was replaced with growth medium containing 10μm NHEJ inhibitor L755,507 (Sigma SML1362). Cells were transfected using 7.5ul TransIT-LT1 (MirusBio) and 2.5ug (ratio of 3:1 transfection reagent to plasmid DNA). For generating knockouts, guide RNA (gRNA) were cloned into constructs containing wildtype (Wt) Cas9 (px330 and px458, both from Addgene) (See Section 2.5 for more detailed information). For generating reporters or endogenous tags, gRNA were cloned into constructs either containing Wt Cas9 or Cas9 knickase (px335 and px459, both from Addgene) (See Section 2.5 for more detailed information). For the 2.5μg of plasmid DNA, 0.5μg CRISPR/Cas9 construct and 2.0μg of pDTA-targeting vectors were combined and then added to 225ul Opti-MEM (Gibco). 7.5ul of TransIT-LT1 was then added to the Opti-MEM-DNA mixture and incubated for 15 min. After transfection, cells were grown in a 37°C humidified incubator with 5% CO_2 for

24h after which the media was replaced with fresh growth medium without NHEJ inhibitor.

After an additional 24h, cells were harvested with 0.05 trypsin and pipetted up and down to make a single cell suspension. The single cell suspension was diluted 1:50, 1:75, 1:100, 1:150, 1:200, and 1:250 in growth media supplemented with either G418 (pX330) (Sigma) or Puromycin (px459) (Gibco). After 4-7 days, colonies were picked into 96-well round bottom plates, trypsinized, and placed into 96-well 0.1% gelatin-coated tissue culture plates containing fresh growth medium. After cells reached confluency, cells were harvested and 75% of cells were used for genomic DNA (gDNA) with the remaining 25% expanded in new 0.1% gelatin-coated 96-well plates in fresh growth medium. gDNA was harvested in cell lysis buffer consisting of 100mM Tris-HCl (Sigma) and 0.2% Triton X-100 (Sigma) and incubated at 50°C for 1h followed by 95°C for 15min. To detect gene editing or targeting, polymerase chain reaction (PCR) primers were designed to flank the insertion site. PCR was performed using HotStarTaq (Qiagen) for 34-40 cycles with annealing temperatures primer dependent. Positive clones were trypsinized from the 96-well plate, expanded, and frozen down. For gene editing and targeting, PCR products were sequenced to confirm deletions. For knock-ins, an additional PCR to confirm positive clones was done using long-range PCR with primers outside the homology arms. Karytopyes were confirmed using in house methods. Briefly, cell pellets were treated with 0.075 M KCl (Sigma), Carnoy's fixative, Giemsa staining, and plated onto glass slides for chromosome counting.

2.2 Mice

Animal husbandry

All animal procedures were in compliance with Home Office regulations. Mice were housed in IVC cages with free access to food and water. Mouse lines were kept on a C57BL/6xCBA mixed background. The Sanger Colony Management System was used to track all mice including mouse pups, which were weaned after 3-4 weeks.

Ear notching and genotyping

2-3 week-old pups were ear notched for genotyping. Ear notches were disassociated with in 100µl PBND lysis buffer (10mM Tris, 50mM KCl, 2.5mM MgCl2, 0.45% Tween, 0.45% NP-40) with proteinase K (200µg/ml) (Invitrogen). Notches were incubated at 55°C for 3h (or overnight) and 95°C for 5min to inactive proteinase K. gDNA was then used as a template for genotyping PCRs using sequence specific primers and HotStarTaq (Qiagen).

Gene	Sequence
GFP1 (FP)	GACGTGAACGGCCACAAGTTCA
GFP2 (RP)	GTGGCGGATCTTGAAGTTCACC
Myo1 (FP)	TTACGTCCATCGTGGACAGC
Myo2 (RP)	TGGGCTGGGTGTTAGTCTTA
Cas9_FP	TTGTTATCAGTAAGGGAGCTGCAGTGGAGTAG
Cas9_Wt_RP	AGGACAACGCCCACACACCAGGTTAG
Cas9_Mut_RP	CAGTTTACCGTAAATACTCCACCCATTGACGTCA

Table 2.1 Primer sequences for genotyping transgenic mice.

Timed matings, embryo collection and dissections

For time pregnancies, matings were set up overnight. Detection of a vaginal plug the next morning was designated as day E0.5. Embryos were harvested at the desired developmental time points: E8.5, E9.5 and E10.5. On the desired day of embryo harvest, pregnant females were sacrificed via Schedule 1 methods, and embryos collected in dissection buffer (DB) made up of PBS +Mg+Cl (Gibco), 10% FCS, 50U/ml penicillin and 50ug/ml streptomycin (Gibco). Dissections were performed in DB with all subsequent steps performed in PBS-Mg-Cl, 10% FCS, 50U/ml penicillin and 50ug/ml streptomycin (PBS-CMF). Embryos from E8.5 to E10.5 were staged by counting somite pairs. Embryos were kept at room temperature during the dissection and then on ice prior to generation of single cell suspensions.

Dissections and generation of cell suspensions

The PAS/AGM and VU vessels were dissected from E8.5, E9.5 and E10.5 embryos in DB using watchmaker forceps and 25 G needles. After pooling, tissues were incubated for 15min at 37°C in PBS-CMF and 0.1% (w/v) type I collagenase (Sigma) and disassociated by pipetting. Single cell suspensions were washed and counted on a Neubauer haemocytometer using trypan blue (Sigma) to assess viability. For flow cytometry, cells were stained with antibodies from Table 2.2 for 10min while on ice. Cells were washed with PBS-CMF and re-suspended in PBS-CMF containing Hoechst 33258 (1:1000 dilution) (Thermo Fisher) before being passed through 0.35uM filters (BD).

Antibody	Conjugation	Supplier	Cat No.
B220	APC	BD Pharmigen	553086
CD11b	APC-Cy7	BD Pharmigen	557657
CD19	PE	BD Pharmigen	557399
CD41	PE	BD Pharmigen	558040
CD45	APC-eflour780	eBioscience	47-0451-82
Flk1	Biotin	eBioscience	13-5821-81
SAV	PE-Cy7	Invitrogen	SA1012
Ter119	PE-Cy7	BD Pharmigen	557853
VEC	APC	eBioscience	17-1441-82

Table 2.2. Antibodies used for Flow Cytometry.

Flow cytometry and cell sorting

Cells were isolated by flow cytometry using a 100µm nozzle at 10 p.s.i. and ~1,000 – 2,000 events per second (Aria III and Fusion 2, Becton Dickinson). Acquired data was analyzed using FACS Diva (BD) and FlowJo (FlowJo, LLC). Instruments were set up using unstained and single stain controls. Sort gates were drawn according to fluorescent minus one (FMO) controls. Gating strategy was based on the following sequence: Hoechst negative (live cell gate), Ter119 negative (non-erythroid gate), VE-Cadherin (VE-Cad) positive (endothelium), CD41 versus CD45 (preI and preII gates) and CD41-CD45- onto 23GFP (non-hemogenic and hemogenic endothelium gates).

2.3 RNA-sequencing

RNA-sequencing library preparation and sequencing

RNA-sequencing libraries were generated from E8.5 – E10.5 mouse embryos using SMARTer (Clontech 634935). 100 cells were sorted in 3.16ul lysis mix (Clontech) and lysed for 3 min at 72°C. SMARTer reverse transcriptase reactions were performed and a 12 cycle PCR was used for preamplification. After preamplification, cDNAs were sheared using a Covaris, and Illumina adaptors were ligated using NEBNext DNA library Prep Master Mix Set for Illumina (Cat E6040), using custom workflow. All libraries were paired end 50bp sequenced on Illumina Hiseq2000 according to manufacturer's instructions. Cells for RNA-seq were isolated by Dr. Gemma Swiers, and library preparation and sequencing was performed in the WTCHG (Oxford).

RNA-sequencing analysis

Following QC analysis with the fastQC package (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), reads were aligned using Tophat using default parameters (Trapnell et al., 2009) against the mouse genome assembly mm9 (UCSC transcripts). Non-uniquely mapped reads and reads that were identified as PCR duplicates using Samtools (Li et al., 2009) were discarded. Gene expression levels were quantified as read counts using the featureCounts function in specified exons only (Liao et al., 2014) from the Subread package (Liao et al., 2013) with default parameters. The read counts were used for the identification of global differential gene expression between specified populations using the edgeR package (Robinson et al., 2010). RPKM values were also generated using the edgeR package. Genes were considered differentially expressed between populations if they had an adjusted *p*-value (FDR) of less than 0.05. The pathway and synexpression analyses were done using in-house modified functions based on the attract package (Mar et al., 2011).

2.4 Chromatin assays

ATAC-seq library preparation and sequencing

ATAC-seq libraries were generated from E8.5 – E10.5 mouse embryos as previously described with minor modifications (Buenestro et al., 2015). Briefly, 1500 cells were sorted into 100% FCS (Gibco), pelleted at 500g for 10min, and lysed in lysis buffer (10mM Tris-HCl (pH 7.5), 10mM NaCl, 3 mM $MgCl_2$, 10% IGEPAL CA-630, 0.1% TritonX-100) for 10min on ice. After lysis, nuclei were spun at 500g for 10min to remove lysis buffer. Nuclei were then gently re-suspended and incubated with 10ul of the tagmentation mix consisting of

0.35ul Tn5 transposome and 5ul of 2X tagmentation buffer and 4.65ul Nuclease free water at 37°C for 30min (Illumina Nextera FC-121-1030). After tagmentation, 1.1ul of 500mM EDTA was added directly into the reaction to quench. 10ul of 50mM $MgCl_2$ was added to replenish $MgCl_2$ after quenching. PCR was performed to amplify the library for 13 cycles using the following PCR conditions: 72°C for 5 min; 98°C for 30 s; and thermocycling at 98° for 10 sec, 63°C for 30 sec and 72°C for 1 min; followed by 72°C for 5 min. After the PCR reaction, libraries were purified using MinElute purification (Qiagen). ATAC-seq library profiles were visualized on a TapeStation (Agilent Technolgoies) using D1000 tape. Libraries were quantified using an illumina-specific library quantification kit (KAPA Biosystems). Samples were sequenced on Illumina NextSeq using a 75-cycle paired end v2 reagent kit (Illumina 1002-141).

ATAC-seq data processing

Fastq files of paired end reads were processed using an in-house customized DNaseI and ChIP-seq pipeline (Hay et al., 2016). Briefly, samples were aligned to mm9 genome using Bowtie using default parameters (version 1.0.0). Over amplified products were excluded by collapsing reads mapping to the exact same genomic position into one single representative read. In order to remove ploidy regions (which are genomic regions that bias high-throughput sequencing experiments because of large copy number differences and normalization issues between the real genome and the genome build), a list of ploidy regions to exclude was constructed after visual inspection in UCSC genome browser of each chromosome. The pipeline generated genome-wide tracks using a sam2bigwig.pl perl script that

produced a read density for each track over a set window size and did so across the genome. Using BigWig, Bowtie alignments were thus converted to genome-wide density tracks with the output displayed in the UCSC Genome Browser (Kent et al., 2002). MACS2 was used to identify peaks in the ATAC-seq data using default parameters (q-value 0.05). Differential accessibility analysis across all populations and between pairwise comparisons was performed using the R package DiffBind using EdgeR with default parameters (Stark et al., 2012). Bed files for each population were generated using Bedtools –intersect were biological replicate peak calls were intersected to generate a .bed file with representative peaks for a given population. A peak (elements) was included if it was in 2/3 or 2/4 biological replicates. For representation in UCSC genome browser, bam files for each biological replicate were merged using Bamtools merge tool, and the same numbers of reads were counted for each population's bam file using a custom resampled.pl perl script (WIMM CBRG). For data representation in differential accessibility heatmaps, conditional quantile normalization was used to normalize for library size and and GC content (Hansen et al., 2012).

Categorization of elements into subsets

For each pairwise comparison, elements were sorted into "newly closed", "decreasing", "newly open", and "increasing" subcategories. -/+ logFC elements were intersected with the list of all elements for one of the populations in the comparison. For example, for the cHE to sHE comparison, +logFC elements were intersected with all elements identified in the cHE population. If a +logFC element was also present in the cHE, then it was classified as "increasing". If a +logFC element was not present in the cHE,

then it was classified as “newly open.” Negative logFC elements decreasing in accessibility from cHE to sHE were intersected with all elements in the sHE population. If a –logFC element was also present in sHE, then it was classified as “decreasing”. If a –logFC element was not present in sHE, then it was classified as “newly closed”.

Motif enrichment, GREAT and LOLA

Motif enrichment analysis was performed using the HOMER findMotifsGenome.pl using the 250bp sequences centered on the chromatin accessible regions compared to randomly generated sequences of the same length as background. Both motif enrichment functionalities (1) Discovering de Novo motifs and (2) Screening for Enrichment of Known Motifs were used. Known Motif analysis output was summarized into Ranked Motif enrichment plots using R. De novo motif analysis was summarized using GraphPad. Motif matrix images from either Known Motif or De Novo Motif analysis were displayed as supplied by the respective output files. Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) was used to calculate functional enrichment for genes near ATAC-seq elements, using default settings. The top 20 pathways from the “mSigDB pathway” were exported and processed in Graphpad. Region set-enrichment analysis was performed on the differentially accessible ATAC-seq elements using LOLA together with its core database of ChIP-seq binding peaks from CODEX (Sanchez-Castillo et al. 2015, Sheffield et al. 2016).

Intersecting RNA-seq and ATAC-seq data

To assign elements to genes, HOMER annotatepeaks.pl was used with default parameters. Multiple elements mapped to one gene were given unique

identifiers (e.g., Runx1_1 (P1 promoter), Runx1_2 (+3 element), Runx_3 (+23 element), etc.). ATAC-seq gene lists were intersected with DEG lists to identify genes in both lists. Each element-gene match was associated with two logFC: (1) logFC for the element, reporting changes in differential accessibility, and (2) logFC for expression, reporting changes in expression for the same comparison. Plots for representing intersecting lists were generated in R. Where more than one element mapped to a gene, the elements were given the same logFC value for the expression and thus appear stacked on top of each other in the plots.

Construction of gene regulatory interaction maps

Gene regulatory interaction matrices were constructed and run in Cytoscape (Version 3.5.1) (Shannon et al., 2003). Gene matrices consisted of rows of elements, with multiple elements mapping to one gene given unique identifiers, and columns containing (expression logFC, accessibility logFC, motifs (#) or ChIP-peak (#)). Network generation in Cytoscape was done using circular layouts. Applying “default filters Edge: shared name” identified all nodes with “shared name” as a source (*e.g.*, Runx1 targets). Applying “default filters Edge: shared name – doesn’t contain” identified nodes that regulated selected node (*e.g.*, genes that targeted Runx1).

2.5 Molecular cloning

Generation of targeting vectors

Construct targeting Runx1 stop construct was synthesized using GeneArt (Invitrogen). Briefly, a 3.7kb 5’ homology arm and a 5.5 3’ homology arm were amplified from E14 gDNA using PCR. PCR primers included RE overhangs to allowing cloning into pUC19 vector (Invitrogen). A GeneARt synthetic

construct consisting of the Runx1exon6-3XFlag-ERt2-STOP was cloned into the pUC19 vector containing the homology arms. Constructs were confirmed by Sanger sequencing. Constructs were used for transfection into undifferentiated E14s with a CRISPR/Cas9 vector expressing gRNA that cut immediately preceding the RUNX1 stop codon.

gRNA design and cloning

Guide RNA (gRNA) sequences (**Table 2.3**) were designed using the MIT CRISPR design tool (http://crispr.mit.edu/) (Hsu et al. 2013). All gRNA were either designed to start with a G or were altered to include a G at the start of the gRNA sequence. All oligonucleotides (oligos) were ordered from Integrated DNA Technologies (IDT) with no modifications and standard de-salting. For cloning, gRNAs were cloned into plasmids using methods described previously (Hsu et al., 2013). Briefly, double stranded gRNA were constructed by designing oligos consisting of the gRNA sequence (forward) and reverse complement sequences (reverse). For cloning into BbsI restriction sites, forward gRNAs were designed with a CAAA sequence added to 5' end, and reverse gRNAs were designed with an AAAC sequence added to 5' end. To anneal oligos, 1.0ul of 100uM forward oligo and 1.0ul of 100uM reverse oligo were added to 8.0ul of ddH20 and were incubated in a thermal cycler for 95°C for 4 min followed by a ramp down to 25°C at 1°C/min. CRISPR expression vectors (pX458 (Addgene plasmid 48138) and pX459 (Addgene plasmid 48139) were digested with BbsI (NEB). The annealed oligo duplex was ligated into the appropriate vector (pX458 or pX459) using T4 DNA ligase (NEB). Positive clones were screened using PCR of bacterial

colonies. Positive clones were confirmed by Sanger sequencing (Source Biosciences), Midi prepped (Qiagen) and used for transfections.

gRNA	Protospacer (PAM)	Forward oligo	Reverse oligo
Runx1_exon4_gRNA_1	CGGATTTGTAAAGACGGTGATGG	caccgCGGATTTGTAAAGACGGTGA	aaacTCACCGTCTTTACAAATCCGc
Runx1_exon4_gRNA_2	CCGAAGTAAGTGAGCCCCCTTGG	caccgCCGAAGTAAGTGAGCCCCCT	aaacAGGGGGCTCACTTACTTCGGc
Runx1_Stop_gRNA	CGATGGCGCTCAGCTCAGTAGGG	caccgCGATGGCGCTCAGCTCAGTA	aaacCCCTACTGAGCTGAGCGCCATCGc
Ikzf1_ex3_gRNA_1	CAGAACTCCAAGAGTGATCGGGG	caccgCAGAACTCCAAGAGTGATCG	aaacCGATCACTCTTGGAGTTCTGc
Ikzf1_ex3_gRNA_2	GAGGACCTGTCCACTACCTCTGG	caccGAGGACCTGTCCACTACCTC	aaacGAGGTAGTGGACAGGTCCTC
Ikzf1_-28_gRNA_5'_1	GTTGTGGATGATCCAGACGACGG	caccGTTGTGGATGATCCAGACGA	aaacTCGTCTGGATCATCCACAAC
Ikzf1_-28_gRNA_5'_2	CACACACACACACCGTCGTCTGG	caccgCACACACACACACCGTCGTC	aaacGACGACGGTGTGTGTGTGTGc
Ikzf1_-28_gRNA_3'_1	TAGCAGAGTTTGATCATTAGAGG	caccgTAGCAGAGTTTGATCATTAG	aaacCTAATGATCAAACTCTGCTAc
Ikzf1_-28_gRNA_3'_2	GGTTATTTCCTGCACACTATAGG	caccGGTTATTTCCTGCACACTAT	aaacATAGTGTGCAGGAAATAACC

Table 2.3. Information on gRNAs used for targeting.

In vitro gene knockouts – vector and edit detection

To create gene knockouts, gRNAs were either designed to target early coding sequences within the gene of interest or targeted to critical sequences (e.g. key amino acid residues or non-coding elements). Multiple gRNAs were designed for each loci. Depending on the application, either single gRNA or two or more gRNA were annealed and cloned into the pSpCas9(BB)-2A-Puro (pX459). To clone multiple gRNAs into a single vector, we used Gibson Assembly to clone up to four gRNAs in tandem. Each gRNA had its own U6 promoter along with the requisite tracrRNA. Briefly, single gRNAs were cloned into pX459 constructs. Sequence-verified clones were then subjected to PCR whereby a U6-gRNA-tracrRNA fragment was amplified including 20-25bp overhangs added on to each 5' and 3' ends (**Table 2.4**). Each PCR product was then PCR purified and added into a Gibson Assembly reaction (NEB) and incubated at 50°C for 1h (Vidigal et al., 2015). The 20-25bp overhang contained homology either to the pX459-BbsI cut vector or to the other PCR fragments. Using this homology, the Gibson Assembly reaction produced constructs with multiple gRNAs in tandem. 2ul of Gibson Assembly was then transformed into chemical competent bacteria using methods described herein.

For pX459-1x gRNA constructs, indels were detected by Surveyor assays (Qiu et al., 2004). In brief, two sets of PCR primers were designed flanking the cut site (**Table 2.5**). The first PCR amplified ~1000bp around the cut site. The second PCR amplified ~600bp surrounding the cut site. To detect heteroduplexes created by indels, second PCR products were subjected to thermal ramping (95°C to 85°C at ramp rate of -2C/sec then 85°C to 25°C at

ramp rate of 0.3°C/sec) followed by T7 Endnuclease I (NEB) digest at 37°C for 60 min. Using 2% percentage agarose gel, indels were detected by the presence of additional bands below the 600bp unedited fragment. For pX459-2x gRNA constructs, gRNAs were designed to remove fragments between 75-150bp by flanking the site of interest. Edits were detected using PCR of gDNA with the 75-150bp being resolved on a 0.8% agarose gel.

Name	Sequence (overhang-gene specific)	Over-hang	Gene specific
FP1	AATTCTGCAGACAAATGGCTCTAGAGGGCCTATTTCCCATG	Vector	gRNA 1
RP1	ATACCTATGCCGGATTGCCGAAAAAAGCACCGACTCGG	gRNA 2	gRNA 1
FP2	CGGCAATCCGGCATAGGTATGAGGGCCTATTTCCCATG	gRNA 1	gRNA 2
RP2	TCGCCACTCGATCCTGCCCGAAAAAAGCACCGACTCGG	gRNA 3	gRNA 2
FP3	CGGGCAGGATCGAGTGGCGAGAGGGCCTATTTCCCATG	gRNA 2	gRNA 3
RP3	GCATATGTGCGTACCTAGCATAAAAAAGCACCGACTCGG	gRNA 4	gRNA 3
FP4	ATGCTAGGTACGCACATATGCGAGGGCCTATTTCCCATG	gRNA 3	gRNA 4
RP4	GTAAGTTATGTAACGGGTACCAAAAAAGCACCGACTCGG	Vector	gRNA 4

Table 2.4: Primers for Gibson Assembly of gRNAs in tandem.

	Name	Sequence
Runx1 3' targeting geno	Runx1_Outside HA_FP	CAAGAGGGATATGGGCAGA
	Runx1_Stop_RP	TTCCCTCCGGGATTCTTCCT
Runx1 Exon 4	Runx1_ex4_Surv_FP	CTGTGTGCATTACATGGTGAAC
	Runx1_ex4_Surv_RP	GCACACTGAACATTGCTGTATATG
Ikzf1 exon 3	Ikzf1_Surv_FP1	CCACCATTAGCCCAGCAGTA
	Ikzf1_Surv_RP1	CACTGGCCGAACTCCTTGAT
	Ikzf1_Surv_FP2	AGGATGTCTAAAAGCCAGTTTGC
	Ikzf1_Surv_RP2	GCAGATCACCCATATCCCAACT
Ikzf1 -28	Ikzf1-28_FP	CCACCATTAGCCCAGCAGTA
	Ikzf1-28_RP	CACTGGCCGAACTCCTTGAT

Table 2.5: Primers for surveyor assays and genotyping of gene targeting (knockins and knockouts).

Lentivirus generation

gRNA used for *in vitro* gene editing were cloned into a pKLV-U6gRNA5(gGFP)-PGK-BFP (Addgene plasmid 67991) lentiviral backbone. gRNAs were altered to include an additional GT on the 3' end to ensure proper ligation into the pKLV vector. pKLV-2xgRNA were constructed using Gibson Assembly as described above. pKLV-gRNA or pKLV-2xgRNA plasmids were then used to make lentiviruses. Briefly, 293FT cells (Thermo Fisher) were plated at 2.5 x10^6 per T75 flask 1 day prior to transfection. 293FT were transfected with lentivrial packaging plasmids (pMD2.G and psPax2) and the pKLV vectors using PEI transfection reagent (Sigma). 48h post transfection, supernatants were harvested, filtered through a 0.45um filter and centrifuged at 20,000rpm at 4°C for 2h. Viral pellets were re-suspended in 100ul of IMDM and either used fresh or immediately frozen and stored at -80°C.

2.6 Ex vivo gene knockouts, lenti infections, and re-aggregation cultures

Cas9 mice harboring a constitutively active Cas9 protein targeted to the Rosa26 locus (Platt et al. 2014) were infected with pKLV lentiviruses expressing gRNA targeting genes of interest. Caudal portions of E8.5 or E9.5 embryos were disassociated using Collagenase (Gibco) treatment at 37°C for 15min. Cells were counted using a haemocytometer to determine number of cells per embryo equivalent (EE). Either 0.5, 1 or 2 EE were centrifuged at 1,000rpm for 5min at 4°C and then resuspended in CMF-PBS. Concentrated viral supernatants and 8 ng/ml of Hexadimethrine bromide (polybrene) (EMB Millipore) were added to the cell suspensions, which were then incubated in a

37°C water bath for 30min. The cell-virus mix was then added to a yellow-tip, sealed with parafilm that contained 100,000 OP9 cells. Tips were centrifuged at 400g for 10 min at room temperature. 0.65-um Durapore membranes were floated onto of IMDM (Invitrogen) supplemented with L-glutamine, Penicillin/streptomycin (50 IU/ml and 50 ug/ml) and cytokines (100ng/ml SCF (Peprotech), 100ng/ml IL-3 (Peprotech), and 100ng/ml Flt3L (Peprotech)). Cells from the yellow tips were transferred as re-aggregates onto the Durapore filters and were grown at the liquid-gas interface. Re-aggregates were harvested at different time points depending on the assay.

2.7 Fluidigm

Microfluidic quantitative RT-PCR.

Multiplex qRT-PCR was performed as previously described (Swiers et al., 2013) using the Fluidigm (Biomark) platform. Briefly, cells were sorted directly into 10 µl RT-PreAmp Master Mix (5 µl CellsDirect 2X Reaction Mix (Invitrogen), 1.2 µl TE buffer (Invitrogen), 3.7ul of RNase-free water and 0.1 µl SUPERase-In (Ambion)). Taqman assays (Applied Biosystems; Table 2.6) were pooled to a final concentration of 0.2X. Reverse transcription and pre-amplification were performed using the following PCR conditions: 50°C for 15 min and 95°C for 5min for reverse transcription; thermalcyling at 95°C for 15 sec and 60°C for 4 min for 22 cycles. Pre-amplified mix was immediately frozen at -20°C. For 48.48 Dynamic Arrays, samples were diluted fivefold using TE buffer (Invitrogen) and analyzed using the following PCR conditions: 95°C for 10 min; thermalcycling at 95°C for 15 s and 60°C for 1 min for 40 cycles (Fluidigm, BioMark). Amplification curves were quality filtered using a threshold < 0.65, and Ct thresholds were set for each assay with the same

thresholds used across all experiments. Ct values were calculated by BioMark Real-time PCR Analysis software (Fluidigm). Data were exported to Excel as .csv files for subsequent analysis.

Gene	Assay id	Notes
Atp5a1	Mm00431960_m1	Housekeeper
Atp5a1	Mm01257366_m1	Housekeeper
Bcl11a	Mm00479358_m1	
Gata2	Mm00492300_m1	
Gfi1	Mm00515855_m1	
Gfi1b	Mm00492318_m1	
Hes1	Mm01342805_m1	
Hey1	Mm00468865_m1	
Hey2	Mm00469280_m1	
Hrpt1	Mm01545399_m1	Housekeeper
Ikzf	Mm01187878_m1	
Ikzf1	Mm01187880_m1	
Myb	Mm00501741_m1	
Pik3cd	Mm00435674_m1	
Pik3cg	Mm00445038_m1	
Ptprc	Mm00448490_m1	
Runx1	Mm01213404_m1	
Sfpi1	Mm00488142_m1	
Tal1	Mm01187033_m1	
Ubc	Mm01201237_m1	Housekeeper

Table 2.6 Taqman assays.

Fluidigm analysis

Raw Ct tables were exported from BioMark Real-time PCR analysis software (Fluidigm) and further processed in Excel. Cells that failed in 3 (or more) of the 4 housekeeping genes were excluded from the analysis. Assays with Ct values in greater than 30% no-template-controls (NTC) were excluded. Assays were run in duplicate, which were averaged prior to analysis. For single cell analysis, all genes that were not expressed were given a Ct of 27. All Ct values were subtracted from the background NTC value of 27 (logEx Ct – sample). An average housekeeping gene Ct (HK Ct) was then calculated from the average of *Atp5a1*, *Hprt* and *Ubc* (LogEx Ct –HK). A 2-deltadeltaCt was calculated by subtracting LogExCtHK from LogExCtSample. Hierarchical clustering, Pearson's correlation, and pseudotime moving average plots were generated in R (http://www.r-project.org). For bulk cell analysis, 25 cell pools were analyzed for 2 or 3 experimental replicates. deltaCt was calculated and then normalized to an average of house keeping genes *Atp5a1*, *Gapdh* and/or *Hprt* to calculate relative mRNA expression values. Histograms were generated in GraphPad Prism7.

Statistical analysis

Significance was established using an unpaired, two-way Student's T test with Welch's correction. Significant p-values were highlighted by asterisks: * $p<0.05$, ** $p<0.005$, *** $p<0.0005$, and **** $p<0.0001$.

2.8 Functional assays:

Mouse colony forming cell (CFU-C)

Colony-forming assays were performed using Methocult M3434 (StemCell Technologies), according to the manufacturer's guidelines. The cells were

plated in duplicate onto 35 x 10mm Cellstar tissue culture dishes (Grenier Bio One) at different cell densities in 1 ml of Methocult media. 1ml of Methocult-cell suspension was dispensed using a 16-gauge blunt-end needle, and cultures were grown at 37ºC for 14 days at which point hematopoietic colony formation was quantified.

OP9 maintenance and co-culture assay for hematopoietic potential

OP9 cells (provided my M. Yoder, University of Indiana) were maintained in alpha-MEM (Gibco) supplemented with 20% FCS (Hyclone) and 0.22% sodium bicarbonate (Gibco). Cells were maintained at strict density ranging from 30-80% with passaging about every 3 days. One day prior to co-culture, OP9 cells were plated at 2.3 $x10^3$ cells per cm^2. Depending on the application, cells were either sorted directly into the plate or into co-culture medium without cytokines. Co-culture medium with 2X cytokines were added in the latter case.

Co-culture media consisted of alpha-MEM (Gibco), supplemented with 10% FCS (Hyclone), and 100uM 2-mercaptoethanol (Sigma). For hematopoietic potential assays, co-culture media is supplemented with 10ng/ml IL7 (Peprotech), 10ng/ml SCF (Peprotech) and 10ng/ml Flt3-ligand (Peprotech). Co-cultures were cultured at 37°C for 8 days with no media changes. Colony formation was detected after 3-4 days of culture. Cultures were inspected at day 4, 6 and 8 to confirm blood cell formation using an Evos FL Cell Imaging system (Thermo Fisher).

For endothelial tube formation potential assays, co-culture media was supplemented with 50ng/ml VEGF and cultured without media changes for 8 days. After 8 days, cells were fixed in 4% PFA and stained with rat anti-mouse

CD31 (BD Pharminogen). Staining was revealed using a Vector Blue Alkaline Phosphatase (AP) substrate kit (Vector Laboratories). Briefly, an anti-rat-HRP and Vector Blue AP reagent labeled CD31 endothelial cells with a blue color.

For B-cell potential assays, co-culture media was supplemented with 10ng/ml IL7 (peprotech), 10ng/ml SCF (peprotech) and 10ng/ml Flt3-ligand (peprotech). Co-cultures were cultured at 37°C for 8 days with media changes every 3 days removing half of culture volume and replacing it with co-culture media with 2X cytokines. After 8 days, supernatant was collected and wells were washed with PBS and placed with supernatant. The cells were then spun down at 1,000 rpm for 5 min and then processed for flow cytometry analysis using the following antibodies: B220, CD19, CD11b, and Ter119 (See Table 2.2 for descriptions of antibodies).

2.9 General assays

Confocal

mESC-derived hematopoietic cells were grown on glass coverslips or on 24-well glass-bottom plates (Ibidi) and were imaged using a Zeiss 550 or Zeiss 880 Inverted confocal microscope. Either 25X, 40X or 63X objectives were used either with or without oil immersion depending on the lens requirements. ZEN software was used to acquire and process images (adding scale bars). Final processing of images was done using Adobe Photoshop and Illustrator.

For staining, cells were incubated with PBS, 0.3% Triton X-100, 3% FCS for 30 min. Primary antibodies were added to the same buffer and cells were incubated with 500ul of primary antibody for 2h at room temperature. Cells were washed in PBS twice and then incubated with secondary antibodies for

1h at room temperature. Cells were washed with PBS, excess PBS was remove, and Vectashield containing DAPI (Vector Labs) or Hoechst 33258 was added.

Antibody	Species	Supplier	Cat. No.	Dilution
CD41	rat	BD Pharm.	553847	1 in 40
CD31	rat	BD Pharm.	557355	1 in 1000
Flag	Mouse	Sigma	F1804	1 in 1000
Runt (Runx1)	rabbit	abcam	ab92336	1 in 200
AlexaFluor488	Phalloidin	Invitrogen	A34055	1 in 500
AlexaFluor488	Phalloidin	Invitrogen	A12379	1 in 1000
AlexaFluor488	Goat anti-rabbit	Invitrogen	A11008	1 in 500
AlexaFluor647	Goat anti-rabbit	Invitrogen	A27040	1 in 500
AlexaFluor555	Goat anti-Mouse	Invitrogen	A32727	1 in 1000
AlexaFluor594	Donkey anti-rat	Invitrogen	A21209	1 in 400

Table 2.7 Antibodies for confocal imaging.

Western Blotting

Single cell suspensions were washed with 1ml ice cold PBS. Pellets were re-suspended in RIPA buffer (10mM Tris-HCl (pH8.0), 1 mM EDTA, 0.5mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, 1 mM PMSF, 1X cOmplete protease inhibitors (Roche)) and incubated on ice for 30 min. After 30 min, cells were vortexed for 30sec and then placed on ice for 20 min. Lysed cells were then spun at 14,000g for 10 minutes at 4°C. Protein concentrations from cell extracts was measured with Protein Assay Dye Reagent (Bio-Rad). Equal amounts of total protein were loaded and separated in a 8-12% bis-tris SDS-PAGE pre-cast gel and transferred to a PVDF membrane (GE). After 1h of blocking with 5% Milk, membranes were incubated with primary antibodies overnight at 4°C. Primary antibodies used included Rabbit anti-Runx1 (Abcam) and mouse anti-Flag (Sigma F1804). Membranes were incubated for 1h with secondary antibodies including HRP-conjugated antibody to rabbit (1:10,000) (Sigma) and HRP-conjugated antibody to mouse (1:20,000) (Sigma). Amersham ECL prime western blotting detection reagent (GE) was used for chemiluminesence detection.

Luciferase

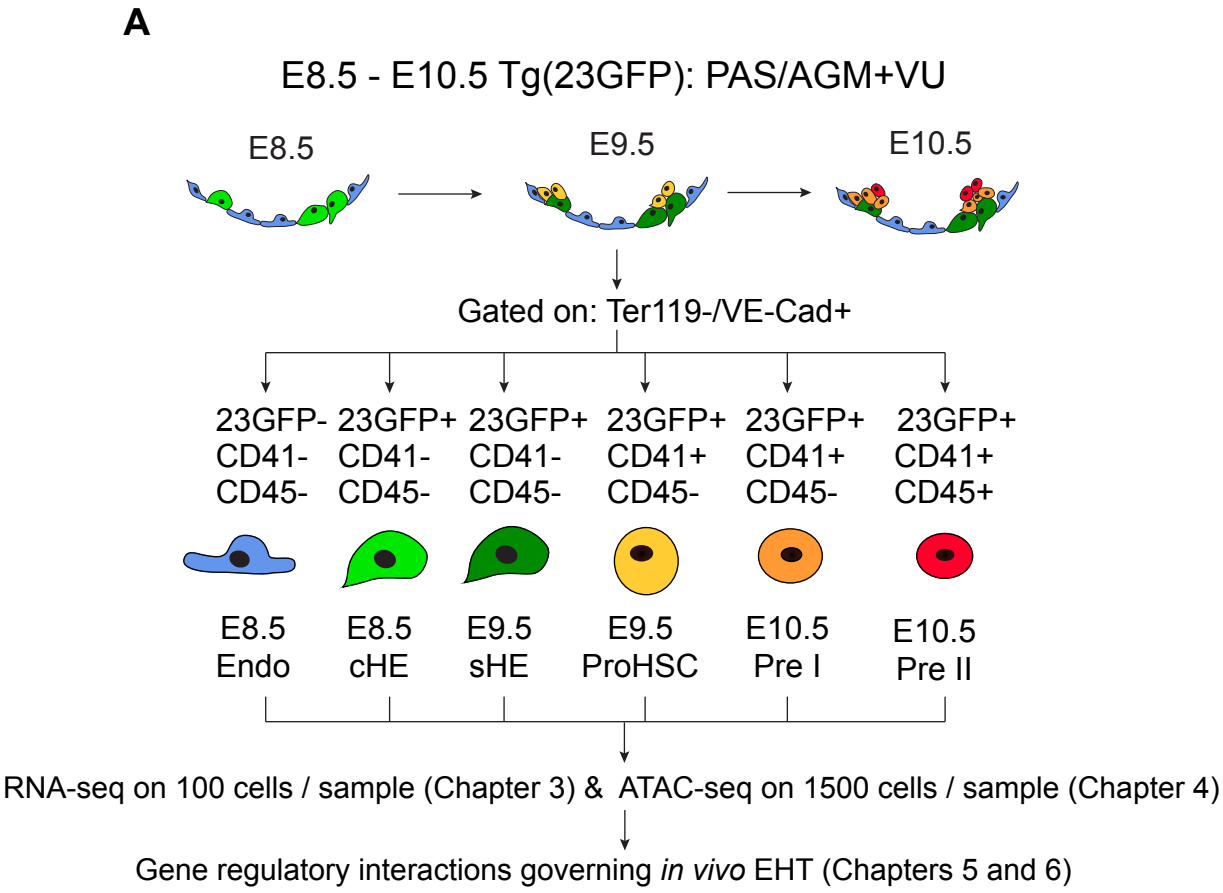
For Luciferase assays, we used a Dual-Luciferase reporter assay protocol (Promega). DNA fragments of interest were cloned into pGL3 expression vectors (Promega) using Gibson assembly either upstream or downstream of the luciferase reporter gene depending on the proximity to the promoter in the native locus for each fragment (e.g., Ikzf1 -28 element cloned upstream of pGL3 promoter). $5.0x10^6$ 416b cells were electroporated with pGL3 vectors and cultured for 24 hours at 37°C. Cells were harvested by centrifuging cells

at 1,000rpm for 5min at room temperature. Cell pellets were washed with 1X PBS and resuspended in 1ml of 1X Passive Lysis Buffer (PLB) and gently rocked for 15min at room temperature. 20ul of each of PLB lysate was added per well into 96-well luminometer plates (Thermo Fisher). A luminometer was programmed to perform a 2-second premeasurement delay followed by a 10-second measurement for each reporter assay. Injectors were set to dispense 100ul of Lucifease Assay Reagent II (LAR II) and Stop & Glo Reagent. Relative luciferase was calculated by subtracting Renilla luciferase activity and Firefly luciferase activity for each well. Each construct was measured in technical duplicates from 3 independent experiments.

Chapter 3: Expression profiling of EHT

Despite seminal studies establishing EHT as the means by which HSCs arise during development, questions remained regarding the point at which cells change their fate and initiate the process of commitment to the hematopoietic lineage. Recent work in the lab using high-throughput functional and transcriptional assays revealed that endothelium undergo an early loss of endothelial potential along with a gradual unfolding of the hematopoietic program (Swiers et al., 2013). This work utilized the *Runx1* +23GFP reporter line and identified two stages of HE in the dorsal aorta and vitelline and umbilical arteries, a competent stage (cHE) and a specifying stage (sHE). These populations are separated by developmental time and differ not only in functional potential but also at the molecular level. In the context of EHT, this study rewrote the timeline of hematopoietic commitment during EHT and showed commitment is initiated in sHE before the proHSC and type I and type II preHSCs stages (Figure 1.4). However, despite the important implications of this study, some of the same questions from the original EHT imaging studies remain unanswered. Mainly, how is EHT regulated at the molecular level?

Here, to map the global molecular changes during EHT with the aim of identifying novel regulators, genome-wide expression profiling by RNA-sequencing (RNA-seq) was performed on endothelium, HE, and HSPCs isolated from the dorsal aorta and vitelline and umbilical arteries of 23GFP transgenic E8.5, E9.5 and E10.5 embryos (Figure 3.1A-B). Quality control and validation were carried out to check the quality of the data. In addition to global analysis of differential expression over all 6 EHT populations, we focused on each transition individually because of the significant hematopoietic potential acquired with each subsequent stage.



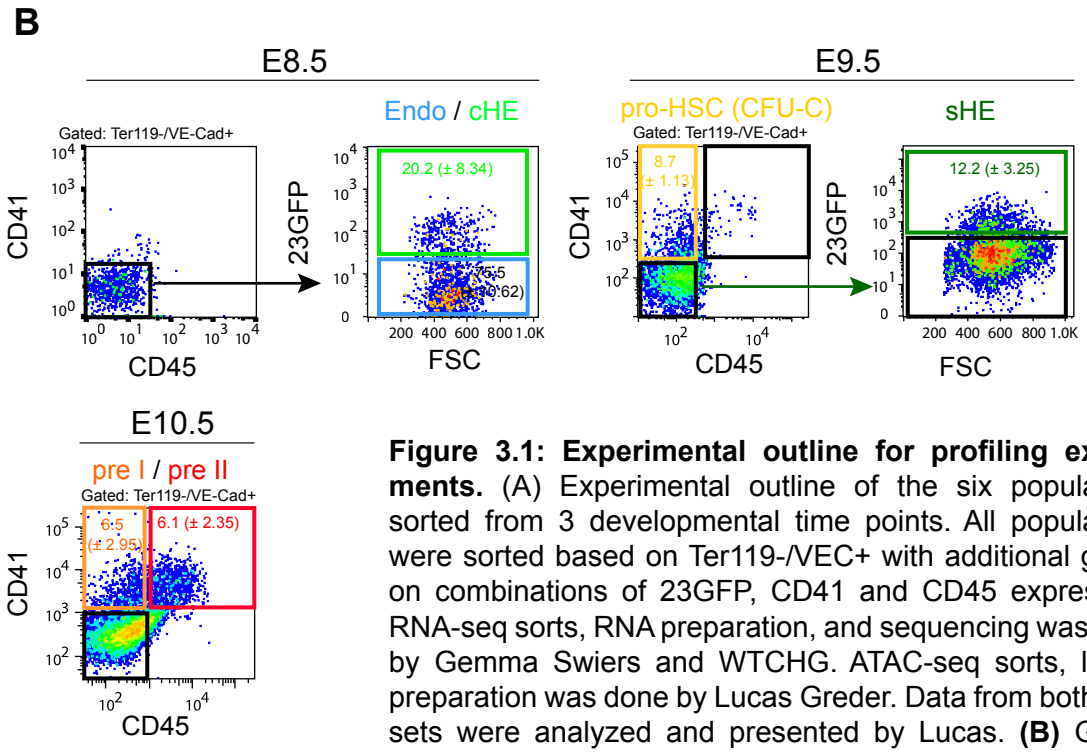


Figure 3.1: Experimental outline for profiling experiments. (A) Experimental outline of the six populations sorted from 3 developmental time points. All populations were sorted based on Ter119-/VEC+ with additional gating on combinations of 23GFP, CD41 and CD45 expression. RNA-seq sorts, RNA preparation, and sequencing was done by Gemma Swiers and WTCHG. ATAC-seq sorts, library preparation was done by Lucas Greder. Data from both data sets were analyzed and presented by Lucas. **(B)** Gating strategy for FACS isolation with scatter plots showing gates and distribution for all six EHT populations: Endothelium (Endo), competent hemogenic endothelium (cHE), pro-HSC/CFU-C (pro-HSC), specified hemogenic endothelium (sHE), pre-HSC Type I (preI) and pre-HSC Type II (preII).

3.1 Expression profiling of EHT by RNA-seq

Sorting, data processing, and quality controls

The RNA-seq sorting and sequencing was done prior to my arrival in the lab. However, I performed all subsequent steps beginning with processing of Fastq files from WTCHG. Analysis was performed with assistance from Dr. Emmanouela Repapi (ER; WIMM Computational Biology Research Group (CBRG)), and more recently by Joe Harman (a bioinformatician in the lab). The analysis presented here was the outcome of a joint effort between ER and myself.

Endo, HE and hematopoietic progenitors from the dorsal aorta and vitelline and umbilical arteries were isolated from E8.5, E9.5, and E10.5 embryos by flow cytometry (Figure 3.1A-B, Appendix Figures 3.1; 3.2 and 3.3) and were processed for expression profiling using RNA sequencing (RNA-seq). The E9.5 Ter119-VEC+CD41+CD45-23GFP+ cells were not a pure population of proHSC but a mix of clonogenic hematopoietic progenitors (CFU-C) and proHSCs. Analysis of these cells using an antibody for CD43 revealed that 16 ± 6 % were CD43- (corresponding to proHSCs) and 80±5% CD43+ (corresponding to CFU-C) (Appendix Figure 3.4) (Rybtsov et al., 2014). In line with this, RNA-seq counts per million values (CPMs) for CD43 revealed low-level expression in this population, indicative of the presence of clonogenic hematopoietic progenitors (Figure 3.2C). Thus, E9.5 CD41+ cells are called “proHSCs”.

For RNA-seq, 100 cell pools per replicate from 3-4 independent biological replicates were obtained and sequence reads were mapped to mouse genome mm9 producing between 5-15 million uniquely mapped reads (Figure

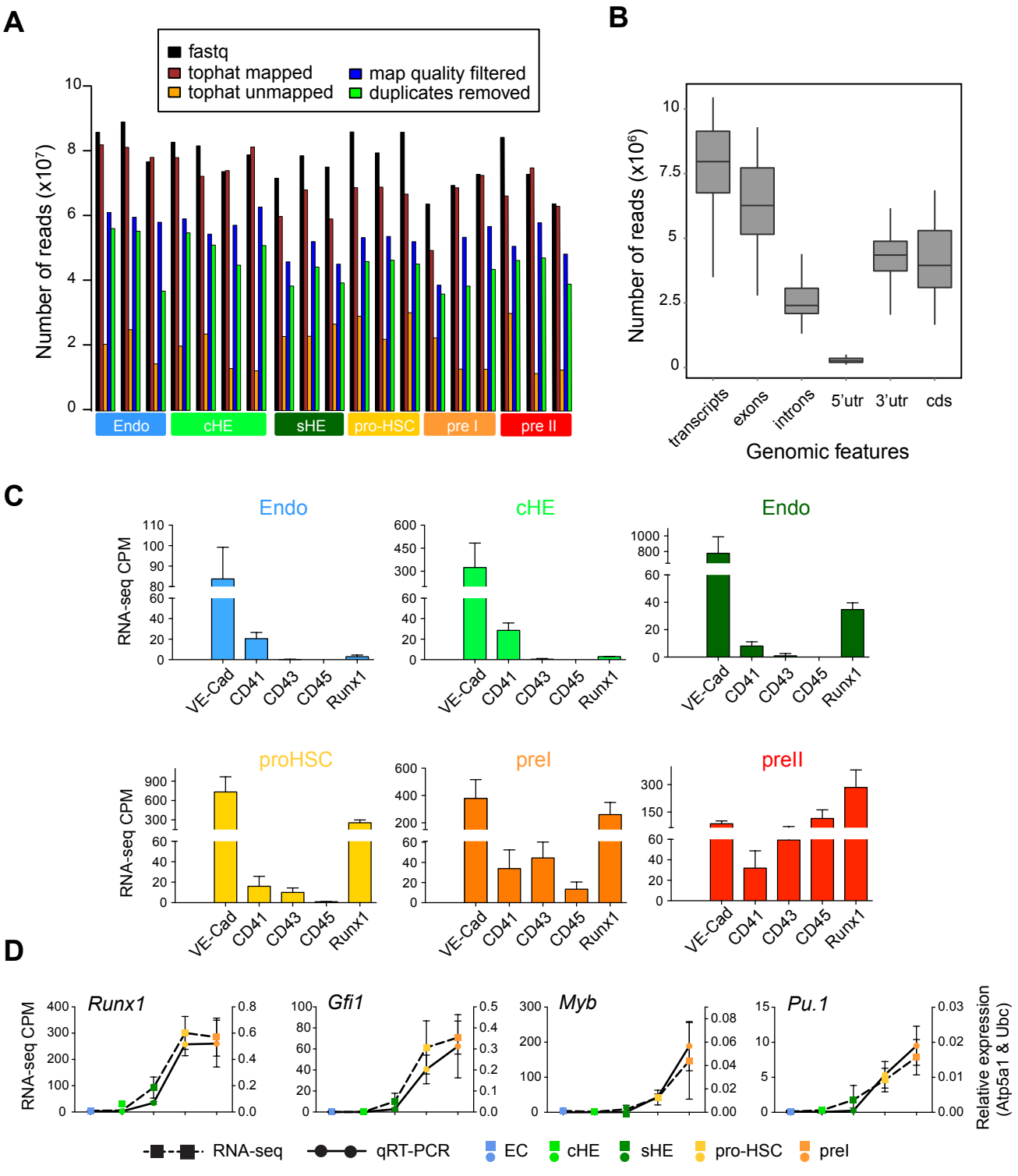


Figure 3.2: RNA-seq data quality control and validation. (A) Distribution of read counts were represented as bar graphs at each level of mapping and filtering for each sample. Fastq files were mapped to mm9 using Tophat, reads were filtered for low quality mapping to include only uniquely mapped reads, and PCR duplicates were removed. (B) Summary statistics represented as box plots for reads mapping to different genomic features including transcritps, exons, introns, 5' and 3' UTRs and coding sequence (cds). (C) Normalized number of mapped fragments for genes encoding the markers used for sorting the six EHT populations. CPMs for Ter119, VE-Cad, CD41 and CD45 are shown (error bars are SD). CPMs for CD43 and Runx1 also included. (D) RNA-seq CPM (dotted line) on the left y-axis and qRT-PCR (solid line) on the right y-axis for genes in same phenotypic populations (colored according to cell type) (error bars are SD).

3.2A). Reads were biased to 3' genomic elements (Figure 3.2B), which was expected given the method of RNA isolation and library preparation. There was overall good correlation between counts per million (CPM) values and cell surface expression for the markers used to sort the cells, *Ter119*, *VE-cadherin*, *CD41*, and *CD45* (Figure 3.2C), serving as an internal control. One exception was *CD41* expression in E8.5 Endo and E9.5 cHE and sHE, which did not materialize into CD41 protein at the cell surface as described before (Figure 3.2C; Swiers et al., 2013). Expression of *Runx1* in E9.5 HE confirms these cells were at the specifying stage (Figure 3.2C). Notably, *Runx1* expression begins at around E9.25 (Swiers et al., 2013) with CPM values correlating with Fluidigm qRT-PCR expression performed on 25 cell pools from the same phenotypic populations (Figure 3.2C-D). CPM values for other genes with known roles in EHT, such as *Myb*, *Gfi1*, and *Pu.1*, also correlated well with qRT-PCR expression performed on the same phenotypic populations, thereby validating the RNA-seq data set (Figure 3.2D)

Differential expression analysis

Analysis and normalization of the six EHT populations with EdgeR (McCarthy et al., 2012) revealed 5,834 (false discovery rate (FDR) < 0.05; Figure 3.3D) differentially expressed genes (DEGs). Principal component analysis (PCA) showed clustering according to cell type and developmental stage (Figure 3.3A), reinforcing the transient nature of HE and EHT generally. KEGG pathway enrichment of the top 100 loadings for principal component 1 (PC1) and principal component 2 (PC2) showed "Hematopoiesis" as a primary driver of variance among populations along PC1 (Figure 3.3B), and metabolic and cycle genes were key drivers of variance along PC2 (Figure 3.3B).

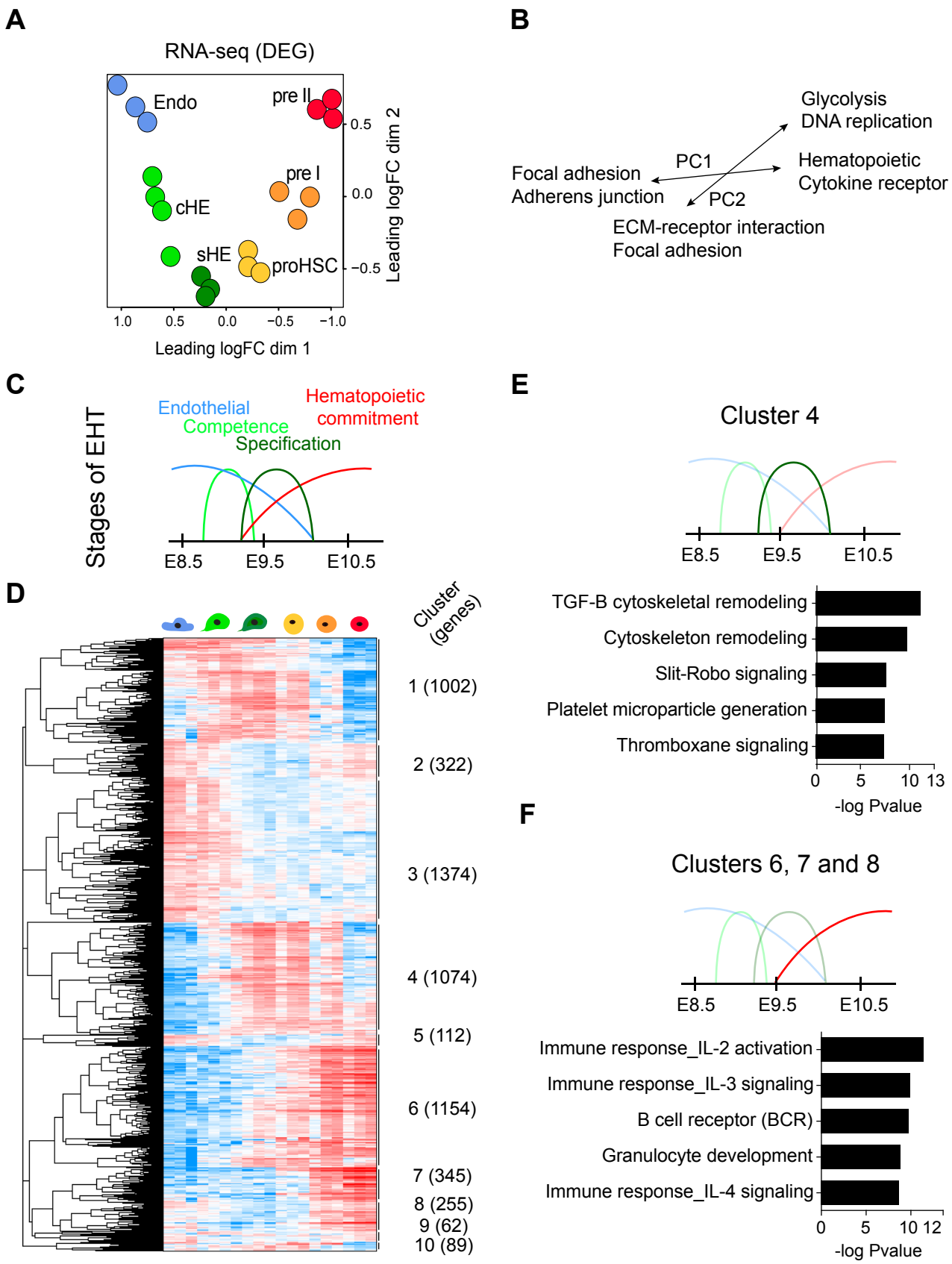


Figure 3.3: RNA-seq data analysis. Legend is on the next page.

Figure 3.3: RNA-seq data analysis. (A) Principal component analysis (PCA) of RNA-seq samples from all biological replicates for each population (all n=3, except cHE n=4) colored according to population. (B) KEGG pathway enrichment for top 100 principal component loadings in each of the four quadrants (+PC1/+PC2, +PC1/-PC2, -PC1/-PC2, -PC1/+PC2). (C) Schematic of stages of EHT as they appear in the RNAseq data. (D) Heatmap showing all 5,834 DEGs and hierarchical clustering of the genes in each EHT population from either n=3 or n=4 biological replicates. (E) Metacore pathway enrichment during the specficiation stage of EHT shown in bar graph of top 5 most enriched based on -log Pvalue. (F) Metacore pathway enrichment during hematopoietic commitment stage shown in bar graph of top 5 most enriched based on -log Pvalue.

Unsupervised hierarchical clustering of all 5,834 DEGs highlighted expression patterns representative of the different stages of EHT: Endothelial, Competence, Specification and Hematopoietic commitment (Figure 3.3C-D; Figure 1.4).

Metacore pathway enrichment of "Specification" genes show over representation of "TGF-β cytoskeletal remodeling" and "Slit-Robo signaling" terms (Figure 3.3E). "Hematopoietic" genes showed over representation of "Immune response" and hematopoietic progenitor specific pathways ("B cell receptor (BCR)" and "Granulocyte development") (Figure 3.3F). Enrichment for "Immune response" matched previous work showing the requirement for innate immunity during HSPC emergence (Li et al., 2014, Espin-Palazon et al., 2014, Sawamiphak et al., 2014). Interestingly, enrichment for "B cell receptor" signaling may offer insight into recent work showing B cell lineage potential at the preHSC stage (Hadland et al., 2017). Moreover, as B cell potential in preHSCs coincides with the onset of *Runx1* expression, it is tempting to postulate that the later may be playing a direct role in conferring B cell potential at this stage during EHT.

Differentially expressed TFs

To identify candidate regulators of EHT, we looked specifically at transcriptional regulators TFs because they are intrinsic regulators of cell fate decisions. Of the 5,834 DEGs, 413 were TFs (Figure 3.4A). Unsupervised hierarchical clustering grouped the 413 differentially expressed TFs (FDR < 0.05) into four expression patterns representative of the different stages of EHT (Figure 3.4A). 130 TFs in Pattern 1, which represented TFs significantly up-regulated in the hematopoietic-fated cells, contained many TFs with known

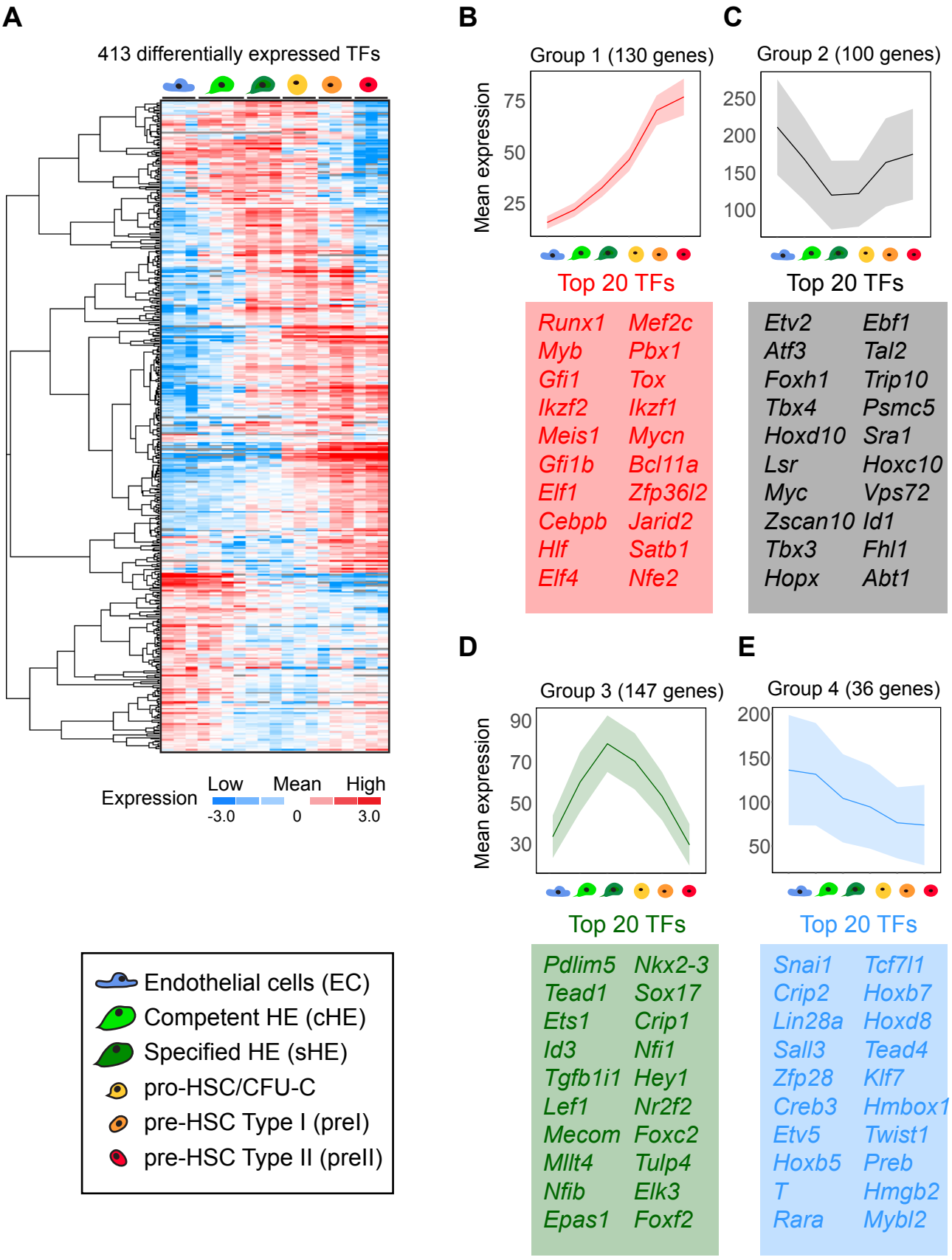


Figure 3.4: Differentially expressed TFs. (A) Unsupervised hierarchical clustering of 413 differentially expressed TFs. Represented here as a heatmap. (B-E) Plots of the four main expression patterns of 413 differentially expressed TFs represented here as means of each gene from all biolgoical replicates and then as mean for all genes with the cluster (shaded line represent SD). Clusters colored according to stage of EHT. Top 20 differentially TFs from each cluster list in order of which they appeared in the list from the hierarchical clustering.

roles in HSC development, including *Runx1*, *Myb*, *Gfi1b*, *Gfi1*, *Sfpi1, Meis1, Lmo2* and *Cebpb*, among others (Figure 3.4B; Appendix Table 3.2). Pattern 1 also included TFs involved in hematopoiesis generally but not previously implicated in EHT including *Eya2*, *Stat5a*, *Stat4*, *Bcl11a*, *Fos*, *Nfe2, Ikzf1, Mllt3, Ikzf2, Lyl1, Mef2c, Nfix* and *Hlf*, among others (Figure 3.4B). Consistent with recent work showing the requirement for Glucocorticoid receptor (GR) in HSC emergence in Zebrafish (Kwan et al., 2016), several genes implicated in GR-mediated transcriptional activation are present in group 1 including *Nr3c1*, *Ncoa2*, *Stat1*, *Sfpi1*, and *Stat5a*. This suggests a similar mechanism may be required for HSC emergence in mice.

The 147 TFs within the "specification phase" cluster (cluster 3) peaked in expression in sHE before being down regulated in pro- and preHSCs (Figure 3.4D; Appendix Table 3.2). These TF included several different developmental pathways including TGF-β (*Tgfb1l1*), Notch (*Hey1*, *Hey2*, *Hes1*), Hippo (*Tead1*, *Tead2*), and beta-catenin (*Ctnnb1*, *Tcf7*, *Tcf7l2*, *Lef1*) (Figure 3.4D; Appendix Table 3.2). Of note was the presence of Hox genes *Hoxb4* and *Hoxa3*. Down regulation of *Hoxa3* matches previous reports showing an inverse relationship with *Runx1* during HSC emergence (Iacovino et al., 2011). On the other hand, the *Hoxb4* expression pattern suggests if Hoxb4 plays a role in HSC emergence in the embryo it occurs prior to preHSC maturation rather than via its role in promoting self-renewal of HSCs (Cellot et al., 2007) or conferring engraftment potential onto hematopoietic progenitors (Kyba et al., 2004). Interestingly, cluster 2 contains genes with the opposite expression pattern from cluster 3 (Figure 3.4C; Appendix Table 3.2). This list includes *Etv2*, which has been implicated in regulation of HSC emergence

(Liu et al., 2015; Sumanas et al., 2008). The presence of *Myc* in this cluster suggests that proliferation, if any, during the specification phase (up in cHE, sHE and proHSC), is independent of Myc. Implications of this finding need further investigation but may provide useful information regarding how to best drive proliferation in *in vitro* cultures.

Finally, the 36 TFs in the "endothelial phase" cluster (cluster 4) had highest expression in Endo before losing expression in each subsequent cell fate transition (Figure 3.4E: Appendix Table 3.2). A portion of these TFs are involved in EMT (*Snai1*, *Snai2* and *Twist1*), highlighting similarities between EHT and EMT (Figure 3.4E). Of note, a candidate tumor suppressor gene, *cysteine-rich intestinal protein 2* (*Crip2*) under goes a 25-fold reduction from 948±67 CPM in Endo to 38±27 in preII. Of note, *Crip2* acts as a transcriptional repressor of NF-κB-mediated pro-angiogenic cytokine expression (Cheung et al., 2011). Given the requirement for NF-κB-mediated signaling in HSC emergence (Espin-Palazon et al., 2014), proper down-regulation of *Crip2* is likely required for full-blown HSC emergence.

Expression of TFs used for reprogramming studies

With one of the major aims of this study being to inform future improved protocols for the generation of HSC *in vitro*, we investigated how current *in vitro* methods compared to *in vivo* HSC formation. Specifically, we focused on the TF cocktails used to reprogram various different starting populations into HSCs, or HSC-like cells because reprogramming has proven to be the most reliable way of generating HSC in a culture dish (Sugimura et al., 2017, Lis et al., 2017, Riddell et al., 2016, Periera et al., 2013). Despite the recent success of Sugimura *et al.* and Lis *et al.*, many questions remain about the

mechanisms underlying de novo generation of HSCs. Given the relative inefficiency of generating HSC using reprogramming and the need to "incubate" the cells *in vivo* after the cells reach a certain "pre-mature" stage *in vitro*, we hypothesized that the cocktail of reprogramming factors may need optimization in order to make the process work entirely *in vitro*. Therefore, we looked in our EHT RNA-seq data to see which of our populations expressed these TFs *in vivo*. We focused on four different reprogramming lists: Sugimura et al., 2017 (Daley Lab); Riddell et al., 2014 (Rossi Lab); Lis et al., 2017 (Rafii lab); and Periera et al., 2014 (Moore Lab) (Figure 3.5). Each study began with a pool of TFs reduced down to a core set of TFs (core TFs are in bold). When expressed in a given cell type, the core TFs conferred hematopoietic potential.

With a huge advance for the field, Sugimura *et al.* generated HE from pluripotent cells and then transplanted these PSC-derived HE into mice to make HSCs. Tetracyline-induced expression of the exogenous factors in conjunction with the *in vivo* microenvironment of the mouse facilitated HSC production from the transplanted PSC-derived HE. In our data set 5 of the 7 core TFs were expressed in proHSC, preI and preII (*Runx1*, *Hoxa5*, *Lcor*, *Sfpi1*, *Hlf*) (Figure 3.5A), while 2 of the 7, *Hoxa9* and *Hoxa10*, were expressed at highest levels in Endo (Figure 3.5A). This suggested that in addition to TFs expressed in hematopoietic cells TFs expressed at the onset of EHT might still be needed at the time of transplantation. Thus, the right combination of TFs needed for making HSCs from pluripotent cells likely includes TFs expressed both at the onset of EHT and during preHSC maturation.

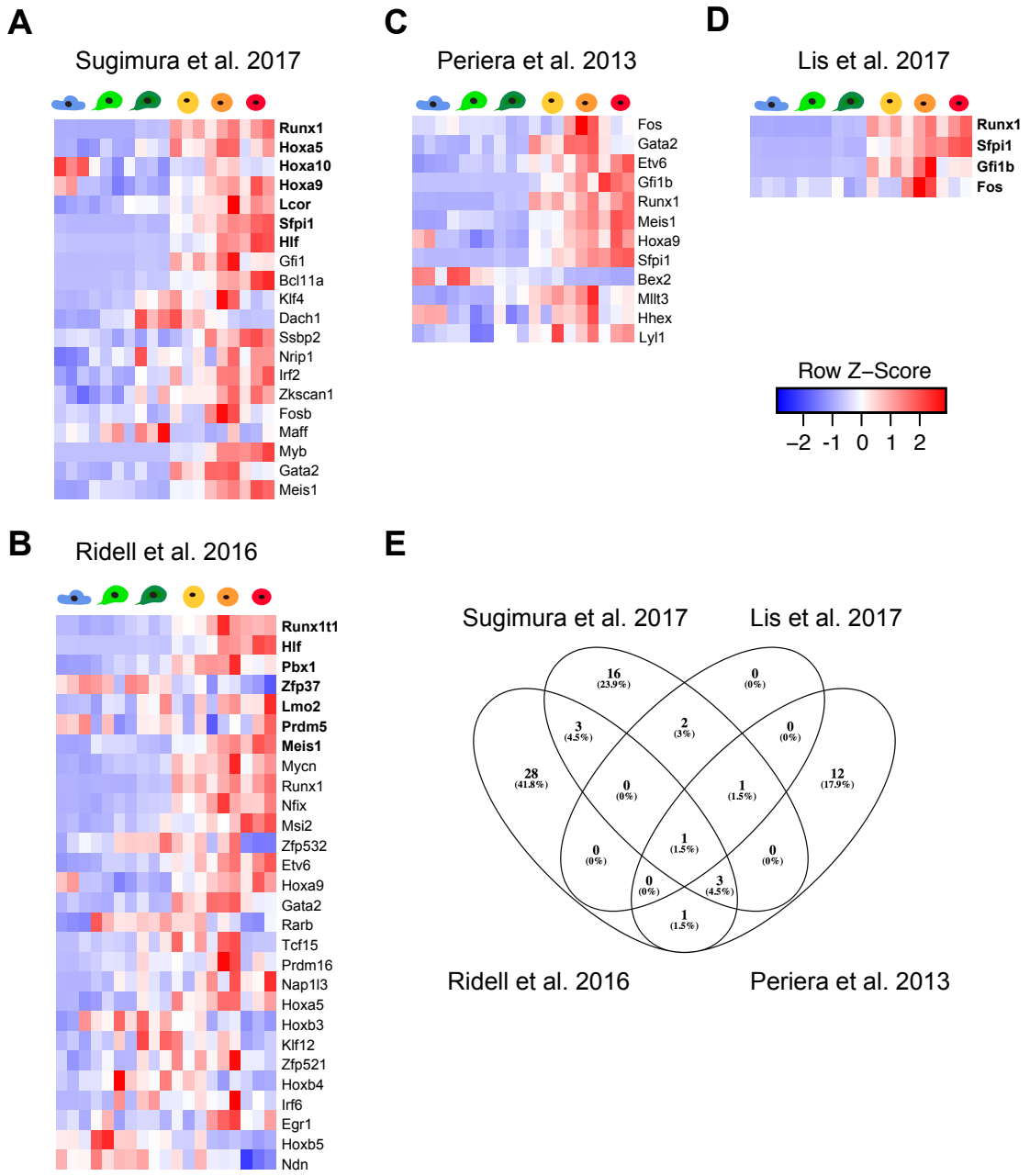


Figure 3.5: Expression of reprogramming TF cocktails during EHT. (A) Heatmap of 26 TFs used for reprogramming in Sugimura et al. 2017. (B) Heatmap of 18 TFs used for reprogramming in Periera et al. 2014. (C) Heatmap of 4 TF used for reprogramming in Lis et al. 2017. (D) Heatmap of 36 TFs from reprogramming in Riddel et al. 2016.

Riddell *et al.* partially reprogrammed cells using exogenous expression of TFs before transplanting these cells into mice. Interestingly, the same principle was at work in both studies – use of the mouse microenvironment to facilitate generation of HSCs. In our data, 6 of the 8 core TFs were expressed only in proHSC, preI and preII (*Runx1t1*, *Hlf*, *Pbx1*, *Lmo2*, *Mycn*, *Meis1*) (Figure 3.5B). The other two TFs, *Zfp37* and *Prdm5*, were expressed in Endo, cHE and sHE (Figure 3.5B). These expression patterns supported the idea that TFs promoting an endothelial or HE-related transcriptional program might still be required when the cells were transplanted into mice. Not surprisingly, when considering other TF cocktails (Lis et al, 2017, Periera et al., 2013) (Figure 3.5C-D) and overlapping the four to see if there is a "core" set of reprogramming factors, the only gene in all four lists is *Runx1* (Figure 3.5E). This again supports the idea that TFs (*e.g.*, *Runx1*) that promote an HE-related transcriptional program are required for in vitro generation of HSCs.

Overall, despite recent advances in *de novo* generation of dHSCs, the process has been hindered by our lack of understanding of how HE differentiate into preHSCs and how preHSCs become HSCs. Analysis of *in vivo* expression patterns of TFs used in reprogramming studies suggested that greater emphasis should be placed on TFs that drive transition from HE to preHSC and from preHSCs to dHSCs. For example, Runx1 was present in each reprogramming cocktail and is required for the transition from HE to preHSC. Moreover, a greater understanding of the Runx1-dependent transcriptional program and Runx1 target genes is critical, especially when considering how Runx1 haploinsufficiency affects dHSC generation (Cai et al.,

2000). In order to identify key TFs, we performed differential gene expression on individual transitions in EHT.

3.2 Expression profiles during early EHT: competence and specification stages

The first clonal analysis of cells undergoing EHT in E8.5 to E10.5 mouse embryos showed that the hematopoietic program is initiated much earlier than previously thought (Swiers et al., 2013). At E8.5, endothelium is separated into non-hemogenic and hemogenic with the latter capable of responding to blood inducing signals (Swiers et al., 2013). This is followed at E9.5 by a gradual loss of endothelial potential and an increase in hematopoietic capacity. In order to understand how these cell fate decisions are regulated at the molecular level, we examined pairwise differential expression for each transition.

Endo to cHE

First, we identified and characterized the differentially expressed genes for Endo and cHE. Comparing Endo and cHE expression profiles, differential gene expression identified 1,323 DEGs with 906 up regulated and 417 down regulated in cHE (FDR < 0.05, Figure 3.6A). Of the 91 differentially expressed TFs, 63 were up regulated and 28 were down regulated in cHE (Figure 3.6B). Up-regulated TFs were associated with Notch, Smad, and TGF-β signaling, which all have been previously implicated at the onset of EHT. Mean expression analysis of the down-regulated genes showed there was a subset of down-regulated genes that increased expression in pro- and preHSCs (Figure 3.6C). Pathway enrichment for the down-regulated genes included ribosome biogenesis, which may be explained by *Runx1* expression, which is

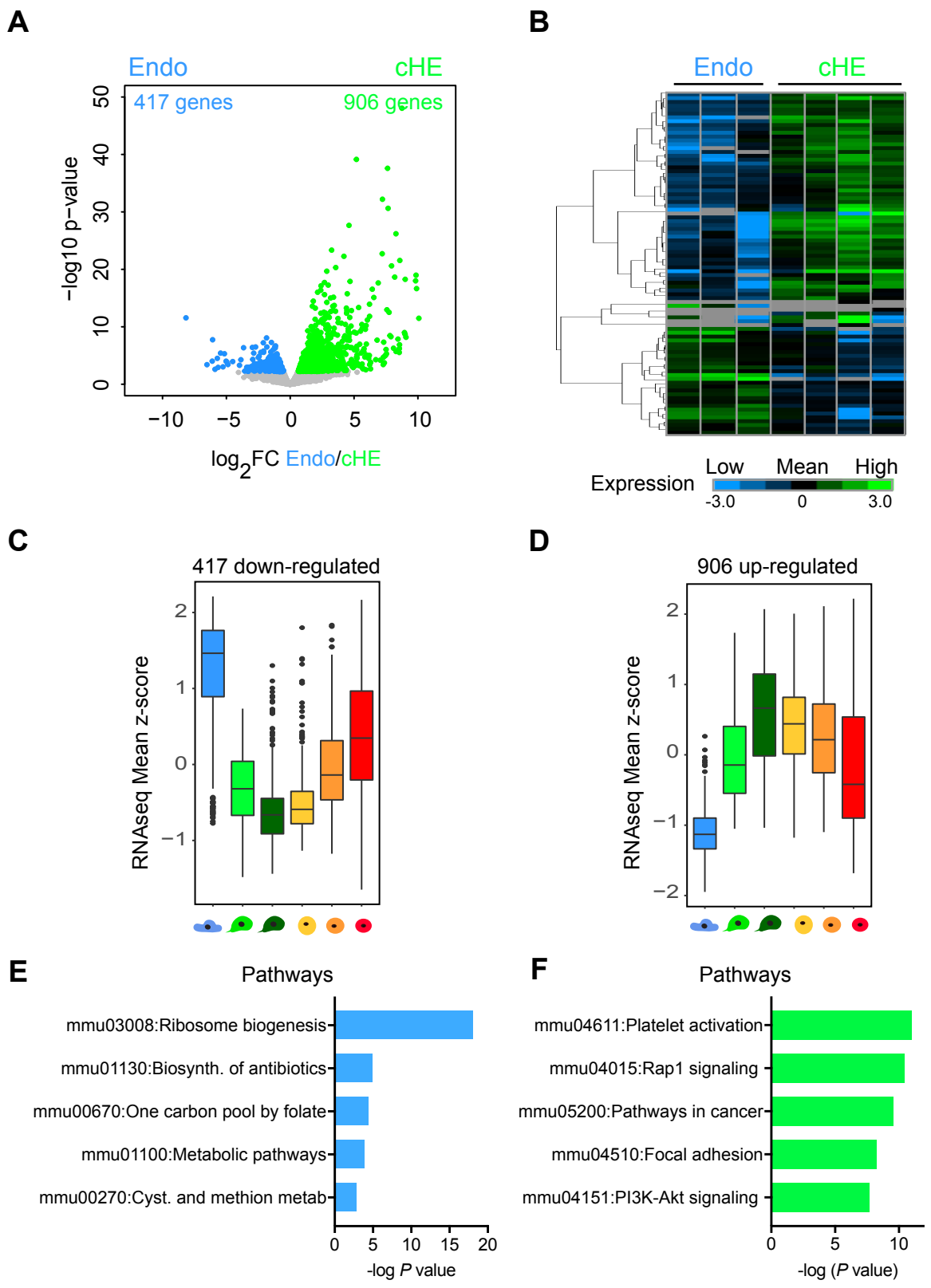


Figure 3.6: RNA-seq data analysis of Endo to cHE transition. (A) MA plot of differential expression for Endo to cHE with down-regulated genes in blue (FDR <0.05) and up-regulated genes in geen (FDR < 0.05) (FDR > 0.05 in grey). (B) Heatmap of 91 differentially expressed TFs. (C) Box plots representing the normalized mean across all EHT populations for all 417 down-regualted genes. (D) Box plot representing normalized mean across all EHT populations for all 906 up-regulated genes. (E) Top 5 KEGG pathways enriched in down-regualted genes represented in a bar graph ranked according to -log Pvalue. (F) Top 5 KEGG pathways enriched in up-regulated genes represented in a bar graph ranked according to -log Pvalue.

highly expressed in preII, has been show to positively regulate ribosome biogenesis (Cai et al., 2015). However, why ribosome biogenesis is enriched in Endo where *Runx1* is not expressed remains unclear. Means expression analysis of the up-regulated genes reveals increased expression in cHE to sHE transition before being down regulated in the remaining transitions (Figure 3.6D). Up-regulated genes are enriched for platelet activation, Rap1 signaling, Focal adhesion, and PI3K-Akt signaling (Figure 3.6F). Of note, PI3K-AKT signaling promotes arteriovenous specification and is therefore required for the initial cell fate decision.

cHE to sHE

Differential gene expression of the cHE to sHE transition identified 1,385 DEGs with 588 up regulated and 797 down regulated in sHE (FDR < 0.05, Figure 3.7A). 99 TFs were differentially expressed including TFs previously implicated in EHT (*Runx1*, *Gfi1*, *Sfpi1*, among others) and TFs with unknown roles, if any, in EHT (*Mef2c*, *Bcl11a*, *Nfia*, among others) (Figure 3.7B). Down-regulated genes were also expressed in Endo but remained down regulated as cells became proHSCs (Figure 3.7C). Interestingly, as the cells became more hematopoietic, these genes tended to slightly increase expression. Down-regulated genes were enriched for “Biosynthesis of amino acids”, “Biosynthesis of antibiotics” and “Ribosome biogenesis” (Figure 3.7E), similar to the pathways enriched in the down-regulated genes in the Endo to cHE transition. The Biosynthesis-related pathways are unexpected but could be explained by cell cycle requirements (Lane et al., 2015). As was the case in the Endo to sHE transition, although “Ribosome biogenesis” was a known target of Runx1 in hematopoiesis (Cai et al., 2015), its negative correlation

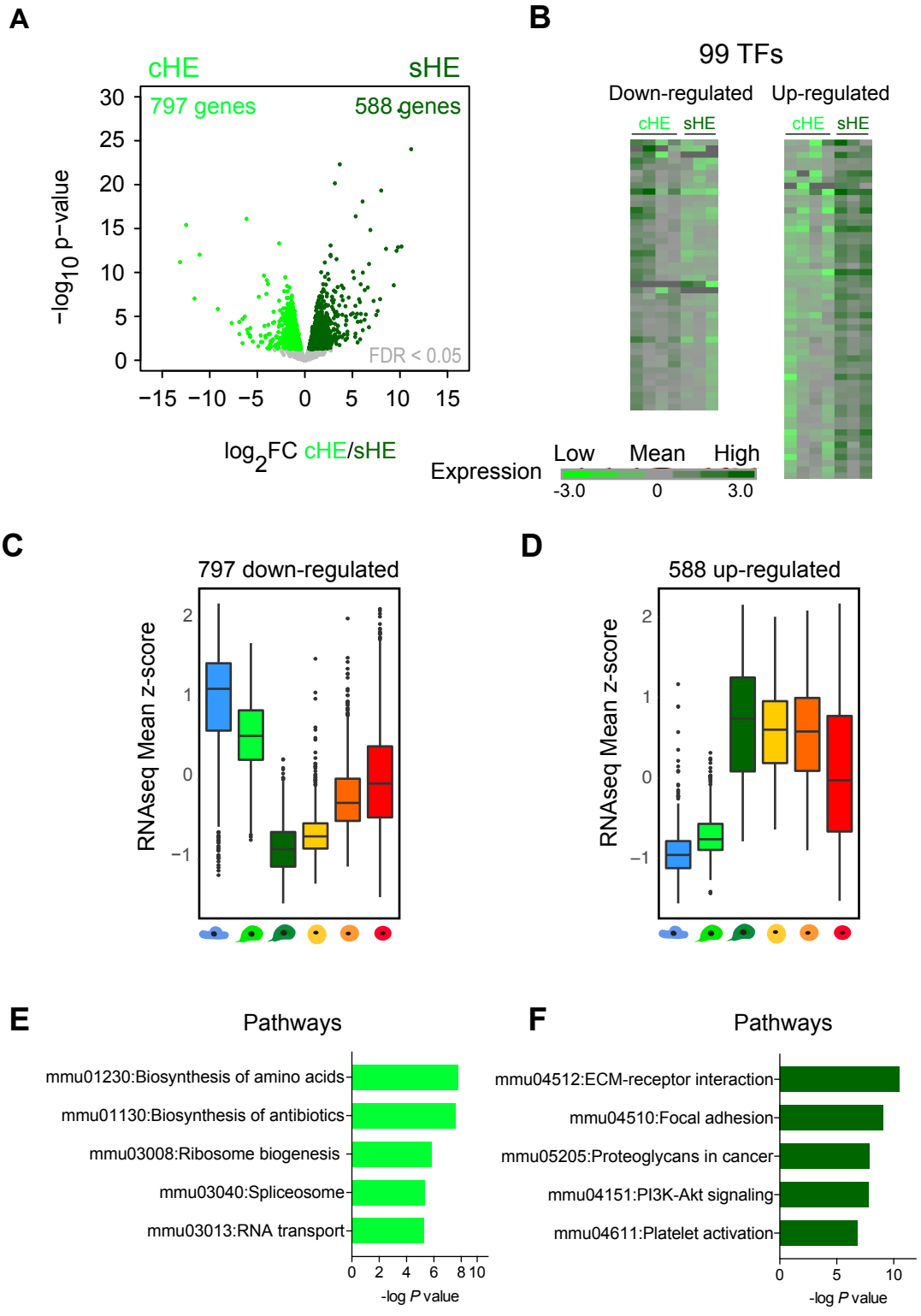


Figure 3.7: RNA-seq data analysis of cHE to sHE transition. (A) MA plot of differential expression for cHE to sHE with down-regulated genes in light green (FDR <0.05) and up-regulated genes in dark geen (FDR < 0.05) (FDR > 0.05 in grey). (B) Heatmap of 99 differentially expressed TFs. (C) Box plots representing the normalized mean across all EHT populations for all 797 down-regualted genes. (D) Box plot representing normalized mean across all EHT populations for all 588 up-regulated genes. (E) Top 5 KEGG pathways enriched in down-regualted genes represented in a bar graph ranked according to -log Pvalue. (F) Top 5 KEGG pathways enriched in up-regulated genes represented in a bar graph ranked according to -log Pvalue.

with Runx1 expression, which increases in expression from 3.1±0.1 to 35±5 CPM, is surprising.

Up-regulated genes maintained similar expression levels throughout the rest of EHT (Figure 3.7D). This supports the idea that the hematopoietic program (*e.g.*, the up-regulated genes) began to be "locked in" during this transition. Up-regulated genes were enriched for focal adhesion genes, highlighting the migration and adhesion changes associated with blood cell formation that have previously been identified (Figure 3.7F) (Lie-A-Ling et al., 2014). Another enriched pathway in up-regulated genes previously implicated in blood cell formation is "proteoglycans in cancer" (Figure 3.7F). Proteoglycans are glycosylated proteins consisting of covalently attached glycosaminoglycans that allow connective tissue to withstand compressional forces (Yanagishita 1993). Heparin sulfate, a major ligand of proteoglycans, has been used *in vitro* to identify a subpopulation of cells with increased hematopoietic potential (Baldwin et al., 2008) and to increase hematopoietic differentiation from mESC (Holley et al., 2011). A third enriched pathway that has been previously identified to be involved in regulating hematopoiesis but has not been investigated in EHT is "PI3K-Akt signaling" (Figure 3.7F). Of note, phosphatidylinositol 3'-kinase (PI3K)-AKT signaling ("PI3K-AKT") is enriched in several different cell fate transitions during EHT. We explored the role of PI3K-AKT signaling during EHT in **Chapter 5**.

Given the significance of the cHE to sHE transition as the point in which commitment to the hematopoietic lineage is "locked in", we sought to understand this transition in more detail. In order to do this, we looked at GO enrichment analysis of all 1,385 DEGs to unravel new biological processes

underlying this transition. The DAVID functional annotation tool (https://david.ncifcrf.gov/summary.jsp) generated biological processes enriched in all DEGs. Using REVIGO to cluster biological processes based on semantic similarity revealed six clusters with log Pvalue < 10^{-4}, including tissue morphogenesis, cell adhesion/movement, positive regulation of developmental process, regulation of cell proliferation, chemotaxis, and cytoskeleton organization (Figure 3.8) (Supek et al., 2011). Many of these have been investigated in EHT in some capacity. However, to our knowledge, little is known about cytoskeletal organization during EHT. Therefore, when combined with the "cytoskeletal remodeling" pathways identified in the global differential expression analysis (Figure 3.3E-F), we hypothesized an active rather than passive role for actin cytoskeletal remodeling during EHT, which is discussed in greater detail in Section 3.4.

3.3 Expression profiles during determination phase of EHT

Similar to the early transitions of EHT, regulatory mechanisms governing cell fate decisions within the determination phase of EHT were also not well understood. To address this gap in our knowledge, we focused on the three pairwise comparisons within this phase of EHT with the aim of identifying putative regulators.

sHE to proHSC

Differential gene expression of the sHE to proHSC transition identified 214 DEGs, 171 up regulated and 43 down regulated in proHSCs (FDR < 0.05, Figure 3.9A). Of the 214, 24 were TFs with many having been previously identified for their role in HSC development or HSC biology, including *Runx1*, *Gfi1*, *Gfi1b*, *Myb*, *Spi1*, *Nfe2*, *Cebpb*, *Hlf1*, *Meis1*, *Gata2*, *Ikzf1* and *Ikzf2*

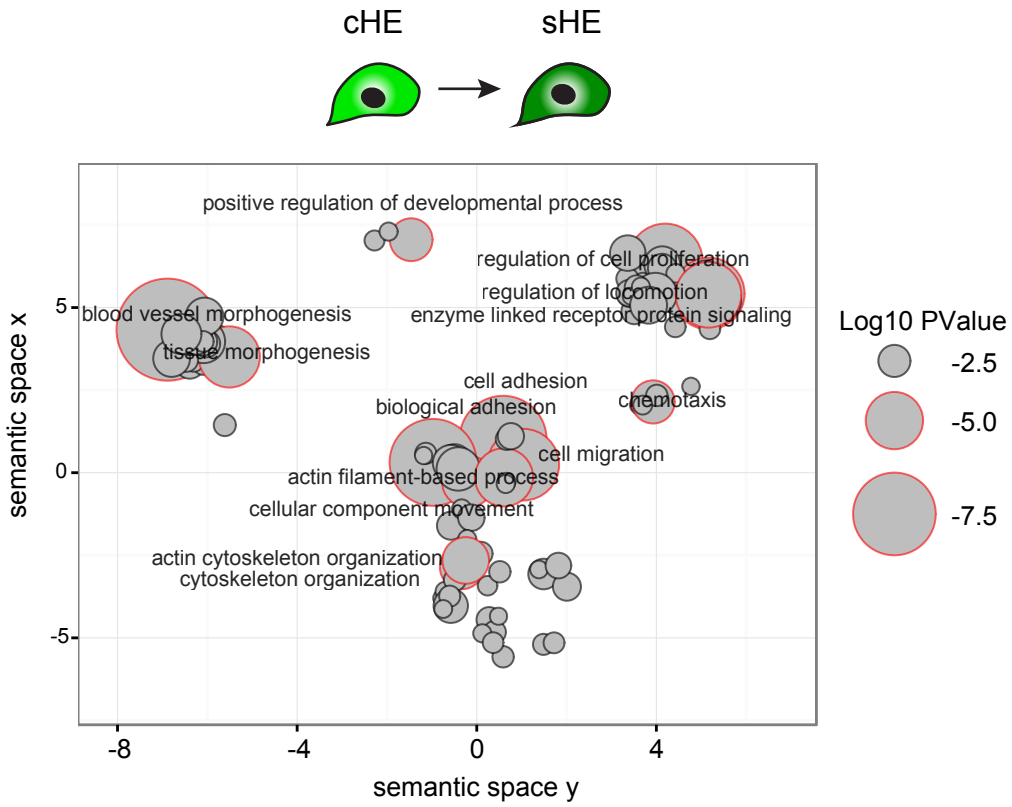


Figure 3.8: Enrichment of GO-terms from all 1,385 DEGs in cHE to sHE transition. Functional enrichment of GO-terms generated by DAVID (https://david.ncif-crf.gov/summary.jsp) was summarized using REVIGO. For enriched GO-terms from all 1,385 DEGs, the non-redunant terms remaining after redundancy reduction were represented in scatterplots in two dimensional space that summarize GO terms' semantic similarities. Similar GO terms remain close together in the plot, but the semantic space has no real meaning. Circle size indicates the p-value for the false discover rate based on the frequence of the GO term in the GO output from DAVID.

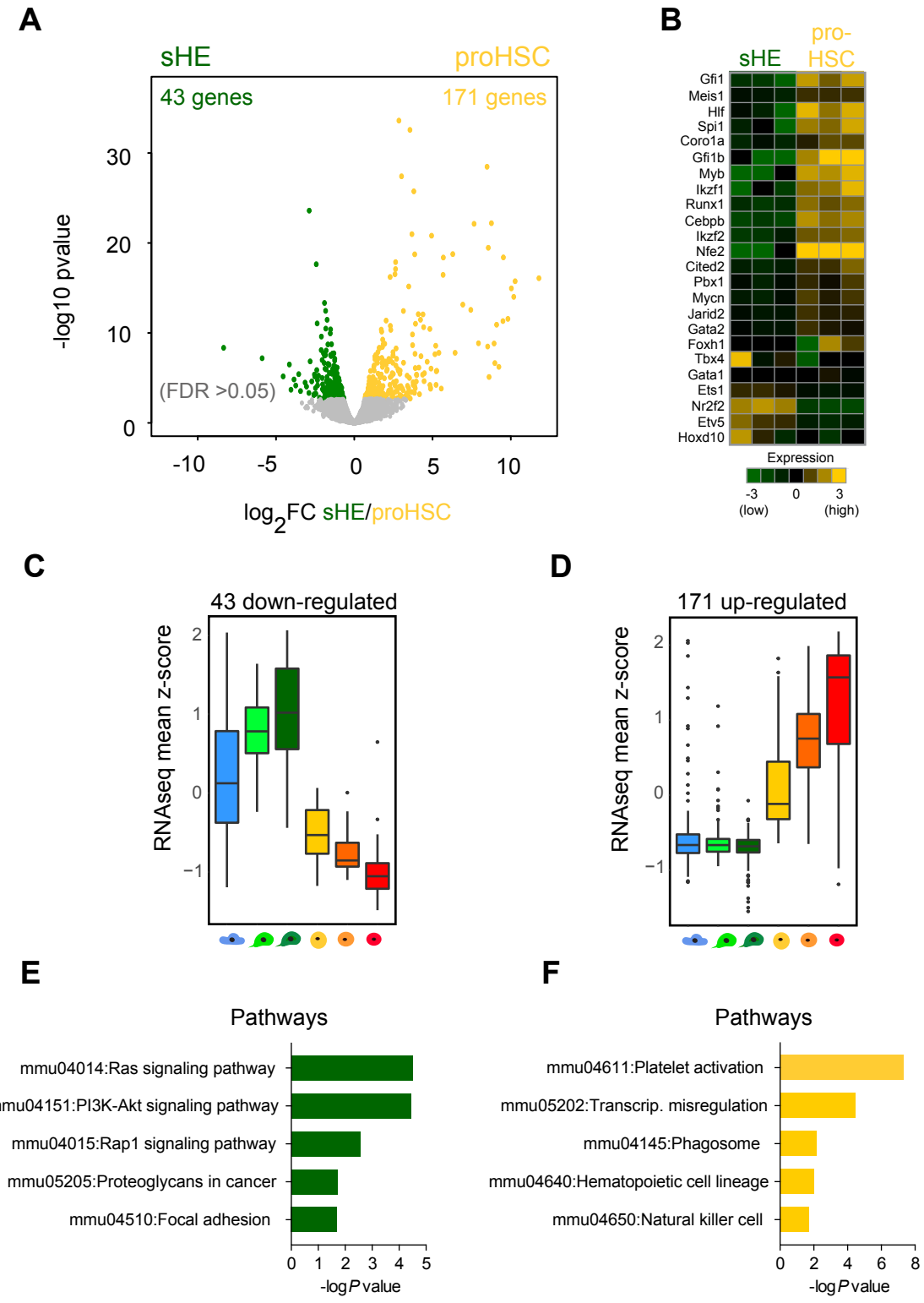


Figure 3.9: RNA-seq data analysis of sHE to proHSC transition. (A) MA plot of differential expression for sHE to proHSC with down-regulated genes in green (FDR <0.05) and up-regulated genes in yellow (FDR < 0.05) (FDR > 0.05 in grey). (B) Heatmap of 24 differentially expressed TFs. (C) Box plots representing the normalized mean across all EHT populations for all 43 down-regualted genes. (D) Box plot representing normalized mean across all EHT populations for all 171 up-regulated genes. (E) Top 5 KEGG pathways enriched in down-regualted genes represented in a bar graph ranked according to -log Pvalue. (F) Top 5 KEGG pathways enriched in up-regulated genes represented in a bar graph ranked according to -log Pvalue.

(Figure 3.9B). Interestingly, several of these genes (e.g., *Ikzf1*, *Ikzf2*, and *Pbx1*) were yet to be investigated fully during EHT. The Ikaros-family is discussed in more detail in Chapters 5 and 6. Another interesting gene, *Jarid2*, a member of the polycomb repressive complex 2 ("PRC2"), was differentially expressed from 94±13 CPM in sHE to 196±25 CPM in proHSC. *Jarid2* was recently identified as an important component of PRC2-dependent coordination HSPC function (Kinkel et al., 2015). However, the function of *Jarid2*, and PRC2 generally, during EHT was not known. Our data suggested that *Jarid2* may be playing a role at the onset of EHT.

To see how the DEGs behaved throughout the whole model of EHT, we looked at mean expressions in both positive and negative LogFc. Interestingly, the 43 down-regulated genes increased in expression in the two early cell fate transitions (Endo to cHE and cHE to sHE) before becoming progressively down regulated in the following two transitions (proHSC to preI and preI to preII) (Figure 3.9C). KEGG pathway analysis of down-regulated genes showed enrichment for "Ras signaling", "PI3K-AKT signaling", "Rap1 signaling", and "Proteoglycans in cancer" (Figure 3.9E).

Means analysis for the 171 up-regulated genes showed no expression prior to this transition and then continued to increase expression in the next two transitions (proHSC to preI and preI to preII) (Figure 3.9D). KEGG pathway analysis of up regulated genes showed enrichment for "platelet activation", "transcriptional mis-regulation" and "hematopoietic cell lineage" (Figure 3.9F). Although enrichment for "platelet activation" may be explained by the presence of CFU-Cs within the proHSC sorting gate (See Appendix Figure

3.4), several genes within this pathway have previously been identified as genes of interest in EHT including *Nfe2* (Solaimani Kartalaei et al., 2015).

proHSC to preI

Differential gene expression of the proHSC to preI transition identified 652 DEGs, 282 up regulated and 370 down regulated in preI (FDR < 0.05, Figure 3.10A). Of the 48 DE TFs, 21 were up regulated and 27 down regulated (Figure 3.10B). Up-regulated TFs consisted of genes previously implicated in HSC biology, including *Fos*, *Ikzf1*, *Nfe2*, *Fosb*, *Myb*, *Bcl11a*, *Runx1t1* and *Hlf*, among others. Up regulated TFs not previously implicated in HSC biology include *Eya2*, *Nfatc2*, and *Per3*, among others. Interestingly, comparing the up-regulated TFs from this transition to the reprogramming TFs (Figure 3.5) reveals overlap for *Vdr*, *Egr1*, *Prdm16*, *Fosb*, *Hlf*, *Bcl11a*, *Runx1t1*, and *Meis1*. This suggests that the key TFs needed for reprogramming experiments are those that are differentially expressed during this transition. The 27 down-regulated TFs included Notch targets (*Hey1* and *Hey2*), Smads (*Smad1*) and EMT-related TFs (*Snai1*), showing the multitude of programs needing to be silenced as cells became more hematopoietic. Down-regulated TFs also included *Sox17*, which negatively regulates EHT (Lizama et al., 2015), and Lin28a, which negatively regulates hematopoietic output in the adult (Chaudhuri et al., 2012).

Mean expression analysis of the down-regulated genes showed upregulation in the preceding transitions but down-regulation in the following transition (Figure 3.10C). Not surprisingly these genes are enriched for pathways involved in focal adhesion and PI3K-AKT signaling both of which need to be down regulated for HSC emergence (Figure 3.10E). Mean expression

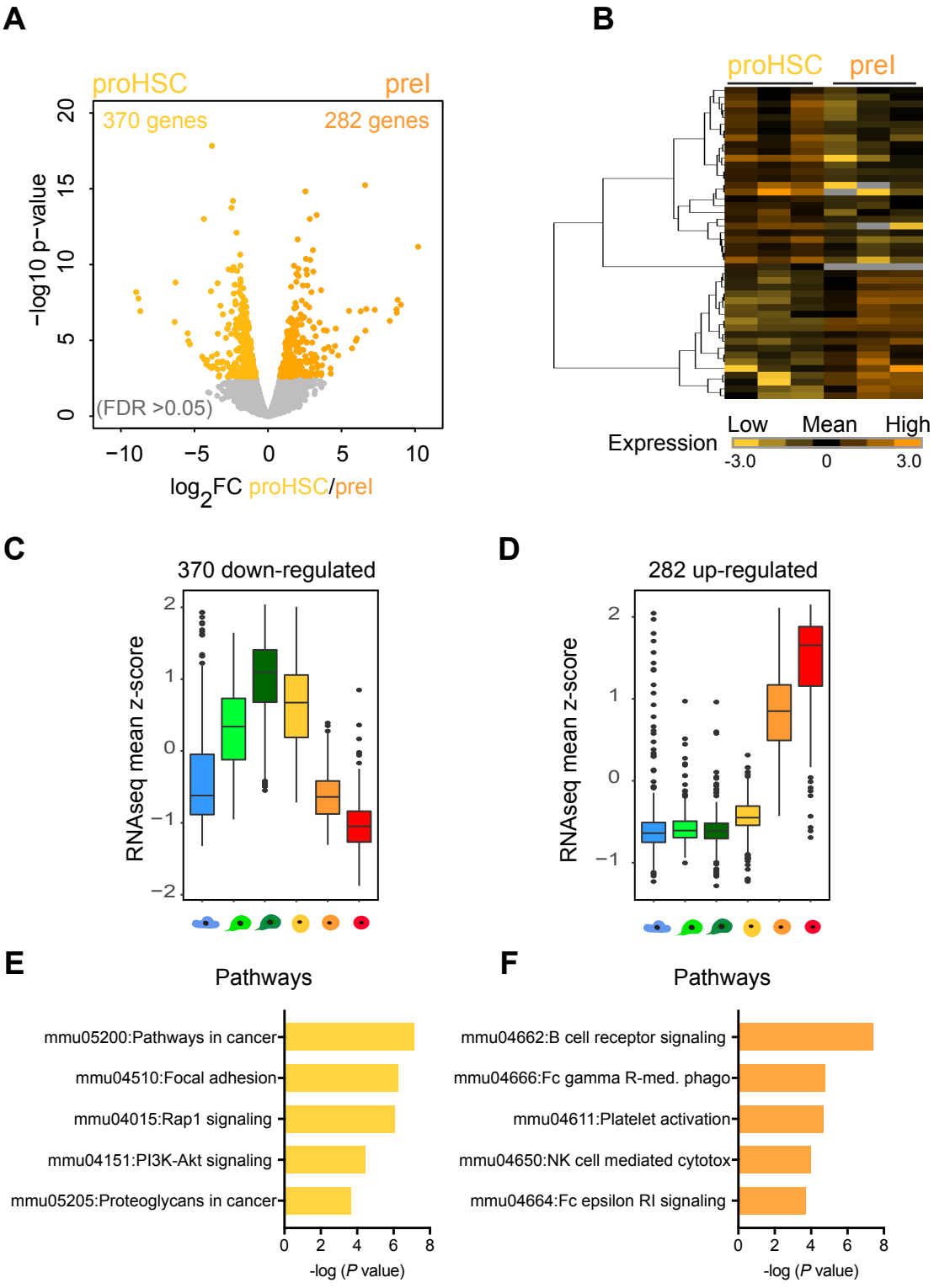


Figure 3.10: RNA-seq data analysis of proHSC to preI transition. (A) MA plot of differential expression for proHSC to preI with down-regulated genes in yellow (FDR <0.05) and up-regulated genes in orange (FDR < 0.05) (FDR > 0.05 in grey). (B) Heatmap of 48 differentially expressed TFs. (C) Box plots representing the normalized mean across all EHT populations for all 370 down-regualted genes. (D) Box plot representing normalized mean across all EHT populations for all 282 up-regulated genes. (E) Top 5 KEGG pathways enriched in down-regualted genes represented in a bar graph ranked according to -log Pvalue. (F) Top 5 KEGG pathways enriched in up-regulated genes represented in a bar graph ranked according to -log Pvalue.

analysis of the up-regulated genes shows that prior to this transition these genes were largely off (Figure 3.10D), which makes sense given the enrichment for hematopoietic-related KEGG pathways (B cell receptor signaling, platelet activation and NK cell mediated cytotoxicity) (Figure 3.10F).

preI to preII

Differential gene expression of the preI to preII transition identified 740 DEGs with 44 up regulated and 696 down regulated in preII (FDR < 0.05, Figure 3.11A). Within the 740 DEGs were 72 TFs (Figure 3.11B). Interestingly, only 2 of the 72 were up regulated, *Satb1* and *Coro1a*. *Satb1* is a chromatin organizer that regulates HSC self-renewal (Will et al., 2013) but can also direct HSCs toward lymphoid lineages (Satoh et al., 2013). Little is known about its role, if any, during EHT. In addition, very little is known about mouse *Coro1a* other than it is an actin binding protein and not previously known for binding DNA. Other coronin proteins have been implicated in nuclear actin dynamics and are therefore closely associated with DNA, which may explain why *Coro1a* is identified as a TF. In zebrafish, *Coro1a* is restricted to expression in myeloid cells (Song et al., 2004), which may explain why Coro1a is expressed in CD45+ preII.

Mean expression of the down-regulated genes shared a near identical pattern to the mean analysis of the previous transition (Figure 3.11C). Therefore, not surprisingly, the same pathways were enriched in the down-regulated genes, including Focal adhesion and PI3K-Akt (Figure 3.11E). Mean expression of the up-regulated genes showed genes are off in all populations except preI and preII, with preII harboring most of the expression (Figure 3.11D). Given the low number of up-regulated genes it was not surprising that pathway

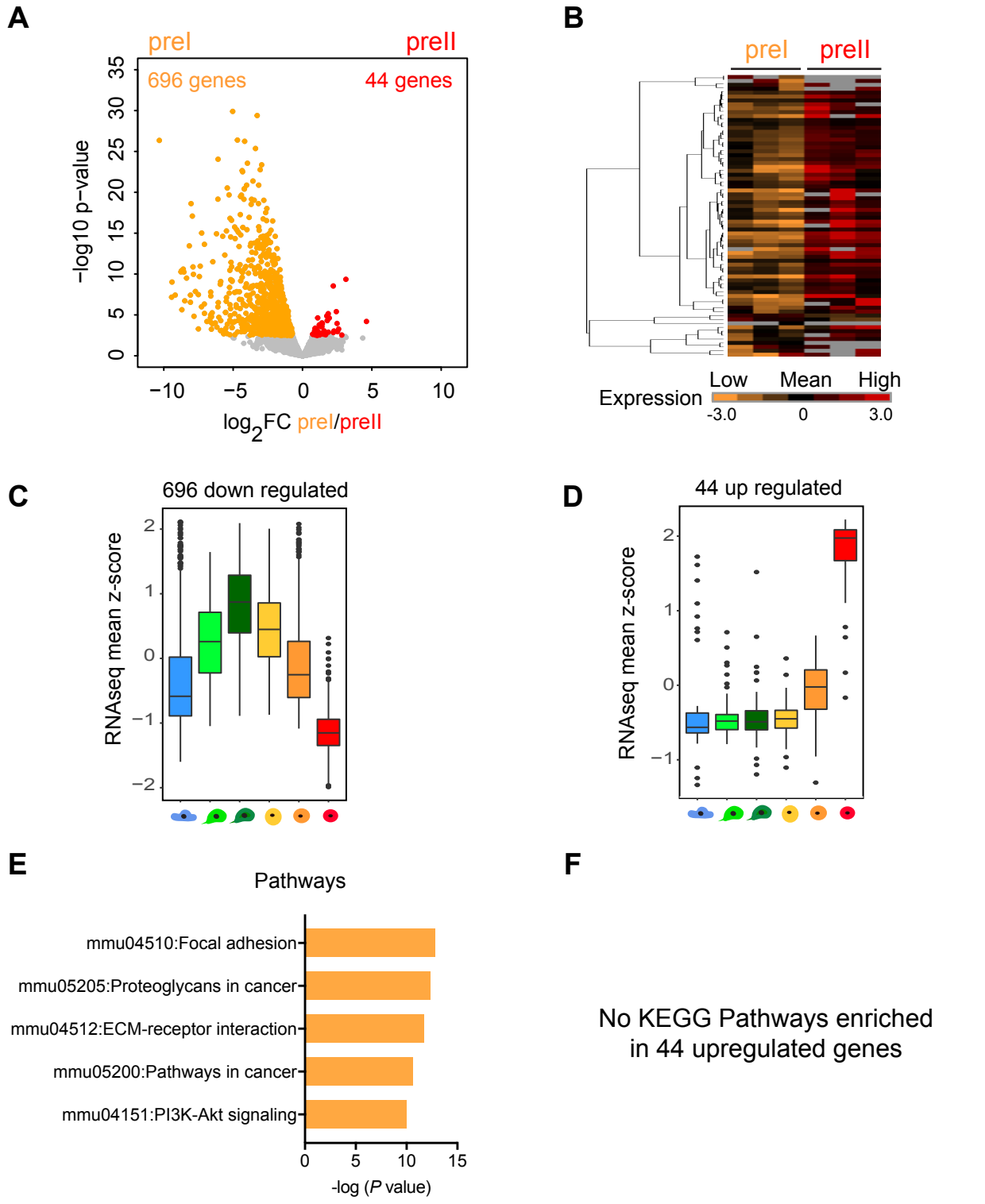


Figure 3.11: RNA-seq data analysis of preI to preII transition. (A) MA plot of differential expression for preI to preII with down-regulated genes in orange (FDR <0.05) and up-regulated genes in red (FDR < 0.05) (FDR > 0.05 in grey). (B) Heatmap of 72 differentially expressed TFs. (C) Box plots representing the normalized mean across all EHT populations for all 696 down-regualted genes. (D) Box plot representing normalized mean across all EHT populations for all 44 up-regulated genes. (E) Top 5 KEGG pathways enriched in down-regualted genes represented in a bar graph ranked according to -log Pvalue. (F) No KEGG pathways enriched in 44 up-regulated genes.

enrichment analysis yielded no significantly enriched pathways in the 44 genes.

3.4 RNA-seq: enriched pathways – Regulation of actin cytoskeleton organization

As mentioned in the Section 3.1, metacore enrichment analysis of clusters of differentially expressed genes across the model of EHT revealed "Specification" genes enriched for pathways involved in "cytoskeletal remodeling" (Figure 3.3E). The same analysis done on the 5 transitions discussed in Sections 3.2 and 3.3 revealed "Regulation of actin cytoskeleton" enrichment in the up-regulated genes 3 of 5 transitions (Appendix Figure 3.5). Moreover, in the global analysis of all 5,834 genes differentially expressed during EHT, we found enrichment for KEGG pathway "regulation of actin cytoskeleton organization" (mmu01648) (Figure 3.12A). Interestingly, 84 of 178 genes associated with this KEGG pathway were differentially expressed during EHT (Figure 3.12B). Previous reports discussed the regulation of cell adhesion and motility during EHT (Lie-A-Ling et al., 2014), but little is known about the regulation of the actin cytoskeleton. Given the absolute requirement for *Runx1* during EHT we examined +/-25kb around TSS of genes associated with the actin reorganization KEGG pathway for Runx1 binding, looking to infer a regulatory interaction. Interestingly, 60 out of 84 genes were bound by Runx1 in both 416B and HPC7 cells (Figure 3.12C) (Wilson et al., 2010, Schutte et al., 2016). This suggested a role for Runx1 in direct regulation of the actin cytoskeleton during EHT. Such a role would provide an additional, indirect mechanism by which Runx1 could affect cell fate, next to its role in the direct regulation of downstream TFs and the endothelial program (Swiers et

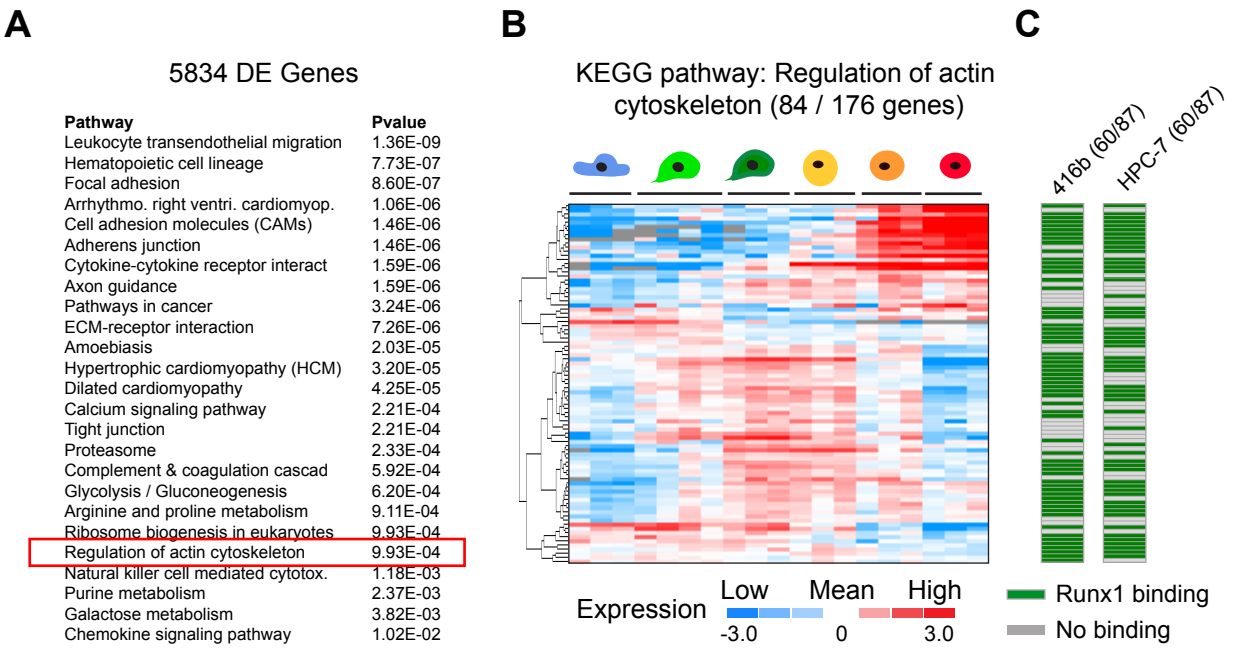


Figure 3.12: Runx1 and the regulation of actin cytoskeleton during EHT. (A) KEGG pathway enrichment for all 5,834 DEGs. (B) Heatmap of KEGG pathway "regulation of actin cytoskeleton" (mmu01648) showing the 87 of 176 differentially expressed genes within the pathway. (C) Heatmap indicating Runx1 binding with +/- 25kb of the TSS of the 87 DEGs in 416b (60 / 87) and HPC-7 (60 / 87) (Schuette et al., 2016; Wilson et al., 2010).

al., 2010, Swiers et al., 2013). Therefore, we set out to establish the relationship between Runx1 and actin reorganization during EHT.

Blocking actin reorganization blocks EHT

First, with little known about regulation of actin cytoskeleton reorganization during EHT, we established to what extent perturbation of actin reorganization affected EHT. Of note, changes in the structure, assembly and generation of actin have been reported to modulate signal transduction as well as gene transcription (Mammoto et al., 2009). We hypothesized that blocking actin assembly using CytochalasinD (Sigma) would block actin reorganization and thus affect signal transduction and gene transcription during EHT. *In vitro* differentiations of mESC cultured in the presence or absence of CytochalasinD (Cyto D) revealed inhibition of actin reorganization blocked EHT (Figure 3.13A-C). Flow cytometry analysis showed 38.6% CD45+ cells in DMSO-treated controls compared to 0.7% in CytoD treated (Figure 3.13C, Appendix Figure 3.6).

The presence of Runx1 protein despite continuous culture with CytoD meant that actin reorganization was downstream of Runx1 (Figure 3.13B). In support, expression of Runx1 targets *Myb* and *Pu.1* in CytoD-treated cells suggested establishment of a Runx1-dependent transcriptional program, despite an overall block in EHT (Figure 3.13D). Washout of CytoD partially released the block, further supporting the idea that actin reorganization was required for EHT (Figure 3.13B and 3.13C). Taken together, this suggests an active role for actin reorganization and actin-related pathways during EHT, rather than it being a consequence of hematopoietic differentiation. Moreover,

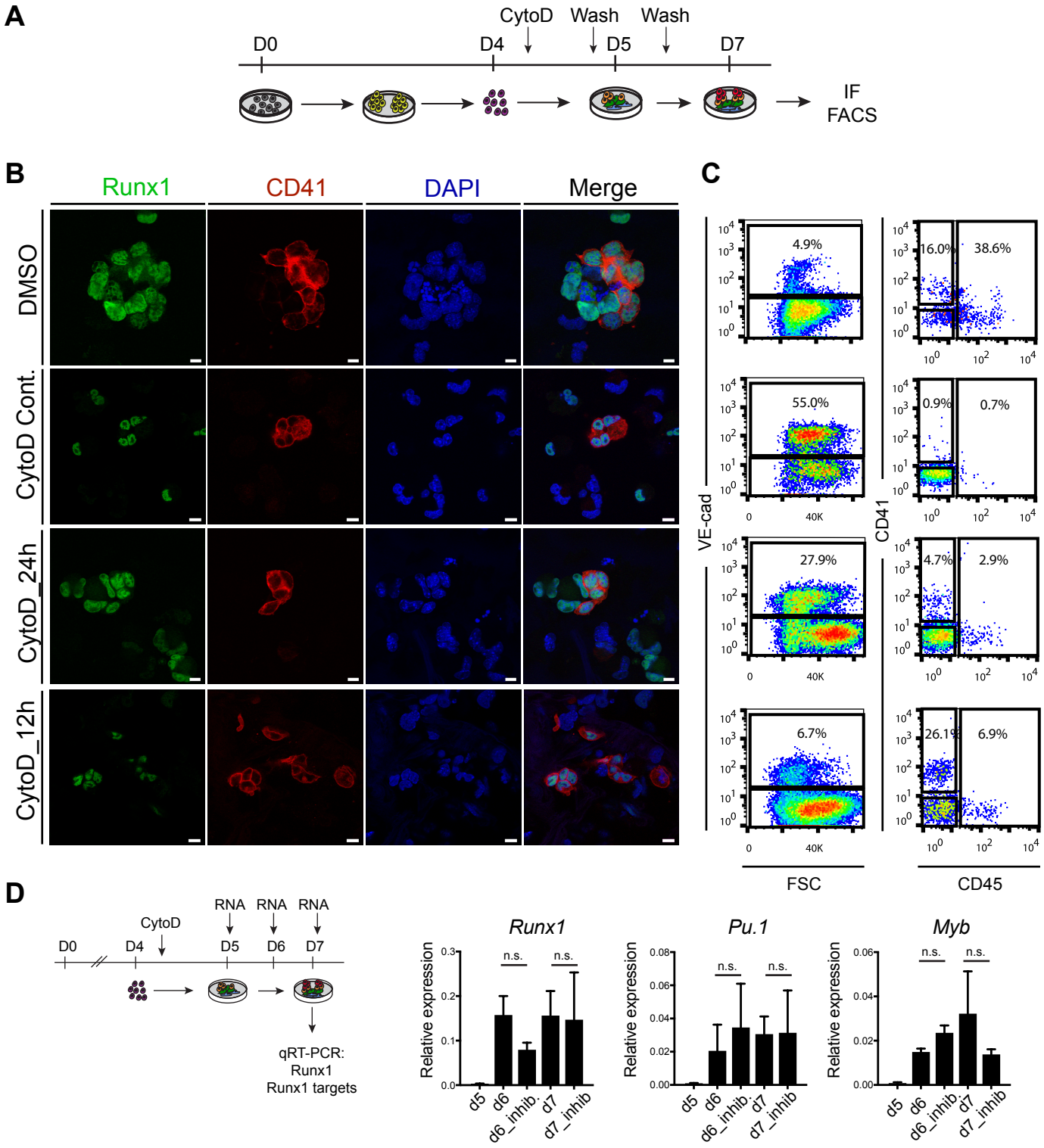


Figure 3.13: Blocking actin reorganization blocks EHT. (A) Schematic of *in vitro* mESC differentiation with addition of Cytochalasin D (CytoD) after plating of Flk1+ mesoderm followed either by continuous culture or washout after 12h or 24h of exposure. (B) Immunofluorescence images of Runx1 (green), CD41 (red) and DAPI (blue) for DMSO controls (untreated), continous treatment with CytoD (CytoD Cont.), 12h treatment followed by washout (CytoD 12h) and 24h treatment followed by washout (CytoD 24h), scale bar = 10um. (C) Representative sort plots of flow cytometry analysis at Day 7 for same conditions as in (B) where VEC+ are gated onto CD41/CD45 plots. Numbers indicate the average events within the sort gates from 2 independent experiments. (D) Schematic of in vitro differentiation with addition of CytoD followed by RNA isolation and qRT-PCR for Runx1, Sfpi1(Pu.1) and Myb. Histograms show average expression of 2 independent experiments (error bars = SD).

expression of Runx1 and its target genes *Myb* and *Pu.1*, suggested that the requirement for actin reorganization was downstream of *Runx1*.

Runx1-associated changes in actin organization and cell shape

To further elucidate the relationship between Runx1 and actin reorganization, we correlated appearance of Runx1 with changes in actin structures. Morphological changes associated with cell rounding are accompanied by a transition from actin stress fibers to cortical actin rings, which is a process observed in EMT (Haynes et al. 2011). With the similarities between EMT and EHT already highlighted (See 1.4), we hypothesized Runx1-mediated actin reorganization, potentially through regulation of an effector of actin reorganization. To determine if actin reorganization was dependent on *Runx1* expression, we performed a time course where we fixed Wt mESC differentiation cultures at 12h intervals. Reorganization of actin to the membrane in CD41+ cells occurred only after the onset of Runx1 at 24h after plating (arrowheads, Figure 3.14). In $Runx1^{-/-}$ mESC, CD41+ cells had pronounced actin stress fibers with no evidence of cortical actin filaments (Figured 3.15) ($Runx1^{-/-}$ are characterized in Appendix Figure 6.6). Overall, these data suggest that actin reorganization was dependent on Runx1.

Runx1 may regulate actin reorganization through direct activation of Gelsolin

Next, we sought a mechanism for Runx1 dependent actin reorganization. We reasoned that Runx1-meidated actin reorganization was accomplished through activation of a target gene(s) specifically involved in regulating actin dynamics. We looked back at the 84 DEGs of the actin reorganization pathway (Figure 3.12B, C) looking specifically for putative target genes of Runx1 (i.e., bound by Runx1) capable of enacting the necessary actin

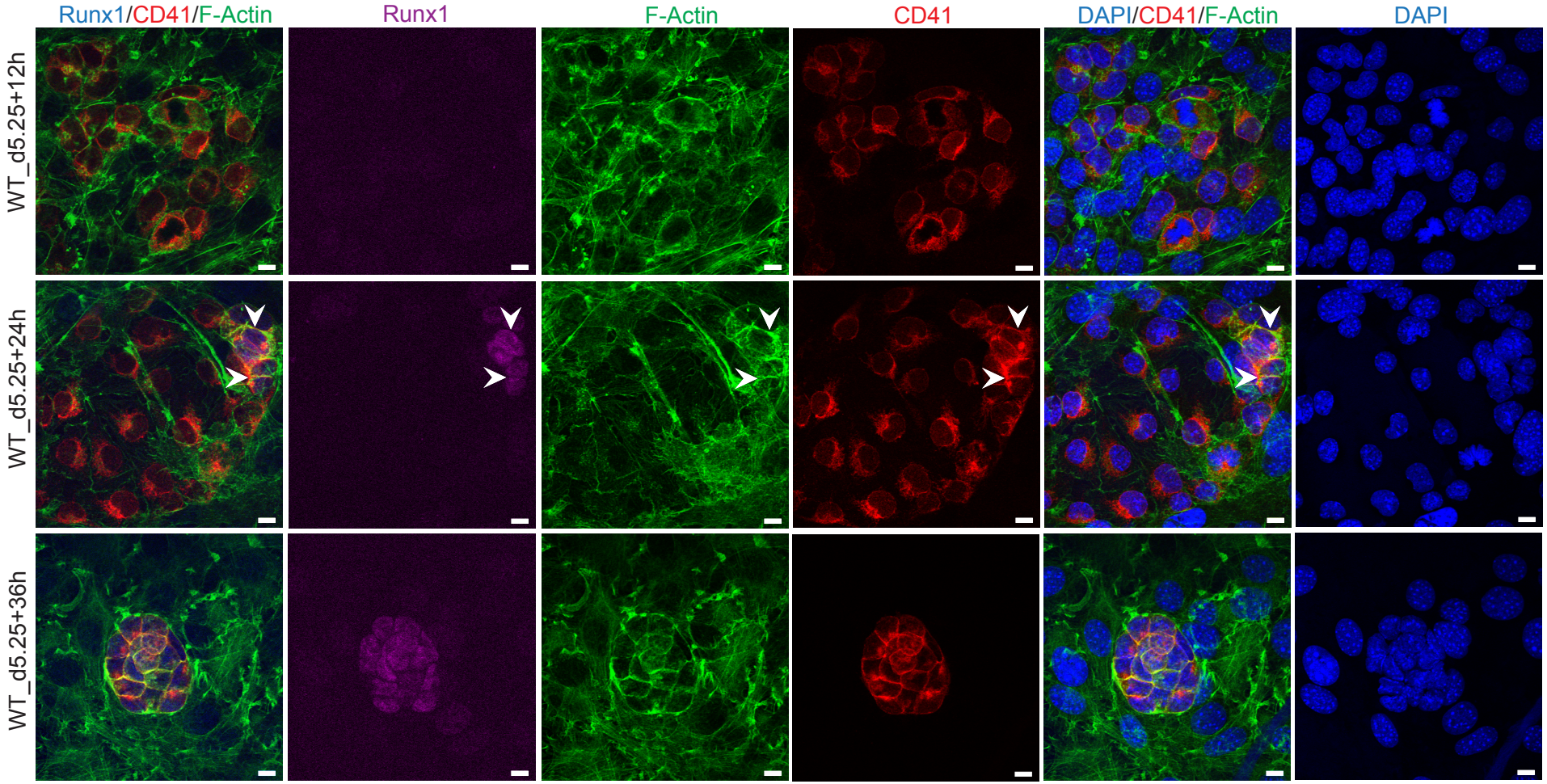


Figure 3.14. Time course of Runx1 expression and F-actin reorganization in Wt mESC differentiation. Top panel: Wt mESC fixed 12h after Flk1+ plating onto OP9s and stained with DAPI (blue), Phalloidin/F-actin (green), anti-CD41 (red), and anti-Runx1 (Magenta) antibodies. Middle panel: Wt mESC fixed 24h after Flk1+ plating onto OP9s and subjected to same staining as above. Arrowheads show F-actin localization with CD41 at the membrane only in cells that are expressing Runx1. Lower panel: Wt mESC fixed 36h after Flk1+ plating and subjected to same staining as above.

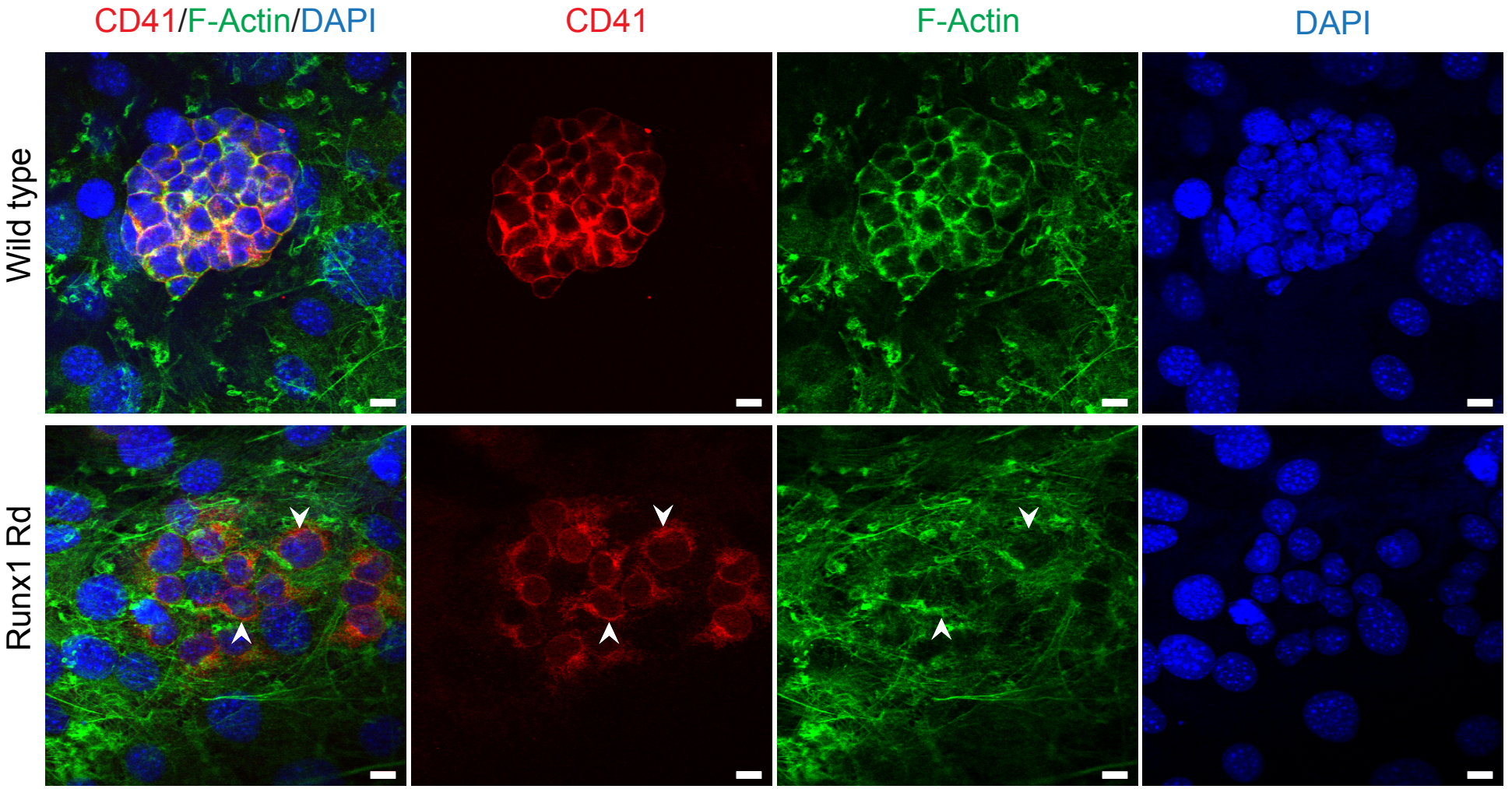


Figure 3.15. mESC differentiation of wild type (Wt) mESC and Runx1(-/-) (Runx1 Rd) mESC.
mESC cells were differentiated according to mESC differentiation in Figure 2.8 and were fixed at d7 of differentiation. Upper planel: Runx1 Rd stained with CD41(red), F-actin (green), and DAPI (blue). Lower panel: Wild type mESC (J1) stained with CD41(red), F-actin (green) and DAPI (blue). Scale bars are 10um. Images are representative of n=3 experiments.

reorganization. Gelsolin (Gsn) was bound by Runx1 and harbored the ability to reorganize actin. *Gelsolin,* a Ca2+ and polyphosphoinositide 4,5-biphosphate (PIP_2)-regulated actin filament severing and capping protein, had previously been implicated in actin remodeling (Sun et al., 1999). *Gelsolin* null fibroblasts exhibited a build up of actin stress fibers (Witke et al., 1995), which was consistent with an inability to remodel actin filaments by severing (Sun et al., 1999). This phenotype was similar to the stress fiber buildup in $Runx1^{-/-}$ mESC (Figure 3.15).

We hypothesized Runx1-dependent Gsn expression regulated actin reorganization during EHT. *Gsn* increased expression as cells became more hematopoietic (Figure 3.16A). A putative enhancer +4 kb downstream of the *Gsn* TSS was bound by Runx1 in ChIP-seq of 416b cells and contained 5 putative Runx sites (Figure 3.16B) (Schutte et al., 2016). Mutation of these sites resulted in loss of Luciferase activity (Figure 3.16C). This insinuates Runx1-dependent activation of *Gsn*. However, additional data was needed to demonstration actin reorganization during EHT was driven by Runx1-dependent *Gsn* expression. Adding to the complexity of how actin is regulated during EHT, single cell qRT-PCR from the same phenotypic populations as was done for the RNA-seq showed *Gsn* was not expressed in all hematopoietic cells (Appendix Figure 3.7A-B). This implies *Gsn* was either not required for EHT or other members of the Gelsolin superfamily were sufficient to drive requisite actin reorganization (Silacci et al., 2004). For example, *Supervilin* (*Svil*) was a putative target of Runx1 based on expression pattern and Runx1-ChIPseq (Figure Appendix 3.8) (Schutte et al., 2016). Svil and Gsn had overlapping expression patterns and therefore could function

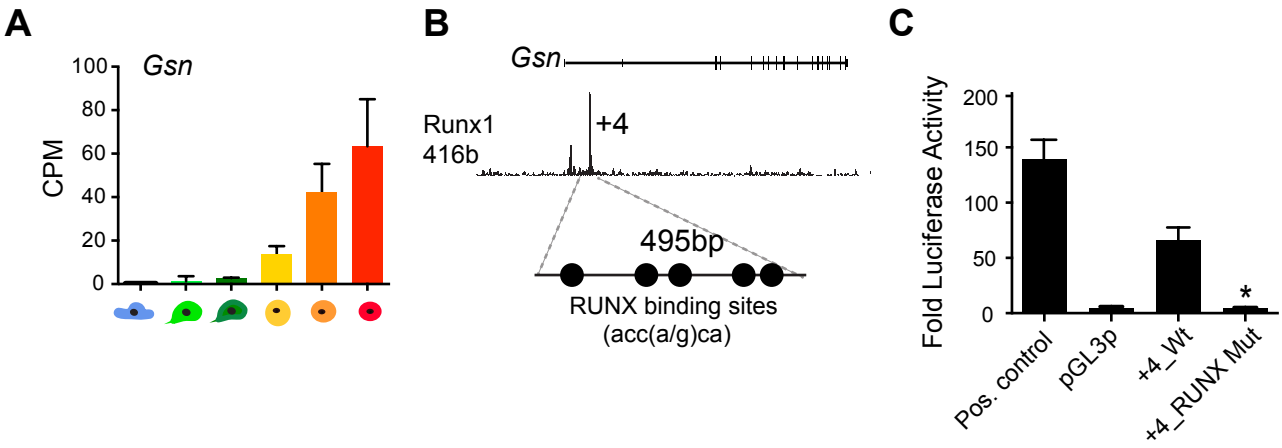


Figure 3.16: Runx1-dependent activation of Gelsolin. (A) *Gsn* CPM for each population during EHT. Error bars represent SD of either 3 or 4 biological replicates. (B) UCSC screen show of *Gsn* locus showing ChIP-seq data in 416b cells (Schuette et al., 2016). (C) Luciferase reporter assay for Gsn+4 cloned into pGL3 and used in a dual luciferase assay (Promega). Positive control is pGL3-Runx1+23. +4_RUNX mut destroyed RUNX sites by substituting purines for pyrimidines to create AC(T/A)(G/C)CA variants compared to ACCACA consensus motifs. Error bars represent SD of 3 biolgical replicates and * indicates pvalue = 0.02.

together, or in place of one another in reorganizing actin during EHT. In summary, actin reorganization during EHT is necessary, Runx1 dependent and may be the product of a Runx1-Gsn signaling cascade. However, in order to fully characterize the regulatory interaction between Runx1, Gsn and actin reorganization, additional genetic and chemical perturbation is necessary.

3.5 Conclusion

In this chapter, we described, validated, and analyzed an RNA-seq data set of the cellular intermediates of EHT. RNA-seq was performed on endothelium, HE, and HSPCs isolated from the dorsal aorta and vitelline and umbilical arteries of 23GFP transgenic E8.5, E9.5 and E10.5 embryos. Correlation between RNA-seq and qRT-PCR for several key genes known to be important for EHT showed the validity of the data. Differential expression analysis revealed stage-specific clusters of gene expression and identified 413 differentially expressed TFs. Pairwise comparisons of early cell fate transitions Endo to cHE and cHE to sHE revealed an underappreciated role for actin cytoskeletal regulation during EHT. Further investigation of cytoskeletal dynamics during *in vitro* EHT revealed actin cytoskeleton reorganization is required for EHT, correlates with Runx1 expression, and may (in part) result from a Runx1-dependent activation of Gelsolin. Finally, analysis of the remaining pairwise comparisons (sHE to proHSC, proHSC to preI and preI to preII) highlighted candidate regulators of EHT (*Bcl11a*, *Ikzf1*, *Ikzf2*, *Jarid2*) and the PI3K-AKT signaling pathway. Next in Chapter 4, we assessed chromatin accessibility using ATAC-seq in the same phenotypic populations as the RNA-seq and used motif enrichment, among other analyses, to shed light on the regulatory mechanisms that underlie the

transcriptional changes discovered in this chapter. In Chapter 5, we combined RNA-seq and ATAC-seq to generate putative gene regulatory interaction maps governing EHT.

Chapter 4: Chromatin profiling of EHT

In Chapter 4, the aim was to identify putative regulatory elements used by differentially expressed TFs (discovered in Chapter 3) to regulate transcriptional programs during EHT. Recent work using embryonic stem cell-derived HSPCs help shed light on the core regulator elements governing hematopoietic specification but did so using an *in vitro* model (Goode et al., 2016). Therefore, this work likely reflects yolk sac and not AGM-derived hematopoiesis (McGrath et al., 2015). Using standard chromatin and TF binding techniques to identify core regulatory elements regulating definitive hematopoiesis in the embryo is not feasible given the low cell numbers available. However, accessible regions of chromatin can be identified reliably even in small cell numbers by ATAC-seq (assay for transposable-accessible chromatin using sequencing) (Figure 4.1A; Buenestro et al., 2013). As regulatory elements bound by TFs are defined by their accessibility to DNaseI (or in this case ATAC-seq) (ENCODE project Consortium, 2012), ATAC-seq can be used to identify putative regulatory elements.

In Chapter 4, we mapped accessible regulatory elements by ATAC-seq in the cellular intermediates of in vivo EHT—the same phenotypic populations as for RNA-seq—with the aim of identifying putative regulatory elements and the mechanisms regulating accessibility (Figure 4.1B). Clustering analysis was used to measure correlation between different populations and to identify elements with shared patterns of accessibility. Differential accessibility analysis identified the most dynamic elements, and mechanisms underlying these dynamics were elucidated using GREAT and motif enrichment analysis. Ultimately, we focused on the mechanisms underlying chromatin accessibility for two key transitions: (1) cHE to sHE and sHE to proHSC.

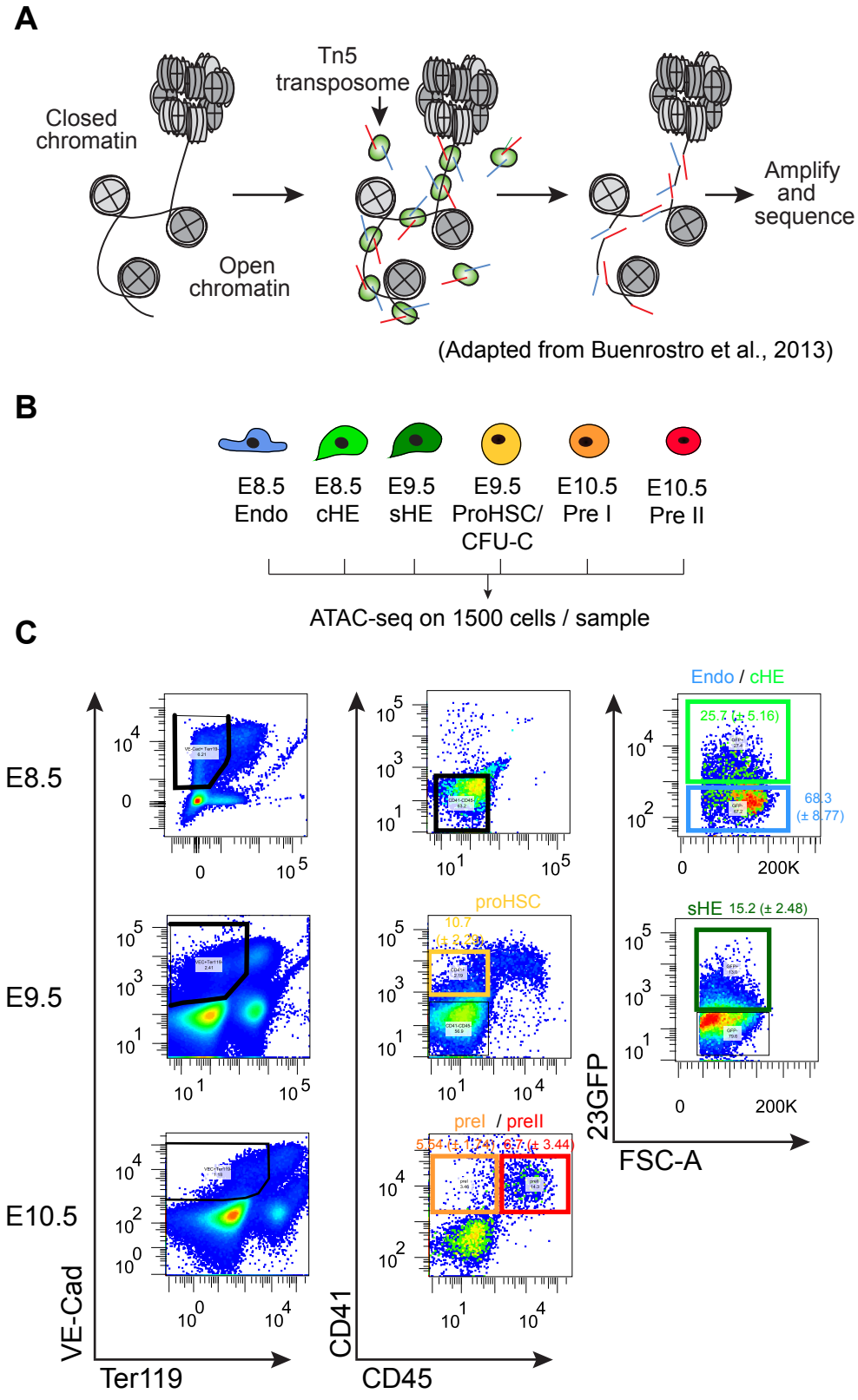


Figure 4.1: ATAC-seq schematic, experimental populations and sort gates. (A) Schematic of ATAC-seq technique (adapted from Buenrostro et al., 2013). (B) Experimental populations sorted for ATAC-seq (all n=4, except proHSC n=3) colored according to population. (C) Gating strategy for FACS isolation with scatter plots showing gates and distribution for all six EHT populations: Endothelium (Endo), competent hemogenic endothelium (cHE), pro-HSC/CFU-C (pro-HSC), specified hemogenic endothelium (sHE), pre-HSC Type I (preI) and pre-HSC Type II (preII).

4.1 Chromatin profiling of EHT

Sorting, sequencing, data processing and quality control

We used ATAC-seq to measure genome-wide chromatin accessibility in the same phenotypic populations that were analyzed for RNA-seq (Figure 4.1A-B and Figure 4.1A-C) (Buenestro et al., 2013). 1500 cells per replicate were sorted from 3-4 independent biological replicates (Figure 4.1B; Appendix Figure 4.1; Appendix Figure 4.2; Appendix Figure 4.3). Of note, samples collected for ATAC-seq were sorted at a different time than the RNA-seq samples. In addition, sorts were done using different equipment and different antibodies, which explains the difference in cell surface profiles between Figure 3.1B and Figure 4.1C. Despite the technical differences, population percentages remained similar (Figure 3.1B and Figure 4.1C). After library preparation, sequencing and processing, replicates were merged to give 20 – 40 million reads per sample (Figure 4.2A). As previously shown for this type of data (Buenestro et al., 2013), ATAC-seq coverage for all populations was concentrated in narrow regions with high signal throughout the genome (Appendix Figure 4.4).

In total, 49,520 elements were accessible in any one of the subsets. All samples showed expected periodicity of insert length (Figure 4.2B; Appendix Figure 4.5A) and were enriched for reads at transcription start sites (TSSs) (Figure 4.2C; Appendix Figure 4.5B). Interestingly, for all populations except cHE, reads were most abundant at promoter regions (Figure 4.2D). As the cells became more hematopoietic there was increased build up of reads at the promoters (Figure 4.2D) rather than at distal elements (±1kb from TSSs), acknowledging that overall number of ATAC elements decreased as well.

A

Stage	Embryos	Populations	Cells Sorted	Unique reads	Called peaks	Distal peaks
E8.5 (6 - 12sp)	59	Endo	1500 (n=4)	42514448	25477	13978
		cHE	1500 (n=4)	36097234	40020	27477
E9.5 (18 - 25sp)	112	sHE	1500 (n=4)	38008738	40886	27495
		pro-HSC	1500 (n=3)	33110892	18833	8746
E10.5 (31 - 36sp)	79	preI	1500 (n=4)	35147756	20333	9593
		preII	1500 (n=4)	23486702	9526	3068

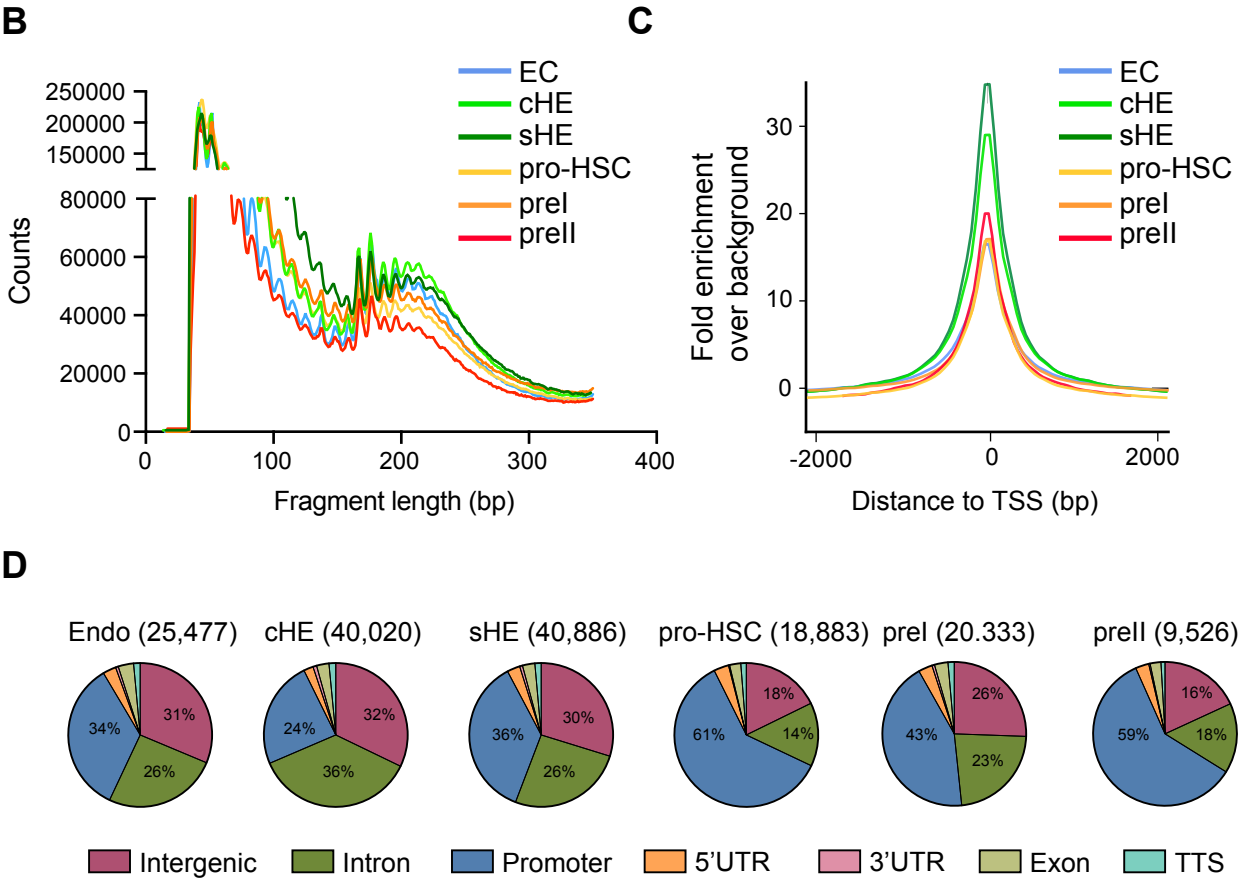


Figure 4.2: ATAC-seq read counts and quality control. (A) Summary table for experiments showing "Stage" of embryos (sp = somite pair); total number of "Embryos" used at each time point for sorting; "Populations"; number of "Cells Sorted" per experiment per population; "Unique reads" after fastq files from each *n* after processing; peaks were considered "Called peaks" if present in n=2 biological replicates; and "Distal peaks" greater than 1,000bp from TSS. (B) Representative insert size distribution showing expected peak at 200bp. (C) Representative plot of aggregate signal around transcription start sites (TSS) show enricment for reads located around TSS. (D) Pie chart showing breakdown of reads mapped to different of genomic features using HOMER's annotatepeaks.pl (5' untranslated region (5'UTR); 3' untranslated region (3'UTR); Transcription termiation sequence (TTS).

Given that distal elements are known hotspots for TF binding (Buecker et al., 2012), a reduction in the proportion of distal elements as the cells became more hematopoietic suggested that chromatin accessibility was centered around a core set of distal, hematopoietic-specific elements at the onset of EHT. As the cells became more hematopoietic, chromatin accessibility at these hematopoietic-associated elements made up a smaller percentage of overall elements.

Cell type specific accessibility

Each cell type contained uniquely accessible chromatin regions mapping to genes previously identified to be part of transcriptional regulation complexes important to EHT (Figure 4.3A-D) (Wilson et al., 2010; Schutte et al., 2016). For example, in the 200 kilobases (kb) of the Runx1 locus, we identified elements that were accessible in at least one cell type (Figure 4.3A), including known enhancers +3, +23, +110 and +204. The previously identified +23 enhancer was accessible in all cell types with increasing accessibility as the cells became more hematopoietic (Figure 4.3A). To further illustrate the robustness of the data, we looked at genomic loci of TFs known to play a part in regulating blood stem cell transcriptional networks: *Erg*, *Tal1* and *Meis1*. For *Erg*, chromatin accessibility during EHT was consistent with previous work showing +65, +75, and +85 cis-regulatory elements as having hematopoietic activity (Figure 4.3B, Schutte et al., 2016). Lack of chromatin accessibility at *Erg* +90 and +149 regions matched previous reports identifying these regions as “hematopoietic inactive” (Schutte et al., 2016). For *Tal1*, accessible chromatin at -4, +19 and +40 cis-regulatory elements matched their “hematopoietic active” statuses (Figure 4.3C) (Schutte et al., 2016). Finally,

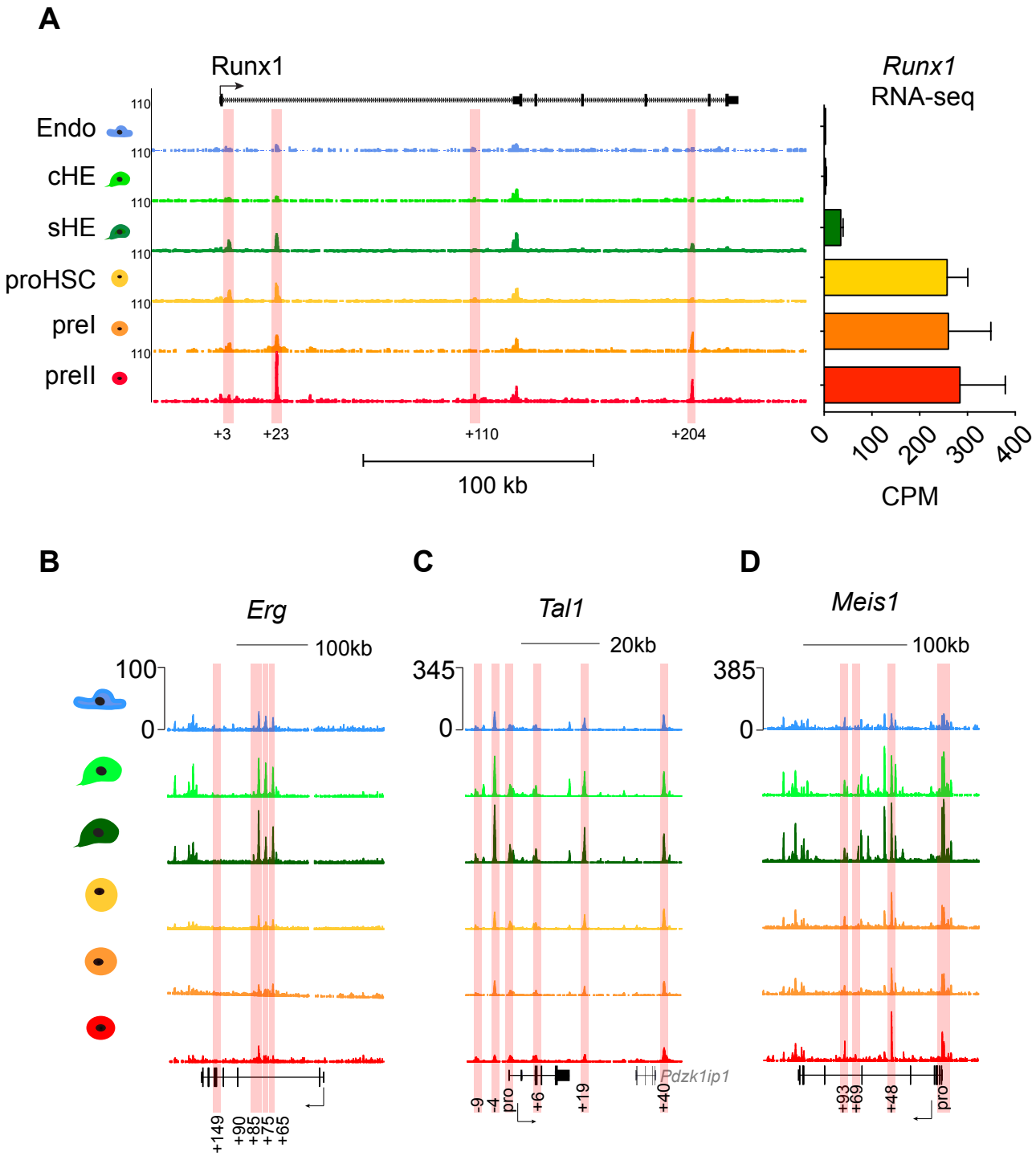


Figure 4.3: UCSC genome browser tracks for examples of cell type specific accessibility. (A) Mean ATAC-seq coverage at *Runx1* (Chr16:92,573,384 - 92,852,234) with RNA-seq gene expression in the same populations. Previously identified regulatory elements indicated by red bars. Scale of 0-110. (B) Mean ATAC-seq for *Erg* (Chr 16:95,533,533 - 95,836,042). Previously identified regulatory elements indicated by red bars. Scale of 0-100 for each population. (C) Mean ATAC-seq for *Tal1* (Chr4:114,718,005 - 114,773,073). Previously identified regulatory elements indicated by red bars. Scale of 0-345 for each population. (D) Mean ATAC-seq for *Meis1* (Chr11: 18,746,372 - 87,229,646). Previously identified regulatory elements indicated by red bars. Scale of 0-385 for each population.

for *Meis1*, only the +48 element had been previously characterized as having hematopoietic activity (Schutte et al., 2016), which matched our data showing increasing accessibility as the cells became more hematopoietic (Figure 4.3D). Interestingly, our data showed previously uncharacterized regions of accessible chromatin around *Meis1* +48 not previously reported in the literature (Figure 4.3D). This was just one example that highlighted the robustness of our data set and the power that it possessed to identify new cis-regulatory elements critical to EHT.

Principal component analysis

Looking at intensity of the signal at all 49,280 regions between all samples, pearson correlation coefficient showed robust clustering into distinct cell type specific groups (Figure 4.4A). Using this method, hematopoietic cells were most similar to each other, whereas the endothelial, cHE and sHE were part of a separate cluster (Figure 4.4A). A closer look using principal component analysis of all 49,280 elements showed chromatin accessibility was not as proficient at classifying cell types as RNA-seq (Figure 3.3A and Figure 4.4B). However, the data still clustered according to cell type with the exception of proHSC and preI overlap (Figure 4.4B).

Brief examination of the 20,730 elements in proHSC and 22,327 elements in preI showed overlap of 17,487 with 3,243 specific to proHSC and 4,840 specific to preI (Appendix Figure 4.6A). Characterization of each set of elements using GREAT showed enrichment of different pathways in each cell-type specific subset and showed enrichment for expected “ubiquitous” pathways in the intersecting elements (Appendix Figure 4.6C). A heatmap of the differentially accessible elements showed a subset of elements in preI_1

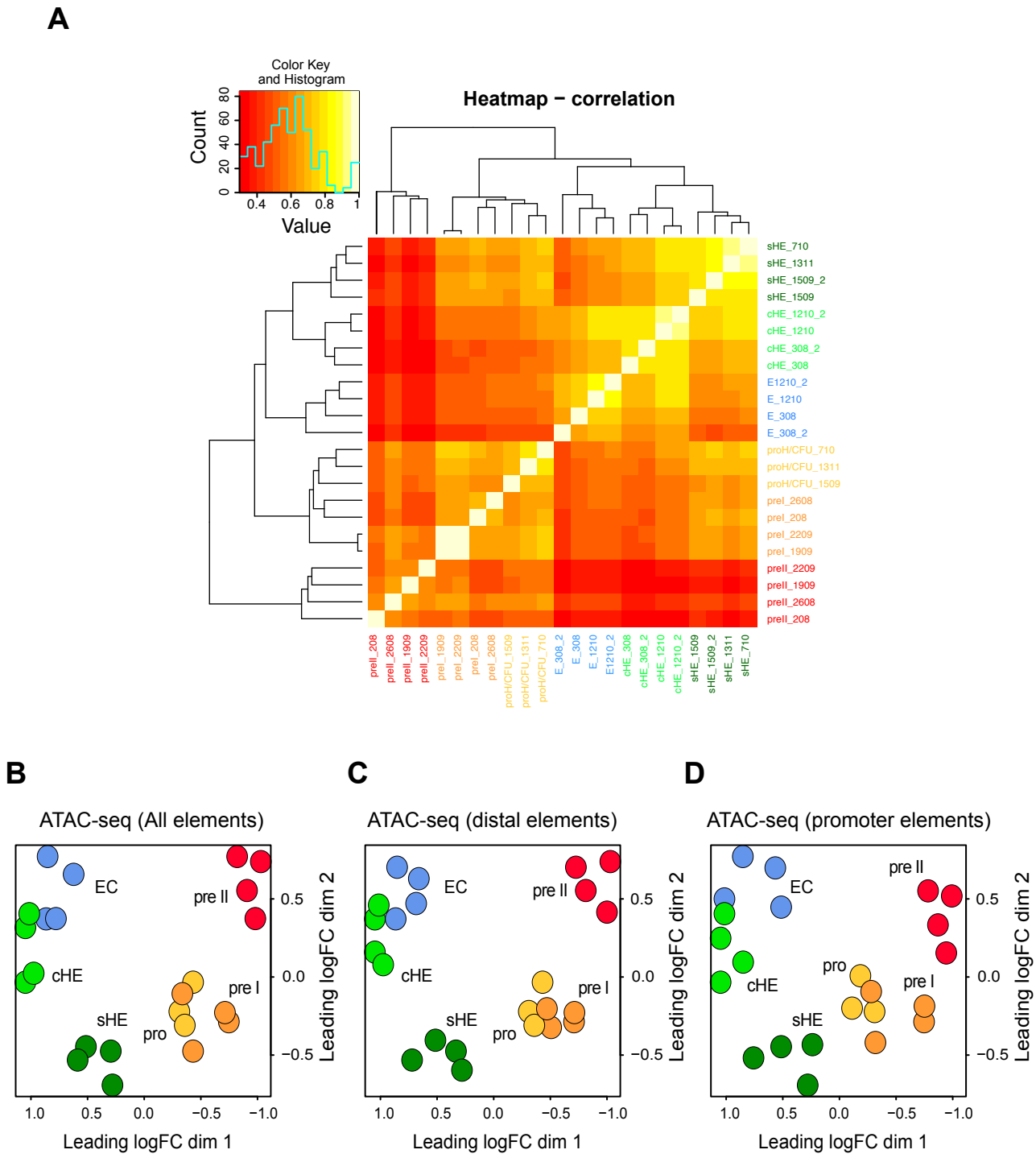


Figure 4.4: ATAC-seq correlation coefficient and principal component analysis. **(A)** Correlation of chromatin accessibility fo all 23 samples. Samples are colored according to populations and include experimental dates. **(B)** Principal component analysis (PCA) of all 49,520 elements from ATAC-seq samples from all biolgoical replicates for each population (all n=3, except cHE n=4) colored according to population. **(C)** PCA of distal elements (>1,000 bp from TSS) from ATAC-seq from all biological replicates for each population. **(D)** PCA of promoter elements (<1,000 bp from TSS)from ATAC-seq from all biological replicates for each population.

and preI_2 samples with read counts similar to the proHSC samples (Appendix Figure 4.6B). However, examination of these elements did not reveal any relevant biological patterns or other explanations for the clustering in the PCA. Thus, it is likely that despite clear differences in transcriptional profiles (Figure 3.3A), this transition did not require fundamental alterations to chromatin accessibility and therefore clustered together in the PCA. Interestingly, proHSC mature into preI and are differentiated by CD43 expression (proHSC CD43^{+}, preI CD43$^{-/low}$) (Rybtsov et al., 2014). Our data suggests maturation of proHSC into preI may begin already at E9.5.

Reports by other groups suggested that when regulatory elements are separated into gene promoters or distal elements (>1,000bp from a TSS), distal elements alone are better able to differentiate cells types (Coerces et al., 2016; Cauchy et al., 2015; Heinz et al., 2010). Comparing PCA plots of distal versus promoter elements, distal elements separate cell types more effectively (Figure 4.4C-D). When you compare clustering based on all elements versus distal elements, clustering appears roughly the same meaning that distal elements are responsible for the accessibility differences (Figure 4.4B-C). In summary, ATAC-seq PCA shows clustering according to developmental time and stage with biological variance driven by distal elements.

4.2 Differentially accessible chromatin: Global and Pairwise

After mapping and characterizing the accessible regulatory elements, we sought to identify which of these elements undergo dynamic changes in accessibility during EHT. Given that regulation of transcriptional programs is controlled, at least in part, through the action of TFs binding to distal

regulatory elements (Buecker et al., 2012), identification of which TFs bound to these elements would help determine the transcriptional program at work during EHT. In order to do this, we first looked globally at differential accessibility in all populations using K-means clustering to separate elements based on shared patterns of accessibility and then investigated those elements for motif enrichment. Following on, we looked specifically at the dynamics in each of the five transitions within our model. Given the importance of early cell fate transitions, cHE to sHE and sHE to proHSC, these transitions were investigated in greater detail in the Section 4.3 and 4.4.

Global differential accessibility

Differential accessibility across all six populations identified 15,295 elements that were largely gene distal with few elements being promoter-proximal (Figure 4.5A). Elements greater than 1,000bp (>10^3bp) from the TSS were considered distal. In all 49,280 elements, ~72% of elements are distal compared to 96% in the 15,295 differentially accessible elements (Figure 4.5A). Thus, changes in chromatin accessibility during EHT occur at distal regulatory elements. This is confirmed by PCA analysis showing the biological variation is driven largely by distal elements (Figure 4.4B-D).

K-means clustering grouped differentially accessible elements according to accessibility patterns. Cumulative distribution functions plots identified an optimal cluster of k = 6 (Appendix Figure 4.7). Clustering revealed cell type (developmental stage) specific elements (Figure 4.5B). In order to gain a mechanistic insight into the formation of the stage specific elements, each cluster was used as input for GREAT (Genomic Regions Enrichment of Annotations Tool) annotation of nearby genes (McLean et al., 2010) and motif

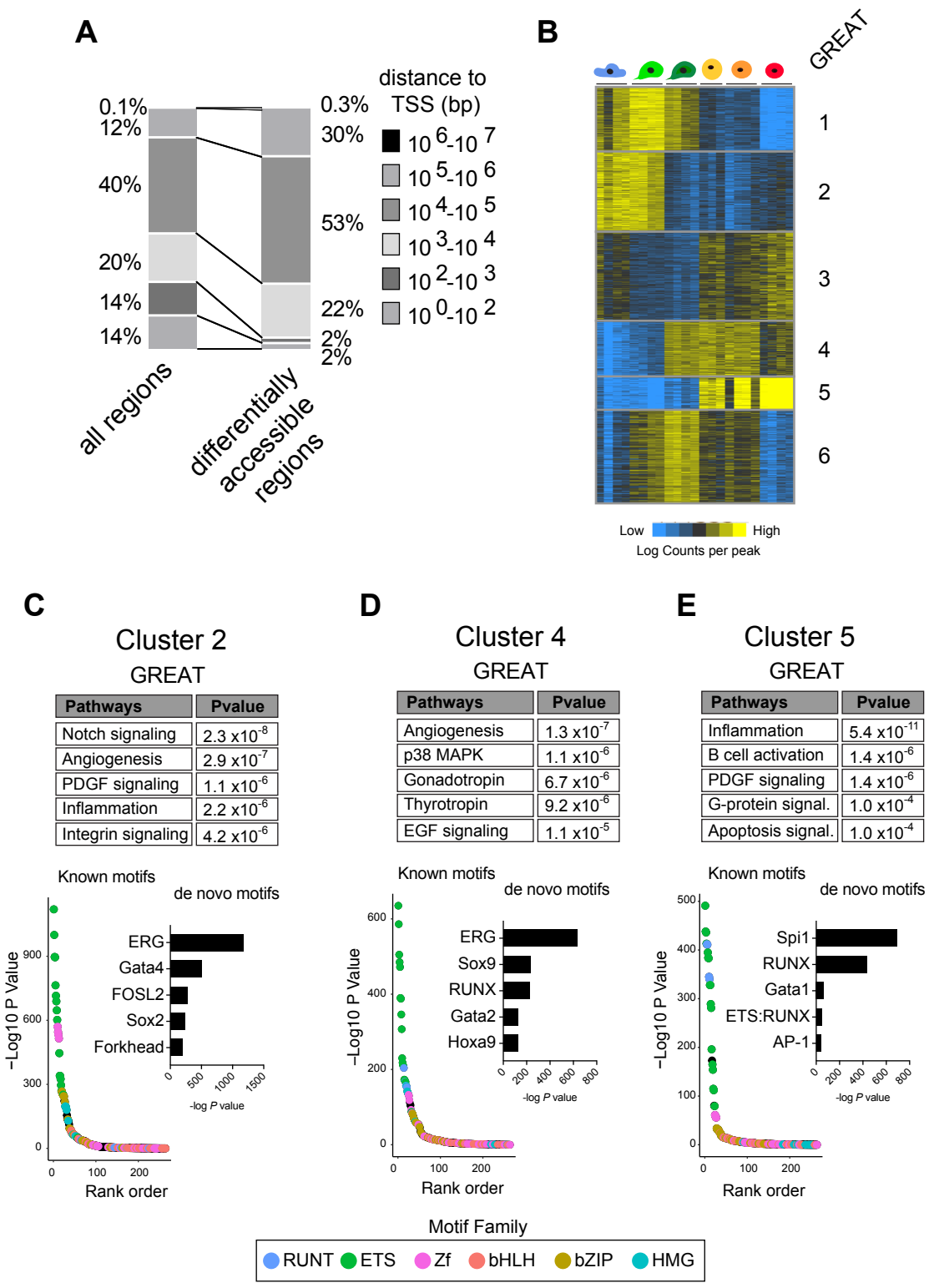


Figure 4.5: Analysis of differential accessibility dynamics. (A) Distance to closest transcription start stie (TSSs) of all 49,520 accessible elements and 15,295 differentially accessible elements. (B) Heatmap of k-means clustering of all 15,295 differentially accessible peaks. (C) Analysis of Cluster 2 showing Top 5 mSigDB pathways from GREAT analysis, known motif enrichment showing enriched motif familiies, and top 5 de novo enriched motifs. (D) Analysis of Cluster 4 showing Top 5 mSigDB pathways from GREAT analysis, known motif enrichment showing enriched motif familiies, and top 5 de novo enriched motifs. (E) Analysis of Cluster 5 showing Top 5 mSigDB pathways from GREAT analysis, known motif enrichment showing enriched motif famil lies, and top 5 de novo enriched motifs.

enrichment of transcription factor motifs using HOMER motif analysis software (Heinz et al., 2010). For further investigation, we selected clusters that represented different stages of EHT (Figure 1.4, Figure 4.5C-F): cluster 2 (endothelial), cluster 4 (specification) and cluster 5 (hematopoietic commitment).

Cluster 2 elements were most accessible in Endo and cHE and lost accessibility as cells transitioned into sHE (Figure 4.5C). GREAT analysis of the 2,941 elements revealed enrichment for “Notch signalling” and “Angiogenesis” (Figure 4.5C). This was consistent with previous studies showing the importance of proper regulation of Notch signalling (Souhihol et al., 2016; Zhang et al., 2015) and arterial-venous specification (Clements and Traver, 2013) during EHT. Cluster 2 motif enrichment revealed the top 5 most enriched known motifs are part of the ETS family (Figure 4.5C) (See Section 2.4; used 250bp centered on ATAC elements). De novo motif enrichment again highlighted the ETS family with ERG as the most enriched (Figure 4.5C). However, de novo enrichment also revealed an increased role for GATA motifs (Zinc Finger family) and the FOSL2 motif (Basic Leucine Zipper family Family (bZip) that also includes AP-1 and FOS) (Figure 4.5C).

Cluster 4 elements were most accessible in sHE, proHSC and preI populations with many of the elements becoming accessible for the first time in sHE (Figure 4.5B). GREAT analysis of the 2,251 elements showed enrichment for “Angiogenesis”, “MAPK”, “Gonadotropin”, “Thyrotropin” and “EGF signaling” (Figure 4.5D). Motif enrichment revealed ERG, Sox9, RUNX, Gata2, and Hoxa9 as the top 5 de novo motifs (Figure 4.5D).

Cluster 5, which consisted of 1,204 elements with greatest accessibility in hematopoietic cells, showed GREAT enrichment for “Inflammation”, “B cell activation” and “PDGF signaling” (Figure 4.5E). Reassuringly, “Inflammation” and “B cell activation” were also pathways enriched in the hematopoietic clusters in our RNA-seq data (Figure 3.3C, 3.3E). Interestingly, recent work showed PDGFra-mediated signaling is required for the initiation of the hematopoietic program and specification of HSCs (Damm et al., 2017). Our work suggests that this mechanism may be conserved in mice. To identify potential drivers of the increase in accessibility, we looked at the differentially accessible elements for enrichment of TF motifs. Known and de novo motif enrichment each reported Sfpi1 and Runx1 as the most highly enriched motifs (Figure 4.5E).

Overall, k-means clustering of differentially accessible elements revealed cell type specific clusters enriched for motifs known to regulate those stages of EHT. Within these clusters, we identified motifs for TFs not previously implicated in EHT. Next, we focused on pairwise differential accessibility.

Pairwise differential accessibility

Global differential accessibility analysis identified unique subsets of differentially accessible elements responsible for different stages of EHT. In order to drill down further into the underlying mechanisms driving chromatin dynamics during EHT, we focused on characterization of elements differentially accessible during stepwise acquisition of the hematopoietic program. Differentially accessibility analysis using DiffBind (Stark et al., 2014) separated elements into +LogFC and –LogFC. Further analysis revealed subgroups within the +/- LogFC elements. –LogFC elements contained

elements that decreased accessibility but remained open (“Decreasing”), and elements that lost accessibility all together (“Newly Closed” or “NC”) (Figure 4.8A). Conversely, +LogFC elements included elements previously accessible that became more accessible (“Increasing”), and elements accessible only in that later population (“Newly Open” or “NO”) (Figure 4.8C). Of note, ATAC-seq was performed on bulk populations, thus changes in accessibility may reflect either an increase cell numbers with accessibility at a particular element or it may reflect increasing accessibility at an element within individual cells. With these considerations in mind, we reasoned that a change resulting in a new element (NO) or the complete loss of an element (NC) had greater significance than an increase or decrease. Similar to the approach we used in the global differential accessibility analysis, we hypothesized that TF motif binding site analysis could identify TFs responsible for these key accessibility changes.

Pairwise differential accessibility analysis revealed greater overall accessibility in the first three transitions when only total number of DAEs were considered (Figure 4.6A-C). The portion of –LogFC to +LogFC favored the +logFC in the first transition (64% +LogFC in Endo to cHE) and -LogFC in the second and third transitions (59% and 61% in cHE to sHE and sHE to proHSC, respectively) (Figure 4.6A-C). The final two transitions experienced similar dynamics regarding percentage of +/- LogFC subsets. For proHSC to preI, 59% were –LogFC (Figure 4.6D). For preI to preII, 59% were +LogFC (Figure 4.6E). Based solely on percentage + or – LogFC, no clear pattern emerged. Thus, we looked more closely at the increasing, NO, decreasing and NC within the LogFC populations.

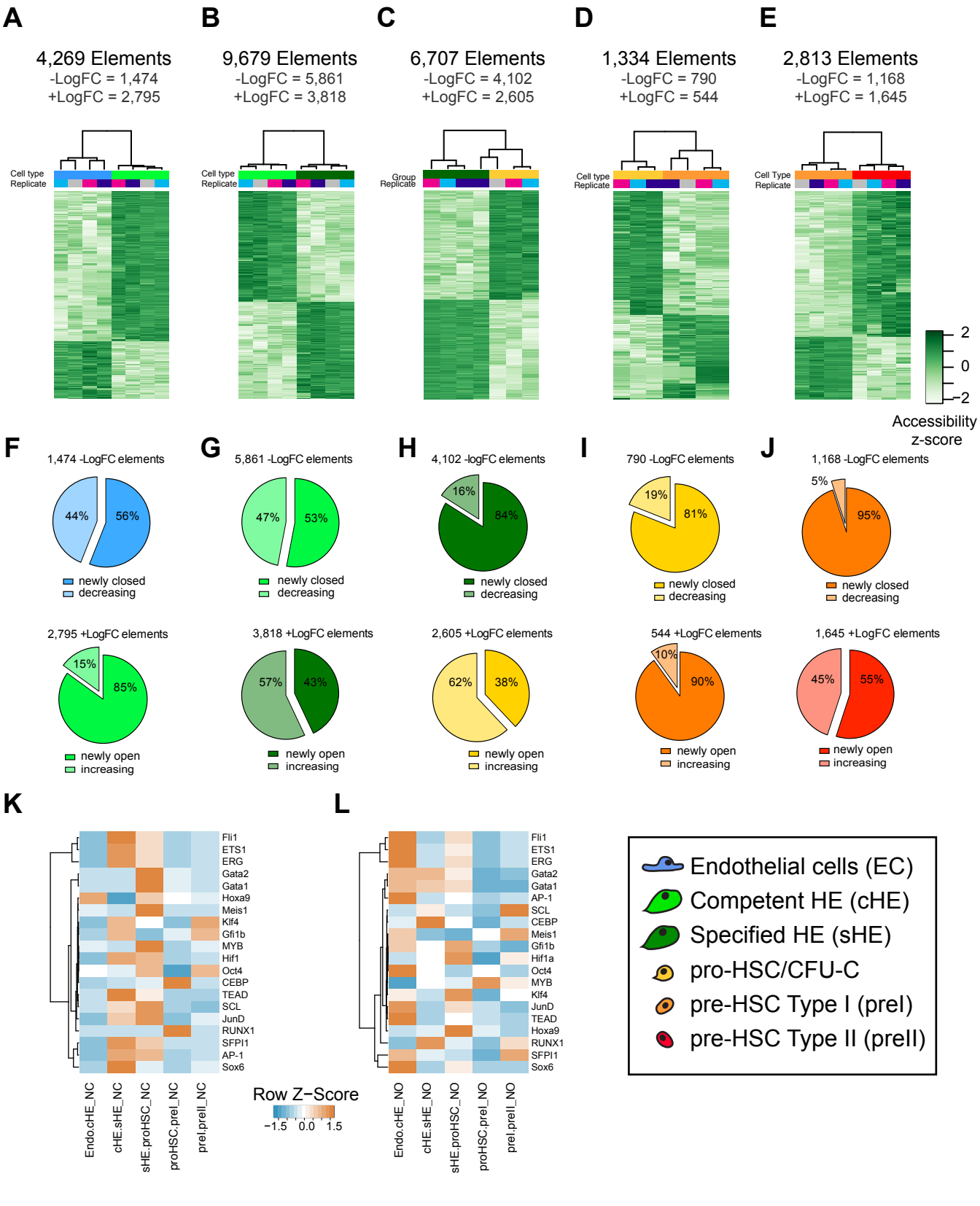


Figure 4.6: Pairwise comparisons for all 5 transitions. (A-E) Heatmap of differentially accessible elements for each pairwise comparisons with increased accessibility in darker green. (F-J) Pie charts of chromatin dynamics for each pairwise comparison. Each comparison included analysis of -LogFC (decreasing in lighter color and newly closed in darker color) and +LogFC (increasing in lighter color and newly open in darker color). (K) TF motifs enriched in newly closed elements in each pair wise comparison. Motifs were selected based on TFs known to be important during EHT. (L) TF motifs enriched in newly open elements in each pairwise comparisons. Motfis were selected based on TFs known to be important during EHT.

Interestingly, an initial wave of Newly Open elements marked the acquisition of the cHE phenotype from Endo. Of the 2,735 +LogFC elements, 85% were accessible for the first time in cHE (Figure 4.6F). This burst of NO elements included loci enriched with motifs for Scl, Cebp and Runx1 (Figure 4.6L).

As discussed previously, the cHE to sHE and sHE to proHSC transitions were key cell fate transitions whose chromatin accessibility dynamics is described in much greater detail in the following sections. In brief, for cHE to sHE DAE, there was no trend toward either NC or NO (Figure 4.6G). In contrast, for sHE to proHSC, 84% of the 4,102 -LogFC elements were NC (Figure 4.6H), suggesting active negative modulation of these elements by some factor. In addition, 38% of the 2,605 +LogFC elements were NO (Figure 4.6H), hinting that the positive changes in accessibility and therefore positive changes in expression came primarily from elements already being positively modulated in sHE.

Despite only 1,334 elements being differentially accessible in the proHSC to preI transition (FDR < 0.05, Figure 4.6D), 85% underwent dramatic changes in accessibility. 81% (640 / 790) of –LogFC elements were NC (Figure 4.6I) and enriched for Cebp and Runx1 motifs (Figure 4.6K). 90% (489 / 544) of +LogFC elements were NO (Figure 4.6I) and enriched primarily for Myb motifs (Figure 4.6L). This suggested that this transition marked a key cell fate decision during EHT, at least at the level of chromatin dynamics.

In the next transition, preI to preII –LogFC were 95% NC (Figure 4.6J). Not surprisingly, the remaining 58 elements were not enriched for any TF motifs or associated with any pathways. However, the 1,110 NC elements were enriched in motifs for Klf4, Gfi1b and Oct4 (Figure 4.6K). 55% of +LogFC

elements were NO (Figure 4.6J) and enriched for Scl, Meis1, Hif1, Runx1, and Sfpi1 (Figure 4.6L). Overall, despite similar percentages of Increasing and NO in the +LogFC elements, the 95% of –LogFC elements classified as NC suggested a dynamic chromatin accessibility landscape during this transition despite both cell types being hematopoietic. Further study of the 1,110 NC elements may identify putative regulators driving this transition.

Overall, for pairwise comparisons that included hematopoietic populations, the –LogFC consisted of a larger percentage of NC elements. This reflected, in part, the active down-regulation of the endothelial program. In contrast, the initial burst followed by fluctuating percentages of NO elements likely reflected the heterogeneity within the hematopoietic populations. For example, with only 1 dHSC present at E10.5, the remaining CD45+ cells likely consisted of other HSPC-fated cell types (Kumularová et al., 2002).

In summary, global analysis identified stage specific DAE representing the different phases of EHT. Analysis of pairwise comparisons provided insights into potential mechanisms underlying chromatin accessibility dynamics during each transition. While each transition had biologically meaningful reasons warranting further investigation, we focused on two transitions in the next two sections: cHE to sHE and sHE to proHSC transitions.

4.3 Chromatin accessibility in cHE and sHE

Recent work from our lab identified cHE as capable of responding to blood inducing signals, and sHE as the population in which cells begin to commit to the hematopoietic lineage (Swiers et al., 2013). However, it remained unclear how hematopoietic commitment was regulated at the molecular level. Working towards a greater understanding of the mechanisms at work during this

transition, differential expression analysis in Chapter 3 identified differentially expressed TFs. Knowing that TFs bind to TSSs and regulatory elements to regulate transcriptional programs, we used pairwise differential accessibility to identify putative regulatory elements and determine enriched TF motifs used by the differentially expressed TFs to govern the transcriptional program regulating this transition.

Differential accessibility

First, we identified and characterized the differentially accessible chromatin in cHE to sHE transition for GREAT and motif enrichment. Comparison of ATAC-seq profiles for cHE and sHE revealed 9,679 differentially accessible elements with 5,861 elements losing accessibility (-logFC) and 3,818 elements becoming more accessible (+logFC) (FDR < 0.05, Figure 4.7A). Elements losing accessibility were enriched near genes involved in Focal adhesion, VEGF/Angiogenesis and HIF2-alpha TF networks, among other pathways (Figure 4.7B). Known motif enrichment plotted according to motif family revealed strong enrichment for the ETS family motifs (Figure 4.7D). De novo motif analysis reported Erg as most enriched (Figure 4.7E). Not surprisingly, Gata4 motifs reiterated the role of the GATA factors during early EHT (Figure 4.7D). Interestingly, de novo enrichment revealed motifs for TFs yet to be investigated during EHT, including Atf3 and Sox6 (Figure 4.7E).

+LogFC elements were enriched for TGF-β-related signaling including TGF-β induced SMAD transcriptional activity (Figure 4.7C). Independent studies showed TGF-β (Monteiro et al., 2016) and Smad1/5 (Zhang et al., 2014) were both critical for HSC emergence. Of note, elements with increasing accessibility that harbor SMAD motifs included the Runx1 +23 enhancer,

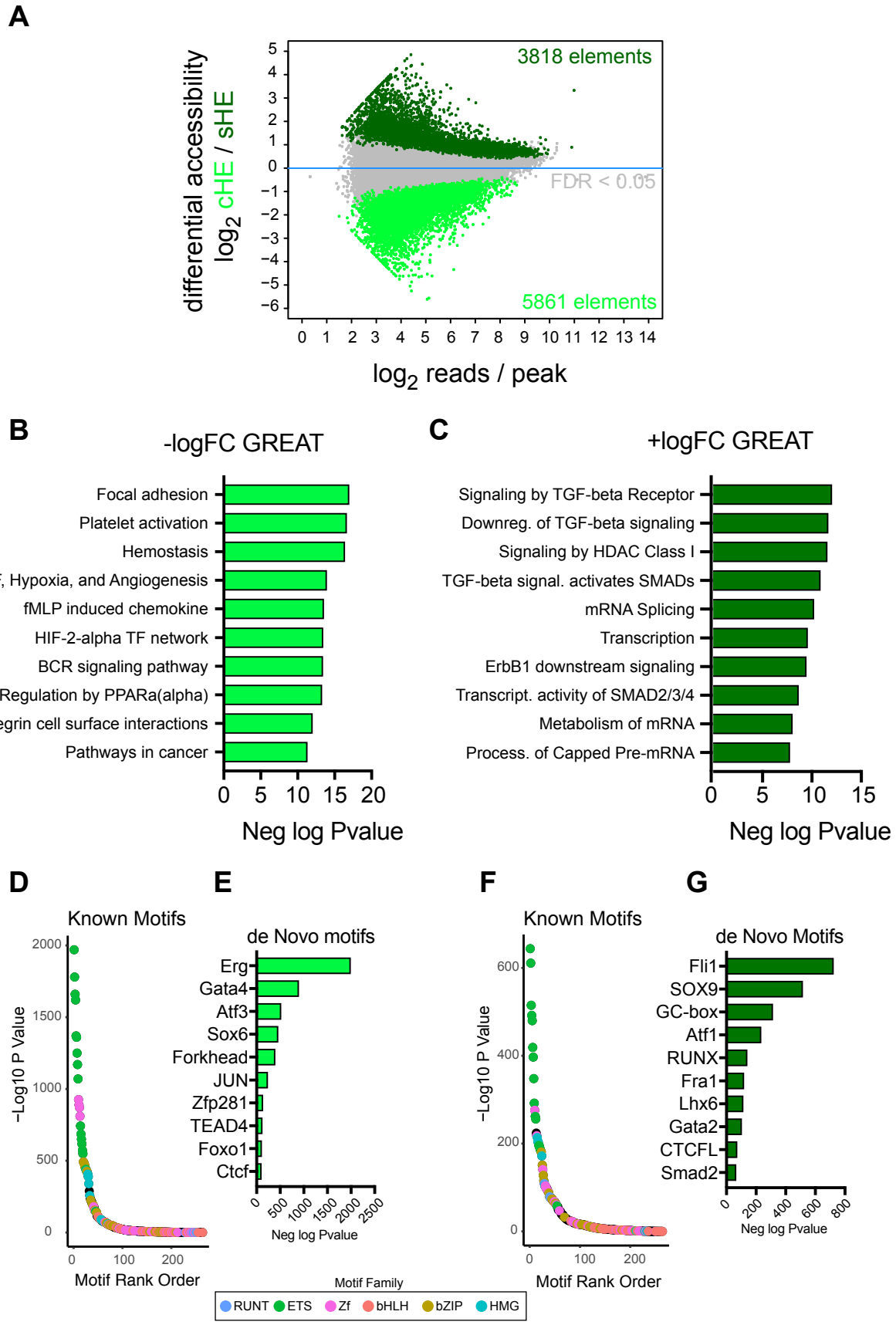


Figure 4.7: Chromatin accessibility during cHE to sHE transition. (A) MA Plot of differential accessibility for ATAC-seq counts in cHE and sHE. (B) Top 10 mSigDB pathways from GREAT analysis of -LogFC elements. (C) Top 10 mSigDB pathways from GREAT analysis of +LogFC elements. (D) Known motif analysis showed motif family enrichment in -LogFC. (E) de Novo motif enrichment of top motifs in -LogFC. (F) Known motif analysis showed motif family enrichment in +LogFC. (G) de Novo motif enrichment of top motifs in +LogFC.

which further supports a link between TGF-beta signaling and Runx1 transcriptional regulation. Known motif enrichment analysis of +logFC elements revealed most significant enrichment for ETS family motifs with Fli1 being the most enriched in both known and de novo enrichments (Figure 4.7F and 4.7G). Fli1 is part of the heptad of TFs that regulate blood cell TF networks (Wilson et al. 2010). Fli1 has been implicated in NFK-B signaling, which has been recently shown to be essential for HSC emergence (Espin-Palazon et al., 2014).

Mechanisms underlying chromatin accessibility dynamics

Next, we sought to further characterize chromatin accessibility dynamics by looking specifically at the subgroups within the + or – logFC. Within the -logFC elements were two groups: 1) elements accessible in cHE but not accessible in sHE (“newly closed”) and 2) elements accessible in cHE but less accessible in sHE (“decreasing”) (Figure 4.8A). Similar dynamics appear in the 3,818 +logFC elements with: 1) elements not accessible in cHE but accessible in sHE (“newly open”), and 2) elements accessible in both populations but more accessible in sHE (“increasing”) (Figure 4.8C). Of the 4,103 –logFC elements 47% lost some accessibility from cHE to sHE (“decreasing”) and 53% lost all accessibility becoming “newly closed” (Figure 4.8A). Across all populations, once decreasing elements lost accessibility the elements remained inaccessible throughout the remainder of EHT (Figure 4.8B). The newly closed elements remained inaccessible from sHE through to preI but then increased in accessibility in preII (Figure 4.8B). Of the 3,818 +logFC elements, 56% were accessible in cHE but increased accessibility in sHE, and 46% were not accessible in cHE but were accessible in sHE (Figure

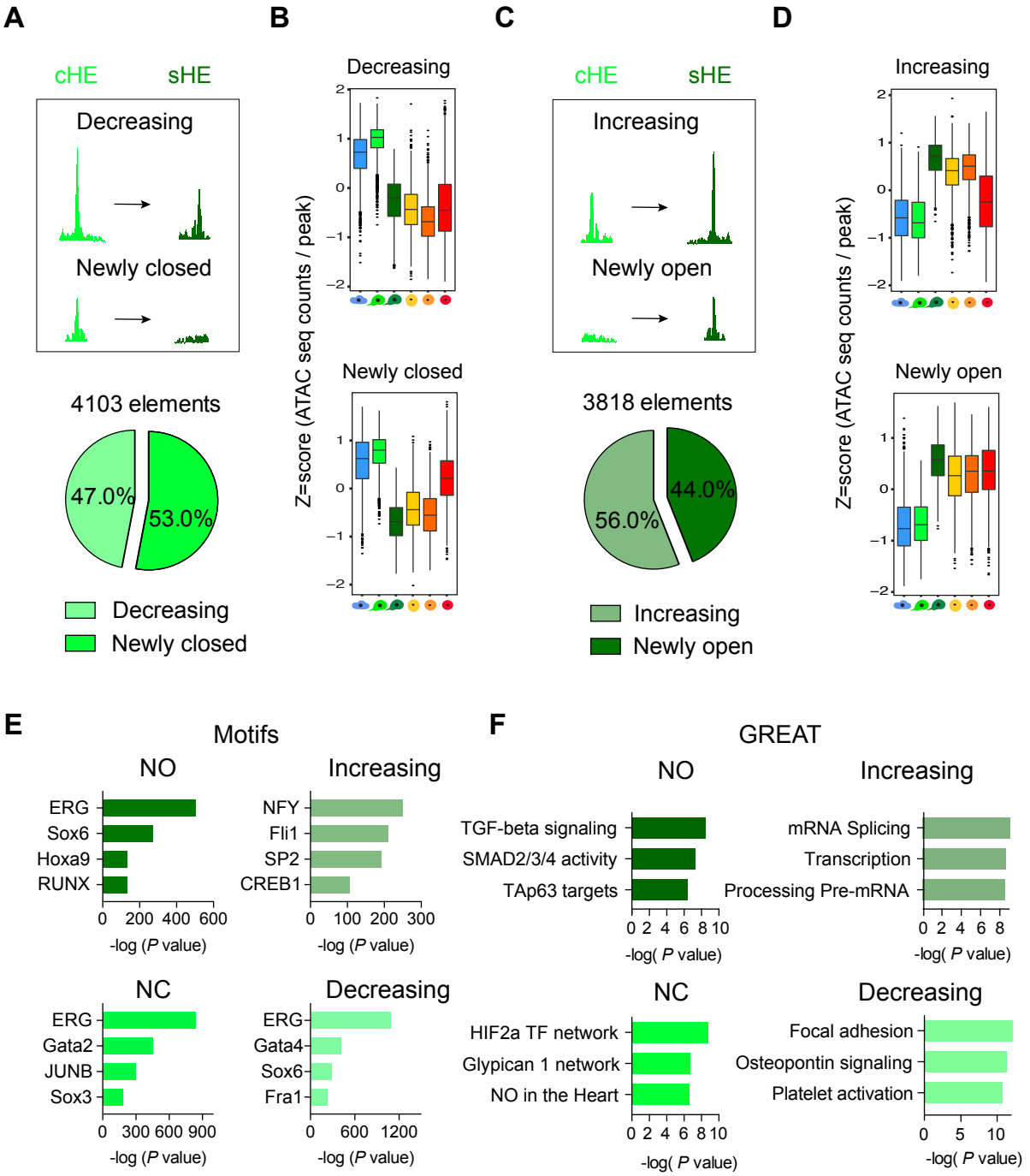


Figure 4.8: Chromatin accessibility subsets for cHE to sHE transition. (A) Schematic of chromatin dynamics including "decreasing" and "newly closed", including pie chart showing breakdown of 4,103 -LogFC elements. (B) Boxplot of ATAC-seq counts per peak from the indicated samples (labeled at bottom) at decreasing and newly closed elements. Box indicates interquartile range. (C) Schematic of chromatin dynamics including "increasing" and "newly open", including pie chart showing breakdown of 3,818 +LogFC elements. (D) Boxplot of ATAC-seq counts per peak from the indicated samples (labeled at bottom) at increasing and newly open elements. Box indicates interquartile range. (E) De novo motif enrichment for each subset showing the top 4 motifs. (F) GREAT analysis of each subset showing the top 3 mSigDB pathways.

4.8C). Across all populations, increasing elements maintained accessibility from sHE through to preI but then lost some accessibility in preII (Figure 4.8D). In contrast, once newly open elements became accessible in sHE, they remained accessible throughout EHT (Figure 4.8D). Examples of elements exhibiting these dynamics can be seen in Figure 4.9.

In order to gain a mechanistic insight into these dynamics, each group of elements was used as input for motif enrichment using HOMER (Heinz et al. 2010) and GREAT annotation (McLean et al. 2010). NC elements were enriched for Erg, Gata2, Junb, Sox3 and c-Jun motifs, and GREAT analysis revealed annotation to genes involved in focal adhesion and VEGF, Hypoxia, and Angiogenesis (Figure 4.8E-F). Analysis of "decreasing" elements showed motif enrichment for ERG, Gata4, Sox6 and Fra1 (Figure 4.8E), and GREAT analysis reported annotation to genes involved in "Focal Adhesion", "Osteopontin signaling" and "Platelet activation" (Figure 4.8F).

NO elements were enriched for binding motifs of Erg, Sox6, Hoxa9, Runx and Creb (Figure 4.8E). The GREAT analysis revealed annotation to genes involved in "TGF-β receptor/signalling" and "Transcriptional activity of SMAD2/3/4" (Figure 4.8F), reiterating the importance for these activities at the onset of EHT. Motif enrichment for "increasing" elements reported NFY, Fli1, Sp2, and CREB1 (Figure 4.8E) and GREAT analysis reported enrichment around genes annotated to mRNA splicing, Transcription and processing pre-mRNA (Figure 4.8F).

4.4 Chromatin accessibility in sHE and proHSC

In our model of EHT, cHE and sHE precede the proHSC (Rybtsov et al., 2014; Figure 1.4). Similar to the cHE to sHE transition, little is known about

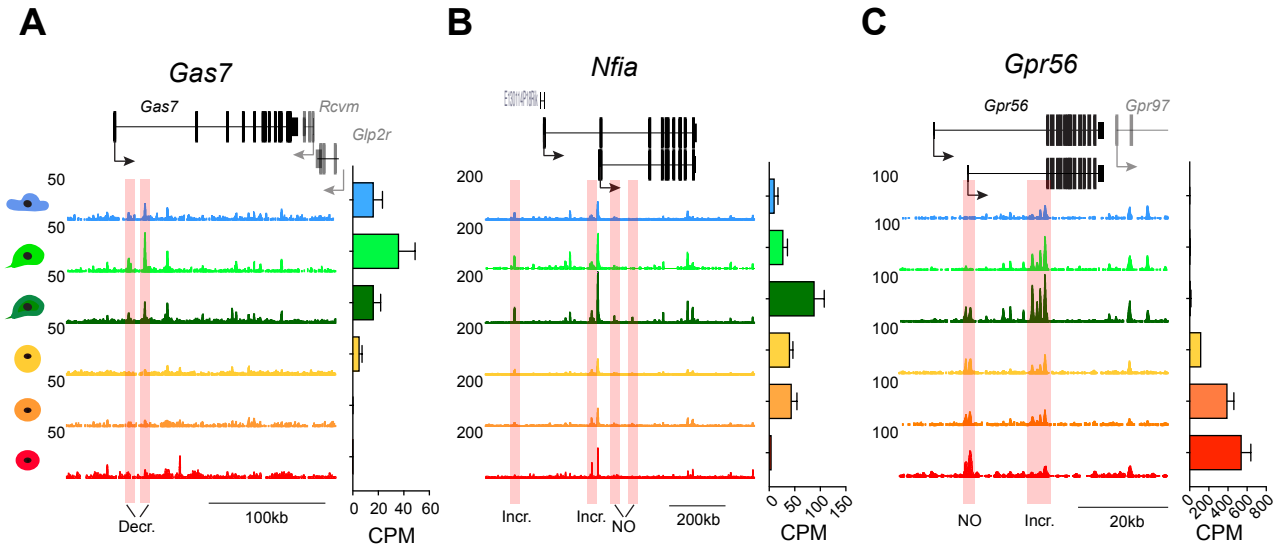


Figure 4.9: Examples of chromatin dynamics and corresponding expression during cHE and sHE transition. (A) Mean ATAC-seq for *Gas7* (Chr11:67,305,320 - 67,535,848) with corresponding gene expression in same populations. (B) Mean ATAC-seq for *Nfia* (Chr:4:97,039,827-97,994,375) with corresponding gene expression in same populations. (C) Mean ATAC-seq for *Gpr56* (Chr8:97,493,394-97,553,013) with corresponding gene expression in same populations.

the underlying mechanisms driving the sHE to proHSC transition. ProHSCs were defined by the expression of CD41, which is a marker of hematopoietic potential (Ferkowicz et al., 2002; Mikkola et al., 2003). In single cell functional assays, CD41+ cells harbored no endothelial tube formation potential while having increased blood formation potential as compared to endothelial populations (Swiers et al., 2013). Given this transition's importance as the stage in which the hematopoietic program is "locked in", we used pairwise differential accessibility to identify putative regulatory elements used by TFs to govern the transcriptional program regulating this transition.

Differential accessibility

Next, we identified and characterized differential chromatin accessibility. Of the 49,520 total elements, 6,707 were differential accessible between sHE and proHSC with 4,102 –logFC elements and 2,605 +logFC elements (FDR < 0.05; Figure 4.10A). Elements losing accessibility were enriched near genes involved in Hif2-alpha TF network, VEGF signaling and focal adhesions (Figure 4.10B). TF motif enrichment revealed the top 10 most enriched known motifs were part of the ETS family (Figure 4.10D). De novo motif enrichment revealed ERG, Sox6 and Atf3 as the top 3 (Figure 4.10E). ERG was a well-known regulator of hematopoiesis (Wilson et al., 2010). On the other hand, little was known about a role, if any, for Atf3 or Sox6 as novel regulators of EHT. Atf3 is expressed highest in Endo before being down regulated as the cells become more hematopoietic.

Elements becoming more accessible (+logFC) were enriched near genes involved in B cell signaling, Hemostasis, Platelet Activation and the Immune system (Figure 4.10C). The most enriched known motifs were part of Zinc

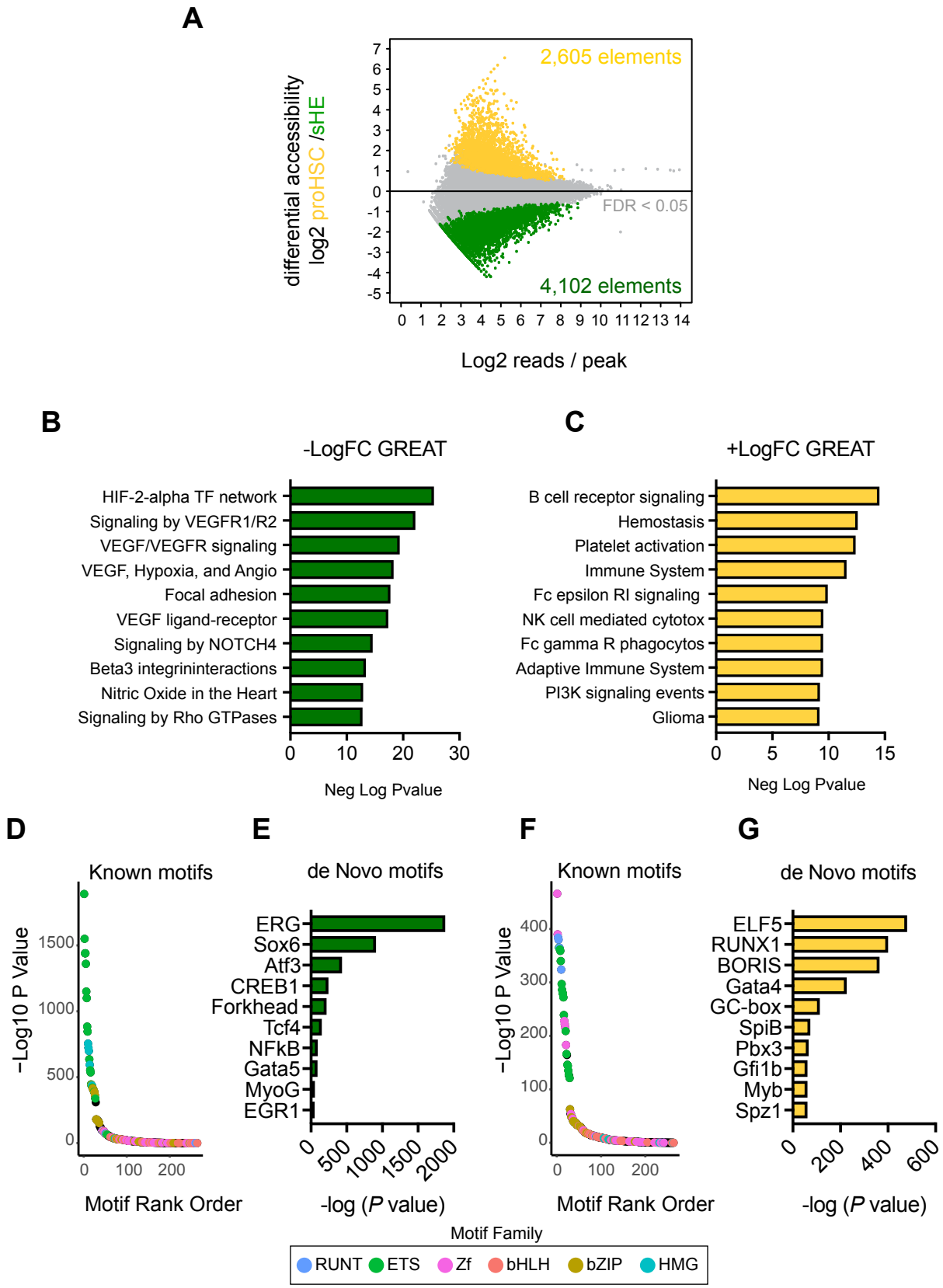


Figure 4.10: Chromatin accessibility during sHE to proHSC transition. (A) MA Plot of differential accessibility for ATAC-seq counts in sHE and proHSC. (B) Top 10 mSigDB pathways from GREAT analysis of -LogFC elements. (C) Top 10 mSigDB pathways from GREAT analysis of +LogFC elements. (D) Known motif analysis showed motif family enrichment in -LogFC. (E) de Novo motif enrichment of top motifs in -LogFC. (F) Known motif analysis showed motif family enrichment in +LogFC. (G) de Novo motif enrichment of top motifs in +LogFC.

finger (Zf), Runt, and Ets families (Figure 4.10F). De novo motif enrichment revealed Elf5, Runx1 and Ctcf as the most enriched (Figure 4.10G). Elf5 is an ETS family TF with a role in regulating keratinocyte differentiation and other epithelium-specific genes (Oettgen et al., 1999). Therefore, its significant enrichment (3.64 $x10^{-2}$ p-value) likely reflects the redundancy across the ETS family especially when considering that the *Elf5* is not differentially expressed when measured in any of the six populations of EHT.

Mechanisms underlying chromatin accessibility dynamics

Analysis of chromatin accessibility dynamics for sHE to proHSC transition revealed the same type of +logFC and -logFC subsets as was found in the cHE to sHE transition—NC, decreasing, NO and increasing. Of the 4,102 -logFC elements 16% lose some accessibility from sHE to proHSC ("decreasing") and 84% became NC (Figure 4.11A). Decreasing elements maintain some accessibility in preI before "closing" in preII (Figure 4.11B), whereas the newly closed elements remain inaccessible as the cells transition from to preI and preII (Figure 4.11B). Of the 2,605 +logFC elements, 61% were accessible in sHE but were more accessible in proHSC (increasing) and 39% were NO in proHSC (Figure 4.11C). For both increasing and NO elements, the pattern of accessibility throughout EHT was similar with increases in accessibility dipping slightly in preI before becoming more accessible in preII (Figure 4.11D).

Mechanistic insight into the chromatin dynamics using TF motif enrichment revealed that NC elements were enriched for Erg, Sox and Batf motifs (Figure 4.11E) and were associated with genes involved in Notch signalling, Oxidative stress and VEGF signaling (Figure 6.3F). Motif enrichment for decreasing

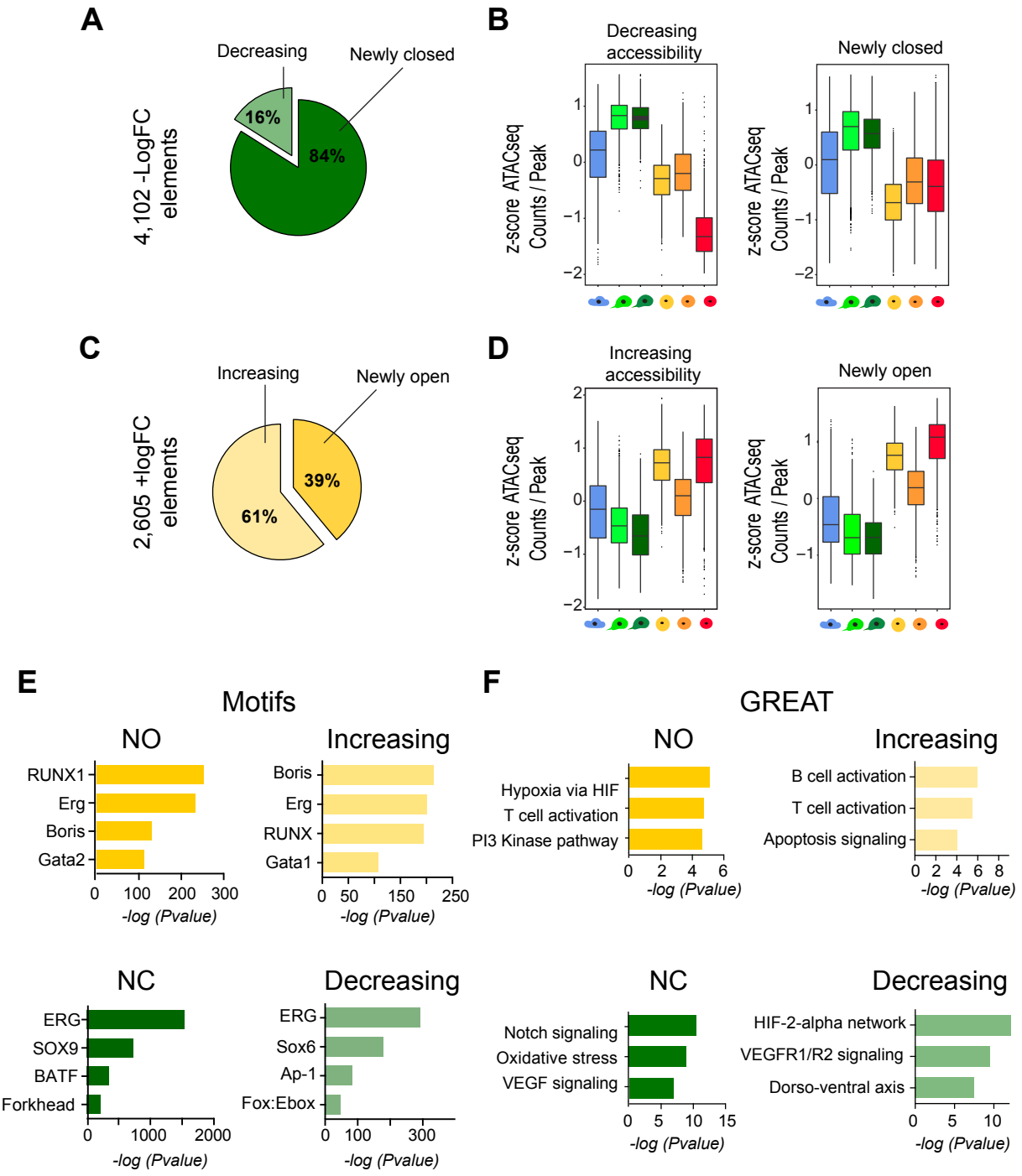


Figure 4.11: Chromatin accessibility subsets for sHE to proHSC transition. (A) Pie chart showing breakdown of 4,102 -LogFC elements into decreasing and newly closed elements. (B) Boxplot of ATAC-seq counts per peak from the indicated samples (labeled at bottom) at decreasing and newly closed elements. Box indicates interquartile range. **(**C) Pie chart showing breakdown of 2,605 +LogFC elements into increasing and newly open elements. (D) Boxplot of ATAC-seq counts per peak from the indicated samples (labeled at bottom) at increasing and newly open elements. Box indicates interquartile range. (E) De novo motif enrichment for each subset shwoing the top 4 motifs. (F) GREAT analysis of each subset shoiwng the top 3 mSigDB pathways.

elements also revealed Erg and Sox but was also enriched for AP-1 (Figure 4.11E). Previous reports show AP-1 negatively affects hematopoietic differentiation toward blood (Obier et al., 2016). Therefore, negative enrichment for AP-1 motifs in the elements losing accessibility suggests that any residual AP-1-related chromatin accessibility still present as the cells transition from sHE to proHSC is being eliminated.

One gene with both decreasing and NC peaks is Vegfc (Figure 4.12A). VEGF signaling, specifically Vegfa, along with Shh regulates angiogenesis, arteriovenous specification and HSC emergence (Clements and Traver, 2013). Vegfc is implicated in angiogenesis but there is little evidence in the literature suggesting that it has a role later on in EHT (e.g., HSC specification). Vegfc expression and chromatin accessibility increases from cHE to sHE and from sHE to proHSC (Figure 4.12A), suggesting that it too might play a role in HSC specification.

NO elements were enriched for binding motifs of Runx1, Erg, and Gata2 (Figure 4.11E) with GREAT analysis reporting hematopoietic-related pathways (Figure 4.11F). This reiterates the importance of Runx1 in facilitating chromatin accessibility at hematopoietic related genes (Lichtinger et al., 2012). Motif enrichment for Increasing elements reported the same motifs as the NO but in a slightly different order (Figure 4.11E). GREAT analysis pulled out B cell activation, in addition to T cell activation (Figure 4.11F). An example of a gene with an Increasing element that was part of the T cell activation pathway was *Plcg2* (Figure 4.12B). An example of a gene with a NO element was *Adap1*, which was involved in B cell activation (Figure 4.12C).

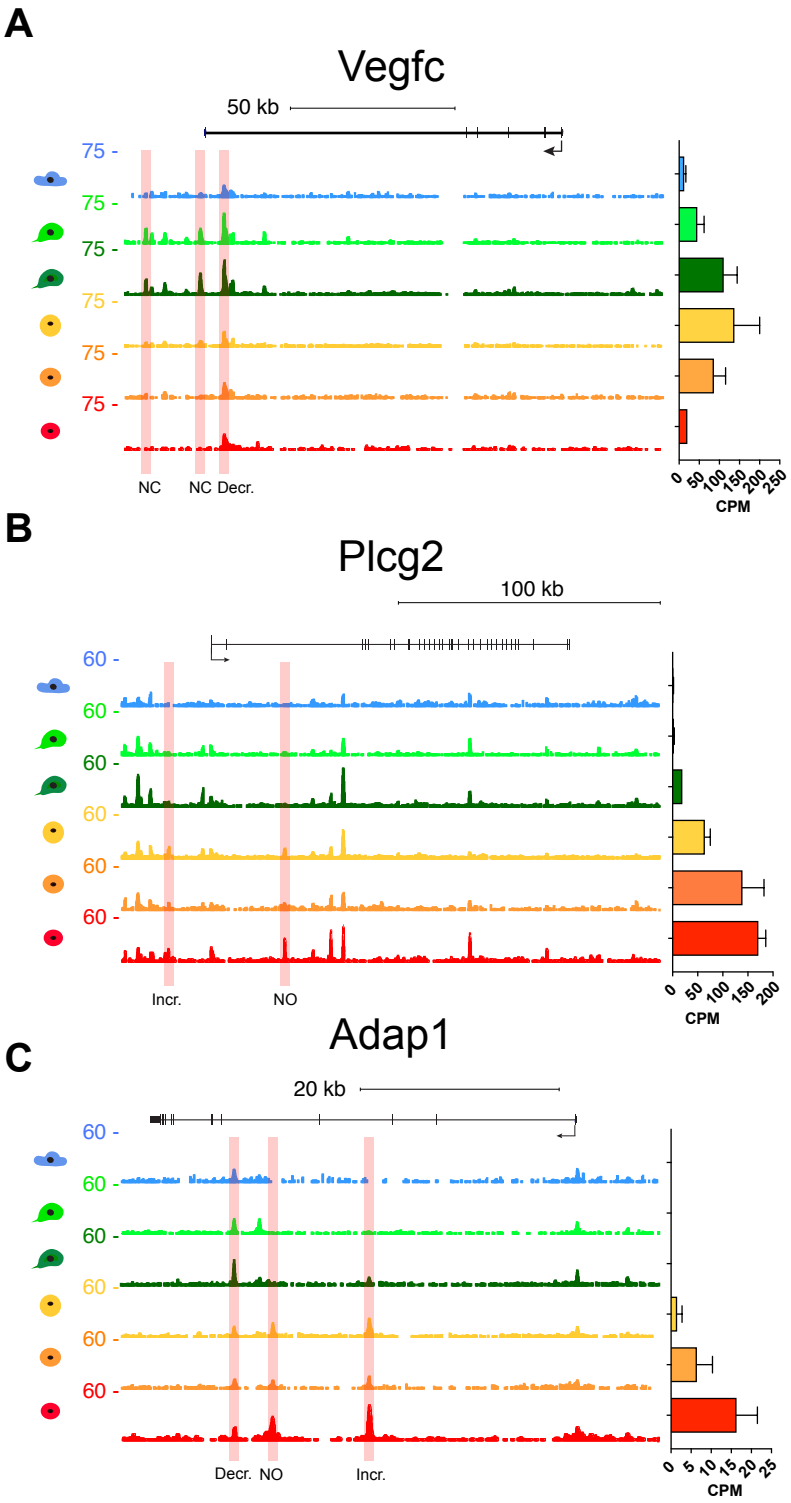


Figure 4.12: Examples of chromatin dynamics during sHE and proHSC transition. (A) Mean ATAC-seq for *Vegfc* (chr8:55,108,424-55,326,269) with corresponding gene expression in same populations. (B) Mean ATAC-seq for *Plcg2* (chr8:119,987,978-120,193,255) with corresponding gene expression in same populations. (C) Mean ATAC-seq for *Adap1* (chr5:139,736,923-139,807,292) with corresponding gene expression in same populations.

4.5 Conclusion and discussion

In this chapter, we gathered, described, validated, and analyzed an ATAC-seq data set of the cellular intermediates of in vivo EHT. ATAC-seq was performed on Endo, HE, and HSPCs isolated from the dorsal aorta and vitelline and umbilical arteries of 23GFP transgenic E8.5, E9.5 and E10.5 embryos. Correlation between all samples showed that biological replicates clustered with other replicates from the same group. Endothelial populations (Endo, cHE and sHE) clustered together away from hematopoietic populations (proHSC, preI and preII). PCA revealed that distal elements were responsible for differences in clustering. Following on, differentially accessible elements were identified as largely gene distal.

Global analysis of differentially accessible elements identified clusters sharing accessibility patterns that reflected different stages of EHT. Pairwise differential accessibility revealed new regulatory elements and motif enrichment identified TFs likely to govern accessibility at those elements. A closer look at chromatin dynamics revealed subsets within differentially accessible elements including newly closed and newly open elements. Focusing on two transitions (cHE to sHE and sHE to proHSC) revealed new regulatory elements and candidate regulators. However, similar to the list of differentially expressed genes in Chapter 3, TF binding motif enrichment independent of other data is a poor predictor of regulatory causation primarily because of redundancy in TF motifs. Therefore, in Chapter 5, we integrated RNA-seq and ATAC-seq to predict causative regulators of accessibility and subsequent changes in gene expression by assigning differentially expressed TFs to motif enrichment (Coerces et al. 2016).

Chapter 5: Gene regulatory interactions during EHT

In Chapters 3 and 4, we performed and analyzed genome-wide expression profiles by RNA-seq and genome-wide chromatin accessibility profiles by ATAC-seq of the cellular intermediates of EHT. In brief, samples were taken from Endo, HE, and HSPCs isolated from E8.5, E9.5 and E10.5 embryos. The aim of Chapter 5 was to identify putative gene regulatory interactions through the integration of the global expression and chromatin profiles. Recent work combining gene expression, chromatin accessibility, histone modification, and transcription factor binding data helped provide a greater understanding of the core regulatory network governing hematopoietic specification (Goode et al., 2016). However, this data was generated from embryonic stem cell-derived blood cells rather then primary embryonic material and thus likely reflected YS and not AGM hematopoiesis (McGrath et al., 2015). Using our integrated data sets, we identified putative gene regulatory interactions and used these interactions to identify novel putative regulators and key downstream targets during *in vivo* EHT. For reasons mentioned previously, we focused on two key transitions, (1) cHE to sHE and (2) sHE to proHSCs transitions. Using regulatory interaction maps, we identified novel interactions and putative candidate regulators of these transitions.

5.1 Strategy for integration of ATAC-seq and RNA-seq

To identify gene regulatory interactions (GRIs) regulating EHT, we focused only on interactions involving TFs because they are the primary regulators of cell fate decisions. Expression profiling in Chapter 3 identified differentially expressed TFs associated with cell fate changes during EHT. Although several of these TFs were known to participate in EHT, the molecular mechanisms describing how they contribute to HSC emergence were not well

understood. Regulation of transcriptional programs is controlled through the action of TFs at transcription start sites (TSSs), but also through TF binding to distal regulatory elements (Buecker et al., 2012). Therefore, we reasoned that putative regulatory elements identified in Chapter 4 served as docking sites for binding of differentially expressed TFs thereby facilitating regulation of the transcriptional programs important for EHT. Based on this reasoning, we looked for differentially expressed TFs and enriched TF binding motifs at ATAC-seq elements both globally, across all six populations in our model of EHT, and in key cell fate transitions.

Across all six populations, expression profiling identified 413 differentially expressed TFs, and motif enrichment of chromatin profiles identified 100+ enriched motifs. Regulatory interactions were inferred based on the integration of expression and chromatin profiles. Specifically, TFs where timing of the appearance of its motif coincided with expression of the TF were identified as putative causative regulators of accessibility and subsequent changes in gene expression (Figure 5.1A) (Coerces et al., 2016). For example, Runx1 was identified as a putative regulator in sHE because it was differentially expressed in sHE and its motif was enriched in the accessible elements in the same population (Figure 5.2).

In addition to global integration, we looked specifically at two key cell fate transitions previously identified for their importance during EHT: (1) cHE to sHE and (2) sHE to proHSC transitions. Here, we used two parallel approaches to infer gene regulatory interactions. The first approach inferred a regulatory interaction from the presence of a TF motif within a differentially accessible element mapped to a differentially expressed gene (Figure 5.1B).

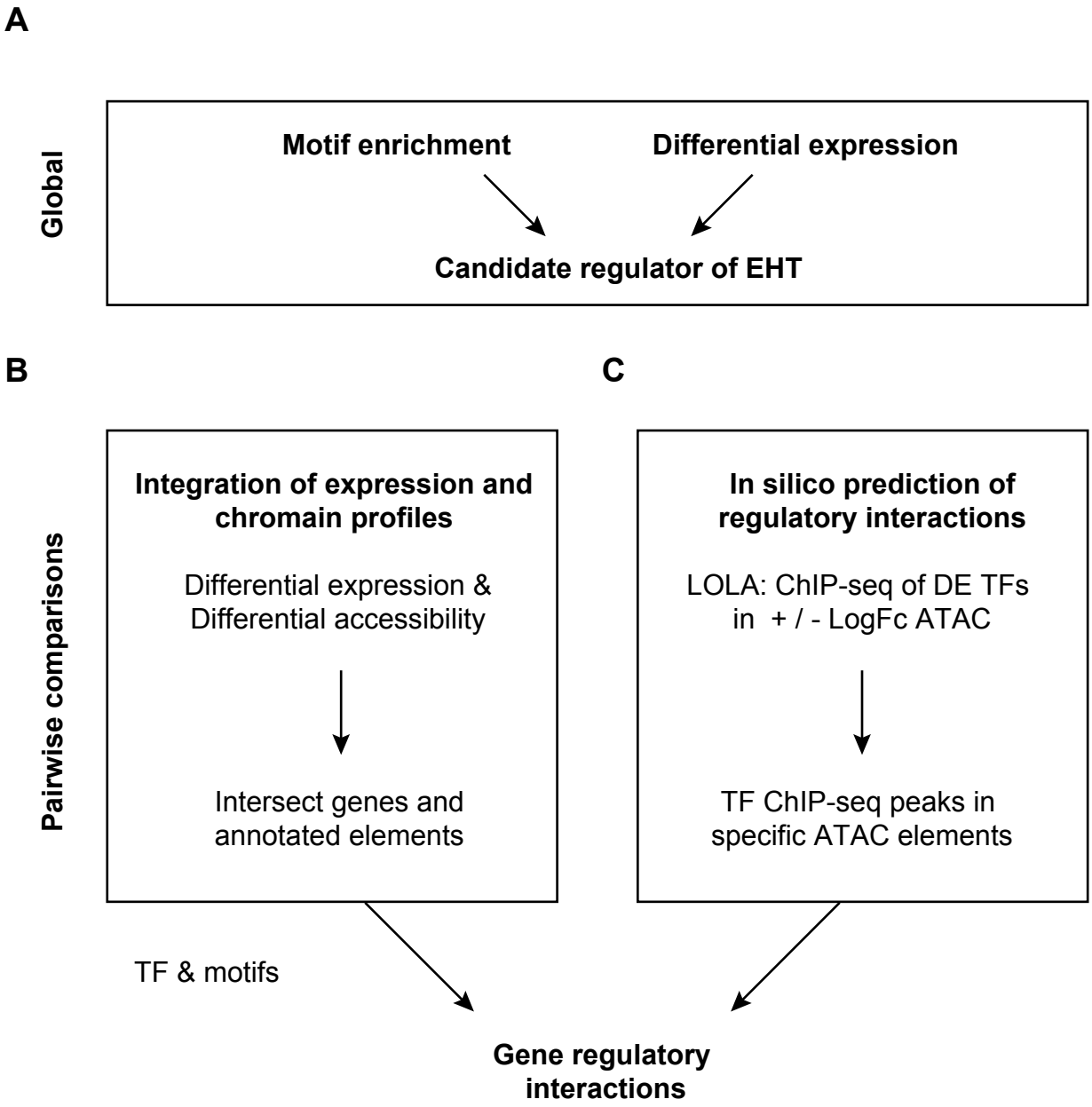


Figure 5.1: Strategy for integration of RNA-seq and ATAC-seq data. (A) Strategy for identifying candidate regulators of EHT through global integration of expression and chromatin profiles. TFs where timing of the appearance of a motif coincided with expression of the TF were identified as putative causative regulators of EHT. (B) Strategy for inferring gene regulatory interactions for key cell fate transitions based on presence of motifs in differentially accessible elements mapping to differentially expressed genes. (C) Strategy for inferring gene regulatory interactions for key cell fate transitions based on in silico predictions of TF binding within differentially accessible elements mapping to differentially expressed genes.

For example, during the cHE to sHE *Runx1* was differentially expressed and the Runx motif was enriched in elements that increased in accessibility from cHE to sHE. From this data, we inferred a putative gene regulatory interaction between Runx1 and any gene associated with an element containing a Runx motif. The second approach inferred a regulatory interaction based on *in silico* predictions of TF binding within differentially accessible elements mapped to a differentially expressed gene (Figure 5.1C). In silico predictions were made using Locus Overlap Analysis (LOLA) of our differentially accessible elements with publicly available ChIP-seq data sets. For example, Runx1 was a differentially expressed TF that had ChIP-seq data within the CODEX database. Regulatory interactions were inferred when a differentially accessible element contained a Runx1 ChIP-seq peak. Both approaches were used to construct gene regulatory interaction maps.

5.2 Global integration of expression and chromatin profiles

We identified differentially expressed TFs from the RNA-seq that also had enriched binding motifs. Motifs were pulled from the JASPAR database (http://jaspar.genereg.net/). Of the 413 differentially expressed TFs, 43 also had enrichment of their corresponding binding motifs (Figure 5.2). Interestingly, we found binding motifs enriched in a stage-specific manner. Several of these TFs have been previously implicated in EHT, including *Sox17*, *Ets1*, *Runx1*, *Sfpi1* (*PU.1*), *Gfi1b*, *Myb* and *Meis1*, among others (Figure 5.2). The timing of the appearance of these motifs coincided with expression of the corresponding TF and were thus classified as putative regulators of EHT. For example, both the binding motifs and expression of Sox17 and Ets1 were enriched in sHE (Figure 5.2). Runx1, Meis1, Gfi1B and

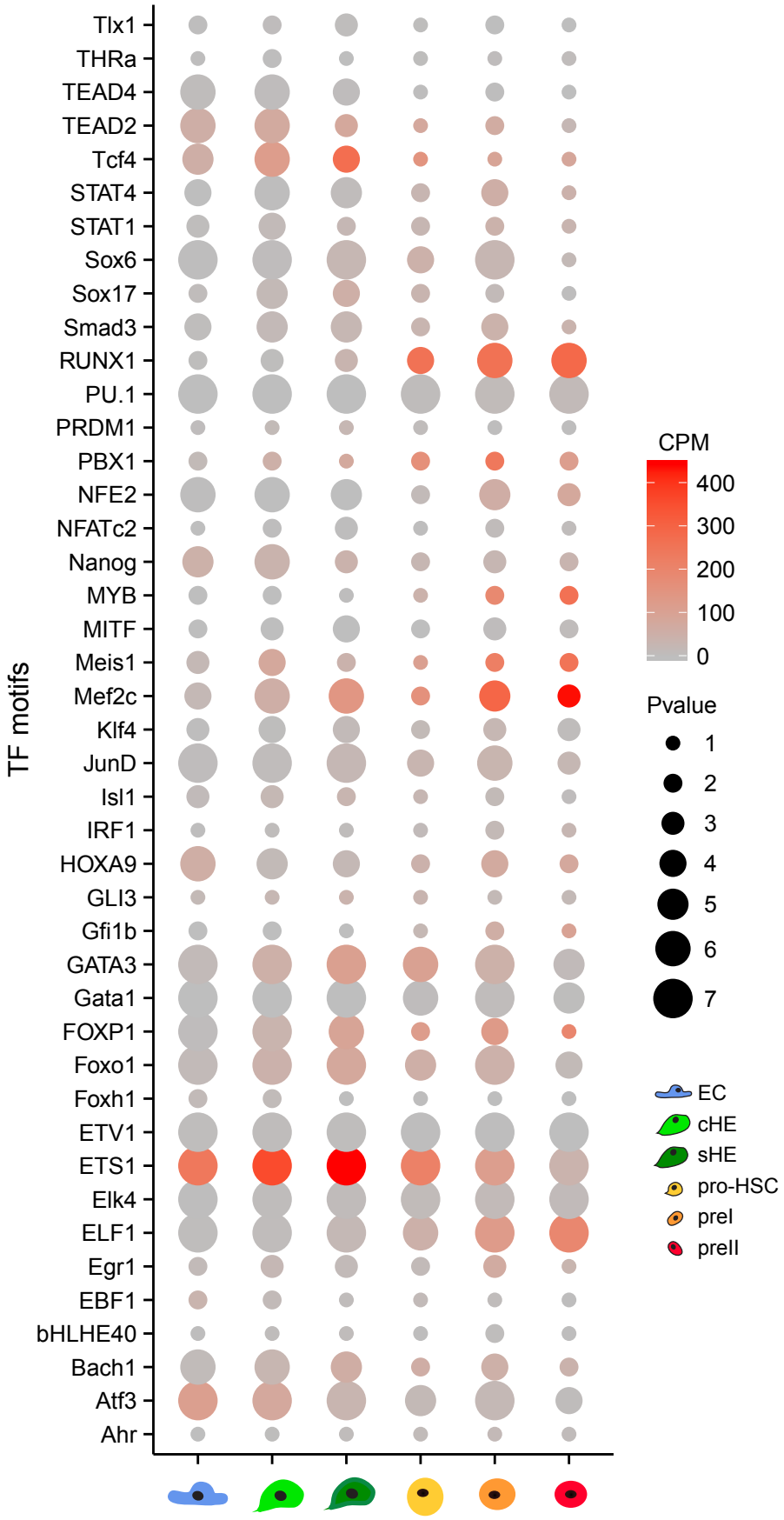


Figure 5.2: Global integration of expression and chromatin profiles. TF motifs enriched in each of the six populations. Only TF motifs where the timing of the appearance of the motif coincided with the expression of their corresponding TF were included. Expression represented as CPM (gray - lowest expression to red - highest expression) and motif enrichment label as -log10 Pvalues with more significant enrichment represented by larger circles.

Myb binding motifs and expression were most enriched in pro- and preHSCs (Figure 5.2). TFs not previously implicated in EHT included Elf1, Klf4, Mef2c, Nfat2c, Pbx1, Sox6 and Stat4. Motif enrichment and expression for Klf4 and Sox6 were most enriched in sHE (Figure 5.2). Binding motifs and expression for Elf1, Mef2c, Pbx1 and Stat4 were strongly enriched in the pro- and preHSCs (Figure 5.2). In cases with likely motif redundancy, for example within the ETS family, strong expression dictated whether genes were considered as putative regulators (Ets1 and Elf1 were not considered because of low expression) (Figure 5.2).

Novel candidate regulator of preHSC maturation – *Stat4*

When differential expression of a TF corresponded with enrichment of the corresponding motif, we called these TFs putative causative regulators of EHT. One of these TFs was Stat4. Before further investigation into its role as a putative regulator during EHT, a look at the *Stat4* genomic locus showed increasing accessibility during EHT at specific elements (Figure 5.3A). Looking specifically at the TSS and elements +3kb and +20kb downstream, we found Runx1 bound to the TSS and +3kb elements in 416b cells (Figure 5.3B) (Schutte et al., 2016). This suggested that Stat4 might be a direct target of Runx1 during EHT. In addition, *Stat4* expression increased as cells became more hematopoietic, and its motif was enriched in sHE and preI populations—identifying it as a putative regulator of EHT (Figure 5.2; Figure 5.3C). Stat4 regulates the long-term repopulating potential of HSCs (Holmfeldt et al., 2016). Therefore, we hypothesized that Stat4 confers long-term repopulation potential onto a subset of preHSC during EHT. Support for our hypothesis came from single cell RNA-seq of cell populations where Stat4 is

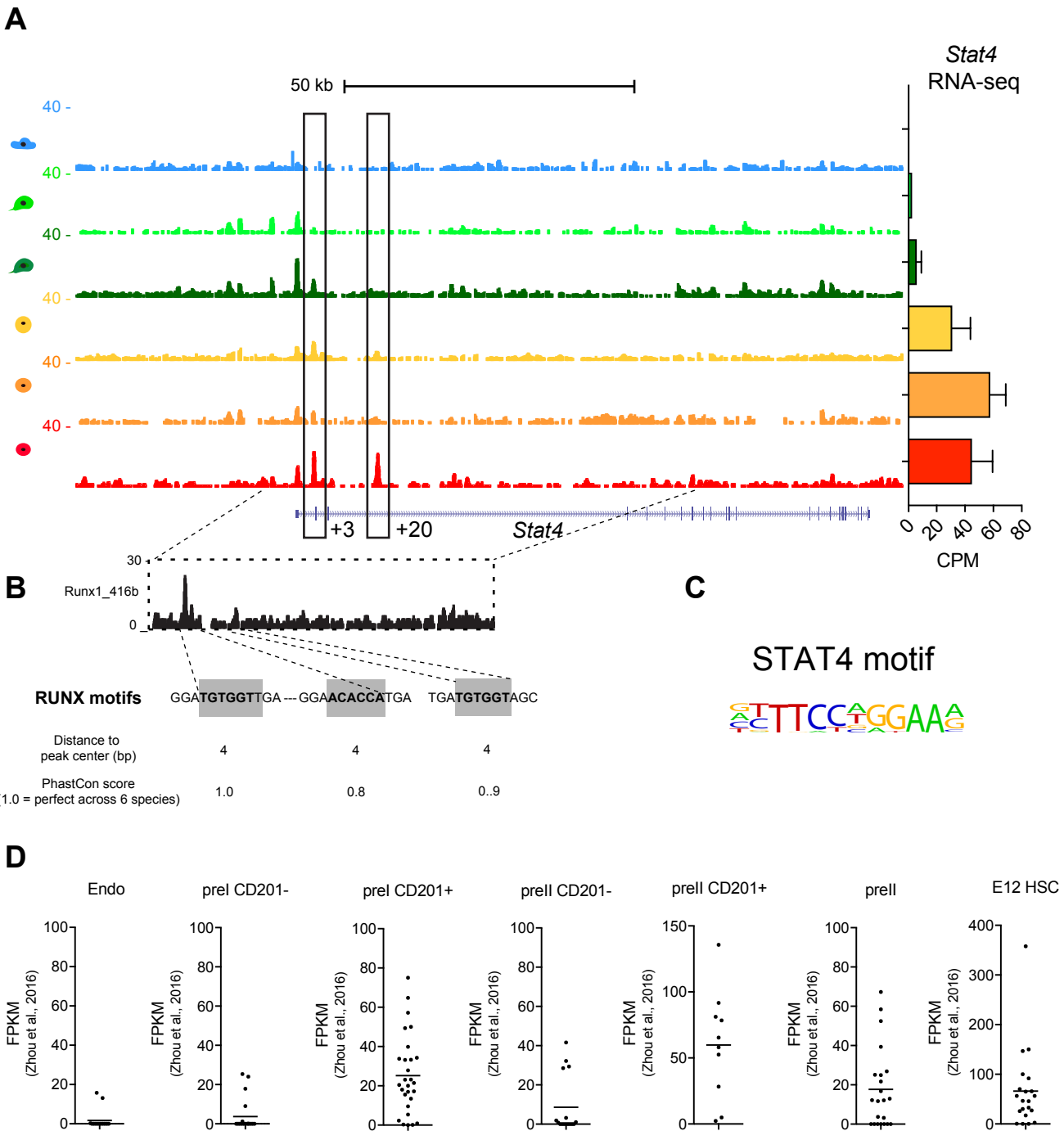


Figure 5.3: Global integration identified Stat4 as a putative target of RUNX1 and a candidate regulator of pro- and pre-HSC maturation. (A) Mean ATAC-seq coverage at *Stat4* (chr1:52,003,249-52,225,866) with RNA-seq gene expression in the same populations. Elements identified as differenitally accessible are label with black bars. (B) RUNX1 ChIP-seq in 416b cells (Schutte et al., 2016) at +3 and +20 elements. RUNX motif position and conservation scores listed below each sequence. (C) STAT4 motif enrichment in each of the five cell fate decisions of EHT. (D) Single cell RNA-seq data from Zhou et al., 2016. Each dot represents a single cell within each population. Line represents mean expression of single cells within each population.

expressed predominantly in CD201+ functional preHSC (Zhou et al., 2016; Figure 5.3D). Zhou and colleagues showed that during mouse EHT CD201 marks functional preHSC at E11 (Zhou et al., 2016). Looking specifically at Stat4 expression showed low or no expression in populations where there is no hematopoietic potential (Endo, preI CD201-, preII CD201-) and higher relative expression in populations containing functional pre-HSCs (Figure 5.3D).

5.3 Gene regulatory interactions governing the cHE to sHE transition

Previous work from the lab separated HE into cHE and sHE with the latter having initiated commitment to the hematopoietic program. Despite identification of these populations, little is known about the molecular events regulating the acquisition of the hematopoietic program in the sHE. We hypothesized that analyzing expression and chromatin dynamics would allow us to construct gene regulatory interaction (GRI) maps thereby providing mechanistic insight to this transition. In Chapter 3, we identified differentially expressed genes and in Chapter 4 we identified accessible elements. Here, we integrate this data to create GRIs using a two-pronged approach: integration of expression and chromatin profiles and *in silico* prediction of TF binding (Figure 5.1B-C). For each approach, we generated a GRI map, which was validated by characterization of regulatory elements known to play a role during EHT. Finally, using these maps, we identified new regulatory interactions.

Integration of expression and chromatin profiles

We integrated ATAC-seq and RNA-seq for the cHE to sHE transition with the aim of identifying putative regulators of this transition and inferring GRIs.

Differentially expressed TFs were identified and those whose corresponding motif was enriched in the differentially accessible elements were identified as putative regulators. Using HOMER, the 9,679 differentially accessible elements were mapped to 5,877 genes based on proximity (Figure 5.4A). The 5,877 genes were intersected with the 1,364 DEGs for the cHE to sHE transition to generate a list of differentially accessible elements assigned to DEGs (Figure 5.4A). This identified 996 elements mapped to 435 differentially expressed genes (Figure 5.4A) (Heinz et al., 2010). Each element-gene pair was associated with two logFC values—one for changes in accessibility and the other for changes in expression. Plotting differential expression versus differential accessibility separated the 996 elements into four quadrants (Figure 5.4B). Two quadrants of particular interest included (1) the quadrant with –logFC expression and –logFC accessibility (the "-/- quadrant") and (2) the quadrant with +logFC expression and +logFC accessibility (the "+/+ quadrant") (Figure 4.5B). We reasoned that a greater understanding of the driving forces behind the accessibility changes within these two quadrants would facilitate identification of key GRIs. Therefore, we performed motif enrichment and GREAT analysis of the elements in these two quadrants to gain a mechanistic insight underlying these changes. Elements in the -/- quadrant were enriched for Gata and Ets motifs (Figure 5.4C) and Ck1/Cdk5 regulation, Skeletal myogenesis, and Integrin cell surface interaction pathways (Appendix Figure 5.1). The +/+ quadrant was enriched for Ets and Runx motifs (Figure 5.4C) and mapped to genes involved in Notch signaling, Integrin cell surface interactions, SMAD2/3 signaling and Itga2b/Itgb3 (CD41/CD61) signaling, (Appendix Figure 5.1). The other quadrants (+/-) and

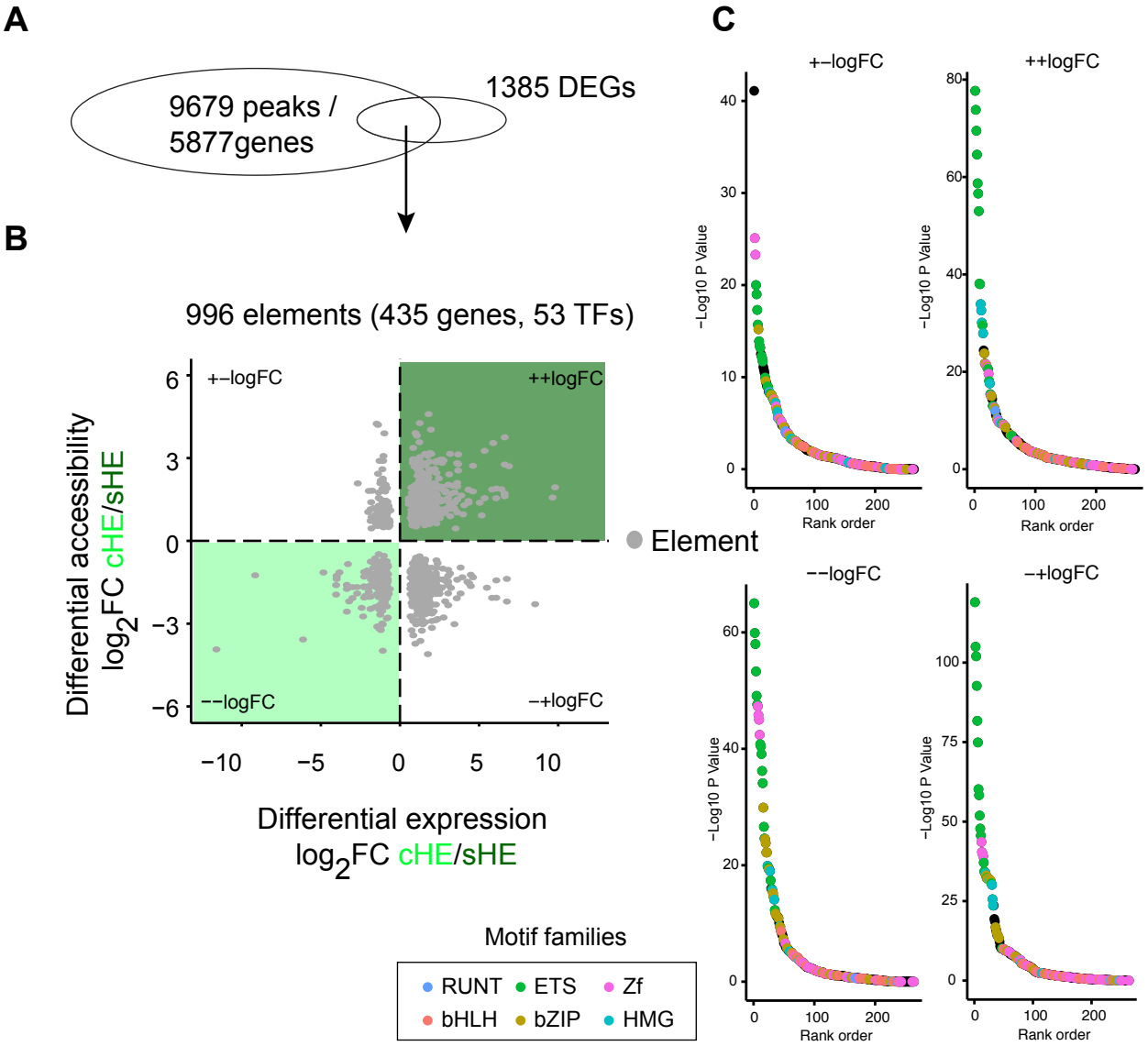


Figure 5.4: Integration of expression and chromatin profiles for cHE to sHE transition. (A) Differentially accessible elements were annotated to nearest gene using annotatepeaks.pl. Intersected gene lists identified elements that mapped to DEG. (B) Scatter plot showing Log2 Fold Change (Log2FC) differentially accessiblity on y-axis for each of 996 elements and Log2 Fold Change (Log2FC) differentially expression for corresponding 435 annotated genes. Dark green ++ LogFC quadrant represented correlation between positive changes in accessibility and positive changes in expression of the mapped gene. Light green -- LogFC quadrant represented correlation between negative changes in accessibility and expression. (C) Motif enrichment for each quadrant using HOMER findgenomemotifs.pl. Motif colored according to families with each motif rdered according to significance of enrichment (-log10 Pvalue).

(-/+) likely represent limitations associated with using proximity as the primary determinant in assigning elements to genes. Not surprisingly, RUNX motifs were enriched in both the -/- and +/+ quadrants because RUNX is known to both down regulate the endothelial program and up regulate the hematopoietic program (Eliades et al., 2016; Lie-A-Ling et al., 2014; Swiers et al., 2013). Our data identifies putative RUNX1-dependent GRIs. For example, as discussed in the following section, RUNX motifs were present in several genes involved in PI3K-AKT signaling.

PI3K-AKT signaling during cHE to sHE transition (and throughout EHT)

PI3K-AKT and MAP/ERK play antagonizing roles in arterial-venous specification (Hong et al., 2006). An arterial fate choice is widely believed to preclude specification to the hematopoietic lineage (Zhang et al., 2014). In addition, recent work suggests that Evi1-pAKT activation of Notch is critical for HSC emergence (Konantz et al., 2016). Downstream of PI3K-AKT, mTor, activation is critical for EHT emergence (Zhou et al. 2016). Thus, in addition to its enrichment in the RNA-seq data, evidence existed suggesting further investigation of PI3K-AKT during EHT.

Various different types of stimuli both intra- and extra-cellular activate PI3K (Vanhaesebroeck et al., 2017). Activated PI3K catalyzes production of phosphatidylinositol-3,4,5-triphosphate (PIP3), which then serves as a second messenger for AKT. AKT then regulates cellular functions including proliferation, growth, survival, transcription and translation (Vanhaesebroeck et al., 2017). According to the Kyoto Encyclopedia of Genes and Genome (KEGG), 345 genes are curated to this pathway (http://www.genome.jp/dbget-bin/www_bget?mmu04151). 142 were differentially expressed across EHT

(Appendix Figure 5.2). Given the vast cellular processes connected to PI3K-AKT signaling, it is not surprising that so many of these genes were differentially expressed. However, a disproportionate amount (48 genes) were differentially expressed during the cHE and sHE transition with 32 up-regulated and 16 down-regulated. The specification of HE coincides with the expression of Runx1 at E8.75 (Swiers et al., 2010, Swiers et al., 2013). Runx1 regulated the PI3K/AKT pathway, at least in part, by regulation of PIK3cd, which is part of the catalytic subunit of the PI3-kinase (Edwards et al., 2009). Therefore, we hypothesized that Runx1 regulation of PI3K is important for EHT.

Increasing expression and chromatin accessibility of *Pik3cd* throughout EHT (Figure 5.5A-B) implicated it as a target of RUNX1 and hinted at a previously un-described role for PI3K/AKT signaling during EHT. In support, single cell gene expression analysis at E9.5 and E10.5 reveals increased expression of *Pik3cd* as cells became more hematopoietic (Figure 5.5C) and was co-expressed with Runx1 as Runx1 is on in hematopoietic cells. This suggested that during EHT Runx1 regulated PI3K/AKT signaling through *Pik3cd*. Moreover, this implied that PI3K/AKT may have a role in EHT apart from its role in arterial/venous specification. Interestingly, a second catalytic subunit of PI3-Kinase, *Pik3cg*, was also differentially expressed and increased in accessibility during EHT (Figure 5.5D-E). Single cell gene expression found *Pik3cg* increased expression as cells became more hematopoietic. This suggested at the onset of EHT, RUNX1 activates at least two of the catalytic subunits of PI3-Kinase via the differentially accessible elements +2kb for Pik3cd and the TSS for Pik3cg (Figure 5.5A and 5.5D).

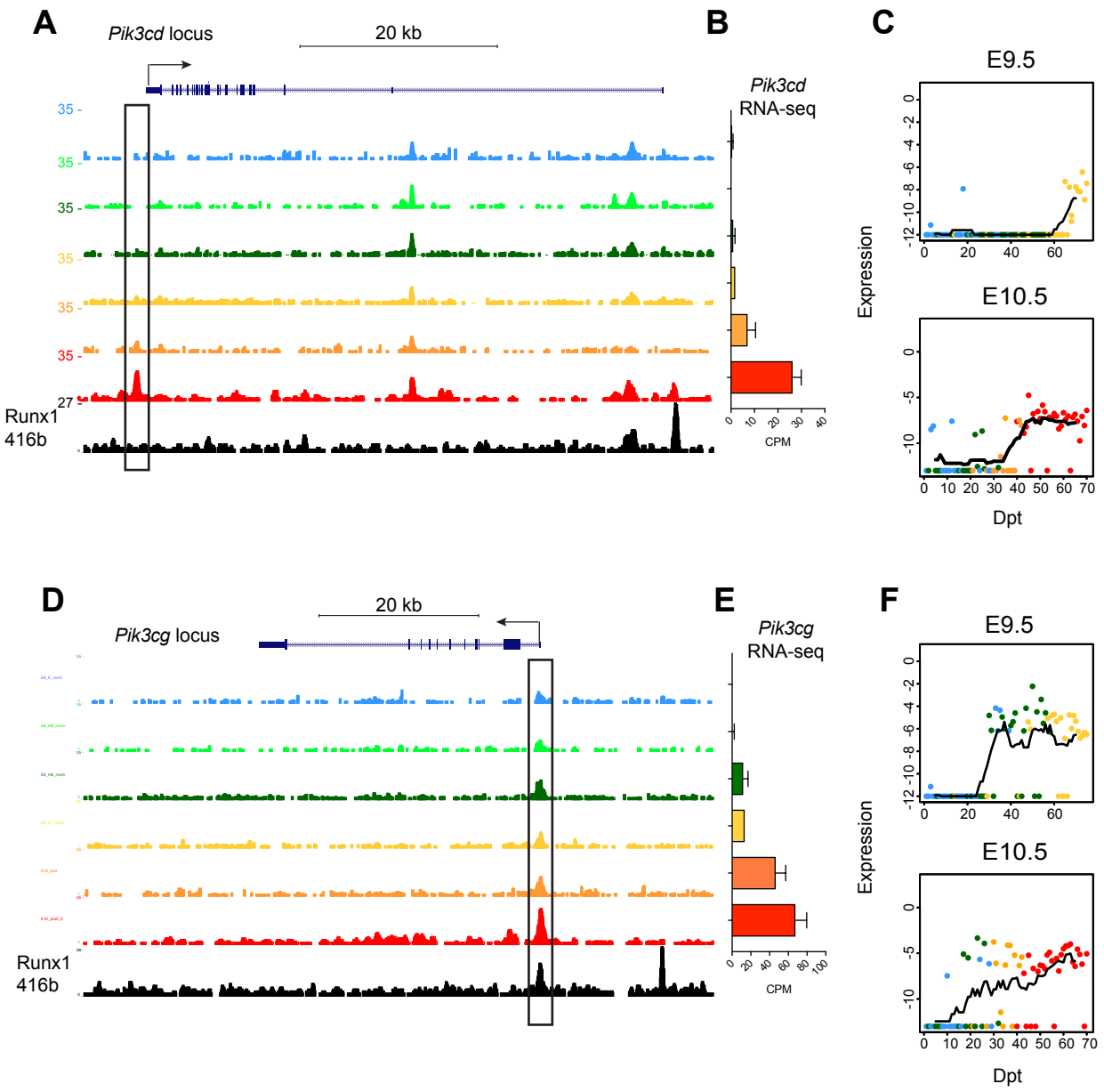


Figure 5.5: PI3K-AKT signaling during EHT. (A-B) Mean ATAC-seq coverage at *Pik3cd* (chr4:149,016,171-149,079,841) with RNA-seq gene expression in the same populations. Elements identified as differenitally accessible are labelled with black bars. RUNX1 ChIP-seq in 416b cells (Schutte et al., 2016). (C) Single cell qRT-PCR for Pik3cd at E9.5 and E10.5 in 23GFP embryos. (D-E) Mean ATAC-seq coverage at *Pik3cg* (chr12:32,858,262-32,893,203) with RNA-seq gene expression in the same populations. Elements identified as differenitally accessible are labelled with black bars. RUNX1 ChIP-seq in 416b cells (Schuette et al., 2016). (F) Single cell qRT-PCR for Pik3cd at E9.5 and E10.5 in 23GFP embryos.

Predicted TF network from integration of expression and chromatin profiles

In order to create GRI maps, we first needed to identify differentially expressed TFs whose corresponding motif was enriched in the differentially accessible elements. Of the 99 differentially expressed TFs within this transition, only 9 coincided with the appearance of their motif. By our definition, these were considered putative causative regulators of the cHE to sHE transition. Expectedly, Runx1 was identified as a putative regulator because of its increased in expression in cHE (3.1±0.1 CPM) to sHE (34.7±4.9) and RUNX motif enrichment from cHE ($-\log_{10}$ pvalue of 0.0312) to sHE ($-\log_{10}$ pvalue of 12.07) (Figure 5.6). Interestingly, Klf4 was differentially expressed (2.93±0.78 CPM to 13.3±4.15) and its corresponding motif was enriched ($-\log_{10}$ pvalue 0.12 to 16.4) (Figure 5.6). This was the first study to identify Klf4 as a putative regulator in EHT. Mef2c was another TF whose differential expression coincided with the appearance of its motif (Figure 5.6). Additional TFs with differential expression coinciding with the appearance of their corresponding motifs included *Tcf4*, *Meis1*, *Jund*, *Foxp1*, and *Atf3* (Figure 5.6).

Next, we identified elements that mapped to TFs and looked specifically for the presence of one of these 9 motifs. From the integration of the expression and chromatin profiles, we identified 996 elements that mapped to 435 genes. Of those 996 elements, 225 mapped to 53 TFs. All of the 225 elements contained at least one of the 9 TF motifs. A regulatory interaction was inferred when a differentially accessible element contained one of the 9 motifs (Figure 5.7A).

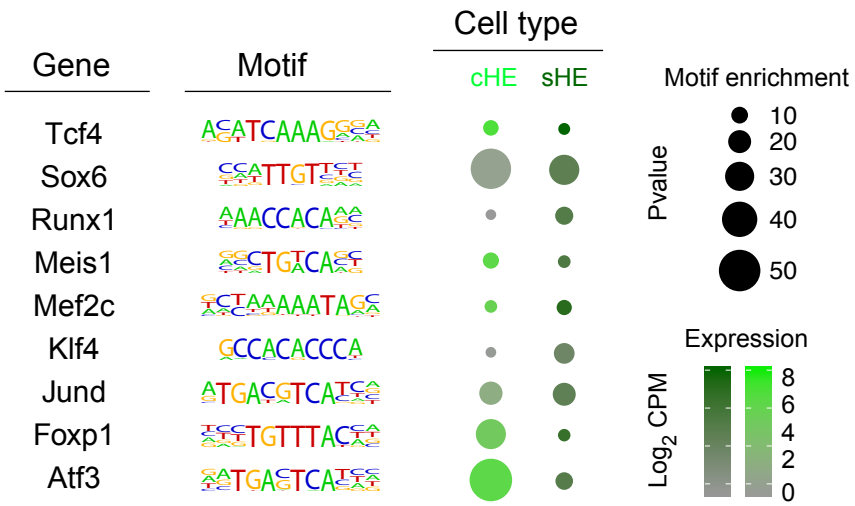


Figure 5.6: 9 TFs and corresponding motifs enriched in cHE to sHE transition. 9 TF motifs where the timing of the appearance of the motif coincided with the expression of their corresponding TF. Expression represented as CPM (gray - lowest expression to light green (cHE) and dark green (sHE) - highest expression) and motif enrichment label as -log10 Pvales with more significant enrichment represented by larger circles.

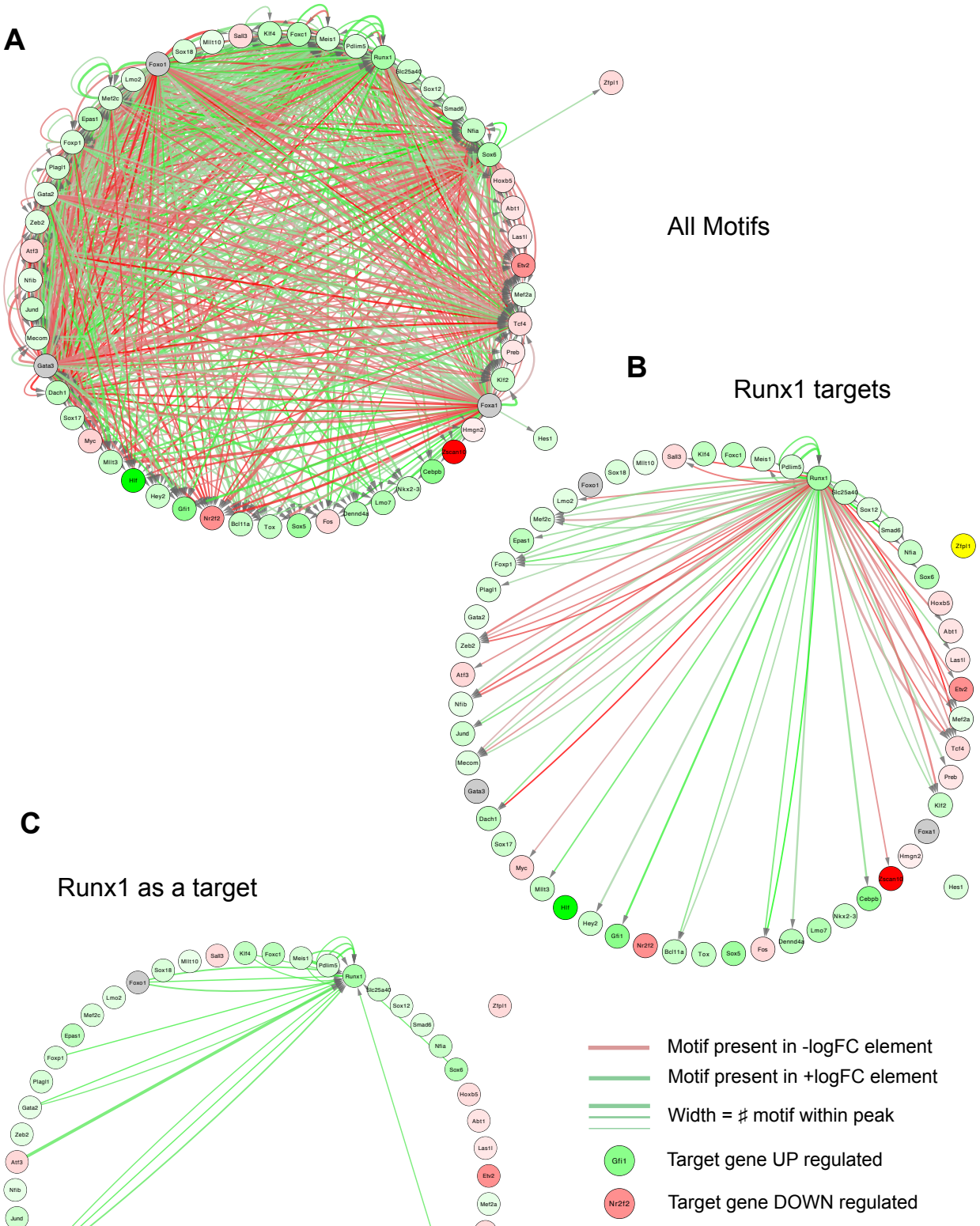


Figure 5.7: Gene regualtory interactions for cHE to sHE based on motifs in differentially accessible elements. (A) GRI based on motfs present in any of the 53 TFs. (B) GRI based on targets of Runx1. (C) GRI based on Runx1 as a target.

In order to parse out information from the GRI, we focused on Runx1 targets or Runx1 as the target (Figure 5.7B-C). Runx motifs were present in elements associated with hematopoietic TFs, including Gfi1, Fos, Cebpb, Etv2, Meis1, JunD, and Mef2c, among others (Figure 5.7B). Runx motifs were also present in elements associated with TFs not previously implicated in hematopoiesis including Kruppel family TFs Klf2 and Klf4 (Figure 5.7B). As mentioned above, the Klf4 motif was also enriched during the cHE to sHE transition (Figure 5.6B). We also looked at *Runx1* as a target because this transition marks the onset of *Runx1* expression (Figure 5.7C). Four elements were identified as differentially accessible including the P1 promoter and the +3, +23 and +204 enhancers. On going work in the lab is using this information to help identify regulators of *Runx1* expression.

Predicted TF network from in silico investigation of ATAC-seq elements

To add confidence to our predicted regulatory interactions, and to identify interactions not identified through integration of our profiling data sets, we performed *in silico* investigation of transcription factor binding within the differentially accessible ATAC elements using ChIP-seq from the CODEX database (Figure 5.1C) (Sanchez-Castillo et al., 2015; Sheffield et al., 2016). 16 of the 53 differentially expressed TFs had CHIP-seq data within the CODEX database. Locus Overlap Analysis (LOLA) for binding of these 16 TFs within the +logFC elements produced statistically significant associations for 15 of 16 data sets (Atf3, Atf4, Cebpb, Elf1, Fos, Gfi1, Jund, Klf4, Lmo2, Mef2a, Meis1, Mllt2, Myc, Nfe2, and Runx1) (Figure 5.8B). In addition to TFs known to play a role in EHT (Cebpb, Gfi1, Lmo2, Meis1, Nfe2, Runx1), we identified new targets Atf4, Mllt2, and Myc and new targets/regulators. Some

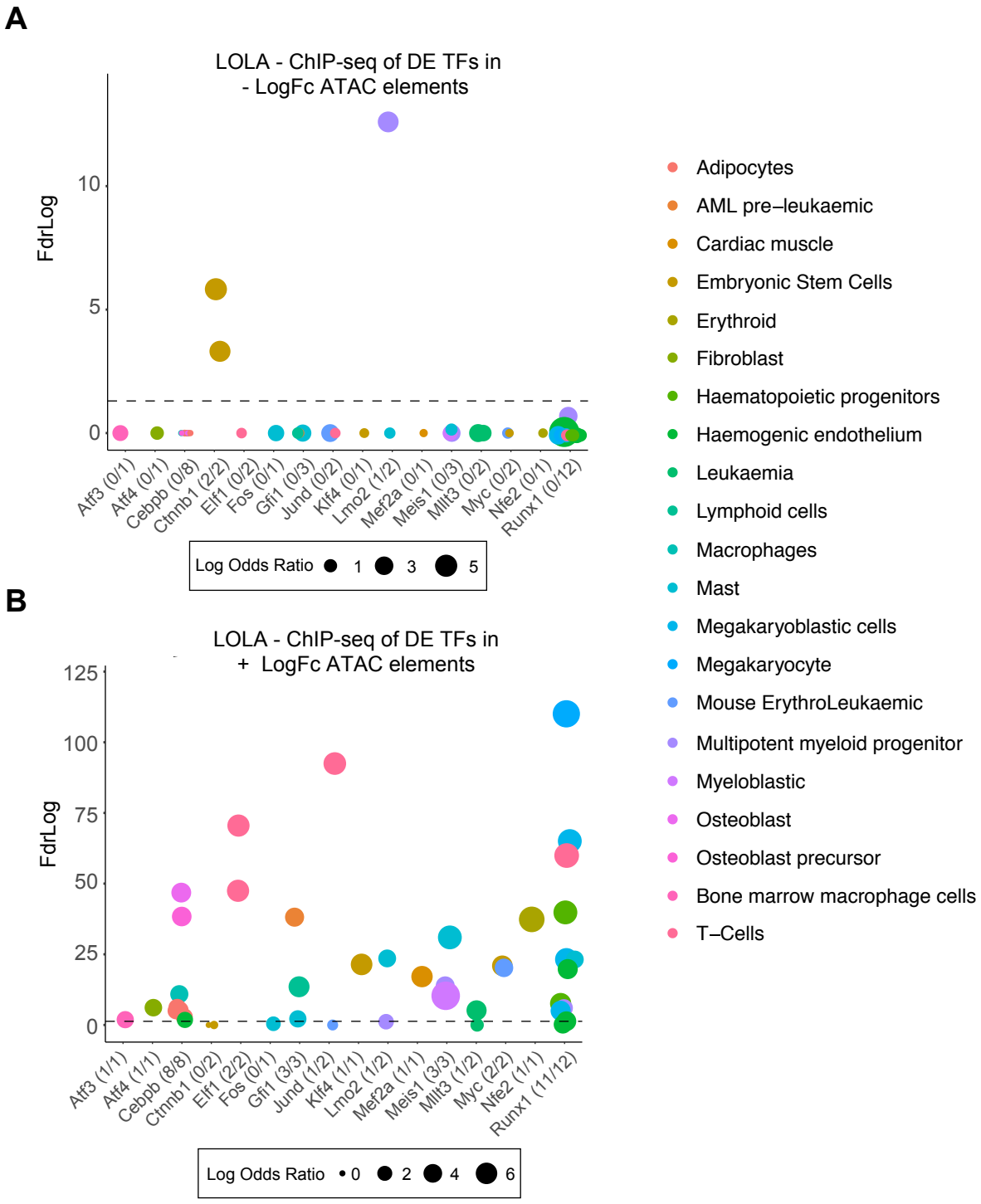


Figure 5.8: LOLA enrichment anlaysis of TF ChIP-seq data for cHE to sHE transition. (A) Results of a LOLA enrichment analysis of -logFC differentially accessible elements from the cHE to sHE transition. Each dot represents one single ChIP-seq experiment with the size relative to the quantiy of enrichment (log odds ratio) and colors indicating cell type used in the respective experiment. The number (x/y) given behind the cell type specify the number of enriched (x) out of all available datasets (y) for the given TF. (B) Results of a LOLA enrichment analysis of +logFC differ entially accessible elements from cHE to sHE transition.

of these had been previously identified in the TF motif analysis or global analysis above (Atf3, Elf1, JunD, Klf4, Mef2a/c). LOLA for the 16 TFs in the –logFC elements revealed statistical significance for only *Ctnnb1* and *Lmo2* (Figure 5.8A). Interestingly, this matched data showing Wnt/β-catenin activation is needed for activation of the hematopoietic program but not the maintenance of HSCs (Ruiz-Herguido et al., 2012).

GRIs were inferred when TF binding (identified from LOLA of ChIP-seq data sets) occurred in one of the 225 differentially accessible element mapped to the 53 differentially expressed TF (Figure 5.9A). Only 32 of 53 TFs contained elements reporting binding by one of the 16 TFs identified above. Because Runx1 had ChIP-seq data sets within CODEX, we could again focus on Runx1 targets (Figure 5.9B). As was seen with the Integration map (Figure 5.7), Runx1 ChiP-seq peaks were present in TFs associated with hematopoiesis, including *Gfi1*, *Sox17*, *Cebpb*, *Jund*, *Tcf4*, *Bcl11a*, *Hoxb5*, and *Lmo2*, among others (Figure 5.9B). Interestingly, while these TFs were implicated in hematopoiesis several of these putative interactions were not previously known, including the connection between Runx1 and Bcl11a, which is explored in greater detail in Figure 5.11. Next, we looked for TFs with ChIP-seq peaks in the differentially accessible elements mapping to *Runx1* (Figure 5.9C). LOLA analysis captured the auto regulatory nature of Runx1 and identified predicted regulation by Gfi1, Meis1, Jund, Cebpb, and Lmo2 (Figure 5.9C). As stated above, on going work in the lab is using this information to identify which, if any, of these TFs regulate *Runx1* expression.

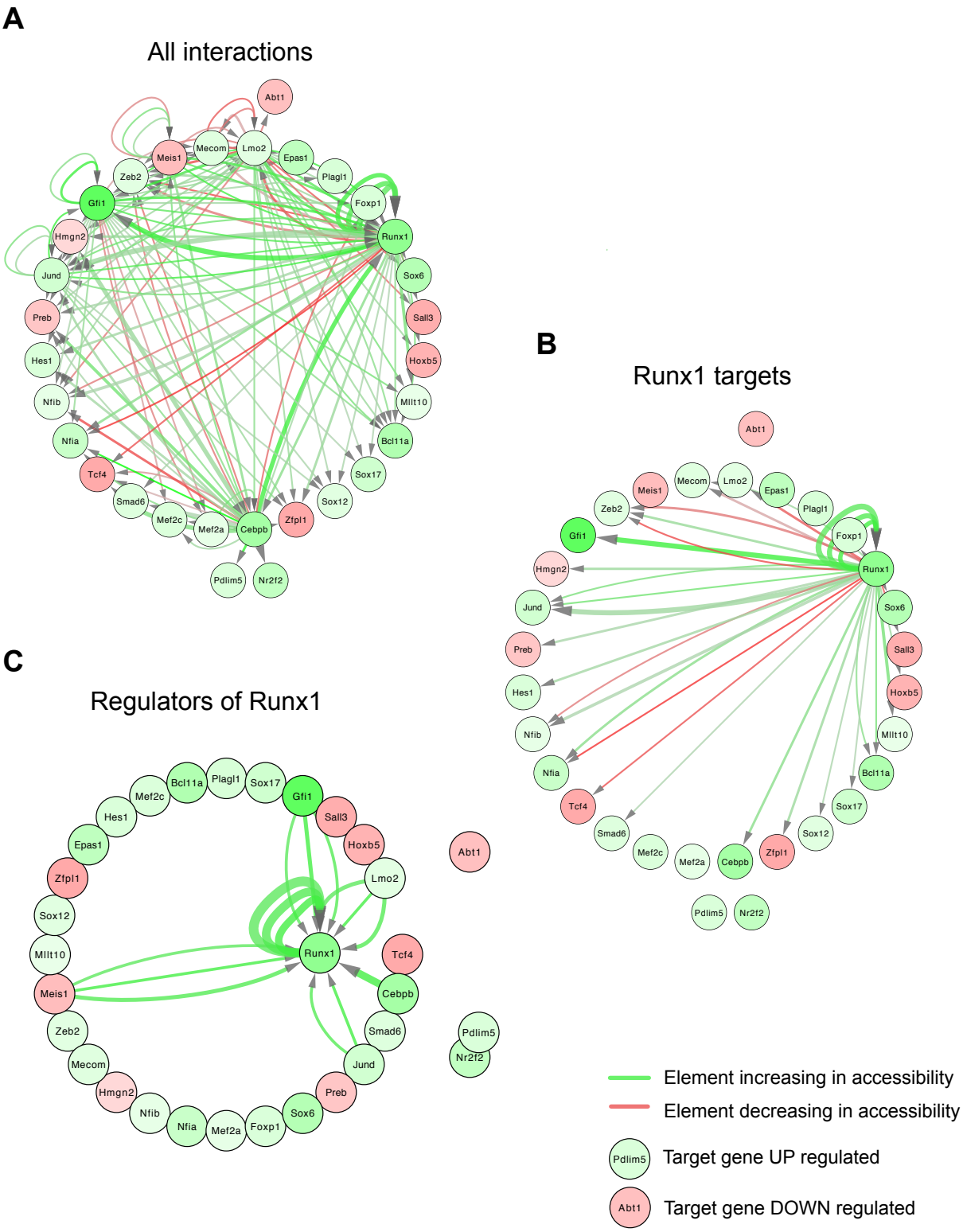


Figure 5.9: Gene regulatory interactions based on LOLA analysis for TF bound to differentially accessible elements for cHE to sHE transition. (A) Interaction map for all 16 TFs with ChIP-seq data within CODEX database (Sanchez-Castillo et al., 2015). (B) Gene interaction map for Runx1 bound elements that indicate candidate Runx1 targets. (C) Gene interaction map TFs bound within one of 4 differenitally accessible elements associated with Runx1.

Overlay of RNA-seq/ATAC-seq integration with LOLA analysis

The validity of both approaches was confirmed by the presence of previously characterized regulatory interactions in both types of interaction maps. For example, in both analysis, we identified an element 33 kilobases (kb) downstream of *Gfi1* that had previously been identified as having activity in the dorsal aorta region (Figure 5.10) (Wilson et al., 2010a). Motif tracks were generated using HOMER scanmotifgenomewide.pl with each track represented by a different symbol (Figure 5.10). Of the 9 motifs identified from integration of expression and chromatin profiles above, only Atf3 and Runx1 were present in *Gfi1*-33. In contrast, LOLA analysis showed binding for Gfi1, Lmo2, Meis1, Myc, and Runx1 (Figure 5.10). Further validation of our approach was supported by data showing Runx1 directly regulating Gfi1 and Gfi1b during HSC emergence (Thambyrajah et al., 2016; Lancrin et al., 2012). Thus, the presence of the Gfi1 -33 element in both interaction maps independently validates each approach. The power of our data was realized upon the identification of new regulatory interactions.

New predicted gene regulatory interaction: Runx1 and Bcl11a

For example, we identified novel regulatory interactions, including new relationships between *Runx1* and *Bcl11a*. *B-cell lymphoma/leukemia 11a* (*Bcl11a*) is a protein that is required for normal lymphopoiesis with *Bcl11a*-/- mice lacking B cells (Liu et al., 2003). However, its precise role, if any, during EHT has yet to be elucidated. Substantial efforts have been made to study the cis-regulatory elements governing *Bcl11a* expression (Canver et al., 2015), including the deletion of two key elements +55 and +62 (Luc et al., 2016). Surprisingly, even with these concentrated efforts, our differential accessibility

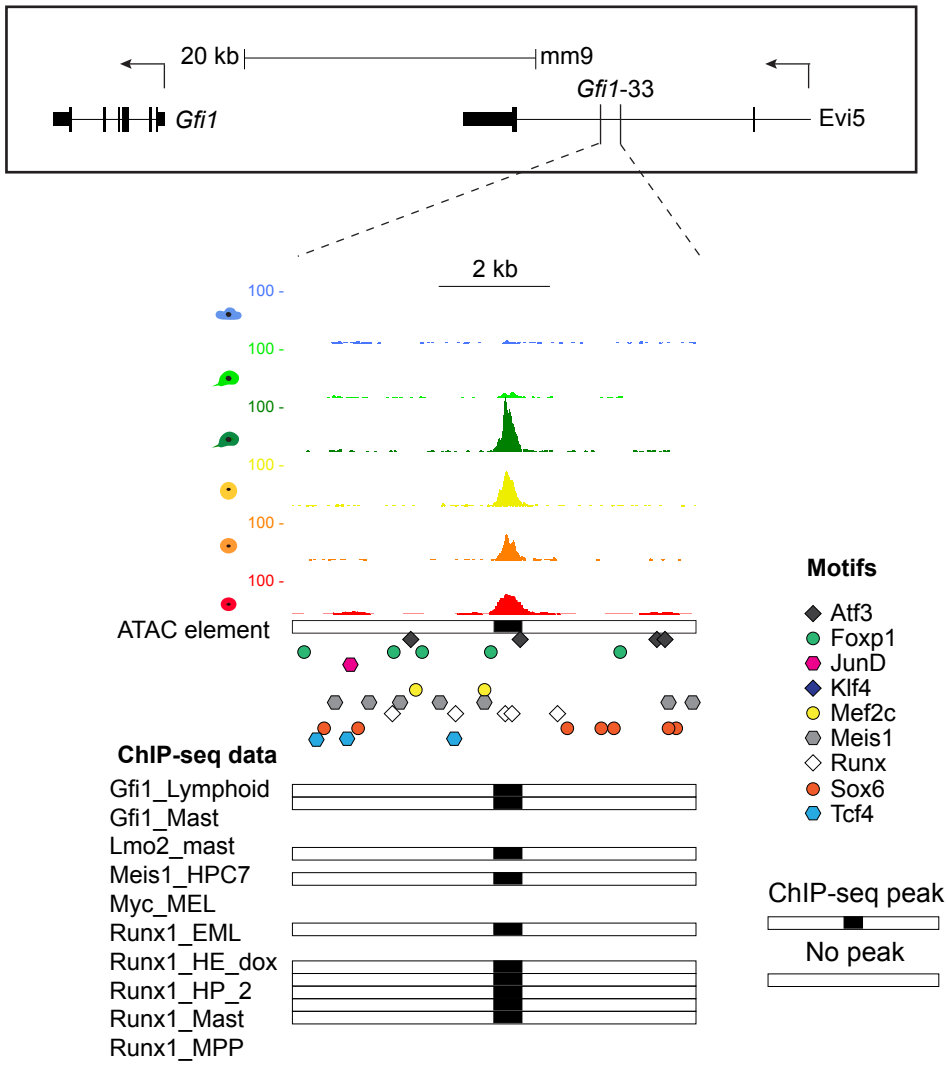


Figure 5.10: Overlay of integrated ATAC-seq/RNA-seq data with in silico ChIP-seq data for cHE to sHE transition. Schematic of *Gfi1* locus with proximity to the -33kb regulatory element. Mean ATAC-seq coverage at *Gfi1* -33kb (chr5:108,138,493-108,190,400). Motifs within this genomic locus labeled with different symbols. ChIP-seq peaks from the LOLA analysis showing the different TFs bound at this locus.

analysis revealed differentially accessibility at new elements near the TSS during the cHE to sHE transition (Figure 5.11). LOLA analysis showed these elements bound by Gfi1, Lmo2, Myc and Runx1 (Figure 5.10). In addition, new elements also appeared as the cells transition from sHE to proHSC, including elements at +40kb and +125kb (Figure 5.11). This highlighted the usefulness of our data in identifying putative GRI as well as new cis-regulatory elements.

5.4 Predicted gene regulatory interactions for sHE to proHSC transition

The transition from sHE to proHSC/CFU-C is phenotypically marked by the acquisition of CD41 (Itgba2) and functionally defined by the near-complete loss of endothelial tube formation potential and increased capacity for hematopoiesis (Swiers et al., 2013). Little is known about the molecular mechanisms underlying these phenotypic and functional changes. As stated previously, Runx1 is critically important for EHT. However, the precise point of Runx1 involvement in EHT has been studied in detail but has not yet been completely resolved (Chen et al., 2009; Tober et al., 2013; Liakhovitskaia et al., 2014). Interestingly, CD41, which marks acquisition of the hematopoietic lineage (Ferkowicz et al., 2003), is a direct target of Runx1 (Tanaka et al., 2012). However, CD41+ embryonic precursors are present in $Runx1^{-/-}$ embyros, suggesting CD41 activation occurs independent of Runx1 (Liakhovitskaia et al., 2014). Incomplete characterization of CD41-CRE mice used to rescue Runx1 expression in the latter study raises questions about whether Runx1 rescue occurred in CD41+ preHSC or in endothelium at E8.5, which is shown to express CD41 (Figure 3.2C). This discrepancy highlights the myriad questions remaining about the precise mechanisms regulating the

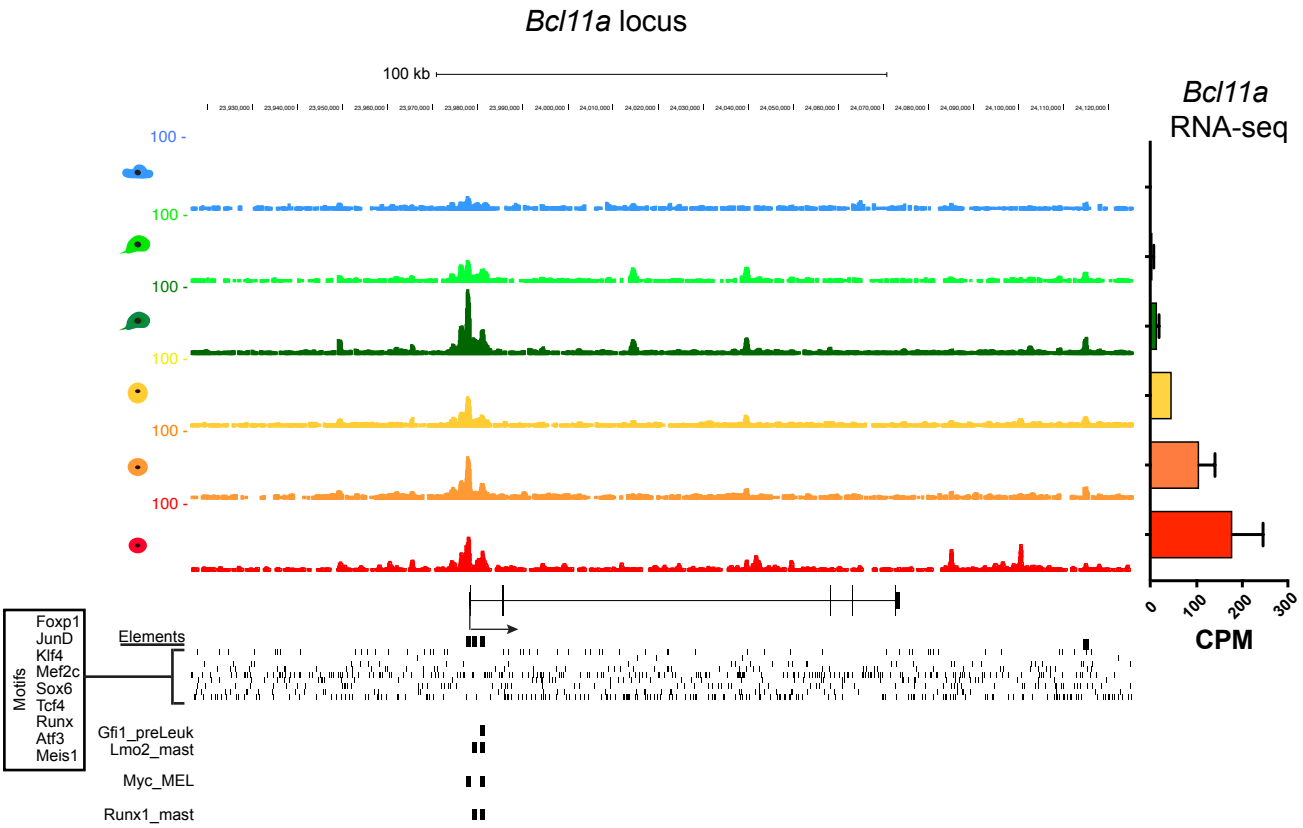


Figure 5.11: Novel regulatory interaction shows *Bcl11a* as key target during cHE to sHE transition. (A) Mean ATAC-seq coverage at *Bcl11a* (chr11:23,957,479-24,096,774) with RNA-seq gene expression in the same populations. Elements identified as differenitally accessible are label with black bars. Location of indicated motifs with black lines. Differenitally accessible elements with bound TF as identified with LOLA shown as thick black bars.

acquisition of the hematopoietic fate. The aim of this section was to help identify potential mechanisms underlying this transition by identifying the gene regulatory interactions. As with the cHE to sHE transition analyzed in Section 5.3, a two-prong approach was used to create gene regulatory interaction maps (Figure 5.1B-C). From these maps, we identified new putative regulatory interactions.

Integration of expression and chromatin profiles

First, we integrated ATAC-seq and RNA-seq for the sHE to proHSC transition with the aim of identifying putative regulators. Putative regulators of this transition were defined as differentially expressed TFs whose corresponding motif was enriched in the differentially accessible elements. Using HOMER, the 6,706 differentially accessible elements were mapped to 4,386 genes (Figure 5.12A). Intersecting these 4,386 genes with the 214 DEGs identified identified 266 differentially accessible elements mapped to 114 DEGs (Figure 5.12A). Each element-gene pair contained one logFC for changes in accessibility of the element and one logFC for changes in expression of the gene associated with the element (Figure 5.12B). A plot of differential expression versus differential accessibility revealed correlation between differential accessibility and differential expression. Positive changes in expression were associated with positive changes in accessibility ("+/+ quadrant"), and vice versa ("-/- quadrant"), suggesting that a large fraction of differential expression is likely attributable to differential accessibility (Figure 5.12B) (Coerces et al., 2016). As was true for the cHE to sHE trantion, the other two quadrants likely represent limitations associated with using proximity as the primary

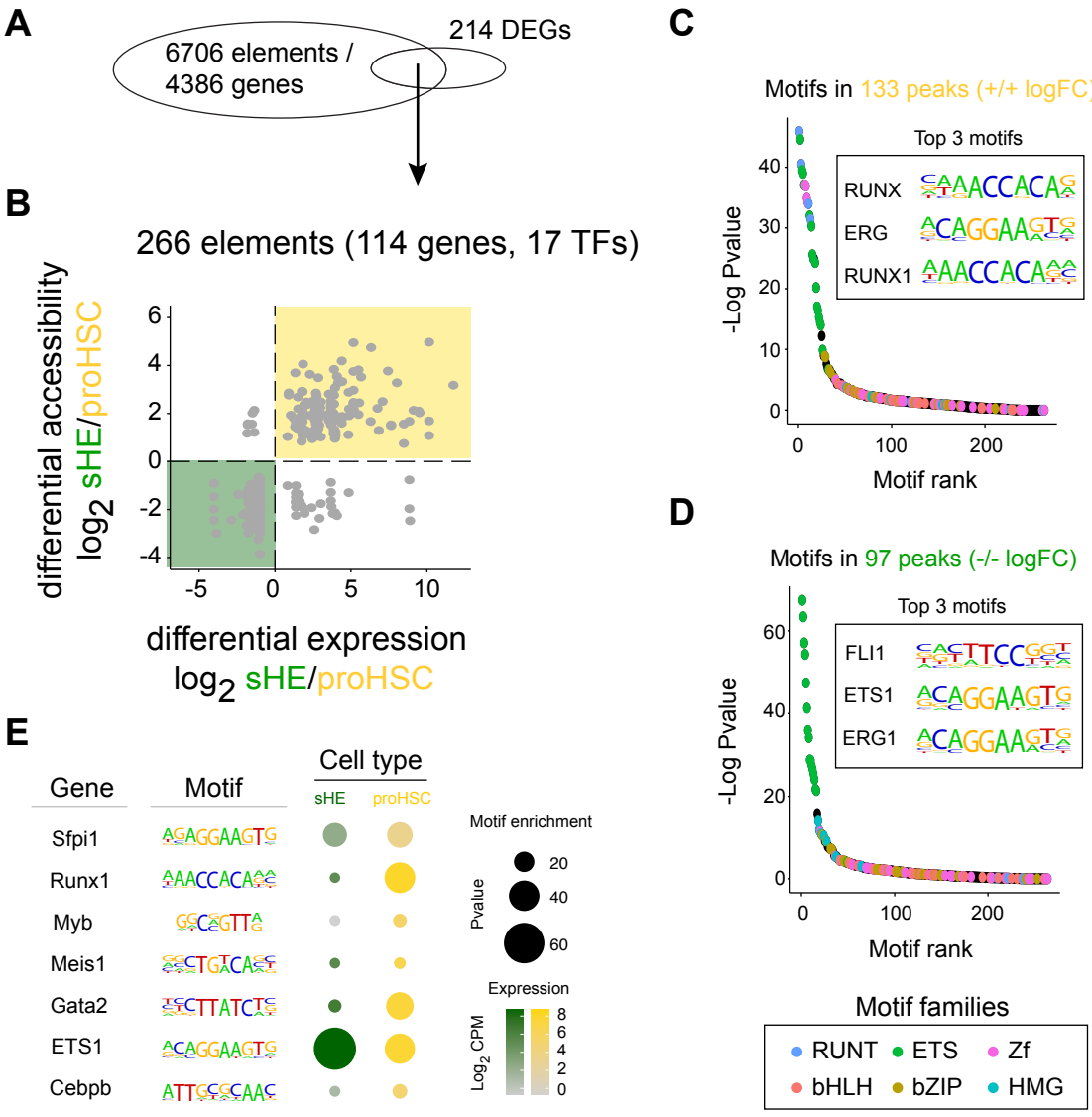


Figure 5.12: Integration of expression and chromatin profiles for sHE to proHSC transition. (A) Differentially accessible elements were mapped to nearest genes using annotatepeaks.pl. Intersected gene lists identified elements that mapped to DEG. (B) Scatter plot showing Log2 Fold Change (Log2FC) differentially accessiblity on y-axis for each of 266 elements and Log2 Fold Change (Log2FC) differentially expression for corresponding 114 annotated-to genes. Dark green ++ LogFC quadrant represented correlation between positivie changes in accessibility and positive changes in expression of mapped gene. Light green -- LogFC quadrant represented correlation between negative changes in accessiblity and expression. (C-D) Motif enrichment for -/- and +/+ quadrants using HOMER findgenomemotifs.pl. Motif colored according to families with each motif ordered according to significance of enrichment (-log10 Pvalue). (E) 7 TF motifs where the timing of the appearance of the motif coincided with the expression of their corresponding TF. Expression represented as CPM (gray - lowest expression to green/yellow - highest expression) and motif enrichment label as -log10 Pvales with more significant enrichment represented by larger circles.

determinant in assigning genes to elements. For further analysis we focused on the +/+ and -/- quadrants.

Motif enrichment analysis of the elements that positively correlated with changes in expression ("+/+ quadrant") showed that the Ets, Runx and Zf (Gata) families were among the top 10 known motifs (Figure 5.12C). De novo motif enrichment returned Runx, Erg and Runx1 as the most enriched (Figure 5.12C; Appendix Figure 5.3A). Motif enrichment analysis of elements that negatively correlated with changes in expression ("-/- quadrant") included only Ets family motifs in the known motifs and Fli1, Ets1 and Erg1 from the de novo enrichment (Figure 5.12D; Appendix Figure 5.3B). In the cHE to sHE transition, Runx motifs appeared in both the -/- and +/+ quadrants compared to this transition where Runx motifs were present only in the +/+. This supported the premise that the unexpected early loss of endothelial tube formation potential in sHE was likely associated with down-regulation of the endothelial program by Runx1 (Swiers et al., 2013). The presence of Runx motifs only in the +/+ quadrant for the sHE to proHSC suggests that Runx1's role in down-regulating the endothelial program was completed by the point at which cells acquired CD41 expression.

Predicted TF network from integration of expression and chromatin profiles

As was true for the cHE to sHE transition, the ultimate aim was to infer regulatory interactions for the underlying TF network. In order to do this, we identified putative regulators (differentially expressed TFs whose corresponding motif is also enriched). Of the 24 differentially expressed TFs within this transition, only 7 coincided with the appearance of their motif. These 7 TFs were deemed putative causative regulators of the sHE to

proHSC transition. For example, Runx1 was differentially expressed (Log_2 CPM from 34.73±4.96 to 257.42±43.33), and the Runx motif was enriched (-log_{10} Pvalue from 0.237 to 40.5) during this transition (Figure 5.12E). Other TFs whose differential expression coincided with the appearance of their corresponding motifs included Sfpi1, Myb, Meis1, Gata2, Ets1, and Cebpb (Figure 5.12E). Each TF except Ets1 increased in both expression and motif enrichment. This is unsurprising given each of these TFs has previously been implicated in hematopoiesis.

Next, we identified elements that mapped to TFs and looked specifically for the presence of one of these 7 motifs. From the integration of the expression and chromatin profiles, we identified 266 elements that mapped to 114 genes. Of those 266 elements, 49 mapped to 17 TFs (Figure 5.13A). All of the 49 elements contained at least one of the 7 TF motifs. A regulatory interaction was inferred when a differentially accessible element contained one of the 9 motifs (Figure 5.13A). We again focused on Runx1 targets or Runx1 as a target because, as stated above, this transition includes significant up regulation of Runx1 coinciding with a "locking in" of the hematopoietic program in the CD41+ proHSCs (Figure 5.13B-C). Runx motifs were present in elements associated with all of the differentially expressed TFs except *Bcl11a*, *Mycn*, *Nr2f2* and *Foxh1* (Figure 5.13B). While several of the TFs with elements harboring Runx motifs were known to have roles in hematopoiesis, there was no previous evidence suggesting a regulatory interaction between *Runx1* and *Ikzf1* or *Ikzf2*. Focusing on Runx1 as a target revealed four differentially accessible elements mapping to Runx1, including the P1 promoter and the +3, +23 and +204 enhancers (Figure 5.13C).

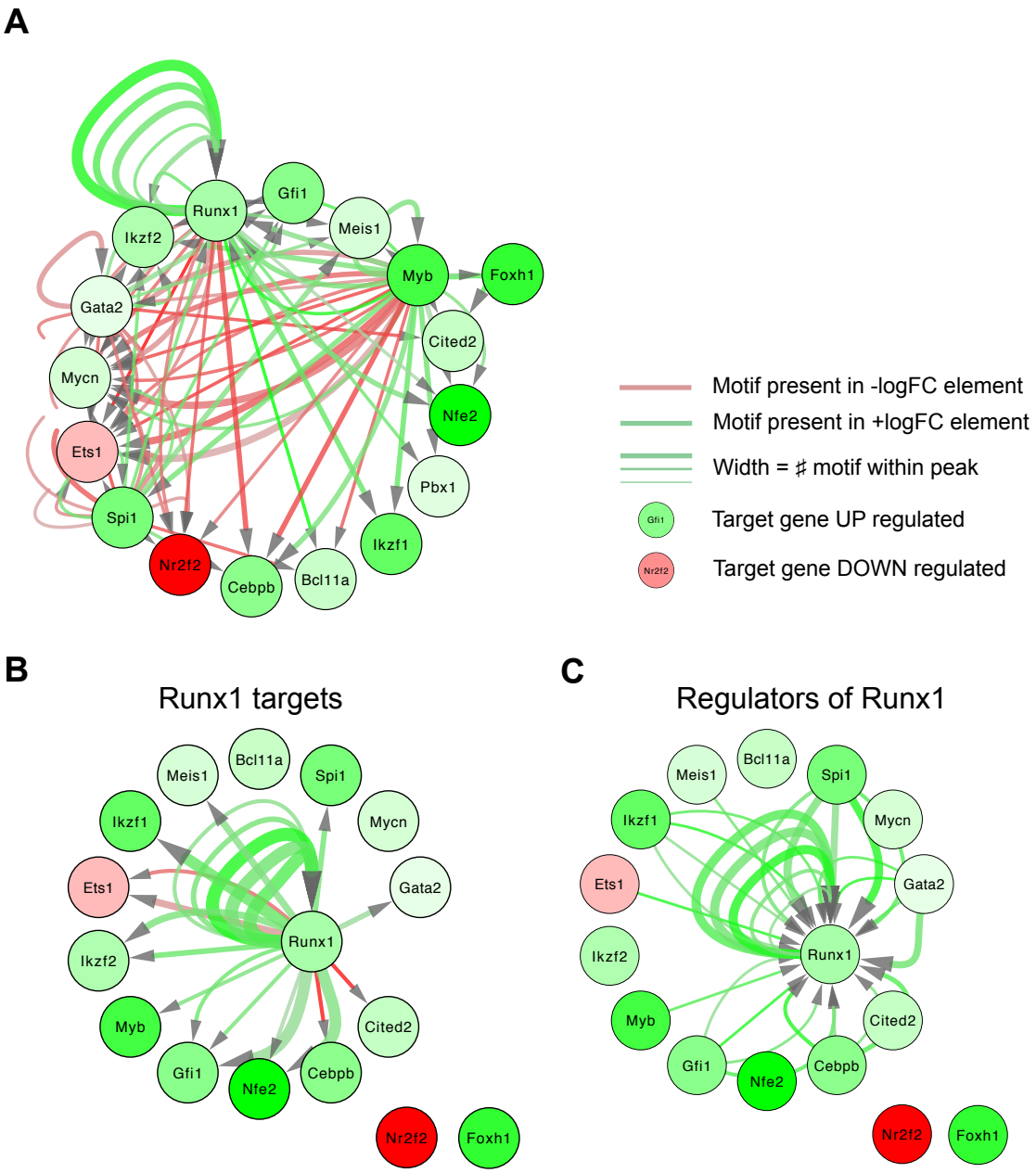


Figure 5.13: Gene regualtory interactions for sHE to proHSC based on motifs in differentially accessible elements. (A) GRI based on motfs present in any of the 17 TFs. (B) GRI based on targets of Runx1. (C) GRI based on Runx1 as a target.

Predicted TF network from in silico investigation of ATAC-seq elements

Following the strategy from Figure 5.1C, we performed *in silico* investigation of transcription factor binding within the differentially accessible ATAC elements using ChIP-seq from the CODEX database (Figure 5.1C). The aim was to predict regulatory interactions with ChIP-seq data rather than motif data alone. We reasoned that this would add confidence to putative regulatory interactions identified from integration alone and identify putative interactions not identified using integration. 12 of the 24 differentially expressed TFs had ChIP-seq data sets in CODEX (Figure 5.14A-B). LOLA for binding of these 12 TFs within -logFC elements produced no statistically significant associations (Figure 5.14A). However, LOLA for binding of the 12 TFs in the +logFC revealed statistically significant association for 8 of the 12, including *Cebpb*, *Gata2, Gfi1, Ikzf1, Meis1, Myb, Runx1, and Sfpi1* binding sites (Figure 5.14B). Interestingly, Gata2, Runx1, Meis1, Sfpi1 and Gfi1 make up 5 of the 10 key regulators of hematopoietic stem and progenitor cells previously identified (Wilson et al. 2010b). Two of the remaining three, Cebpb and Myb, were also implicated in blood cell formation, with the role of Cebpb recently being elucidated (Goode et al., 2016). Ikzf1, despite its known role in hematopoiesis, its role, if any, during EHT was not well understood.

Gene regulatory interactions were inferred when TF binding (identified from LOLA of ChIP-seq data sets) occurred in one of the 49 differentially accessible elements mapped to the 17 differentially expressed TFs (Figure 5.15A). All 17 TFs contained at least one element that was bound by one of the 12 TF ChIP-seq data sets (Figure 5.15A). Looking more closely at Runx1

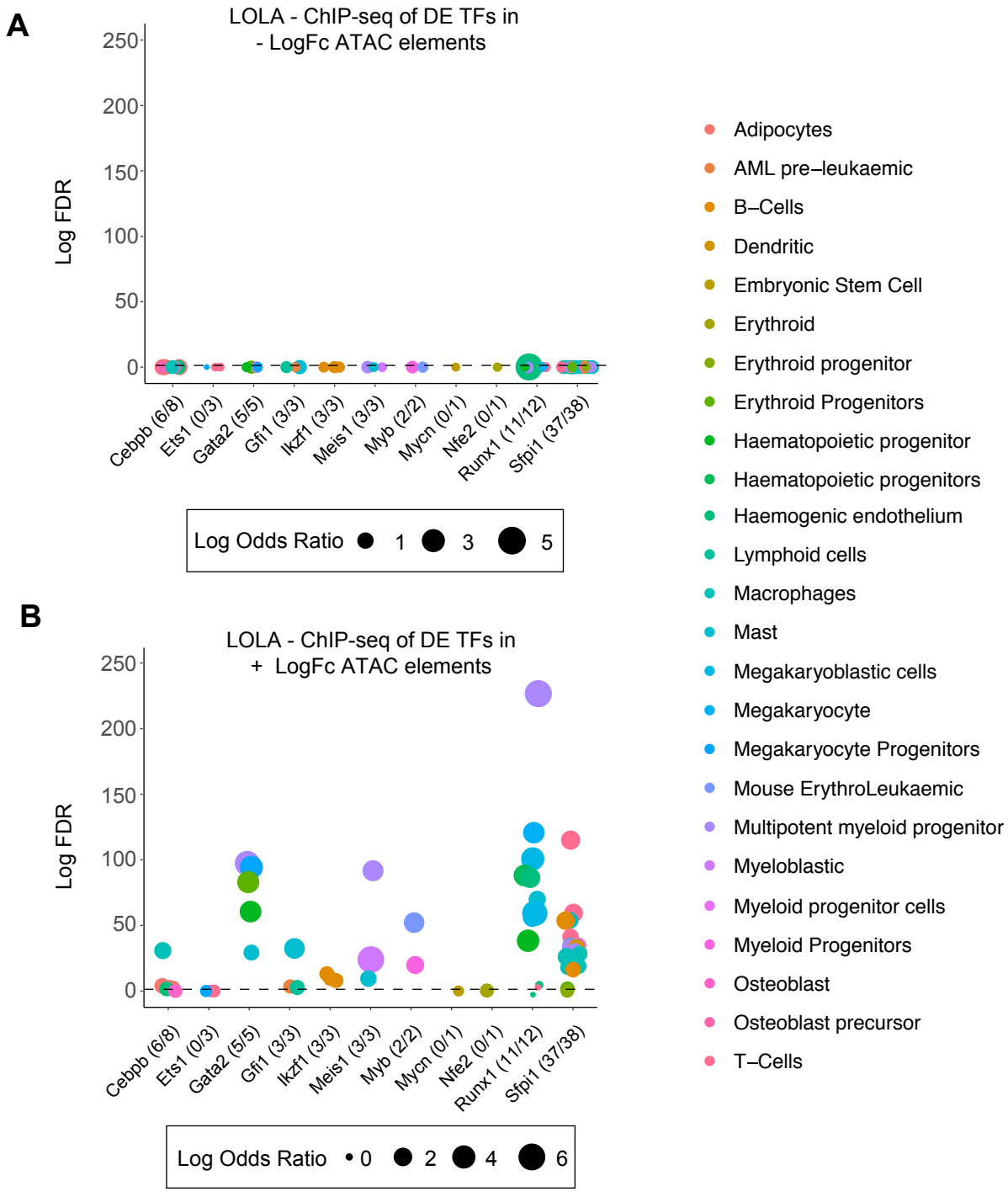


Figure 5.14: LOLA enrichment anlaysis of TF ChIP-seq data in sHE to proHSC transition. (A) Results of a LOLA enrichment analysis of -logFC differentially accessible elements from the sHE to proHSC transition. Each dot represents one single ChIP-seq experiment with the size relative to the quantiy of enrichment (log odds ratio) and colors indicating cell type used in the respective experiment. The number (x/y) given behind the cell type specify the number of enriched (x) out of all available datasets (y) for the given TF. (B) Results of a LOLA enrichment analysis of +logFC differentially accessible elements from sHE to proHSC transition.

targets, all but 4 TFs were bound by Runx1 (*Foxh1*, *Myb*, *Cited2*, *Sfpi1*). Unfortunately, a technical limitation of using proximity to map elements to genes resulted in the assignment of -14kb element of *Sfpi1* to a different TSS. Huang et al. showed Runx1 binding to the -14kb element was critical for hematopoietic expression of *Sfpi1* (Huang et al., 2008). Runx1 ChIP-seq data showed binding at the -28kb element of *Ikzf1*. Of note, Ikzf1 was also identified in the regulatory interaction maps generated using the integration approach (Figure 5.15B). We explored this novel putative regulatory interaction in greater detail in the following sections and in the next chapter. Next, we looked for TFs with ChIP-seq peaks in the differentially accessible elements mapping to *Runx1* (Figure 5.15C). This interaction map showed that the LOLA analysis captured the auto regulatory nature of Runx1 and identified predicted regulation by Gata2, Sfpi1, Ets1, Myb, and Ikzf1 (Figure 5.15C). As stated above, on going work in the lab is using this information to identify which, if any, of these TFs regulate *Runx1* expression.

Overlay of RNA-seq/ATAC-seq integration with LOLA analysis

As was the case for the cHE to sHE transition, both methods of identifying gene regulatory interaction (Figure 5.1B-C) were validated by the presence of regulatory interactions previously implicated in hematopoiesis (Figure 5.16A). For example, we identified the promoter of *Sfpi1* that had previously been identified as having hematopoietic activity (Schutte et al., 2016) (Figure 5.16A). This element contained Ets1 and Sfpi1 motifs and according to CODEX was bound by Cebpb, Gata2, Gfi1, Ikzf1, and Runx1 (Figure 5.16A).

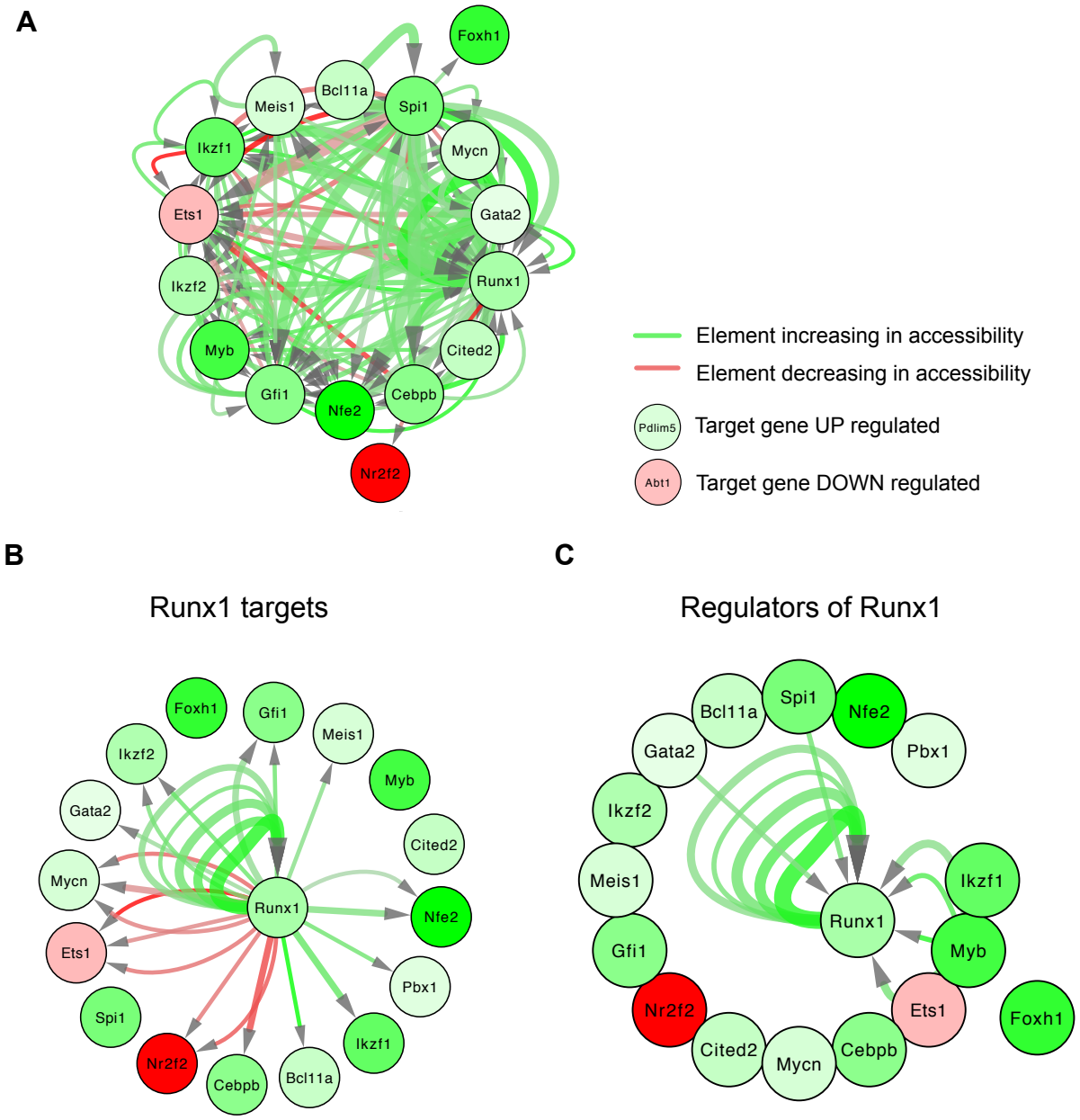


Figure 5.15: Gene regulatory interactions based on LOLA analysis for TF bound to differentially accessible elements in sHE to proHSC transition. (A) Interaction map for all 12 TFs with ChIP-seq data within CODEX database (Sanchez-Castillo et al., 2015). (B) Gene interaction map for RUNX1 bound elements that indicate candidate Runx1 targets. (C) Gene interaction map TFs bound within one of 4 differenitally accessible elements associated with RUNX1.

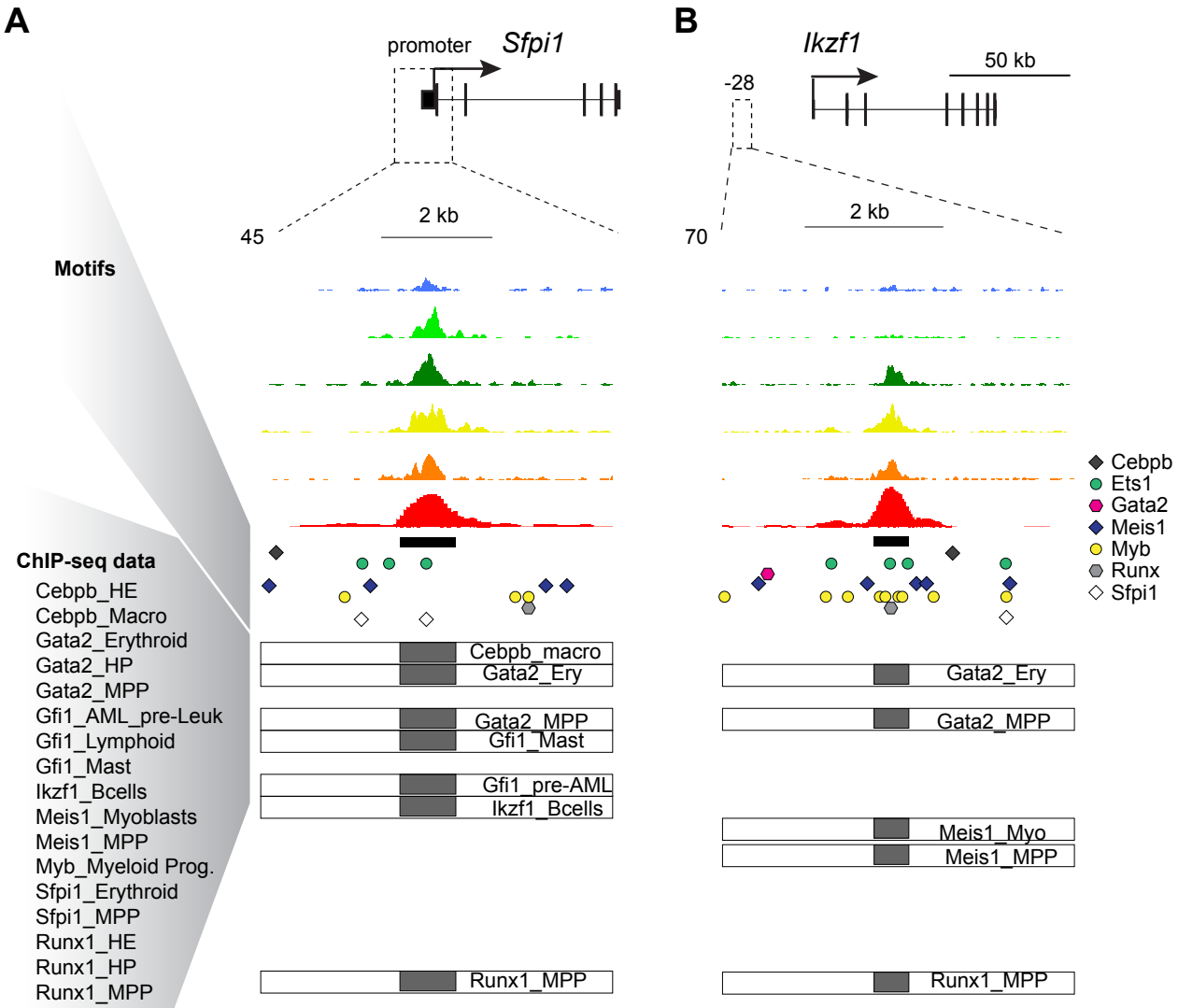


Figure 5.16: Overlay of integrated ATAC-seq/RNA-seq data with in silico ChIP-seq data for sHE to proHSC transition. (A) Schematic of *Sfpi1* locus with proximity to the TSS. Mean ATAC-seq coverage at *Sfpi1* TSS (chr2:90,932,214-90,960,653). Motifs within this genomic locus labeled with different symbols. ChIP-seq peaks from the LOLA analysis showing the different TFs bound at this locus. **(B)** Schematic of *Ikzf1* locus with proximity to the -28 regulatory element. Mean ATAC-seq coverage at *Ikzf1* -28 (chr11:11,548,438-11,712,062). Motifs within this genomic locus labeled with different symbols. ChIP-seq peaks from the LOLA analysis showing the different TFs bound at this locus.

We also identified novel regulatory interactions including new relationships between Runx1 and *Ikzf1* and *Ikzf2*. *Runx1* and *Ikzf1* were connected via RUNX motifs and RUNX1 ChIP-seq binding in an element 28kb upstream of the *Ikzf1* TSS (Figure 5.16B). This element contained Ets1, Myb and Runx motifs and was bound by Gata2, Meis1 and Runx1 (Figure 5.16B). This novel relationship was explored in greater detail in Chapter 6. *Runx1* and *Ikzf2* were connected via two putative cis-regulatory elements that lie >150kb downstream of the TSS (Figure 5.17). *Ikzf2* expression pattern alone suggested that it was a target of Runx1 with expression increasing from 20.79±7.86 in sHE to 132.70±14.19 in proHSC (Figure 5.17). The +170 element is accessible throughout EHT and bound by Runx1, Sfpi1, Gfi1, and Gata2 (Figure 5.17). +190 increased in accessibility as cells became more hematopoietic and could be bound by Runx1, Gata2 and Cebpb (Figure 5.17). Much like *Ikzf1*, *Ikzf2* had not previously been directly implicated in EHT and therefore warranted further study.

5.5 Conclusions and discussion

Our aim for this chapter was to identify gene regulatory interactions regulating EHT. We looked for TFs where the timing of the appearance of its motif coincided with expression of the TF. We hypothesized that these TFs were putative causative regulators of accessibility and subsequent changes in gene expression. First, we looked for regulators globally, across all six populations in our model of EHT. Integrating global expression (Chapter 3) and chromatin (Chapter 4) profiles revealed that of the 413 differentially expressed TFs 43 also had enrichment of their corresponding binding motifs. The presence of TFs with known roles in EHT validated this approach and gave confidence to

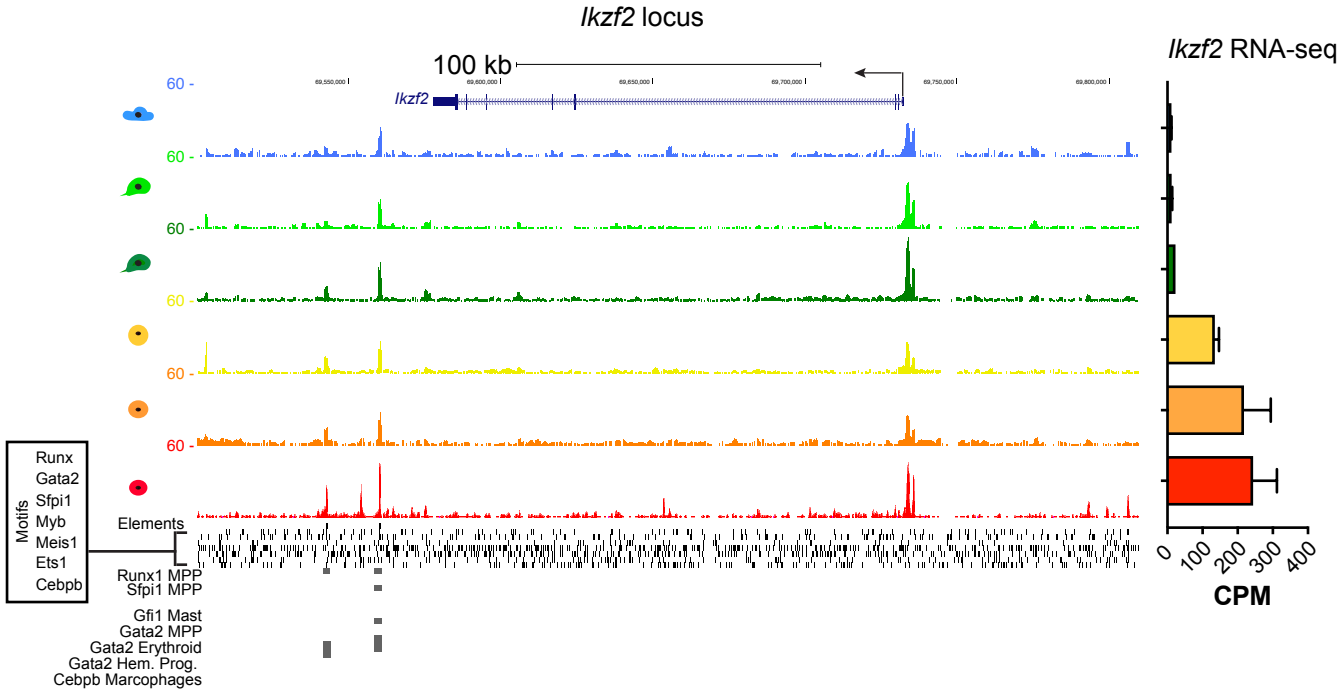


Figure 5.17: Novel regulatory interaction shows *Ikzf2* as key target during sHE to proHSC transition. (A) Mean ATAC-seq coverage at *Ikzf2* (chr1:69,481,064-69,829,254) with RNA-seq gene expression in the same populations. Elements identified as differenitally accessible are label with black bars.

the identification of novel putative regulators. Here, Stat4 was both identified as a putative direct target of Runx1 and as a regulator of pre-HSC maturation.

Second, we focused on two key cell fate transitions during EHT: (1) cHE to sHE and (2) sHE to proHSC transitions. Again, we looked for putative causative regulators of accessibility and subsequent changes in gene expression. Using a two-prong approach, we first integrated expression (Chapter 3) and chromatin (Chapter 4) profiles to identify GRI maps based on the presence of TF motifs in differentially accessible elements mapping ("mapped elements") to differentially expressed TFs. The second prong built upon the first by looking within the "mapped elements" for the presence of TF binding as indicated by ChIP-seq peaks as a proxy for TF binding potential. Each approach facilitated generation of GRI maps for each transition and led to identification of novel regulatory interactions.

For the cHE to sHE transition, of the 99 differentially expressed TFs only 9 also had enrichment of their corresponding binding motifs. Looking specifically at differentially accessible elements associated with TFs, we inferred a regulator interaction when one of the 9 TF motifs was present in a differentially accessible element. The resulting GRI map contained known interactions including those between Runx1 and the Gfi1 genes (Lancrin et al., 2012). In addition, it also identified Klf4 and Stat6 as putative regulators of this transition. This also revealed putative regulator interactions between Runx1 and the PI3K-AKT pathway, suggesting direct Runx1 regulation of effector genes.

For the second prong, we looked for the presence of TF binding (based on ChIP-seq data from the literature) within the differentially accessible elements.

Regulatory interactions were inferred when a TF was capable of binding to a differentially accessible element associated with a differentially expressed TF. Gene regulatory interaction maps based on this approach revealed a putative novel regulatory interaction between Runx1 and Bcl11a.

For the sHE to proHSC transition, based on the first prong, of the 24 differentially expressed TFs only 7 also had enrichment of their corresponding binding motifs. Using the same logic for inferring a regulatory interaction, we constructed GRI maps. From these maps, we identified putative regulatory interactions between Runx1 and Ikzf1 and Ikzf2. GRI maps based on the second prong, reiterated the putative regulatory interaction between Runx1 and Ikzf1 and Ikzf2. In Chapter 6, we explore the regulatory interaction between Runx1 and Ikzf1.

Chapter 6: Runx1-Ikzf1-Notch

In Chapter 5, we generated gene regulatory interaction maps by integrating chromatin and expression profiles from the cellular intermediates of in vivo EHT. From these maps, we identified novel interactions as well as candidate regulators of EHT. One such interaction occurred in an element 28kb upstream of the *Ikzf1* TSS where RUNX motifs implicated Runx1 in *Ikzf1* transcriptional regulation. Runx1 is critically required for the transition from VE-Cadherin+ HE to definitive hematopoietic cells (North et al., 1999; Chen et al., 2009, Boisset et al., 2010; Kissa et al., 2010). Therefore, it is important to understand the Runx1-dependent transcriptional program. Examination of the putative regulatory interaction between Runx1 and Ikzf1 revealed a novel Runx1-Ikzf1-Notch regulatory interaction that likely facilitates generation of definitive hematopoietic cells. Runx1-Ikzf1-Notch interrogation through perturbation revealed complex dynamics. Overall, identification of this interaction highlighted the power of our data to help understand the molecular mechanisms underlying EHT.

6.1 Identification and characterization of the Ikzf1 -28 cis-element

From our predicted gene regulatory interactions governing the sHE to proHSC transition, we identified a candidate cis-regulatory element 28kb upstream of the TSS of *Ikzf1* ("*Ikzf1*-28" or "-28") (Figure 5.13 and 5.15). The *Ikzf1*-28 element showed extensive sequence conservation across different mammalian species, suggesting a role as a gene regulatory element (Figure 6.1). Using chromosome conformation capture, Next Generation Capture-C (Davies et al., 2015), to provide high-resolution promoter-enhancer interaction maps at single-locus resolution, we confirmed that the -28 element strongly

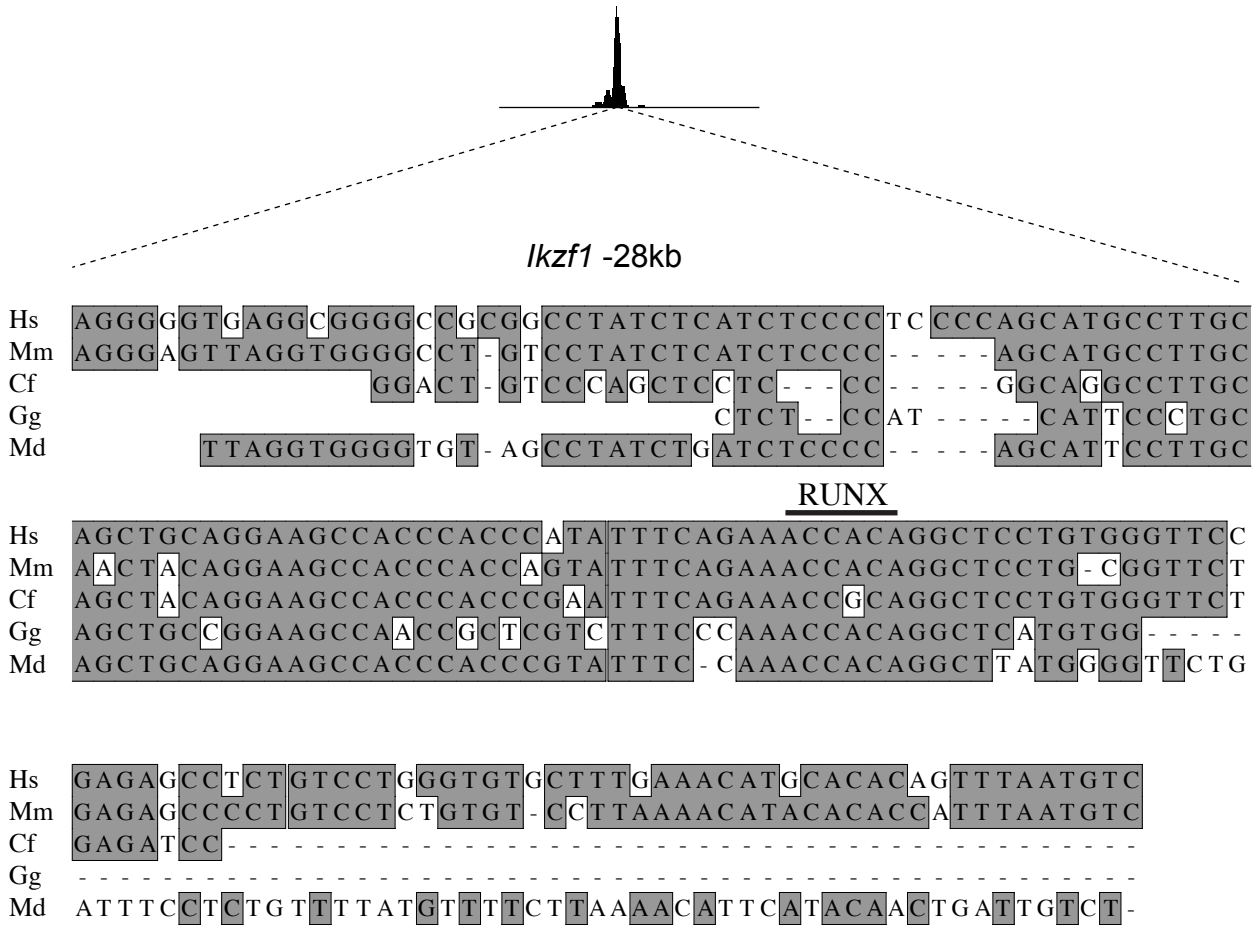


Figure 6.1: Six species alignment of the central 120bp of the -28 Ikzf1 CRE. A 65bp stretch is conserved down to Opossum with 71% identity. Predicted consensus TF-binding site for Runx1 is indicated. Hs indicates *Homo sapiens*; Mm, *Mus musculus*; Cf, *Canis familiariris*; Gg *Gallus gallus*; Md *Monodelphis domestica*.

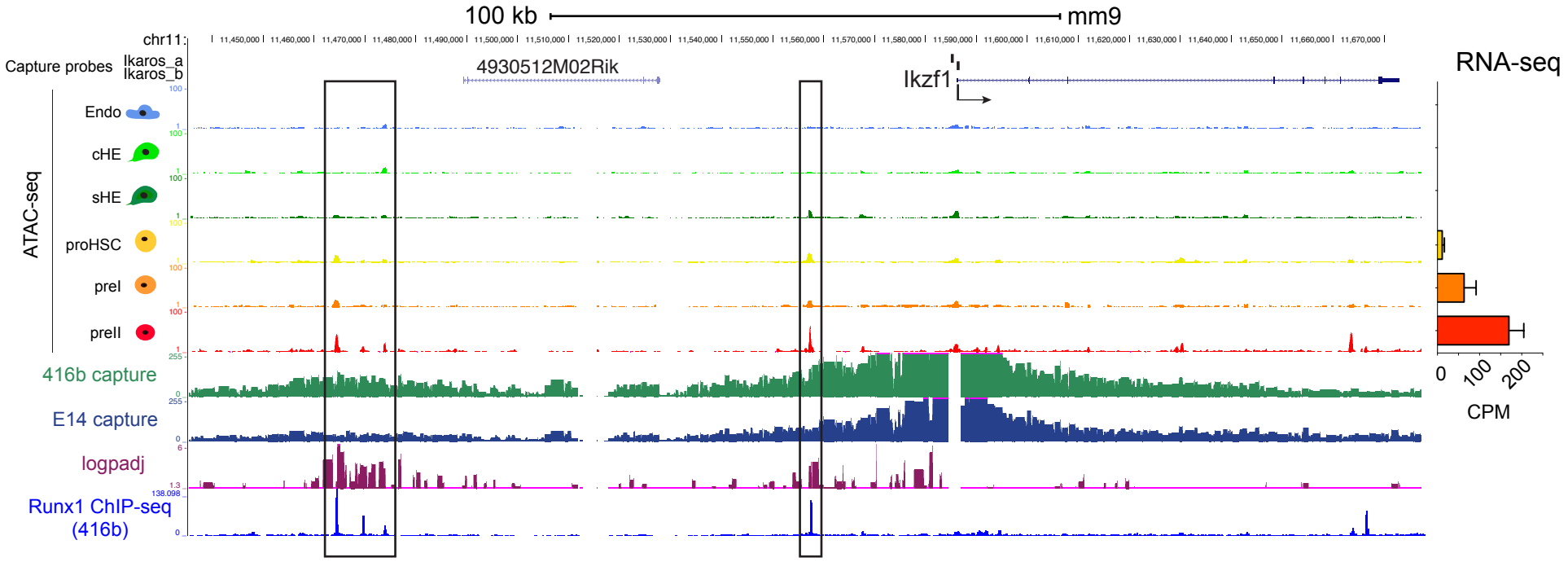


Figure 6.2: ATAC-seq, Capture-C and Runx1 ChIP-seq at *Ikzf1* locus. Mean ATAC-seq coverage at *Ikzf1* (chr11:11,422,091-11,714,751) with RNA-seq gene expression in the same populations. Capture-C using a capture probed designed to the Ikzf1 promoter (dark black bars). Capture in 416b cells (green), capture in undifferentiated E14 mESC (blue), and log adjusted p-values after subtraction of 416b from background E14 (purple). Runx1 ChIP-seq data from Schutte et al., 2016. Capture-C data was generated Dominic Owens (DPhil student in de Bruijn laboratory).

interacts with the *Ikzf1* promoter in 416b cells (Figure 6.2). Interestingly, combining the Capture-C data with Runx1 ChIP-seq data also from 416b cells (last row in UCSC screenshot in Figure 6.2; Schutte et al., 2016) and our ATAC-seq data suggested that Runx1 may facilitate this interaction thereby driving expression of *Ikzf1* (Figure 6.2).

-28 regulates *Ikzf1* expression during mESC differentiation

Before testing to see if Runx1 facilitates expression of *Ikzf1* through the -28, we established whether the 700bp underlying the -28 element contributed to *Ikzf1* expression *in vitro* and *ex vivo*. To test this in an *in vitro* model of EHT, the 700bp of the -28 was deleted from mESC using CRISPR/Cas9 (Figure 6.3A; Appendix Figure 6.1A). Introduction of a single construct containing Wt Cas9 and two gRNAs designed to flank the -28 resulted in both homozygous and heterozygous deletion (Appendix Figure 6.1A, Appendix Figure 6.4B). Sanger sequencing confirmed the deletion (Appendix Figure 6.1B). Multiple homozygous -28 ("-$28^{-/-}$") and heterozygous knockout clones ("-$28^{+/-}$") were expanded and differentiated according to the protocol described in Pearson et al. (Appendix Figure 6.2A-B; Pearson et al., 2015). After day 7 of differentiation, cells were harvested and sorted for analysis by qRT-PCR (Figure 6.3C; Appendix Figure 6.2C-E). Flow cytometry profiles for CD41 and CD45 showed no difference in CD41loCD45+ cells between Wt, -$28^{+/-}$ or -$28^{-/-}$ (Figure 6.3C). However, cells with at least one mutated -28, exhibited an increase $CD41^{hi}$ cells, which was accompanied by a decrease in the $CD41^{lo}$ population (Figure 6.3C). In addition to flow cytometry, RNA was harvested from CD41lo/CD45- and CD41lo/CD45+ to confirm deregulation of *Ikzf1* using qRT-PCR analysis. In -$28^{-/-}$ cells, *Ikzf1* RNA was down regulated by >2-fold in

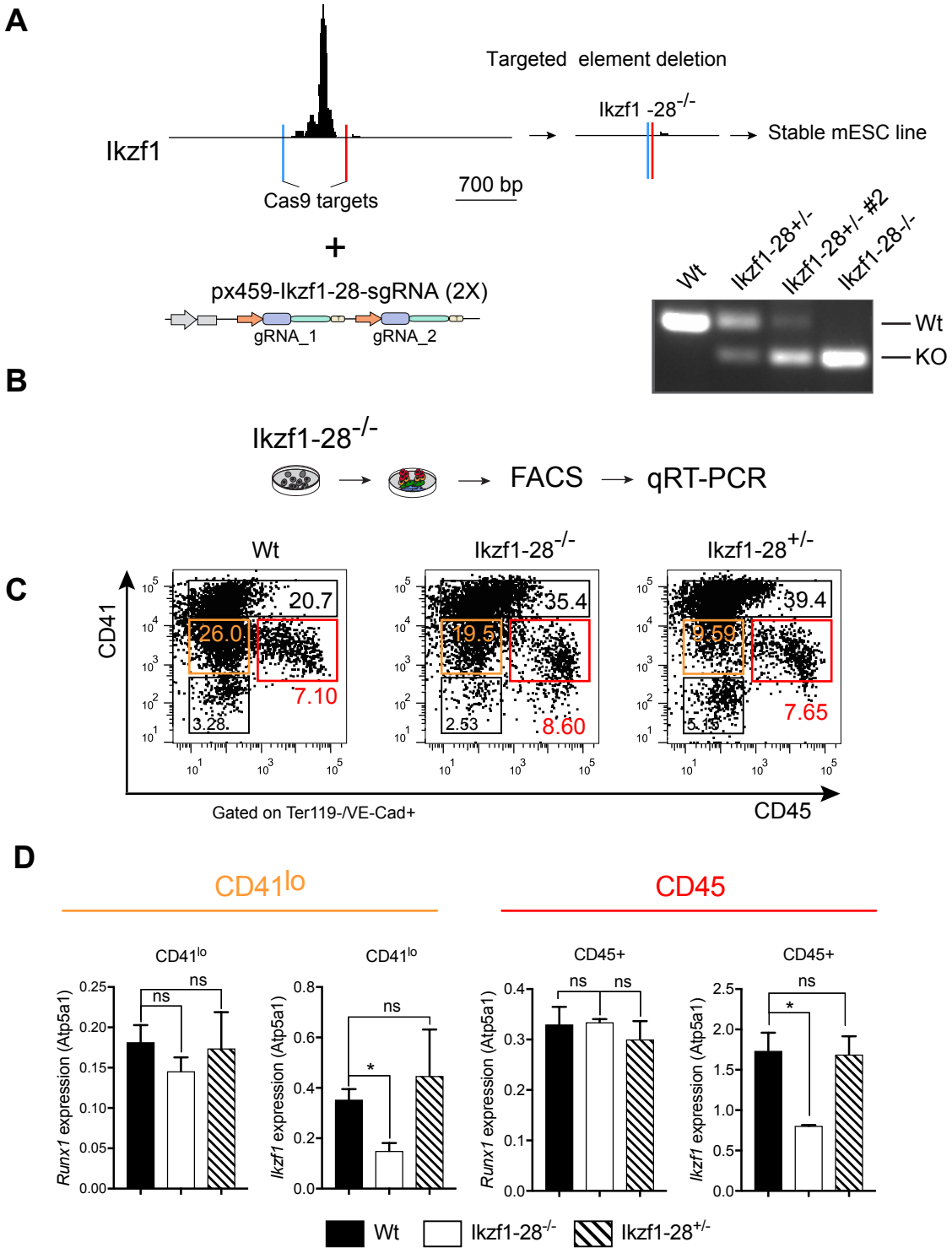


Figure 6.3: Deletion of Ikzf1-28 in mESC. (A) Schematic for targeting Ikzf1-28 using CRISPR/-Cas9 with two gRNAs flanking 700bp of the Ikzf1-28. Gel image confirming heterozygous and homozygous deletions of -28. (B) Experimental outline for assessing Ikzf1-/- mESC. (C) Flow cytometry profiles for Wt, Ikzf1-28+/- and Ikzf1-28-/-. (D) qRT-PCR of 25 sorted cells from CD41lo and CD45 populations at day 7 of differentiation. Generated from 3 indpendent experiments (n=3).

in both CD41lo and CD45+ cells (Figure 6.3D). Interestingly, there was no significant difference between *Ikzf1* levels in heterozygous lines compared to the Wt control (Figure 6.3D). *Runx1* expression was consistent across all lines (Figure 6.3D). Thus, the *Ikzf1*-28 element contributed to proper regulation of *Ikzf1* during *in vitro* EHT.

Ikzf1-28 regulates *Ikzf1* expression *in vivo*

Next, we examined in primary embryonic tissue whether loss of the -28 element affected *Ikzf1* expression during *ex vivo* EHT. In order to do this, we combined CRISPR/Cas9 gene editing with re-aggregation cultures (Figure 6.4A; Taoudi et al., 2008). For *ex vivo* perturbation assays, we utilized a mouse model constitutively expressing Cas9-P2A-GFP (Cas9) from the Rosa26 locus (Platt et al., 2014). Caudal portions from Cas9 embryos were disassociated into single cells suspensions, infected with lentiviruses expressing gRNAs, and cultured *ex vivo* as re-aggregates (Figure 6.4A; Appendix Figure 6.3A-C). Analysis of Cas9-expressing mice confirmed good expression of Cas9 in PAS/AGM by flow cytometry (Appendix Figure 6.3D).

For gRNAs, we re-used gRNAs from the -28 *in vitro* deletion in mESC and designed an additional set of gRNAs flanking the -28 element (Appendix Figure 6.4B; Table 2.4). Control gRNAs targeting a “safe harbor site” were also generated (“Control gRNA”). To generate an Ikzf1$^{-/-}$ whole gene knockout to serve as a positive control for the enhancer knockout, gRNA were designed targeting exon 3 (a 5’ critical exon) to maximize the chances of creating a deleterious Indel (“*Ikzf1*$^{-/-}$ exon 3 gRNAs”) (Appendix Figure 6.4, Vakoc et al., 2014; Table 2.4).

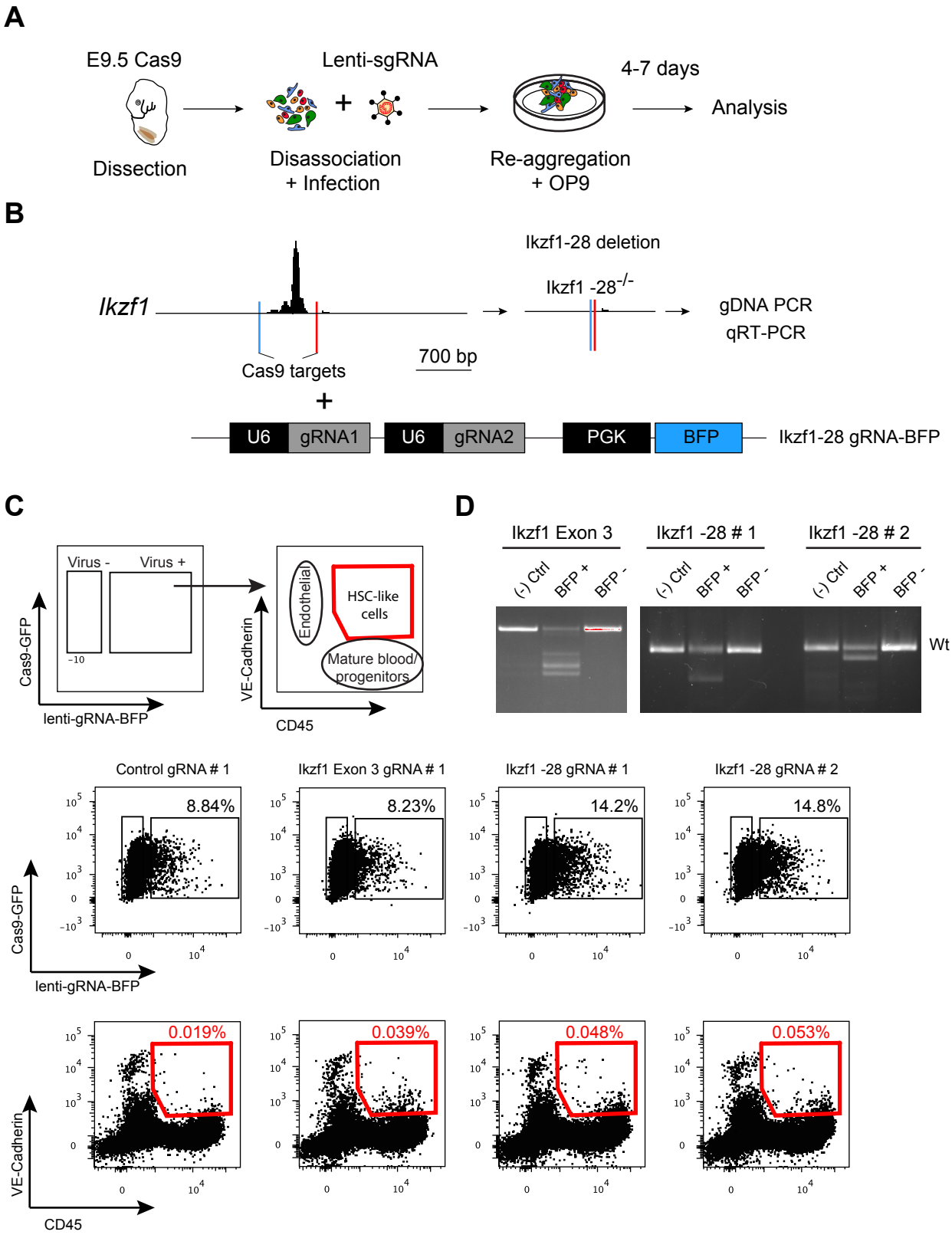


Figure 6.4: Ex vivo targeting of Ikzf1-28 element. Legend on following page.

Figure 6.4: Ex vivo targeting of Ikzf1-28 element. (A) Experimental outline for lentiviral infection of constituitive Cas9-GFP expressing embryp followed by re-aggregation co-culture with OP9. (B) Schematic for targeting Ikzf1-28 using CRISPR/Cas9 with two gRNAs flanking 700bp of the Ikzf1-28. Same gRNAs used for in vitro targeting were cloned into the lentiviral gRNA expression plasmid (U6 - promotoer for gRNA, PGK - promoter for BFP (blue fluorescent proetin). An additional pair of gRNAs flanking Ikzf1-28 were also cloned as an additional experimental set of gRNAs. (C) Flow cytometry profiles for analyzing Cas9-lenti-targeted re-aggregates. Control gRNAs targeted tetR element not present in genome. Ikzf1 exon 3 gRNA #1 served as a positive control for Ikzf1 gene knockout. Two sets of gRNA targeting Ikzf1-28 labeled as Ikzf1 -28 gRNA #1 and #2.

After 7 days in culture following infection (day 0), cultures were analyzed for the presence of VEC+CD45+ HSPCs (Taoudi et al., 2008). Caudal portions treated with Ikzf1$^{-/-}$ gRNAs (targeting exon 3) showed increased numbers of HSPCs compared to cells infected with control gRNAs (Figure 6.4C; Appendix Figure 6.5). Increased numbers of HSPCs in Ikzf1$^{-/-}$ gene knockouts matched data for Ikzf1-null mice where HSPCs accumulated because of a block in B and T cell differentiation (Nichogiannopoulou et al., 1999). Caudal portions infected with either set of -28-targeting gRNAs both recapitulated the same phenotype as the gene knockout—an accumulation of HSPC (Figure 6.4C). Deletion was confirmed by PCR of gDNA from sorted populations (Figure 6.4D). This suggested that in our ex vivo system the -28 element was critical for proper regulation of *Ikzf1* expression.

6.2 Runx1 regulation of Ikzf1 via binding to -28 element

Given the requirement of the -28 for proper regulation of *Ikzf1*, we wanted to know to what extent, if any, Runx1 was responsible for the activity of the -28 element. In order to do this, we first looked at *Ikzf1* expression in the absence of Runx1 in *in vitro* and *ex vivo* EHT. Second, we examined the effect of RUNX1 transcription factor binding site (TFBS) mutation on -28 activity in Luciferase assays.

Runx1 regulation of Ikzf1 during *in vitro* mESC differentiations

First, we looked at *Ikzf1* expression in the absence of Runx1 in *Runx1*$^{-/-}$ mESC. *Runx1*$^{-/-}$ were generated by targeting *Runx1* exon 4 with two gRNAs flanking key amino acids R174 and R177 (Figure 6.5A; Appendix Figure 6.6) (Tahirvo et al., 2001). The aim was to create phenotypes similar to that of Runx1$^{-/-}$ mice (Okuda et al. 1996; Wang et al., 1996). After 7 days of

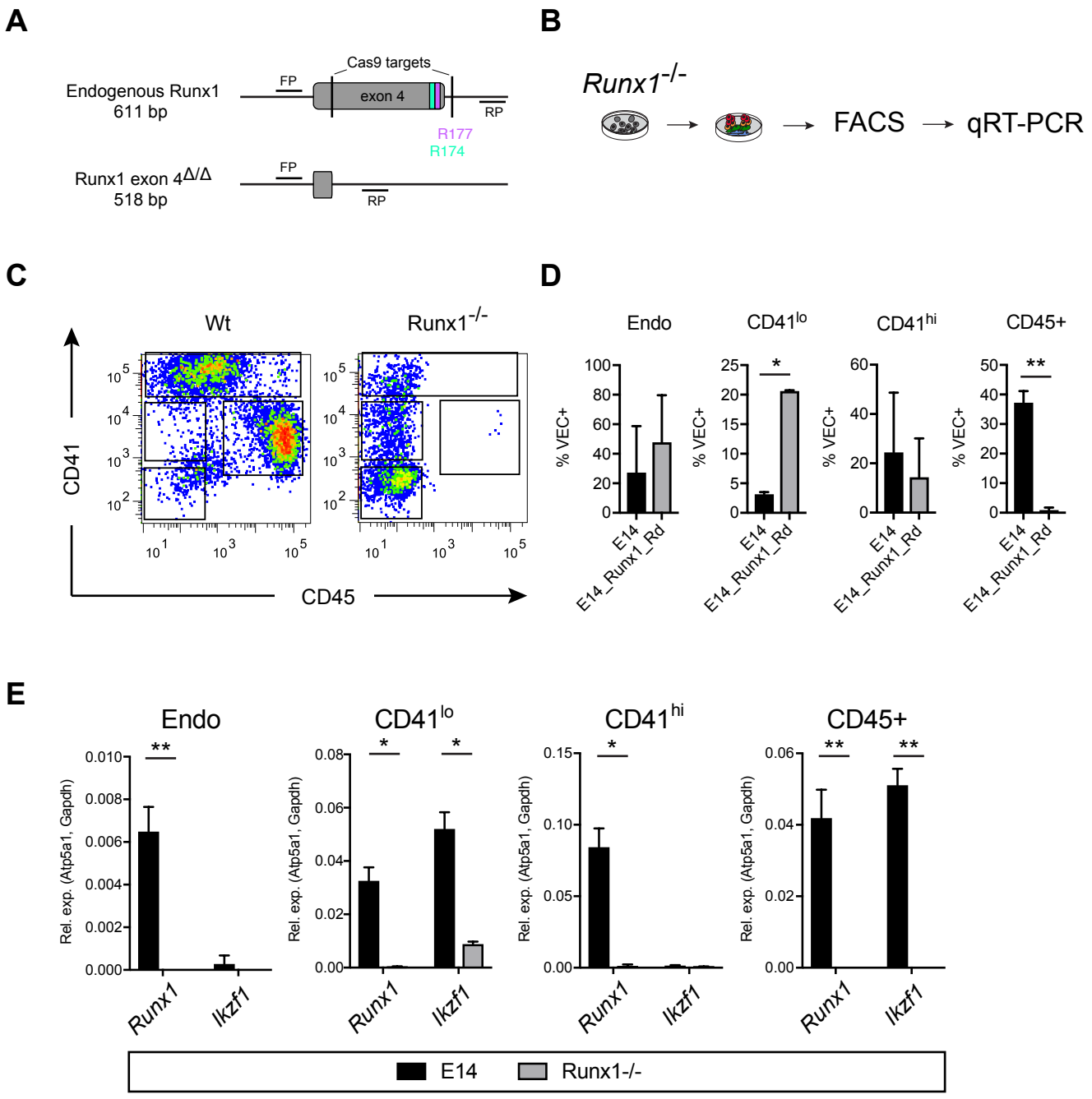


Figure 6.5: Ikzf1 expression in the Runx1-/-. (A) Strategy for using CRISPR/Cas9 to target exon 4 of Runx1 specifically deleting amino acids Arginine (R) 174 and R177. Forward primer (FP) and Reverse primers (RP) used to confirm targeting. (B) Experimental outline for assessing *Ikzf1* expression in Runx1-/- mESC. (C) Flow cytometry profiles of Wt vs. Runx1-/- mESC. (D) Quantification of flow cytometry profiles from C (* indciates pvalue < 0.05, ** indicates pvalue < 0.005). (E) qRT-PCR of 25 cells from sorted populations. Endo defined as Ter119-VE-Cad+; CD41lo, Ter119-VE-Cad+CD41loCD45-; CD41hi, Ter119-VE-Cad+CD41hiCD45-; CD45+, Ter119-VE-Cad+CD41loCD45+ (* indicates pvalue < 0.05, ** indicates pvalue < 0.005). All data is from 3 independent experiments.

differentiation, 25 CD41^{lo} cells were sorted from Wt and *Runx1*$^{-/-}$ mESC and expression was measured using qRT-PCR (Figure 6.5B). As expected, deletion of Runx1 blocked the emergence of CD45+ blood cells (Figure 6.5C-D, Appendix Figure 6.6C). The presence of CD41^{lo} cells despite the absence of functional Runx1 protein was a phenomenon seen by other groups (Figure 6.5C-D; Liakhovitskaia et al., 2014). qRT-PCR confirmed loss of *Ikzf1* expression in mESC lacking functional Runx1 (Figure 6.5E).

Runx1-dependent activation of *Ikzf1* was conserved in Zebrafish. Mutant zebrafish harboring a Runx1$^{W84X/+}$ mutation exhibited dose-dependent effects on blood cell formation (Sood et al., 2010). Work done by Florian Bonkhofer (Patient Lab) in Runx1$^{W84X/+}$ zebrafish showed a dose-dependent effect on *Ikzf1* (Appendix Figure 6.7). In situ hybridization (ISH) analysis at 33 hpf for *Ikzf1* revealed a correlation between *Ikzf1* expression and Runx1$^{W84X/+}$ mutant genotypes (Appendix Figure 6.7). 100% of Homozygous mutants exhibited low Ikzf1 levels (Appendix Figure 6.7). Heterozygous mutants exhibited near Mendelian ratios of *Ikzf1* expression (31% high, 58% medium, 11% low) (Appendix Figure 6.7). Finally, wt fish were classed as having either medium or high *Ikzf1* expression (Appendix Figure 6.7). This suggested that Runx1 activation of *Ikzf1* expression is conserved in Zebrafish, providing further support for Runx1-dependent activation of *Ikzf1* during EHT.

Runx1 regulation of Ikzf1 during ex vivo EHT

Next, to look at how the absence of Runx1 affected *Ikzf1* expression during EHT in the mouse, we utilized the CRISPR/Cas9-based ex vivo perturbation assay described above (Figure 6.6A). The same gRNA used to target *Runx1* in vitro were cloned into the pKLV lentiviral expression vector and were used

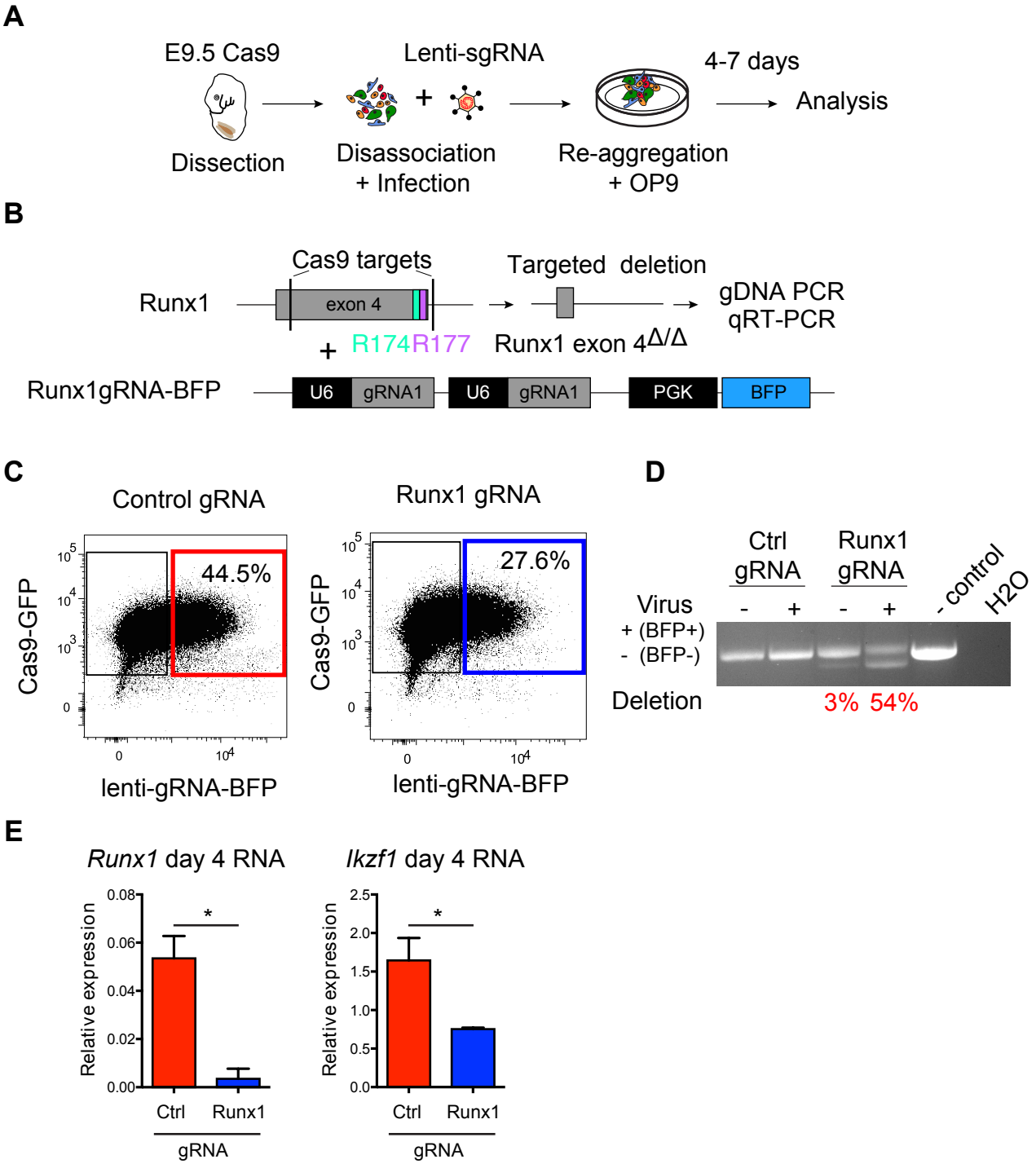


Figure 6.6: Ikzf1 expression in the absence of Runx1 ex vivo. (A) Experimental outline for lentiviral infection of constituitive Cas9-GFP expressing embryos followed by re-aggregation co-culture with OP9. (B) Strategy for ex vivo targeting exon 4 of Runx1, including key amino acids Arginine (R) 174 and R177. Same gRNAs used for in vitro targeting (Figure 6.5) were cloned into the lentiviral gRNA expression plasmid (U6 - promoter for gRNA, PGK - promoter, BFP - blue fluorescent proetin). (C) Flow cytometry profiles for control gRNA and Runx1 gRNA infected Cas9 caudal portions. Virus positive cells are GFP+BFP+. (D) PCR of gDNA to confirm deletion. BFP- (no virus) and BFP+ (virus+) for both control gRNA (Ctrl gRNA) and Runx1 gRNA. Uninfected embryonic material used a PCR control. (E) qRT-PCR of 25 cells from sorted populations. (* indicates pvalue < 0.05) All data is from 3 independent experiments.

to infect single cell suspension of E9.5 caudal portions of Cas9 embryos (Figure 6.6A-B; Table 2.4). After 4 days in culture, flow cytometry revealed robust infection by both control and Runx1-gRNA lentiviruses (Figure 6.6C). gDNA PCR of sorted populations of virus+ (BFP+) cells from each condition confirmed ~50% knockout efficiency at the *Runx1* locus (Figure 6.6D). Analysis of Ikzf1 expression using qRT-PCR on pools of 25 sorted cells showed virus-infected cells had significantly lower levels of *Ikzf1* than uninfected cells (Figure 6.6E), supporting the *in vitro* observations that Runx1 regulated *Ikzf1*. Of note, although targeting Runx1 resulted in a partial block of EHT, this system failed to completely recapitulate the *Runx1*$^{-/-}$-associated block as evidenced by the presence of virus-infected cells within the CD45 compartment (Appendix Figure 6.8A-E). This was likely due to cells already expressing *Runx1* prior to genomic inactivation or from cells where genomic alterations resulted in incomplete inactivation of *Runx1.* In either scenario, prior to genetic inactivation, cells are likely to have enough Runx1 to make it through the developmental window in which Runx1 is required (Chen et al., 2009; Tober et al., 2013). Residual Runx1 left after targeting may also explain *Ikzf1* expression levels in cells treated with *Runx1* gRNA (Figure 6.6E). Overall, despite the caveats of targeting *Runx1* in the Cas9-re-aggregate system, the data supported the idea of Runx1-dependent activation of *Ikzf1* during EHT.

Runx1 dependent activity of Ikzf1-28 cis-element

Finally, we examined the -28 element's ability to drive reporter gene expression in luciferase assays. The -28 contained a conserved RUNX motif at the center of the ATAC element (Figure 6.7A). To see if -28 activity was

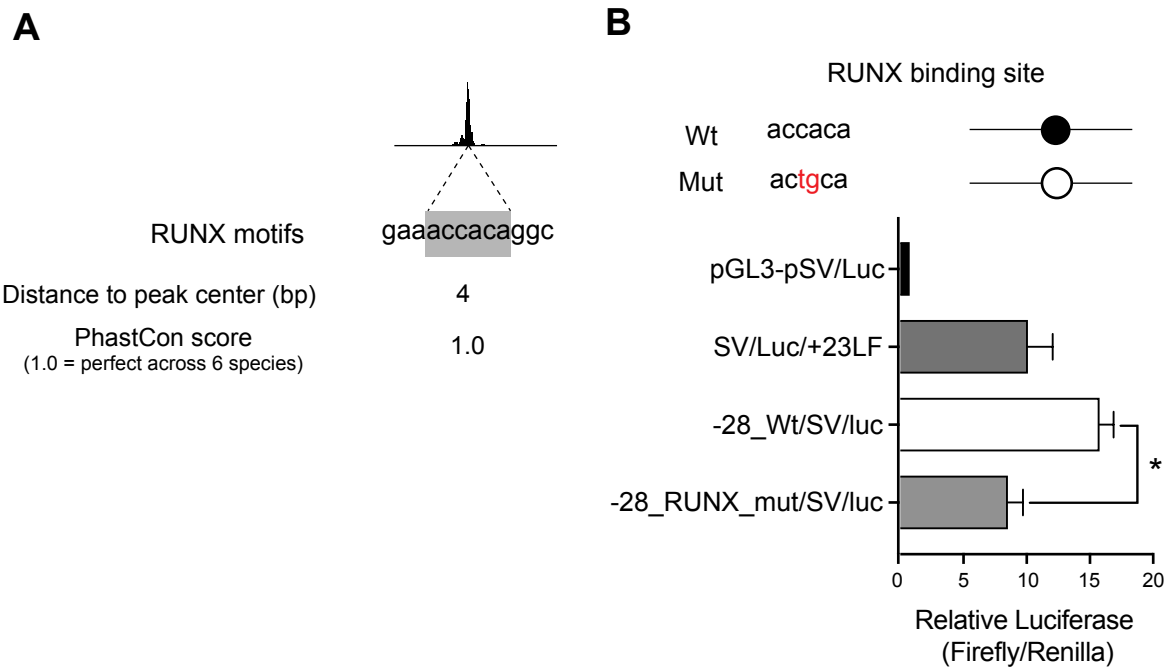


Figure 6.7: Runx motifs in Ikzf1-28 tested tested in luciferase asssay. (A) PhastCon conservation score for RUNX motif 4bp from center of ATAC-seq peak. (B) Luciferase reporter constructs were electroporated into 416b, which express low by sufficient levels of Runx1 protein. Vectors included a positive control construct carrying the Runx1 +23 enhancer fragment (Nottingham et al., 2007), a reporter construct carrying the -28 regulatory element (-28_Wt/SV/Luc) and a construct carrying the -28 element with a mutated RUNX motif (-28_RUNX_mut/SV/Luc).

dependent on this RUNX motif for its activity, luciferase reporter constructs for both the Wt RUNX and RUNX mutant sequences were electroporated into 416B cells. Enhancer activity was significantly reduced (p=0.02) in cells with the RUNX mutant construct (Figure 6.7B), suggesting that at least part of luciferase activity was dependent on Runx1 binding to the RUNX motif.

Summary

In summary, we validated a previously unrecognized regulatory interaction between RUNX1 and *Ikzf1*. Combining enhancer-promoter interaction mapping with deletion of the -28 in vitro and ex vivo, we showed that the -28 interacts with the *Ikzf1* promoter and is critical for its expression. Perturbation of *Runx1* in vitro and ex vivo reduced *Ikzf1* expression thereby establishing *Ikzf1* as a target of RUNX1. Using data from the literature showing RUNX1 binding to the -28 in 416b cells (Figure 6.2; Schutte et al., 2016) combined with luciferase assays showing the importance of intact RUNX motifs for -28 activity, suggested *Runx1*-dependent regulation of *Ikzf1* expression.

6.3 Dynamic regulation of Notch signaling in vivo

Given the importance of *Runx1* during EHT and the identification of its putative target gene *Ikzf1*, we sought a mechanism for how the Runx1-Ikzf1 regulatory interaction regulated EHT. As outlined in earlier Chapters, the sHE to proHSC transition marked a key transition during EHT. *Runx1* expression first appeared in sHE (Figure 3.6 and 3.7) followed by initiation of *Ikzf1* expression in proHSC (Figure 3.7). Moreover, both increased expression from proHSC to preI (Figures 3.15). We reasoned that key regulator events occurring during these two transitions were, in part, product of the *Runx1-Ikzf1* regulatory axis. Specifically, we hypothesized that *Runx1* was acting

through *Ikzf1* to facilitate transcriptional regulation in favor of the hematopoietic lineage.

Runx1-Ikzf1-Notch hypothesis

Revisiting the GREAT analysis at newly open (NO) and newly closed (NC) ATAC elements for the sHE to proHSC transition (Figure 4.11), we noticed enrichment for Notch signaling at NC elements (Figure 6.8A-C). GREAT analysis of NC elements during the proHSC to preI transition identified Notch signaling as the only enriched pathway (Figure 6.8D-F). As stated in Chapter 4, we reasoned that complete loss of accessibility represented the most significant changes in accessibility and were therefore of greatest interest. Several studies have shown the requirement for proper regulation of Notch signaling during HSPC emergence (Reviewed in Clements and Traver 2013). Combining this idea with the increased expression of *Runx1* and its putative target gene *Ikzf1*, we postulated a *Runx1-Ikzf1* gene regulatory interaction that regulated Notch during the sHE to proHSC and proHSC to preI transitions.

Runx1 has previously been shown to bind to *Notch1* cis-elements and potentially act in its regulation (Lichtinger et al., 2012). However, to our knowledge, little evidence existed for direct regulation of Notch transcriptional targets by Runx1. On the other hand, Ikzf1 has been shown to regulate *Hes1* (Hairy enhancer of split 1), a Notch transcriptional target, in T-cells (Kathrein et al., 2008). Combining Runx1's regulation of *Notch1* with Ikzf1's regulation of *Hes1* hinted that the Runx1-Ikzf1 regulatory interaction regulated the *Notch1-Hes1* signaling axis critical for the commitment phase of EHT (Lee et al., 2013).

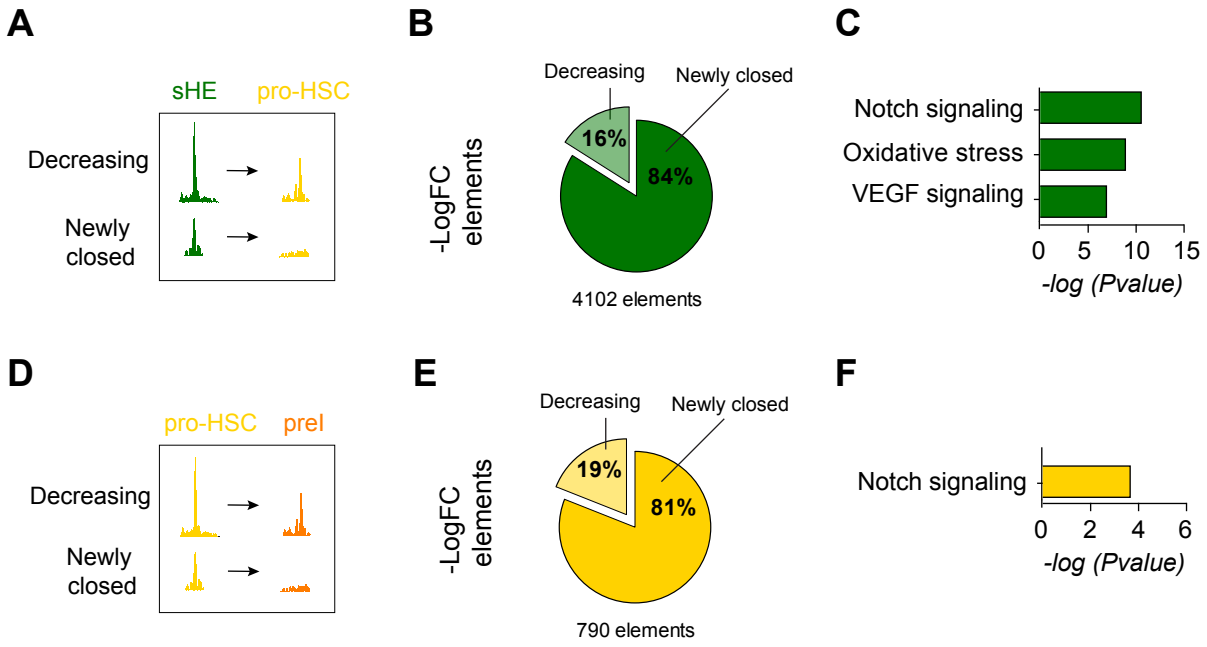


Figure 6.8: Chromatin dynamics during sHE to proHSC and proHSC to preI transitions revealed enrichment of Notch signaling components. (A) Schematic of decreasing and newly closed elements in the sHE to proHSC transition. (B) Percentage of all -logFC elements during the sHE to proHSC transition (4,102 -logFC elements) that were decreasing or newly closed. (C) GREAT analysis showed newly closed elements enriched around genes associated with Notch signaling. (D) Schematic of decreasing and newly closed elements in the proHSC to preI transition. (E) Percentage of all -logFC elements during the proHSC to preI transition (790 -logFC elements) that were decreasing or newly closed. (F) GREAT analysis for newly closed elements showed significant enrichment for only one pathway: Notch signaling.

Notch expression and chromatin profiles during EHT

Prior to testing the hypothesis, we noted that although numerous studies have elegantly described the roles for Notch signaling components during EHT very little is known about Notch expression patterns during early EHT. RNA-seq of Notch components during early EHT showed a general expression pattern of up-regulation from EC to cHE to sHE followed by down-regulation in either proHSC or preHSC I, with near complete silencing in preHSC II (Figure 6.9A). This was true for Notch transcriptional targets (*Hes1*, *Hey1*, *Hey2*, *Gata2, Gata3*, and *Ctnnd1*), Notch receptors (*Notch1*, *Notch3*, and *Notch4*) and Notch ligands (*Dll4*, *Jag1)*, suggesting tight transcriptional regulation was required for all Notch components during EHT (Figure 6.9B).

Analysis of chromatin accessibility at *Hey1*, *Hes1*, *Notch1*, *Notch4*, and *Dll4* showed regions accessible in cHE and sHE were less accessible in hematopoietic cells (Figure 6.9B). For example, within the *Hey1* region regions such as the -2.5kb and -18.5kb had higher signal in sHE then in proHSC (6.9B). Another example within the *Dll4* locus showed accessibility at the promoter was higher in sHE then in proHSC (Figure 6.9B). This was true for other regions within the *Dll4* locus but also for *Notch1* and *Notch 4* loci (Figure 6.9B).

In order to see if Runx1 and Ikzf1 could bind to the accessible chromatin regions and subsequently influence expression, we overlaid Runx1 and Ikzf1 ChIP-seq data onto the ATAC-seq data (Figure 6.9B). Several examples showed binding in elements that changed accessibility. For example, within the *Hes1* locus, the -1.2kb region had higher signal in sHE than proHSC and could be bound by Runx1 and Ikzf1 (Figure 6.9B). The *Hes1* TSS had a

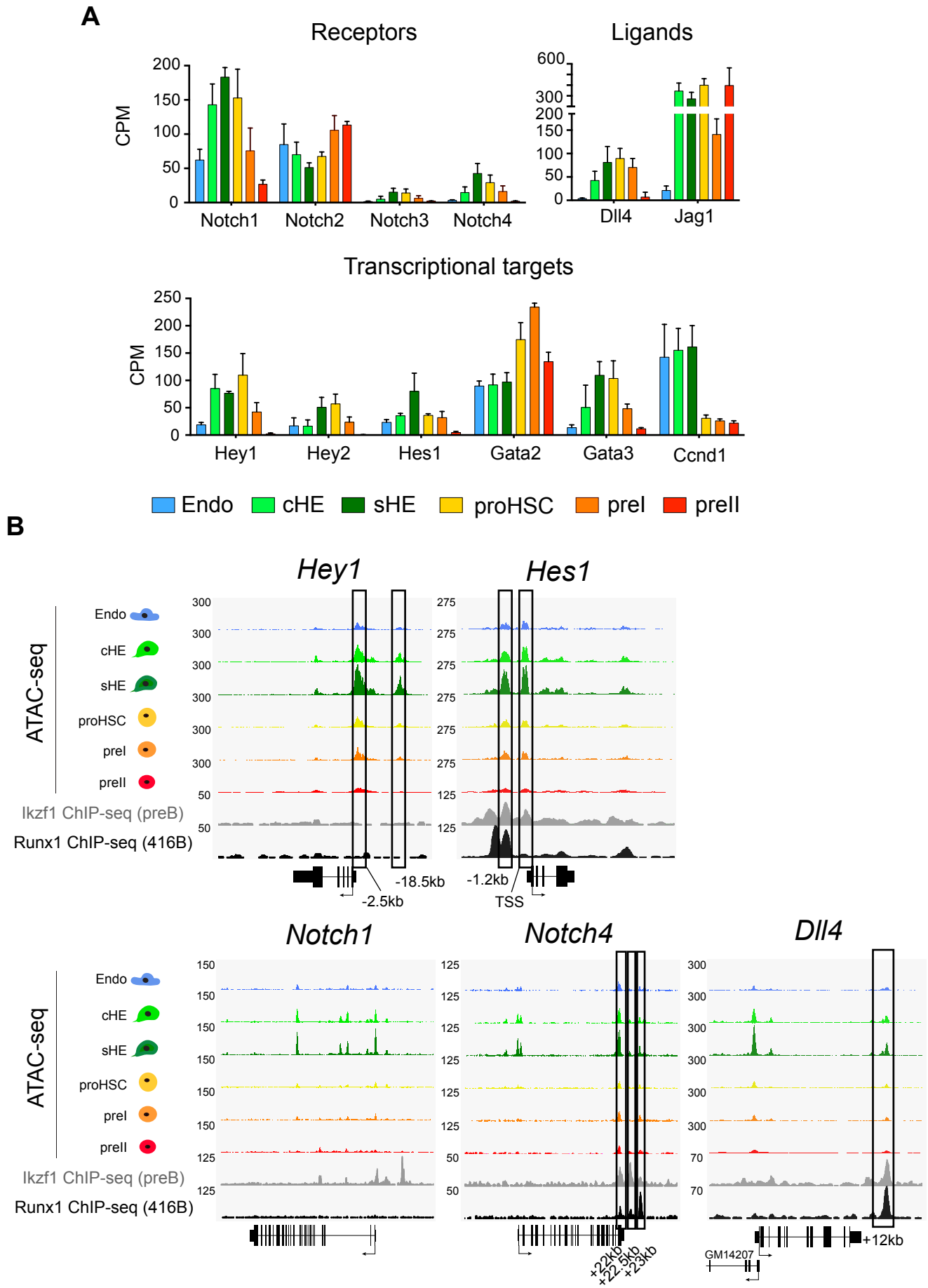

A
Receptors
Ligands
CPM
Notch1
Notch2
Notch3
Notch4
Dll4
Jag1
Transcriptional targets
Hey1
Hey2
Hes1
Gata2
Gata3
Ccnd1
Endo
cHE
sHE
proHSC
preI
preII
B
Hey1
Hes1
ATAC-seq
Ikzf1 ChIP-seq (preB)
Runx1 ChIP-seq (416B)
-18.5kb
-2.5kb
-1.2kb
TSS
Notch1
Notch4
Dll4
+22kb
+22.5kb
+23kb
GM14207
+12kb

Figure 6.9: Expression and chromatin profiles for Notch components during EHT. (A) RNA-seq CPM for receptors, ligands and transcriptional targets. (B) UCSC genome browser views of ATAC-seq profiles for Notch signaling components. Runx1 ChIP-seq (Schutte et al., 2016). Ikzf1 ChIP-seq (Vidal et al., 2013).

similar accessibility pattern but could be bound only by Ikzf1 (Figure 6.9B). In an example from the *Dll4* locus, a +12kb element had higher accessibility in sHE than proHSC and could be bound by both Runx1 and IKZF1 (Figure 6.9B). Lastly, at the *Notch1* TSS higher accessibility in sHE than in proHSC was associated with potential binding by Ikzf1 but not Runx1 (Figure 6.9B). There was no clear correlation between changes in accessibility and binding by either Runx1 or Ikzf1. This may be attributable to the ChIP-seq and ATAC-seq data sets coming from different cell types. Despite the lack of correlation, the specific examples discussed above may be useful in interpreting how, if at all, a Runx1-Ikzf1 interaction regulates Notch signaling components. If Runx1-Ikzf1 does regulate Notch during EHT, regulation likely consists of multiple mechanisms whereby each act independently but also in a synergistic manner.

6.4 sc-qRT-PCR for *Runx1*, *Ikzf1* and Notch target during in vivo EHT

In order to test the Runx1-Ikzf1-Notch regulatory interaction hypothesis *in vivo*, gene expression analysis was performed on single sorted cells isolated from the same phenotypic populations as was done for the RNA-seq and ATAC-seq, excluding E8.5 (Figure 6.10A; Appendix Figure 6.9). Single-cell gene expression analysis was performed on 121 cells for 11 genes (4 house keepers, *Runx1*, *Gfi1*, *Gfi1b*, *Myb*, *Hes1*, *Hey1*, *Ikzf1*) (Figure 6.10B). As expected, *Runx1* expression was highest in pro- and preHSCs (Figure 6.10B-C). *Runx1* transcriptional targets *Gfi1*, *Gfi1b* and *Myb* were included as controls to see how these genes behaved in relation to *Runx1*, *Ikzf1* and Notch targets. As expected, *Myb* was expressed in all pro- and preHSCs (Figure 6.10B). Interestingly, at E9.5, *Gfi1* and *Gfi1b* were almost mutually

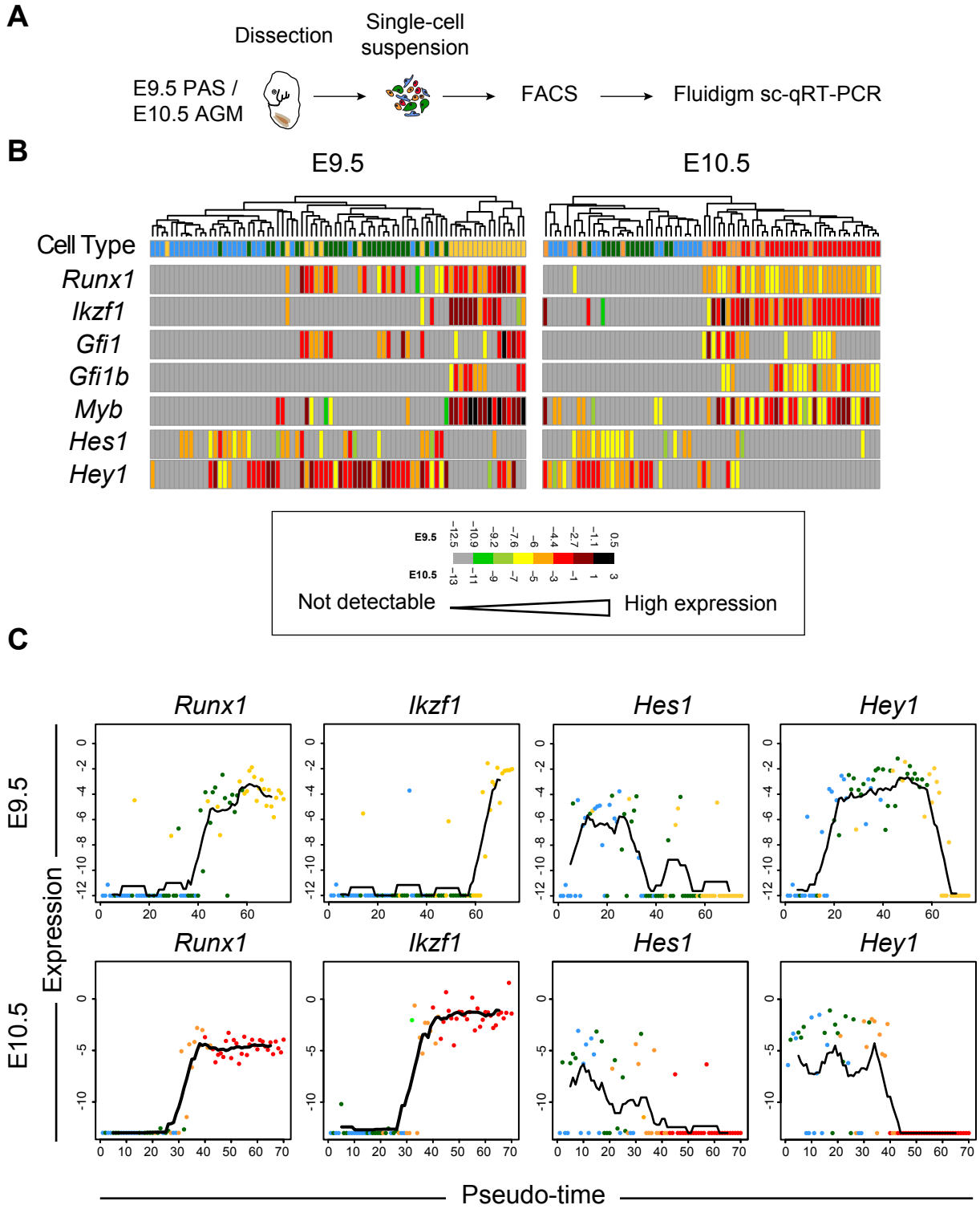


Figure 6.10: Single cell gene expression in EHT cellular intermediates at E9.5 and E10.5. (A) Experimental outline for single cell qRT-PCR . (B) Heatmaps showing expression of 11 genes in single cells for E9.5 Endo (Ter119-VE-Cad+CD41-CD45-23GFP-), sHE (Ter119-VE-Cad+CD41-CD45-23GFP+) and pro-HSC (Ter119-VE-Cad+CD41loCD45-) and E10.5 Endo (Ter119-VE-Cad+CD41-CD45-23GFP-), HE (Ter119-VE-Cad+CD41-CD45-23GFP+), preI (Ter119-VE-Cad+CD41loCD45-23GFP+) and preII (Ter119-VE-Cad+CD41loCD45+23GFP+). The data are ΔCt to the mean of Atp5a1 and Ubc reference genes. (C) Pseudotime plots for genes *Runx1*, *Ikzf1*, *Hes1* and *Hey1* at E9.5 and 10.5. Cells are colored according to cell surface profile (Endo - blue, sHE - green, proHSC - yellow, preI - orange, and preII - red). The data are Δ Ct to the mean of Atp5a1 and Ubc reference genes .

exclusive with different subsets of proHSCs expressing either gene (Figure 6.10B). At E10.5, when ordered using pseudo-time, *Gfi1* was expressed only in the earliest pre-HSCs compared to *Gfi1b* expressed in the later preHSC (Figure 6.10B).

Ikzf1 was expressed only in pro- and pre-HSCs (Figure 6.10C). At E9.5, *Ikzf1* was expressed in a subset of *Runx1* expressing cells suggesting that it is in proHSCs where RUNX1 activated *Ikzf1* expression (Figure 6.10B). At E9.5, Notch transcriptional targets *Hey1* and *Hes1* showed expression in EC and sHE but were down regulated in most of the proHSCs (Figure 6.10C). Providing support for our hypothesis, notch transcriptional targets *Hey1* and *Hes1* were completely down regulated in cells co-expressing *Runx1* and *Ikzf1* (Figure 6.10B). Single cell expression data from EHT at E11 confirmed co-expression of Runx1 and Ikzf1 concomitant with a down-regulation in Notch transcriptional targets (Appendix Figure 6.10; Zhou et al., 2016). In summary, single cell gene expression supported the hypothesis that *Runx1* and *Ikzf1* negatively regulate Notch targets during EHT.

6.5 Investigation of Runx1-Ikzf1-Notch regulatory interaction

To identify the mechanism by which *Runx1* and *Ikzf1* regulate Notch, we sought to perturb expression of each independently and in combination to test affect on Notch signaling. Ideally, we sought a system whereby perturbation of at least one of the genes would be controllable. Therefore, we developed a mESC differentiation system whereby Runx1 was fused to an ERt2 domain, which facilitated control over subcellular localization of Runx1. Using this tool, in combination with knocking out *Ikzf1* expression using CRISPR/Cas9, we identified a complicated regulatory interaction whereby Runx1 and Ikzf1 work

independently, as well as synergistically, to regulate different Notch components during in vitro EHT.

Runx1-ERT2 fusion to study Runx1-dependent events

Identification of Runx1's role *in vitro* was studied by fusing an ERt2 (estrogen receptor's ligand binding domain) to the C-terminus of endogenous *Runx1* (Figure 6.11A). ERt2-fused proteins were sequestered to the cytoplasm in the absence of 4-hydroxytamoxifen (4-OHT). Upon addition of 4-OHT, translocation to the nucleus and subsequent DNA binding occurs within minutes for some ERt2-fused proteins (Liang et al. 2017). We targeted a 3X-Flag-ERt2 into the 3' end of Runx1 immediately preceding the stop codon (Appendix Figure 6.11A-B). Targeting was confirmed by long range PCR with primers outside the homology arms (Appendix Figure 6.11C). Targeting in mESC was facilitated by use of a small molecule to inhibit non-homologous end joining that enhanced targeting efficiency resulting in 15.2% homozygous and 36.4% heterozygous clones (Appendix Figure 6.11D).

To access stringency of ERt2 control over the cellular location of Runx1, targeted mESC were differentiated and assessed for hematopoietic output in the presence or absence of 4-OHT (Figure 6.11B; Appendix Figure 6.12B). After day 7, confocal microscopy confirmed that in the absence of 4-OHT Runx1 was restricted to the cytoplasm (Appendix Figure 6.12A). .As expected, sequestered Runx1 was unable to activate or repress its transcriptional targets *Sfpi1* and *Myb* as measured by qRT-PCR on CD41lo cells sorted from the differentiation (Appendix Figure 6.13). Ultimately, this prevented generation of CD45+ hematopoietic cells in differentiations where

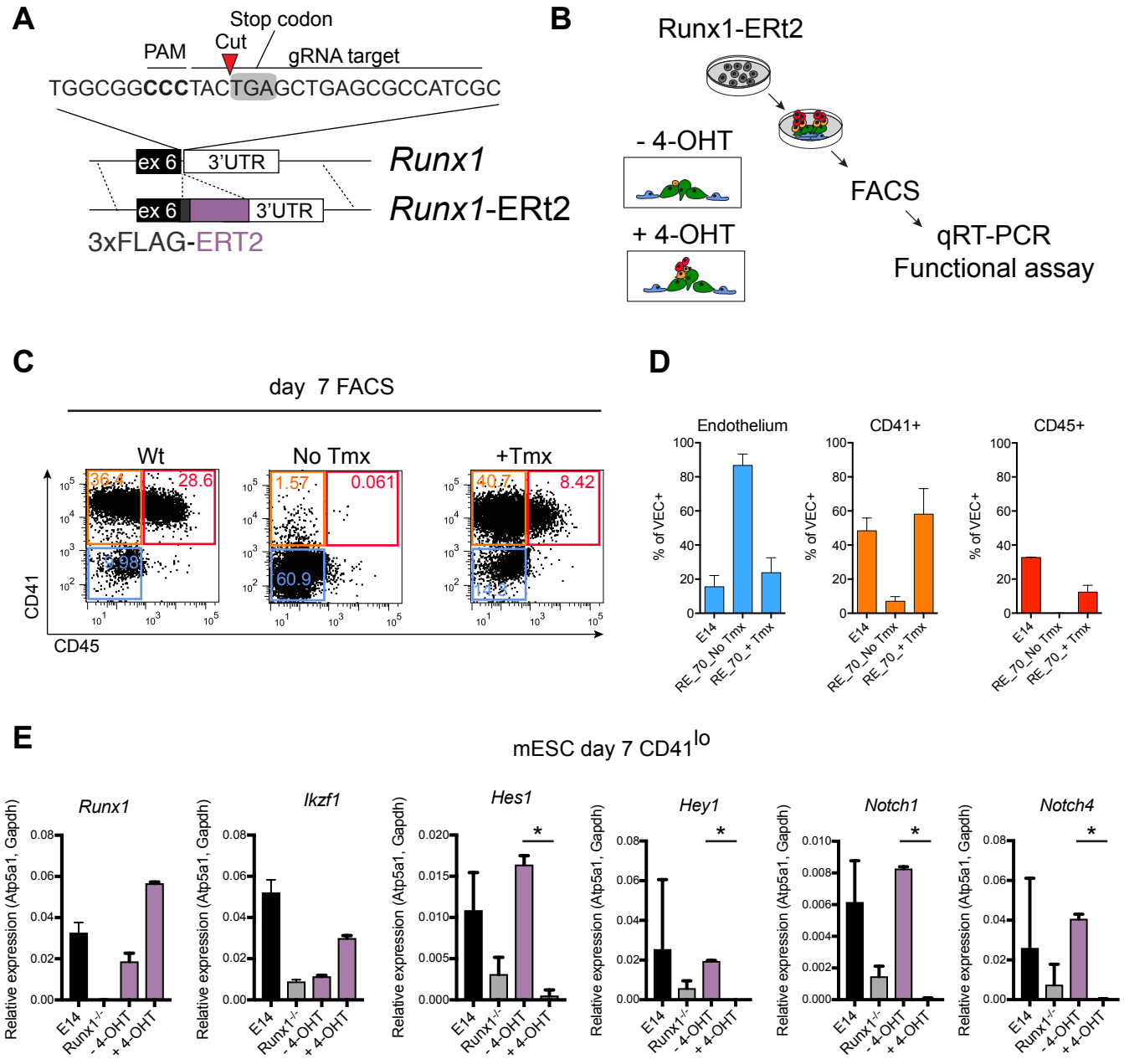


Figure 6.11: Runx1-ERt2 requires 4-OHT for proper control of downstream targets. (A) Targeting strategy inserting ERt2 domain just prior to stop codon in Runx1 locus. (B) Experimental strategy for assessing 4-OHT control of Runx1 cellular location. (C) Flow cytometry profile of Wt, Runx1-ERt2+4-OHT and Runx1-ERt2-4-OHT. (D) Quantification of flow cytometry profiles from C. (E) qRT-PCR of 25 cells from sorted CD41lo, Ter119-VE-Cad+CD41loCD45- for each gene. (* indicates pvalue < 0.05) All data is from 2 independent experiments.

4-OHT was not present (Figure 6.11C-D). Expression of hematopoietic genes upstream of Runx1, including *Tal1* and *Gata2*, showed differentiations proceeded towards the hematopoietic lineage before being halted by Runx1-ERt2 sequestration to the cytoplasm. Co-culture of sorted VE-Cad+ cells from Runx1-ERt2 on OP9 in the presence of hematopoietic cytokines confirmed that 4-OHT controls functional hematopoietic potential (Appendix Figure 6.12C). *Runx1* RNA was detectable in the absence of 4-OHT but was expressed at lower levels when compared to Wt or cells treated with 4-OHT, which was likely due to the strong auto-regulatory nature of its expression (Figure 6.11E). In conclusion, Runx1-ERt2 provided a validated, controlled tool for perturbing Runx1 activity during EHT.

RUNX1-ERt2 controls Ikzf1 and Notch targets

To access RUNX1's role in regulating Notch during EHT, we differentiated RUNX1-ERt2 mESC and measured *Ikzf1* and Notch expression with and without 4-OHT. Addition of 4-OHT corresponded to an increase in *Ikzf1* expression, providing further support that *Ikzf1* is a downstream target of *Runx1* (Figure 6.11E). When 4-OHT was present, *Ikzf1* expression was similar to Wt levels (Figure 6.11E). Similarly, Notch signaling components were mis-regulated in the absence of 4-OHT (Figure 6.11E) but remained at WT in the presence of 4-OHT. For example, *Hes1*, *Hey1*, *Notch1* and *Notch4* were all up regulated in CD41+lo cells in the absence of 4-OHT, meaning that Runx1 translocation to the nucleus was required for proper down regulation (Figure 6.11E).

Generation of Ikzf1-/- mESC line

To determine if Runx1 directly regulated the notch components or did so through, or in synergy with, Ikzf1, we knocked out *Ikzf1* in the Runx1-ERt2 mESC line and measured expression of Notch components in the presence and absence of 4-OHT at day 7 of mESC differentiation. If both *Runx1* and *Ikzf1* were needed for proper regulation of Notch, then we would expect mis-regulation of Notch in the Runx1-ERT2::Ikzf1$^{-/-}$ mESC line in the presence of 4-OHT. To generate Ikzf1 knockouts, we targeted exon 3 because of its presence in all *Ikzf1* isoforms, and data from the literature showing targeting of early protein coding domains increased the chance of gene knockout using CRISPR/Cas9 (Vakoc et al. 2014) (Figure 6.12A). Using constructs expressing 2 sgRNA, we generated Ikzf1-/- knockouts in Runx1-ERt2 mESC using the same 2 gRNAs that were generated for ex vivo targeting but cloned into a pX459 vector (Appendix Figure 6.4A; Table 2.4). Surveyor assays were used to detect targeted mESC clones (Appendix Figure 6.15). Deletions were confirmed by Sanger sequencing (Figure 6.12B), and those with predicted deleterious mutations were selected for further characterization (Figure 6.12C).

The lack of suitable antibodies for Ikzf1 limited the ability to characterize the Runx1-ERt2::Ikzf1$^{-/-}$ mESC. Moreover, the targeting strategy deleted the 3' end of exon, which may have induced exon skipping or other isoform-related problems. Without knowing which RNA was being transcribed, selecting appropriate qPCR primers proved difficult. Using either Ikzf1_778 or Ikzf1_882 yielded different expression levels (Figure 6.12D; Table 2.7). As a result, we relied on the ability to produce B cells as an indicator of the

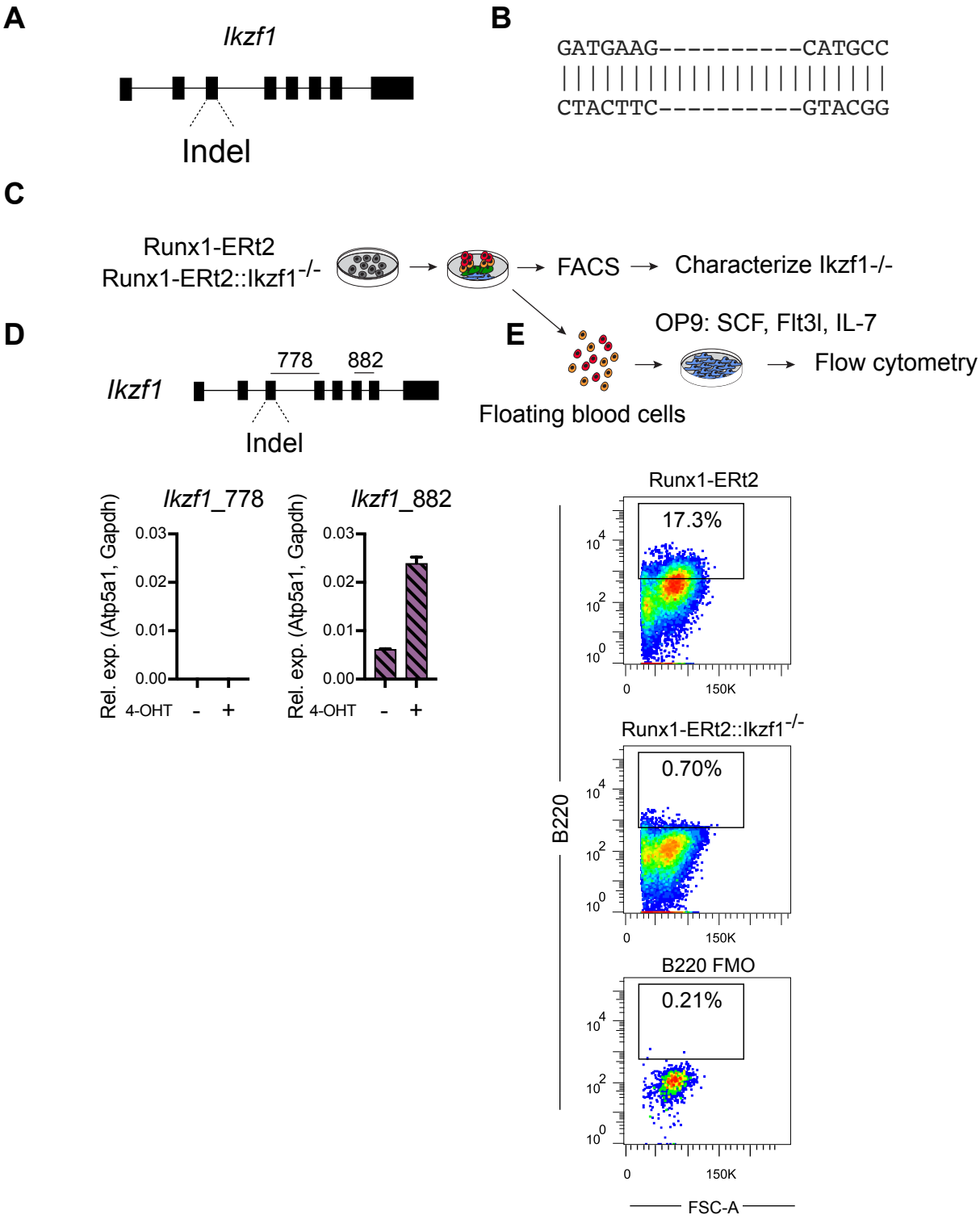


Figure 6.12: Generation and characterization of Runx1-ERt2::Ikzf1-/- mESC. (A) Targeting strategy for creating Runx1-ERt2::Ikzf1-/- mESC. (B) Sanger sequencing track confirmed deletion. (C) Experimental outline for characterizing and analyzing targeted mESC. (D) qRT-PCR of 25 cells from sorted CD41lo, Ter119-VE-Cad+CD41loCD45- for two different Ikzf1 probes. Ikzf1_778 confirmed deletion but Ikzf1_882 confirmed one isoform was still expressed. (E) Flow cytometry profile of non-adherent blood cells from mESC differentiation of Runx1-ERt2 +4-OHT and Runx1-ERt2::Ikzf1-/- +4-OHT taken after 7 days of differentiation plus an additional 7 days of OP9 + cytokine co-culture. Gates were drawn according to FMO for B220.

presence of functional Ikzf1 protein (Nichogiannopoulou et al., 1999). Runx1-ERt2::Ikzf1$^{-/-}$ mESC were differentiated for 7 days and then co-cultured on OP9s in the presence of cytokines (SCF, IL-7, Flt3l) (Figure 6.12E). We observed significant decreases in B220+ cells in Runx1-ERt2::Ikzf1$^{-/-}$ mESC with 0.70% compared to Runx1-ERt2 mESC with 17.3% (Figure 6.12E), suggesting that very little, if any, functional Ikzf1 protein was being produced. Validated RUNX1-ERt2:: Ikzf1$^{-/-}$ lines were then subjected to differentiations with and without 4-OHT.

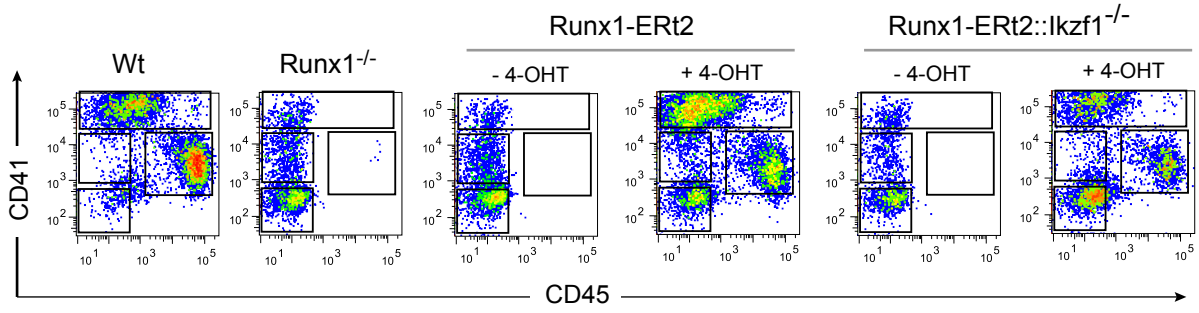
Notch regulation is dependent on Runx1 and Ikzf1

Next, we assessed Notch signaling components in the presence/absence of Runx1 or Ikzf1, or both (Figure 6.13A). After 7 days of differentiation, 25 cells of CD41+lo cells from each wt, Runx1$^{-/-}$, Runx1-ERt2 and Runx1-ERt2::Ikzf1$^{-/-}$ were sorted and subjected to qRT-PCR (Figure 6.13A). No significant differences in cell surface phenotypes were observed for Runx1-ERt2 and Runx1-ERt2::Ikzf1$^{-/-}$ mESC (Figure 6.13B). This supported earlier work targeting *Ikzf1* in Cas9 re-aggregates that showed no negative impact on production of CD45+ HSPCs. Here, analysis of gene expression profiles showed that in the presence of Runx1 (+4-OHT) but the absence of Ikzf1, both *Hes1* and *Notch1* were misregulated (Figure 6.13C). Both *Hey1* and *Notch4* were properly regulated regardless of the presence of *Ikzf1* (Figure 6.13C). Gene expression profiles for *Tal1* and *Gata2* showed the transcriptional program upstream of *Runx1* was not interrupted by the absence of 4-OHT (Appendix Figure 6.13). Given the presence of *Tal1* and *Gata2* in the differentiations, this suggests that neither of these genes was capable of activating *Ikzf1* (Appendix Figure 6.13). As a positive control for

A



B



	Wt	Runx1$^{-/-}$	Runx1-ERt2 -4-OHT	Runx1-ERt2 +4-OHT	Ikzf1$^{-/-}$ -4-OHT	Ikzf1$^{-/-}$ +4-OHT
Endo	27.3±31.4	47.8±31.8	34.2±15.7	22.1±22.5	51.0±25.1	45.4±33.2
CD41^{lo}	3.1±0.3	20.6±0.14	36.8±3.2	8.4±7.0	13.6±0.1	6.15±0.3
CD41^{hi}	24.4±25.2	14.3±15.8	13.6±8.3	36.15±34.9	11.55±9.12	25.5±22.48
CD45	37.2±3.9	0.9±0.8	0.0±0.0	22.6±6.9	0.1±0.0	14.0±7.3

C

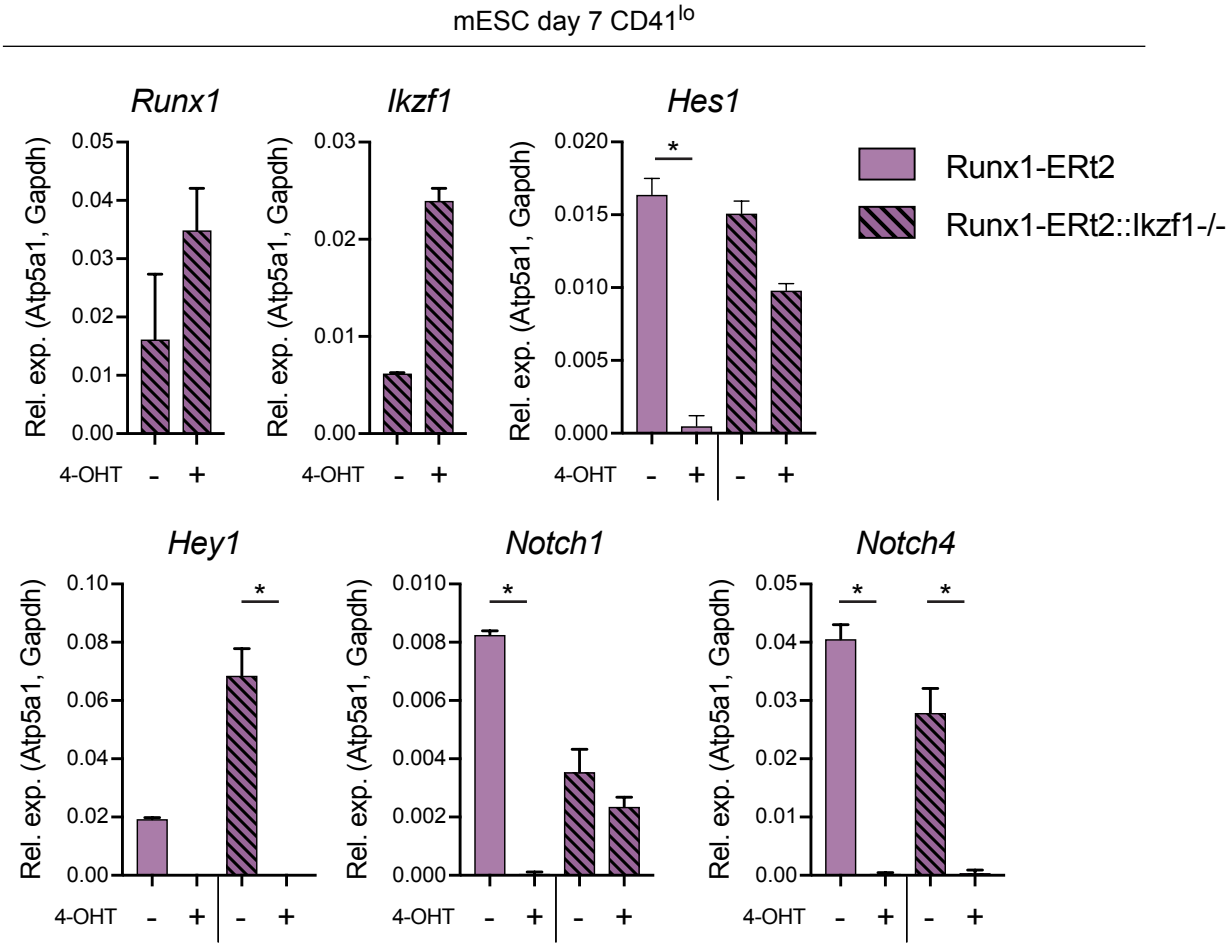


Figure 6.13: Runx1-ERt2 requires 4-OHT for proper control of downstream targets. (A) Experimental strategy for assessing Runx1-ERt2::Ikzf1-/- +/- 4-OHT. (B) Flow cytometry profile of Wt, Runx1-/-, Runx1-ERt2 +/- 4-OHT and Runx1-ERt2::Ikzf1-/- +/-4-OHT. Quantification of flow cytometry profiles with +/- standard deviation. (C) qRT-PCR of 25 cells from sorted CD41lo, Ter119-VE-Cad+CD41loCD45- for each gene. (* indicates pvalue < 0.05). All data is from 2 inde pendent experiments.

Runx1 sequestration, expression profiles for Runx1 targets *Sfpi1* and *Myb* show that in the absence of 4-OHT Runx1 is not capable of activating its transcriptional targets (Appendix Figure 6.13).

Taken together, this suggests the Runx1-Ikzf1 regulatory axis used different mechanisms to control different subsets of Notch signaling during EHT. With *Hes1* and *Notch1*, the Runx1-Ikzf1 axis used a Coherent type 3 feedforward loop (FFL) whereby Runx1 partially regulated *Hes1* and *Notch1* in the absence of IKZF1, but both Runx1 and Ikzf1 were required for complete down regulation (Alon et al. 2007). Complete down-regulation of *Hey1* and *Notch4*, regardless of functional Ikzf1 suggests that Runx1 is sufficient to regulate these notch components during EHT. On going work in the lab aims to further define these relationships.

6.6 Discussion

In Chapter 5, we identified a regulatory interaction between *Runx1* and *Ikzf1* via an element -28kb upstream of *Ikzf1*. In this Chapter, we validated -28-dependent regulation of Ikzf1 expression during in vitro and ex vivo EHT. -28 activity was dependent, in part, on RUNX motifs. Perturbation of *Runx1* expression or cellular location resulted in an absence of *Ikzf1* expression. Taken together, this supports the hypothesis that Runx1-dependent regulation of Ikzf1 occurs, at least in part, through the -28 element.

Runx1-Ikzf1-Notch regulatory cascade

Further analysis of chromatin dynamics and differential expression first discussed in Chapters 3 and 4 revealed Notch signaling enrichment during key stages of EHT. Given the requirement for Runx1 and previous reports identifying Ikzf1 as a regulator of Notch signaling, we hypothesized that the

Runx1-Ikzf1 axis regulated Notch signaling during EHT. Observations using sc-qRT-PCR revealed a correlation between *Runx1* and *Ikzf1* expression and decreased expression of Notch transcriptional targets. Genetic perturbation of both *Runx1* and *Ikzf1*, independently and in tandem, revealed a complex regulation of Notch signaling during in vitro EHT. *Runx1* alone was sufficient to drive regulation *Hey1* and *Notch4*. However, In the case of *Hes1* and *Notch1*, both *Runx1* and *Ikzf1* were needed for proper down regulation.

Chapter 7: General discussion

7.1 Summary of results

The aim of my thesis was to further our understanding of the molecular mechanisms underlying EHT by identifying putative regulators governing gene regulatory interactions during EHT. We hypothesized that integrating genome-wide expression (RNA-seq) and chromatin (ATAC-seq) profiles of the cellular intermediates undergoing EHT would facilitate identification of candidate regulators. After validation and characterization of expression profiles, we identified ~6,000 genes differentially expressed throughout EHT (**Chapter 3**). Of particular interest were the 413 differentially expressed TFs because they are the primary regulators of cell fate decisions. TFs bind TSSs and regulatory elements in order to control transcriptional programs. Thus, the 413 TFs were the core regulators controlling definitive hematopoiesis in the embryo. In order to establish how these TFs regulated EHT, we needed TF binding data for key regulatory elements. Unfortunately, methods for establishing TF binding data required cell numbers not feasible given the low cell numbers available from embryonic material. Alternatively, we used ATAC-seq to map accessible regulatory elements in the cellular intermediates of EHT (**Chapter 4**). Elements were annotated to nearby genes and analyzed for enriched TF motifs. Of particular interest were the TFs whose timing of the appearance of the motif coincided with expression of the TF. These TFs were classified as putative regulators of EHT.

Integration of chromatin and expression profiles facilitated the generation of gene regulatory interactions for two key cell fate decisions: cHE to sHE and sHE to proHSC (**Chapter 5**). Analysis of these regulatory interactions identified putative causative regulators of accessibility and subsequent

changes in gene expression, including *Bcl11a*, *Mef2c*, *Ikzf1*, and *Ikzf2*. Each of these genes had been previously implicated in hematopoiesis but little was known about their roles, if any, during EHT. Looking closely at *Ikzf1*, we identified a -28 element required for proper *Ikzf1* expression (**Chapter 6**). Interestingly, -28 activity was Runx1-dependent. Further investigation identified Notch signaling as a downstream target of the novel Runx1-Ikzf1 regulatory axis. Finally, interrogation of this regulatory axis using genetic perturbation revealed a complex interaction whereby *Runx1* and *Ikzf1* work both individually and synergistically to regulate Notch during *in vitro* EHT.

7.2 Molecular mechanisms of EHT – putative regulators and GRIs

Attempts to better understand the molecular mechanisms underlying EHT have focused on: (1) analyzing gene expression data from *in vivo* EHT or (2) identifying core regulatory networks of *in vitro* blood cell formation through integration of gene expression and TF binding data sets. Up until now, technical hurdles—mainly limitations on starting cell numbers of in vivo EHT populations—have hindered studies looking to identify the core regulatory networks of in vivo EHT. By circumventing these limitations, our data represents a unique tool from which we, and the field can use to better understand the gene regulatory interactions at the heart of in vivo EHT.

<u>In vivo expression profiling</u>

Elegant studies from other groups characterized Endo, HE, preHSCs and HSC from in vivo EHT using genome wide expression profiling (Zhou et al., 2016; Solaimani Karalaei et al., 2015; Li et al., 2015; Mckinney-Freeman et al., 2012). However, these studies had limitations due either to the embryonic stage at which cells were collected, E10 or later, (Zhou et al., 2016, Li et al.,

2016; Solaimani Karalaei et al., 2015) or the failure to separate populations before profiling (Mckinney-Freeman et al., 2012). Our RNA-seq data set was the first to capture EHT at the onset (E8.5). Moreover, utilizing the +23GFP enhancer reporter mouse allowed us to separate cHE from sHE (Swiers et al., 2013). This transition was marked not only by expression of Runx1 but also by the acquisition of the hematopoietic program and loss of endothelial tube formation potential. Other data sets capture HE at E10 or later and are therefore not capturing these populations. Moreover, recent work from Daley and colleagues showed HE specification was key for making HSC from PSC (Sugimura et al., 2017). Thus, our ability to study HE specification during the cHE to sHE transition gives novelty, and importance, to our data set.

In vitro regulatory networks

Recent work using embryonic stem cell-derived HSPCs help shed light on the core regulator elements governing hematopoietic specification but did so using an *in vitro* model (Goode et al., 2016). Therefore, this work likely reflected yolk sac and not AGM-derived hematopoiesis (McGrath et al., 2015). With one of the aims of this study to make discoveries that will ultimately aid in the development of protocols for generation of dHSCs ex vivo, we reasoned that comparing in vitro GRIs to our in vivo GRI would facilitate this aim. First, we needed to identify putative regulators and GRI of in vivo EHT.

Integration of expression and chromatin profiles identified new regulators and GRIs

Various other studies have looked at in vivo EHT, revealing candidate regulators of EHT (Zhou et al., 2016; Solaimani Karalaei et al., 2015; Mckinney-Freeman et al., 2012). Interestingly, in nearly all of the

aforementioned studies, genes found to be novel regulators of EHT were also critical for adult hematopoiesis. In Chapter 5, global integration of profiling data revealed *Stat4* as both a candidate target gene of Runx1 but also as the potential way by which the first HSCs acquire their long-term (LT) repopulation potential (Holmfeldt et al., 2016). Genetic ablation of RUNX binding sites within Stat4 regulatory elements followed by measurement of Stat4 expression would begin to test this predicted regulatory interaction. In order to test whether *Stat4* confers LT repopulation potential, we could sort Stat4+ preHSCs or dHSCs followed by transplantation into irradiated recipients. Repopulation in the Stat4+ but not the Stat4- cells would show that Stat4 is associated with LT potential. Further work would be needed to assess whether Stat4 confers this potential and by what mechanism.

In Chapter 5, looking specifically at the cHE to sHE and sHE to proHSC transitions, we identified other candidate regulators of EHT, including *Bcl11a*, *Mef2c*, *Ikzf1*, and *Ikzf2*. All of which have been previously implicated for their role in adult hematopoiesis (prior to their identification in this study). Although further work validated Ikfz1 as a both a target of Runx1 and a regulator of lymphoid potential during EHT (Chapter 6), the caveats of our approach discussed below made it difficult to confidently select additional targets where the relationships were more speculative based on our data. Additionally, our approach left out key interactions, including the Runx1-Sfpi1-14 (Huang et al., 2008). Thus, we sought additional ways of identifying targets using our data as a “short list”.

CRISPR/Cas9-based pooled library screen to identify new regulators

In order to facilitate identification of additional putative candidates in a systematic manner, we could adopt a custom screen using a CRISPR/Cas9 lentiviral library (Koike-Yusa et al., 2014; Parnas et al., 2015) in mouse ESC differentiation cultures (Figure 7.1). Genome-wide screens require starting with enough cells to eliminate potential bias during selection. The relative inefficiencies of mESC differentiations means that preventing bias is difficult, forcing cell number calculations to start with expected numbers at the end of the differentiation and then work backwards to identify the number of starting cells. This ensures gRNAs lost during differentiation aren't lost to technical artifacts. For inefficient mESC differentiations this means billions of starting cells, which is not feasible. Therefore, we could start with our list of differentially expressed TFs as a small pool from which we could design a custom library thereby reducing the number of starting cells needed.

The custom library would be cloned into lentiviral vectors expressing both the sgRNA and Cas9 (lenti-sgRNA). Mouse ESCs would be transduced, selected, and then differentiated along the hematopoietic lineage using a protocol adapted from Pearson et al. (Pearson et al., 2015). At selected time points throughout the differentiation, specific EHT populations would be sorted and genomic DNA isolated and assessed for sgRNA abundance by deep sequencing. sgRNAs targeting critical TFs would result in those cells failing to progress through EHT (e.g. sgRNA targeting Runx1). Therefore, loss of a sgRNA would likely identify a TF required for one or more stages of EHT. Assessment of sgRNA abundance in each population along the developmental pathway would allow for identification of TFs required for each

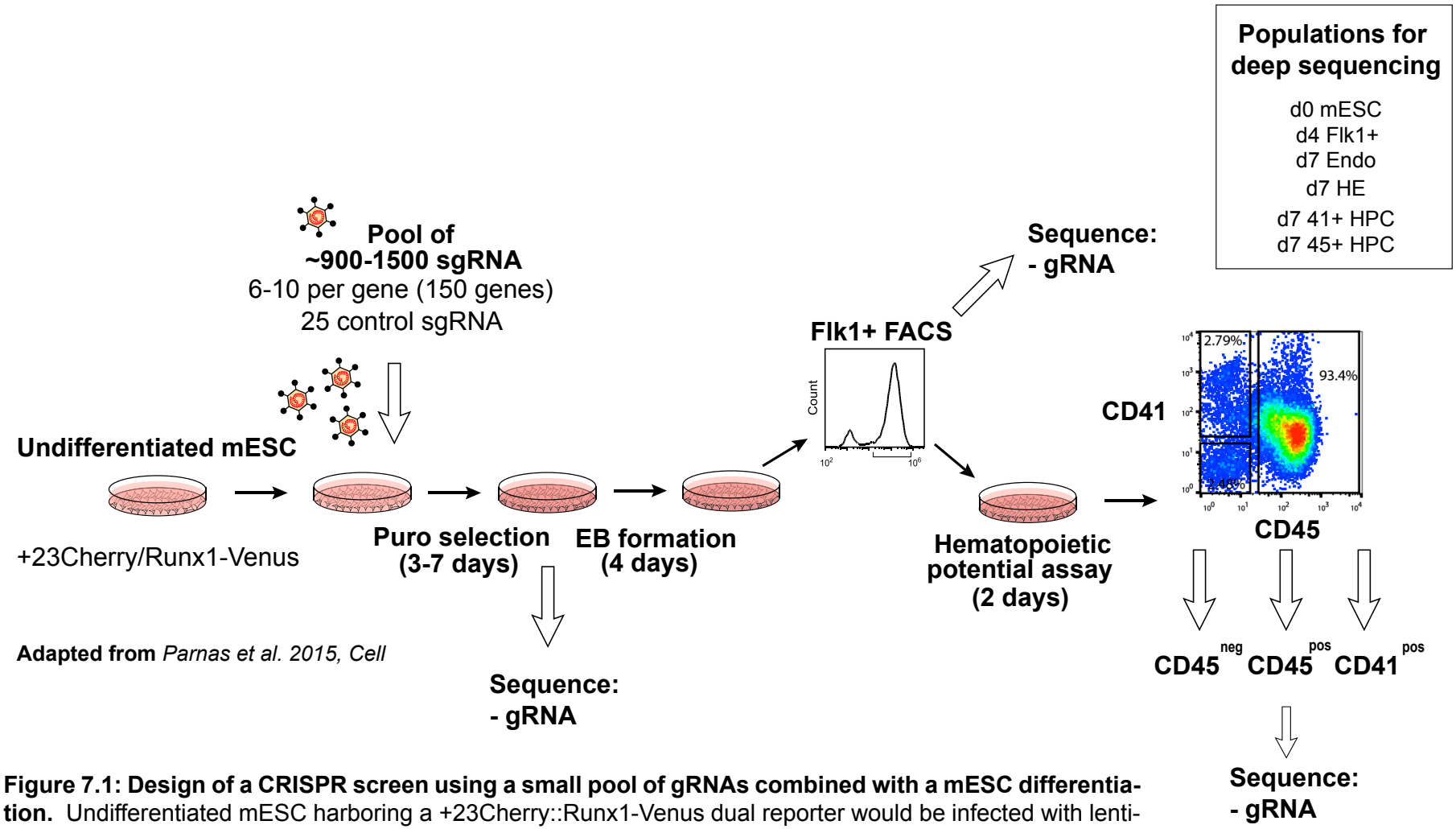


Figure 7.1: Design of a CRISPR screen using a small pool of gRNAs combined with a mESC differentiation. Undifferentiated mESC harboring a +23Cherry::Runx1-Venus dual reporter would be infected with lentivirus expressing selected gRNAs and Cas9. gRNAs would be selected based on a short list of putative regulators as identified by the profiling studies described herein. gRNA controls targeting essential and non-essential genes would be used to assess gRNA dropout at the end of the differentiation. After selection, gDNA would be isolated to check for presence of gRNA sequences to check how technical limitations affect gRNA distribution. This would be done throughout the differentiation to account for any bias. Next, mESC would be differentiated towards the hematopoietic lineage. Af the end of the differentiation, populations would be sorted and deep sequenced to check for the abundance of gRNAs. gRNA drop out would indicate targeting of an essential. gene--potentially identifying a novel regulator of in vitro EHT.

cell fate transition. Ultimately, this screening approach combined with single-cell RNA-seq, called Perturb-seq, would allow perturbations to be evaluated for impact on gene expression in single cells (Dixit et al., 2016).

Comparing AGM EHT to YS (or in vitro) EHT

In this study an emphasis was placed on understanding how HSPC (including dHSCs) emergence occurs in vivo in the AGM. EHT, the process by which HSPCs and dHSCs emerge from the underlying endothelium, occurs at multiple sites, including the YS. However, the YS does not produce dHSCs (Medvinsky et al., 1993; Medvinksky and Dzierzak, 1996; Muller et al., 1994). Therefore, it is critical to understand the differences between EHT at these sites in order to determine what is unique about the AGM that facilitates dHSC production. From a translational medicine perspective this point is even more important given that in vitro hPSC differentiations more closely resemble YS EHT rather then AGM EHT (McGrath et al., 2015). Thus, comparing AGM EHT with YS EHT and YS EHT with mESC EHT may facilitate identification of the missing pieces needed to make de novo dHSCs in vitro.

Ongoing work in the lab set out to profile YS EHT by capturing EC, HE and EMP at E9.5 from 23GFP embryos. Only the ATAC-seq data had been fully generated at the time this thesis was written. Preliminary examination of the chromatin accessibility profiles highlighted key differences with the AGM chromatin profiles. For example, TGF-beta and Notch signaling were enriched in AGM only elements (Figure 7.2). The next step is to integrate chromatin and expression profiles, which is currently ongoing in the lab, with the aim of identifying gene regulatory interactions. By identifying the key missing gene regulatory interactions and then adding these factors into mESC in vitro

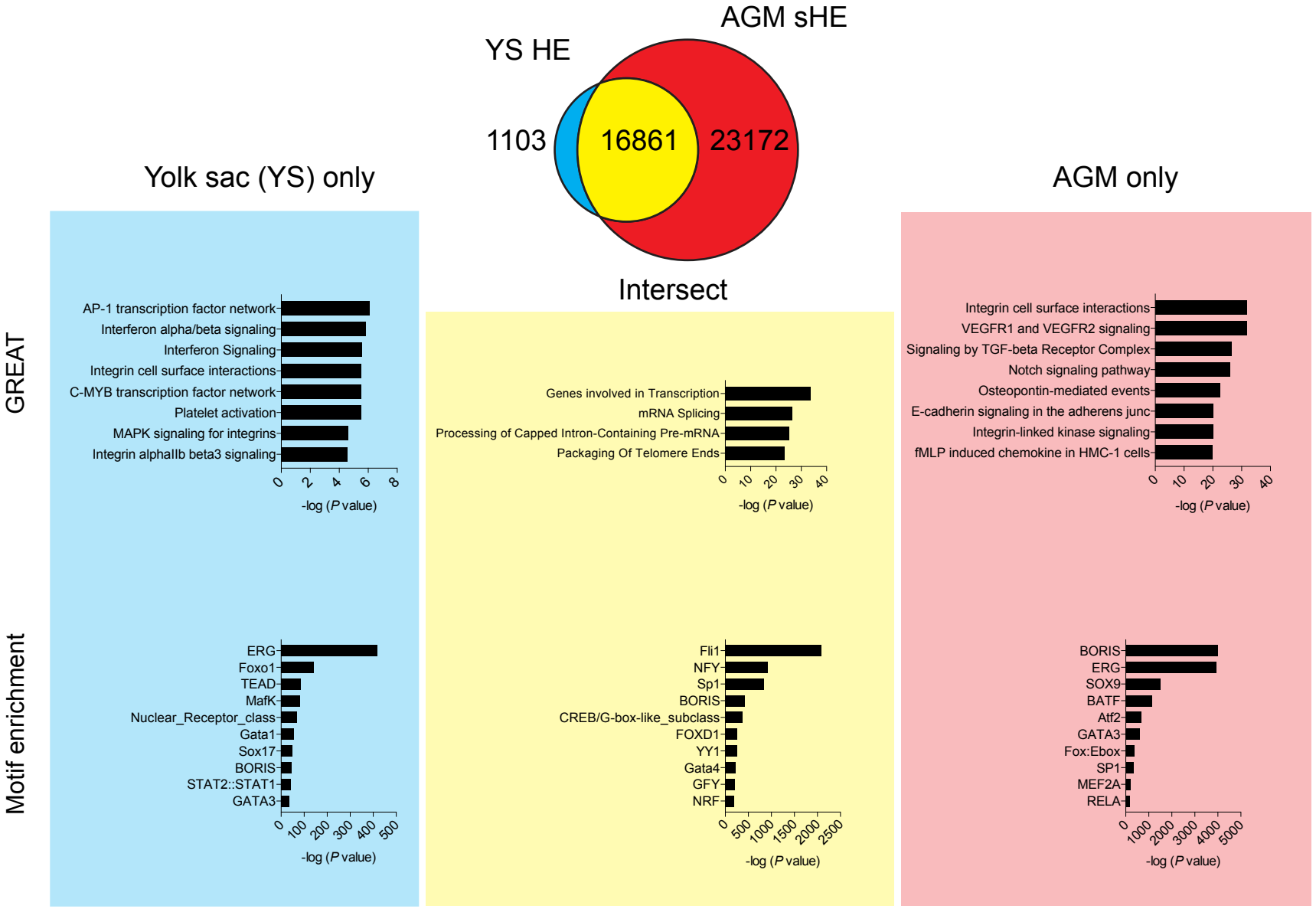


Figure 7.2: Comparison of ATAC-seq for E9.5 HE from the YS and E9.5 sHE from the AGM. Peaks were intersected using Bamtools -intersect. Venn diagram shows number overlapping elements and number of elements specific to each population. GREAT analysis and motif enrichment were done on each subset. Top 10 (or top 4) mSigDB pathways were shown for each subset with -logPvalue on the x-axis. Top 10 enriched motifs represented as -log Pvalue.

differentiations or into ex vivo cultures of YS-derived cells may facilitate the de novo generation of dHSCs.

Resolving differences between in vitro and in vivo Notch requirements

It is important to note that Notch signaling requirements differ between in vitro and in vivo EHT, as was detected with preliminary analysis of YS versus AGM ATAC-seq. This may explain the differences seen between in vivo sc-qRT-PCR and genetic perturbations in vitro. For example, in vivo, it appears that *Runx1* is sufficient to regulate *Hes1*, which contrasts with in vitro data suggesting both *Runx1* and *Ikzf1* are needed. In order to obtain a mechanistic understanding of these differences and whether and how they are important, we performed the same profiling analysis in extra-embryonic yolk sac from E9.5. Analysis of this data is currently ongoing. Previous reports showed that mESC differentiation cultures are more similar to YS EHT than AGM EHT (McGrath et al., 2015). Thus, comparing the YS profiling data to the AGM profiling data generated here may shed light on the mechanisms by which *Runx1*, and a Runx1-Ikzf1 axis, regulate Notch signaling during EHT.

7.3 Caveats of GRIs generated from expression and chromatin profiling

To identify candidate regulators of EHT and uncover underlying GRIs, TFs where timing of the appearance of the motif coincided with expression of the TF were identified as putative causative regulators of accessibility and subsequent changes in gene expression (Coerces et al., 2016). However, there were several limitations to consider when interpreting integration of our expression and chromatin profiles. First, we mapped accessible elements to genes using HOMERs annotatepeaks.pl. The primary consideration within this command is proximity to nearest TSSs. Therefore, elements were mapped to

the nearest TSS regardless of evidence for interaction between elements and nearby promoters. For example, the -14kb element of *Sfpi1* (PU.1) is important for proper expression in adult hematopoiesis (Huang et al., 2008). In terms of genomic location it is closer to the TSS of a neighboring gene and was therefore not assigned to *Sfpi1*. To address this limitation in a systematic manner, we could use chromosome conformation capture (e.g., Capture-C), to provide promoter-enhancer interaction maps using the differential accessible elements as the capture point (Hughes et al., 2014; Davies et al., 2015). This would not provide absolute proof of a biologically relevant interaction (*i.e.*, an interaction that leads to changes in expression) but would provide evidence for assigning an element to a gene(s). Adding an "interaction score" based on Capture-C data would provide key additional information to build into our gene regulatory interaction maps.

A second limitation involves TF motif enrichment. To identify motifs enriched in a subset of elements, we used HOMERs findmotifsgenome.pl. The command looks for motifs from the JASPAR database (http://jaspar.genereg.net/) in the input elements but also looks for enrichment in a background set of genomic elements. The JASPAR database uses published ChIP-seq data to identify a consensus sequence for each motif. This limits the database to TF whose consensus motifs can be identified. Unfortunately, only ~250 TFs meet this criteria, meaning only 10% of the estimated 2,000+ TFs. For example, we identified Ikzf1 as a target of Runx1 because of the presence of RUNX motifs and Runx1 ChIP-seq in the -28kb element, which we identified through the integration of our data sets. However, despite multiple ChIP-seq data sets for Ikzf1 (Vidal et al., 2013;

Schjerven et al., 2017), its consensus motif is not included in the JASPAR database. Therefore, we only came across Ikzf1's potential regulatory role using literature mining and in-house analysis of the Ikzf1 ChIP-seq data sets. One way to address this problem would be to eliminate the use of motif enrichment and focus only on elements with evidence of TF binding from ChIP-seq experiments. Given the low cell number from embryonic sources, one technique that could aid in this respect is ChIP-mentation, which combines ChIP with a Tn5 transposase to label sites of TF binding (Schmidt et al., 2015). This would facilitate the identification of TF binding in starting cells as low as 10,000.

Despite these limitations, our approach was validated by identification of known key regulatory elements. For example, Gfi1-33 was identified as a differentially accessible element containing multiple bound TFs according to CODEX and containing several more TF motifs (Figure 5.8) (Wilson et al., 2010a). In summary, our approach was validated by identification of a known (Gfi1-33) regulatory interaction and discovery of a novel (Ikzf1-28) regulatory interactions highlighted the usefulness of our data to the field.

7.4 Runx1 as direct regulator of cytoskeleton organization

In Chapter 3, expression profiling identified a crucial role for actin cytoskeleton reorganization during EHT. Further investigation revealed reorganization to be Runx1-dependent. This additional, indirect mechanism by which Runx1 can affect cell fate was surprising. It is known that changes in structure, assembly and generation of actin can affect gene transcription (Olson and Nordheim 2011; Mammoto et al., 2009). Thus, Runx1 may regulate cell fate decisions not only through its activation and repression of downstream TFs, but also

through modulation of effector proteins capable of quick, dramatic changes to actin cytoskeleton organization (e.g., actin capping and severing protein, Gelsolin).

In a similar vein, we speculated that Runx1 was either directly interacting with the actin cytoskeleton or forming complexes with other proteins that directly interacting with the actin cytoskeleton. This direct interaction would facilitate changes in actin structure or assembly and ultimately lead to modulation in gene transcription (Olson and Nordheim 2011; Mammoto et al., 2009). A previous report showed a cytoplasmic role for Runx1 whereby it inhibited NF-kB signaling through interaction with a kinase subunit in the cytoplasm (Nakagawa et al., 2011). Runx1 is known to have direct protein-protein interaction with actin when pulled down in IP assays (Yu et al., 2012). Taken together, this suggests a cytoplasmic role for Runx1 whereby it directly interacts with actin or actin-binding proteins to drive reorganization of the actin cytoskeleton. Additional work is needed to test this hypothesis. The new Runx1-ERt2 generated as a part of this study would facilitate this work and could prove to be a useful tool to the field for studying Runx1.

7.5 Concluding remarks

In summary, based on integration of expression and chromatin profiles, we identified gene regulatory interactions regulating the endothelial-to-hematopoietic transition. This data represents the first genome-wide profiling of cells at the onset of EHT. Identification of candidate regulators of EHT using integration of expression and chromatin profiles discovered *Ikzf1* as both a key target and key regulator of EHT. Characterization of a Runx1-Ikzf1 regulatory interaction showed independent and synergistic regulation of Notch

signaling during in vitro EHT. As the requirement for Notch signaling differs between in vitro (YS) EHT and AGM EHT, manipulation of this pathway may provide the necessary signaling cues needed for ex vivo generation of dHSCs. In order to quickly and efficiently identify additional regulators, the list of candidates could be screened using CRISPR/Cas9 small-pooled gRNA libraries. Combining these screens with single cell approaches would facilitate further understanding of the molecular mechanisms regulating the endothelial-to-hematopoietic transition.

References

Abedin, M. J., Nguyen, A., Jiang, N., Perry, C. E., Shelton, J. M., Watson, D. K., & Ferdous, A. (2014). Fli1 Acts Downstream of Etv2 to Govern Cell Survival and Vascular Homeostasis via Positive Autoregulation. *Circulation Research*, *114*(11), 1690–1699.

Ayllon, V., Bueno, C., Ramos-Mejia, V., Navarro-Montero, O., Prieto, C., Real, P. J., Romero, T., et al. (2015). The Notch ligand DLL4 specifically marks human hematoendothelial progenitors and regulates their hematopoietic fate, *29*(8), 1741–1753. doi:Original Article.

Baldwin, R. J., Dam, ten, G. B., van Kuppevelt, T. H., Lacaud, G., Gallagher, J. T., Kouskoff, V., & Merry, C. L. R. (2008). A Developmentally Regulated Heparan Sulfate Epitope Defines a Subpopulation with Increased Blood Potential During Mesodermal Differentiation. *Stem cells (Dayton, Ohio)*, *26*(12), 3108–3118. doi:10.1634/stemcells.2008-0311

Beck, D., Thoms, J. A. I., Perera, D., Schutte, J., Unnikrishnan, A., Knezevic, K., Kinston, S. J., et al. (2013). Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood*, *122*(14), e12–e22. doi:10.1182/blood-2013-03-490425

Bee, T., Ashley, E. L. K., Bickley, S. R. B., Jarratt, A., Li, P. S., Sloane-Stanley, J., Gottgens, B., & de Bruijn, M. F. T. R. (2009a). The mouse Runx1 +23 hematopoietic stem cell enhancer confers hematopoietic specificity to both Runx1 promoters. *Blood*, *113*(21), 5121–5124. doi:10.1182/blood-2008-12-193003

Bee, T., Liddiard, K., Swiers, G., Bickley, S. R. B., Vink, C. S., Jarratt, A., Hughes, J. R., Medvinsky, A., et al. (2009b). Alternative Runx1 promoter usage in mouse developmental hematopoiesis. *Blood Cells, Molecules, and Diseases*, *43*(1), 35–42.

Bee, T., Swiers, G., Muroi, S., Pozner, A., Nottingham, W., Santos, A. C., Li, P. S., Taniuchi, I., & de Bruijn, M. F. T. R. (2010a). Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis. *Blood*, *115*(15), 3042–3050. doi:10.1182/blood-2009-08-238626

Bee, T., Swiers, G., Muroi, S., Pozner, A., Nottingham, W., Santos, A. C., Li, P. S., Taniuchi, I., & de Bruijn, M. F. T. R. (2010b). Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis. *Blood*, *115*(15), 3042–3050. doi:10.1182/blood-2009-08-238626

Bertrand, J. Y., Chi, N. C., Santoso, B., Teng, S., Stainier, D. Y. R., & Traver, D. (2010). Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature*, *464*(7285), 108–111. doi:10.1038/nature08738.

Boiers, C., 2013. Carrelha, J., Lutteropp, M., Luc, S., Green, J.C., et al. Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. Cell Stem Cell 13(5), 535–548.

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*(7285), 116–120. doi:10.1038/nature08764.

Bos, F. L., Hawkins, J. S., & Zovein, A. C. (2015). Single-cell resolution of

morphological changes in hemogenic endothelium. *Development*, *142*(15), 2719–2724. doi:10.1242/dev.121350.

Bravo, J., Li, Z., Speck, N.A., & Warren A.J.. (2001). The leukemia-associated AML1 (Runx1)—CBF beta complex functions as a DNA-induced molecular clamp. Nature Structural Biology 8(4), 371–378.

Brou, C., Logeat, F., Gupta, N., Bessia, C., LeBail, O., Doedens, J. R., Cumano, A., et al. (2000). A Novel Proteolytic Cleavage Involved in Notch Signaling. *Molecular Cell*, Molecular Cell, *5*(2), 207–216. doi:doi: 10.1016/S1097-2765(00)80417-7

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position, *10*(12), 1213–1218. doi:Article

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

Butko, E., Pouget, C., & Traver, D. (2016). Complex regulation of HSC emergence by the Notch signaling pathway. *Developmental biology*, *409*(1), 129–138.

Cai, X., Gao, L., Teng, L., Ge, J., Oo, Z., Kumar, A., Gilliland, D. ., et al. (2015). Runx1 Deficiency Decreases Ribosome Biogenesis and Confers Stress Resistance to Hematopoietic Stem and Progenitor Cells. *Cell stem cell*, Cell Stem Cell, *17*(2), 165–177. doi:doi: 10.1016/j.stem.2015.06.002

Cai, Z., de Bruijn, M., Ma, X., Dortland, B., Luteijn, T., Downing, J. R., & Dzierzak, E. (2000). Haploinsufficiency of AML1 Affects the Temporal and Spatial Generation of Hematopoietic Stem Cells in the Mouse Embryo. *Immunity*, Immunity, *13*(4), 423–431. doi:doi: 10.1016/S1074-7613(00)00042-X

Caprioli, A., Jaffredo, T., Gautier, R., Dubourg, C., & Dieterlen-Lièvre, F. (1998). Blood-borne seeding by hematopoietic and endothelial precursors from the allantois. *Proceedings of the National Academy of Sciences*, *95*(4), 1641–1646.

Carpenter, L., Malladi, R., Yang, C. T., French, A., Pilkington, K. J., Forsey, R. W., Sloane-Stanley, J., et al. (2011). Human induced pluripotent stem cells are capable of B-cell lymphopoiesis. *Blood*, *117*(15), 4008–4011. doi:10.1182/blood-2010-08-299941

Cauchy, P., James, S., Zacarias-Cabeza, J., Ptasinska, A., Imperato, M., Assi, S. A., Piper, J., et al. (2015). Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature. *Cell Reports*, *12*(5), 821–836.

Cellot, S., Krosl, J., Chagraoui, J., Meloche, S., Humphries, R. K., & Sauvageau, G. (2007). Sustained In Vitro Trigger of Self-Renewal Divisions in Hoxb4(hi)Pbx1(lo) Hematopoietic Stem Cells. *Experimental hematology*, *35*(5), 802–816.

Chaudhuri, A. A., So, A. Y.-L., Mehta, A., Minisandram, A., Sinha, N., Jonsson, V. D., Rao, D. S., et al. (2012). Oncomir miR-125b regulates hematopoiesis by targeting the gene Lin28A. *Proceedings of the National Academy of Sciences*, *109*(11), 4233–4238.

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 EP –. doi:10.1038/nature07619

Cheung, A. K. L., Ko, J. M. Y., Lung, H. L., Chan, K. W., Stanbridge, E. J., Zabarovsky, E., Tokino, T., et al. (2011). Cysteine-rich intestinal protein 2 (CRIP2) acts as a repressor of NF-κB–mediated proangiogenic cytokine transcription to suppress tumorigenesis and angiogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(20), 8390–8395.

Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C., & Keller, G. (1998). A common precursor for hematopoietic and endothelial cells. Development, *125*(4), 725-732.

Ciau-Uitz, A., Walmsley, M., & Patient, R. (2000). Distinct Origins of Adult and Embryonic Blood in Xenopus. Cell, *102*(6), 787–796.

Clements, W. K., & Traver, D. (2013). Signalling pathways that control vertebrate haematopoietic stem cell specification. Nature Reviews Immunology, *13*, 336-348.

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

Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., Snyder, M. P., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nature Genetics, *48*, 1193-1203.

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

Crisan, M., Solaimani Kartalaei, P., Neagu, A., Karkanpouna, S., Yamada-Inagawa, T., Purini, C., Vink, C. S., et al. (2016). BMP and Hedgehog Regulate Distinct AGM Hematopoietic Stem Cells Ex Vivo. Stem Cell Reports, *6*(3), 383–395.

Cumano, A., Dieterlen-Lièvre, F., & Godin, I. (1996). Lymphoid Potential, Probed before Circulation in Mouse, Is Restricted to Caudal Intraembryonic Splanchnopleura. Cell, *86*(6), 907–916.

Damm, E. W., & Clements, W. K. (2017). Pdgf signalling guides neural crest contribution to the haematopoietic stem cell specification niche. Nature Cell Biology, *19*, 457-467.

Dancis, J., Jansen, V., Brown, G., Gorstein, F., & Balis, M. (1977). Treatment of hypoplastic anemia in mice with placental transplants. Blood, *50*(4), 663-670.

Dancis, J., Jansen, V., Gorstein, F., & Douglas, G. W. (1968). Hematopoietic cells in mouse placenta. American Journal of Obstetrics & Gynecology, *100*(8), 1110–1121.

de Bruijn, M. F. T. R., Speck N.A., Peeters, M., and Dzierzak, E. (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. EMBO J, *19*(11), 2465–2474.

de Bruijn, M. F. T. R., 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*(5), 673–683.

de Bruijn, M. F. T. R., Peeters, M. C. E., Luteijn, T., Visser, P., Speck, N. A., & Dzierzak, E. (2000). CFU-S(11) activity does not localize solely with the aorta in the aorta-gonad-mesonephros region. *Blood*, *96*(8), 2902-2904.

de Pater, E., Kaimakis, P., Vink, C. S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P. S., et al. (2013). Gata2 is required for HSC generation and survival. The Journal of Experimental Medicine, *210*(13), 2843–2850.

Dieterlen-Lievre, P. F. (1975). On the origin of haemopoietic stem cells in the avian embryo: an experimental approach. J Embryol Exp Morphol, *33*(3), 607–619.

Drake, C. J., & Fleming, P. A. (2000). Vasculogenesis in the day 6.5 to 9.5 mouse embryo. Blood, *95*(5), 1671–1679.

Duarte et al., Hirashima, M., Benedito, R., Trindade, A., Diniz, P., et al., (2004). Dosage-sensitive requirement for mouse Dll4 in artery development. Genes & Develop. 18(20), 2474–2478.

Dumont, D. J., Fong, G.-H., Puri, M. C., Gradwohl, G., Alitalo, K., & Breitman, M. L. (1995). Vascularization of the mouse embryo: A study of flk-1, tek, tie, and vascular endothelial growth factor expression during development. Developmental Dynamics, *203*(1), 80–92.

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. PNAS, *104*(52), 20838–20843.

Edlund, S., Landström, M., Heldin, C.-H., & Aspenström, P. (2002). Transforming Growth Factor-β–induced Mobilization of Actin Cytoskeleton Requires Signaling by Small GTPases Cdc42 and RhoA. Molecular Biology of the Cell, *13*(3), 902–914.

Eilken, H. M., Nishikawa, S.-I., & Schroeder, T. (2009). Continuous single-cell imaging of blood generation from haemogenic endothelium. Nature, *457*(7231), 896–900.

Eliades, A., Wareing, S., Marinopoulou, E., Fadlullah, M. H., Patel, R., Grabarek, J., Plusa, B., Lacaud, G., & Kouskoff, V. (2016). The Hemogenic Competence of Endothelial Progenitors Is Restricted by Runx1 Silencing during Embryonic Development. Cell Reports, *15*(10), 2185–2199.

Ema, M., Yokomizo, T., Wakamatsu, A., Terunuma, T., Yamamoto, M., & Takahashi, S. (2006). Primitive erythropoiesis from mesodermal precursors expressing VE-cadherin, PECAM-1, Tie2, endoglin, and CD34 in the mouse embryo. Blood, *108*(13), 4018–4024.

Espín-Palazón, R., Stachura, D. L., Campbell, C. A., García-Moreno, D., Cid, N. D., Kim, A. D., Candel, S., et al. (2014). Proinflammatory signaling regulates hematopoietic stem cell emergence. Cell, *159*(5), 1070–1085.

Feil, R., Wagner, J., Metzger, D., & Chambon, P. (1997). Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains.

Biochemical and Biophysical Research Communications, *237*(3), 752–757.

Ferkowicz, M. J. (2003). CD41 expression defines the onset of primitive and definitive hematopoiesis in the murine embryo. *Development*, *130*(18), 4393–4403.

Ferreiros-Vidal, I., Carroll, T., Taylor, B., Terry, A., Liang, Z., Bruno, L., Dharmalingam, G., et al. (2013). Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation. Blood, *121*(10), 1769–1782.

Fischer, A., Schumacher, N., Maier, M., Sendtner, M., & Gessler, M. (2004). The Notch target genes Hey1 and Hey2 are required for embryonic vascular development. *Genes & Development*, *18*(8), 901–911.

Galić, Z., Kitchen, S. G., Kacena, A., Subramanian, A., Burke, B., Cortado, R., & Zack, J. A. (2006). T lineage differentiation from human embryonic stem cells. PNAS, *103*(31), 11742–11747.

Gama-Norton, L., Ferrando, E., Ruiz-Herguido, C., Liu, Z., Guiu, J., Islam, A. B. M. M. K., Lee, S.-U., et al. (2015). Notch signal strength controls cell fate in the haemogenic endothelium. Nature Communications, *6*, 8510.

Gekas, C., Dieterlen-Lièvre, F., Orkin, S. H., & Mikkola, H. K. A. (2005). The Placenta Is a Niche for Hematopoietic Stem Cells. Developmental Cell, *8*(3), 365–375.

Gering, M., & Patient, R. (2005). Hedgehog Signaling Is Required for Adult Blood Stem Cell Formation in Zebrafish Embryos. Developmental Cell, *8*(3), 389–400.

Godin, I. (2016). Chapter 8 – Ontogeny of the Hematopoietic System. Kaufman's Atlas of Mouse Development. 111-120.

Goldman, D. C., Bailey, A. S., Pfaffle, D. L., Masri, Al, A., Christian, J. L., & Fleming, W. H. (2009). BMP4 regulates the hematopoietic stem cell niche. Blood, *114*(20), 4393–4401.

Guiu, J., Shimizu, R., D'Altri, T., Fraser, S. T., Hatakeyama, J., Bresnick, E. H., Kageyama, R., et al. (2013). Hes repressors are essential regulators of hematopoietic stem cell development downstream of Notch signaling. The Journal of Experimental Medicine, *210*(1), 71–84.

Hadland, B. K., Varnum-Finney, B., Mandal, P., Rossi, D., Poulos, M. G., Butler, J. M., Rafii, S., et al. (2017). A Common Origin for B-1a and B-2 Lymphocytes in Clonal Pre- Hematopoietic Stem Cells. Stem Cell Reports, *8*(6), 1563–1572.

Haynes, J., Srivastava, J., Madson, N., Wittmann, T., & Barber, D. L. (2011). Dynamic actin remodeling during epithelial–mesenchymal transition depends on increased moesin expression. Molecular Biology of the Cell, *22*(24), 4750–4764.

Heinz, S., Romanoski, C. E., Benner, C., Allison, K. A., Kaikkonen, M. U., Orozco, L. D., & Glass, C. K. (2013). Effect of natural genetic variation on enhancer selection and function Nature, *503*, 487–452.

Holley, R. J., Pickford, C. E., Rushton, G., Lacaud, G., Gallagher, J. T., Kouskoff, V., & Merry, C. L. R. (2011). Influencing Hematopoietic Differentiation of Mouse

Embryonic Stem Cells using Soluble Heparin and Heparan Sulfate Saccharides. J Biol Chem., *286*(8), 6241–6252.

Hoogenkamp, M., Krysinska, H., Ingram, R., Huang, G., Barlow, R., Clarke, D., Ebralidze, A., et al. (2007). The Pu.1 Locus Is Differentially Regulated at the Level of Chromatin Structure and Noncoding Transcription by Alternate Mechanisms at Distinct Developmental Stages of Hematopoiesis. Molecular and Cellular Biology, *27*(21), 7425–7438.

Hoogenkamp, M., Lichtinger, M., Krysinska, H., Lancrin, C., Clarke, D., Williamson, A., Mazzarella, L., et al. (2009). Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. Blood, *114*(2), 299–309.

Huang, G., Zhang, P., Hirai, H., Elf, S., Yan, X., Chen, Z., Koschmieder, S., et al. (2008). PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. Nature Genetics, *40*, 51–60.

Iacovino, M., Chong, D., Szatmari, I., Hartweck, L., Rux, D., Caprioli, A., Cleaver, O., et al. (2010). HoxA3 is an apical regulator of hemogenic endothelium. Nature Cell Biology, *13*(1), 72–78.

Jaffredo, T., Gautier, R., Eichmann, A., & Dieterlen-Lievre, F. (1998). Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. Development, *125*(22), 4575.

Kabrun, N., Buhring, H. J., Choi, K., Ullrich, A., Risau, W., & Keller, G. (1997). Flk-1 expression defines a population of early embryonic hematopoietic precursors. Development, *124*(10), 2039.

Kaimakis, P., de Pater, E., Eich, C., Solaimani Kartalaei, P., Kauts, M.-L., Vink, C. S., van der Linden, R., et al. (2016). Functional and molecular characterization of mouse Gata2-independent hematopoietic progenitors. Blood, *127*(11), 1426–1437.

Kataoka, H., Hayashi, M., Kobayashi, K., Ding, G., Tanaka, Y., & Nishikawa, S.-I. (2013). Region-specific Etv2 ablation revealed the critical origin of hemogenic capacity from Hox6-positive caudal-lateral primitive mesoderm. Experimental hematology, *41*(6), 567–581.e9.

Kennedy, M., D'Souza, S. L., Lynch-Kattman, M., Schwantz, S., & Keller, G. (2007). Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. *Blood*, 109(7), 2679–2681.

Kim, P. G., Nakano, H., Das, P. P., Chen, M. J., Rowe, R. G., Chou, S. S., Ross, S. J., et al. (2015). Flow-induced protein kinase A–CREB pathway acts via BMP signaling to promote HSC emergence. The Journal of Experimental Medicine, *212*(5), 633–648.

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*(21), 4691.

Kinkel, S. A., Galeev, R., Flensburg, C., Keniry, A., Breslin, K., Gilan, O., Lee, S., et al. (2015). Jarid2 regulates hematopoietic stem cell function by acting with polycomb

repressive complex 2. Blood, *125*(12), 1890–1900.

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

Kobayashi-Osaki, M., Ohneda, O., Suzuki, N., Minegishi, N., Yokomizo, T., Takahashi, S., Lim, K.-C., et al. (2005). GATA Motifs Regulate Early Hematopoietic Lineage-Specific Expression of the Gata2 Gene. Molecular and Cellular Biology, *25*(16), 7005–7020.

Krebs, L. T., Xue, Y., Norton, C. R., Shutter, J. R., Maguire, M., Sundberg, J. P., Gallahan, D., et al. (2000). Notch signaling is essential for vascular morphogenesis in mice. Genes & Development, *14*(11), 1343–1352.

Kumano, K., Chiba, S., Kunisato, A., Sata, M., Saito, T., Nakagami-Yamaguchi, E., Yamaguchi, T., et al. (2003). Notch1 but Not Notch2 Is Essential for Generating Hematopoietic Stem Cells from Endothelial Cells. Immunity, *18*(5), 699–711.

Kumaravelu, P., Hook, L., Morrison, A. M., Ure, J., Zhao, S., Zuyev, S., Ansell, J., et al. (2002). Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonization of the mouse embryonic liver. Development, *129*(21), 4891.

Kwan, W., Cortes, M., Frost, I., Esain, V., Theodore, L., Liu, S., Budrow, N., et al. (2016). The Central Nervous System Regulates Embryonic HSPC Production via Stress-Responsive Glucocorticoid Receptor Signaling. Cell Stem Cell, *19*(3), 370–382.

Kyba, M., Perlingeiro, R. C. R., & Daley, G. Q. (2002). HoxB4 Confers Definitive Lymphoid-Myeloid Engraftment Potential on Embryonic Stem Cell and Yolk Sac Hematopoietic Progenitors. Cell, *109*(1), 29–37.

Lam, E. Y. N., Hall, C. J., Crosier, P. S., Crosier, K. E., & Flores, M. V. (2010). Live imaging of Runx1 expression in the dorsal aorta tracks the emergence of blood progenitors from endothelial cells. Blood, *116*(6), 909–914.

Lamouille, S., Xu, J., & Derynck, R. (2014). Molecular mechanisms of epithelial–mesenchymal transition, *15*(3), 178–196.

Lancrin, C., Mazan, M., Stefanska, M., Patel, R., Lichtinger, M., Costa, G., Vargel, O., et al. (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. Blood, *120*(2), 314–322.

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*(7231), 892–895.

Lawson, K. A., Meneses, J. J., & Pedersen, R. A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development, *113*(3), 891–911.

Lawson, N. D., Scheer, N., Pham, V. N., Kim, C.-H., Chitnis, A. B., Campos-Ortega, J. A., & Weinstein, B. M. (2001). Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development*, *128*(19),

3675–3683.

Lawson, N. D., Vogel, A. M., & Weinstein, B. M. (2002). Sonic hedgehog and vascular endothelial growth factor act upstream of the Notch pathway during arterial endothelial differentiation. Developmental Cell, *3*(1), 127–136.

Lee, D., Park, C., Lee, H., Lugus, J. J., Kim, S. H., Arentson, E., Chung, Y. S., et al. (n.d.). ER71 acts downstream of BMP, Notch, and Wnt Signaling in blood and vessel progenitor specification. Cell Stem Cell, *2*(5), 497–507.

Lengerke, C., Schmitt, S., Bowman, T. V., Jang, I. H., Maouche-Chretien, L., McKinney-Freeman, S., Davidson, A. J., et al. (2008). BMP and wnt specify hematopoietic fate by activation of the Cdx-Hox pathway. Cell Stem Cell, *2*(1), 72–82.

Leslie, J. D., Ariza-McNaughton, L., Bermange, A. L., McAdow, R., Johnson, S. L., & Lewis, J. (2007). Endothelial signalling by the Notch ligand Delta-like 4 restricts angiogenesis. Development, *134*(5), 839–844.

Li, Y., Esain, V., Teng, L., Xu, J., Kwan, W., Frost, I. M., Yzaguirre, A. D., et al. (2014). Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell production. *Genes & Development*, *28*(23), 2597–2612.

Liakhovitskaia, A., Gribi, R., Stamateris, E., Villain, G., Jaffredo, T., Wilkie, R., Gilchrist, D., et al. (2009). Restoration of Runx1 expression in the Tie2 cell compartment Rescues definitive hematopoietic stem Cells and extends life of Runx1 knockout animals until birth. Stem cells, *27*(7), 1616–1624.

Liakhovitskaia, A., Rybtsov, S., Smith, T., Batsivari, A., Rybtsova, N., Rode, C., de Bruijn, M., et al. (2014.). Runx1 is required for progression of CD41(+) embryonic precursors into HSCs but not prior to this. Development, *141*(17), 3319–3323.

Lie-A-Ling, M., Marinopoulou, E., Li, Y., Patel, R., Stefanska, M., Bonifer, C., Miller, C., et al. (2014). RUNX1 positively regulates a cell adhesion and migration program in murine hemogenic endothelium prior to blood emergence. *Blood*, *124*(11), e11–e20.

Lis, R., Karrasch, C. C., Poulos, M. G., Kunar, B., Redmond, D., Duran, J. G. B., Badwe, C. R., et al. (2017). Conversion of adult endothelium to immunocompetent haematopoietic stem cells. Nature, *545*(7655), 439–445.

Liu, F., Li, D., Yu, Y. Y. L., 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 reports, *16*(5), 654–669.

Lizama, C. O., Hawkins, J. S., Schmitt, C. E., Bos, F. L., Zape, J. P., Cautivo, K. M., Borges Pinto, H., et al. (2015). Repression of arterial genes in hemogenic endothelium is sufficient for haematopoietic fate acquisition. Nature Communications, *6*, 7739.

Lu, S.-J., Li, F., Yin, H., Feng, Q., Kimbrel, E. A., Hahm, E., Thon, J. N., et al. (2011). Platelets generated from human embryonic stem cells are functional in vitro and in the microcirculation of living mice. Cell Res. *21*(3), 530–545.

Lux, C. T., Yoshimoto, M., McGrath, K., Conway, S. J., Palis, J., & Yoder, M. C.

(2008). All primitive and definitive hematopoietic progenitor cells emerging before E10 in the mouse embryo are products of the yolk sac. Blood, *111*(7), 3435–3438.

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Research, *40*(10), 4288–4297.

McGrath, K. E. (2003). Circulation is established in a stepwise pattern in the mammalian embryo. Blood, *101*(5), 1669–1675.

McGrath, K. E., Frame, J. M., Fromm, G. J., Koniski, A. D., Kingsley, P. D., Little, J., Bulger, M., et al. (2011). A transient definitive erythroid lineage with unique regulation of the -globin locus in the mammalian embryo. *Blood*, *117*(17), 4600–4608.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., et al. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature Biotechnology, *28*, 495–501.

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

Medvinsky, A., & Dzierzak, E. (1996). Definitive Hematopoiesis Is Autonomously Initiated by the AGM Region. Cell, *86*(6), 897–906.

Medvinsky, A., Rybtsov, S., & Taoudi, S. (2011). Embryonic origin of the adult hematopoietic system: advances and questions. Development, *138*(6), 1017–1031.

Mikkola, H. K. A. (2002). Expression of CD41 marks the initiation of definitive hematopoiesis in the mouse embryo. Blood, *101*(2), 508–516.

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

Minegishi, N., Ohta, J., Yamagiwa, H., Suzuki, N., Kawauchi, S., Zhou, Y., Takahashi, S., et al. (1999). The Mouse GATA-2 Gene is Expressed in the Para-Aortic Splanchnopleura and Aorta-Gonads and Mesonephros Region. Blood, *93*(12), 4196.

Miura, S. Davis, S., Klingensmith, J., and Mishina, Y. (2006). BMP signaling in the epiblast is required for proper recruitment of the prospective paraxial mesoderm and development of the somites. Development, *133*(19), 3767–3775.

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

Monteiro, R., Pinheiro, P., Joseph, N., Peterkin, T., Koth, J., Repapi, E., Bonkhofer, F., et al. (2016). Transforming Growth Factor β drives hemogenic endothelium programming and the transition to hematopoietic stem cells. Developmental Cell, *38*(4), 358–370.

Moore, M. A. S., & Metcalf, D. (1970). Ontogeny of the Haemopoietic System: Yolk

Sac Origin of In Vivo and In Vitro Colony Forming Cells in the Developing Mouse Embryo*. British Journal of Haematology, *18*(3), 279–296.

Mucenski, M. L., McLain, K., Kier, A. B., Swerdlow, S. H., Schreiner, C. M., Miller, T. A., Pietryga, D. W., et al. (1991). A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. Cell, *65*(4), 677–689.

Mukouyama, Y.-S., Chiba, N., Hara, T., Okada, H., Ito, Y., Kanamaru, R., Miyajima, A., et al. (2000). The AML1 transcription factor functions to develop and maintain hematogenic precursor Cells in the embryonic aorta–gonad–mesonephros region. Developmental biology, *220*(1), 27–36.

Mukouyama, Y.-S., Chiba, N., Mucenski, M. L., Satake, M., Miyajima, A., Hara, T., & Watanabe, T. (1999). Hematopoietic cells in cultures of the murine embryonic aorta-gonad-mesonephros region are induced by c-Myb. Current Biology, *9*(15), 833–836.

Müller, A. M., Medvinsky, A., Strouboulis, J., Grosveld, F., & Dzierzakt, E. (1994). Development of hematopoietic stem cell activity in the mouse embryo. Immunity, *1*(4), 291–301.

Nakagawa, M., Shimabe, M., Watanabe-Okochi, N., Arai, S., Yoshimi, A., Shinohara, A., Nishimoto, N., et al. (2011). AML1/RUNX1 functions as a cytoplasmic attenuator of NF- B signaling in the repression of myeloid tumors. Blood, *118*(25), 6626–6637.

Nichogiannopoulou, A., Trevisan, M., Neben, S., Friedrich, C., & Georgopoulos, K. (1999). Defects in hemopoietic stem cell activity in *Ikaros* mutant mice. J Exp Med., 190(9), 1201-1214.

Nishikawa, S. I., Nishikawa, S., Hirashima, M., Matsuyoshi, N., & Kodama, H. (1998a). Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. Development, *125*(9), 1747.

Nishikawa, S.-I., Nishikawa, S., Kawamoto, H., Yoshida, H., Kizumoto, M., Kataoka, H., & Katsura, Y. (1998b). In vitro generation of lymphohematopoietic cells from endothelial cells purified from murine embryos. Immunity, *8*(6), 761–769.

Nishimura, M. (2004). VWRPY motif-dependent and -independent roles of AML1/Runx1 transcription factor in murine hematopoietic development. Blood, *103*(2), 562–570.

North, T. E., de Bruijn, M. F. T. R., Stacy, T., Talebian, L., Lind, E., Robin, C., Binder, M., et al. (2002). Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. Immunity, *16*(5), 661–672.

North, T., Gu, T. L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marin-Padilla, M., et al. (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. Development, *126*(11), 2563.

Nottingham, W. T., Jarratt, A., Burgess, M., Speck, C. L., Cheng, J. F., Prabhakar, S., Rubin, E. M., et al. (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. Blood, *110*(13), 4188–4197.

Obier, N., Cauchy, P., Assi, S. A., Gilmour, J., Lie-A-Ling, M., Lichtinger, M., Hoogenkamp, M., et al. (2016). Cooperative binding of AP-1 and TEAD4 modulates

the balance between vascular smooth muscle and hemogenic cell fate. Development, *143*(23), 4324–4340.

Oettgen, P., Kas, K., Dube, A., Gu, X., Grall, F., Thamrongsak, U., Akbarali, Y., et al. (1999). Characterization of ESE-2, a Novel ESE-1-related Ets transcription factor that is restricted to glandular epithelium and differentiated keratinocytes. J Biological Chemistry, *274*(41), 29439–29452.

Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G., & Downing, J. R. (1996). AML1, the Target of Multiple Chromosomal Translocations in Human Leukemia, Is Essential for Normal Fetal Liver Hematopoiesis. Cell, *84*(2), 321–330.

Olson, E. N., & Nordheim, A. (2010). Linking actin dynamics and gene transcription to drive cellular motile functions. Nature reviews Molecular Cell Biology, *11*(5), 353–365.

Palis, J., McGrath, K., & Kingsley, P. (1995a). Initiation of hematopoiesis and vasculogenesis in murine yolk sac explants. Blood, *86*(1), 156–163.

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*(22), 5073–5084.

Pereira, C.-F., Chang, B., Qiu, J., Niu, X., Papatsenko, D., Hendry, C. E., Clark, N. R., et al. (2013). Induction of a Hemogenic Program in Mouse Fibroblasts. Cell Stem Cell, *13*(2), 205–218.

Pimanda, J. E., Ottersbach, K., Knezevic, K., Kinston, S., Chan, W. Y. I., Wilson, N. K., Landry, J.-R., et al. (2007). Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. PNAS, *104*(45), 17692–17697.

Qiu, P., Shandilya, H., D'Alessio, J.M., O'Connor, K., Durocher, J., & Gerard, G.F. (2004). Mutation detection using surveyor nuclease. BioTechniques *36*, 702–707.

Reese, D. E., Hall, C. E., & Mikawa, T. (2004). Negative Regulation of Midline Vascular Development by the Notochord. Developmental Cell, *6*(5), 699–708.

Riddell, J., Gazit, R., Garrison, B., Guo, G., Saadatpour, A., Mandal, P., Ebina, W., et al. (2014). Reprogramming Committed Murine Blood Cells to Induced Hematopoietic Stem Cells with Defined Factors. Cell, *157*(3), 549–564.

Robb, L., Lyons, I., Li, R., Hartley, L., Köntgen, F., Harvey, R. P., Metcalf, D., et al. (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. PNAS, *92*(15), 7075–7079.

Robert-Moreno, À., Guiu, J., Ruiz-Herguido, C., López, M. E., Inglés-Esteve, J., Riera, L., Tipping, A., et al. (2008). Impaired embryonic haematopoiesis yet normal arterial development in the absence of the Notch ligand Jagged1. EMBO Journal, *27*(13), 1886–1895.

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. Developmental Cell, *11*(2), 171–180.

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*(3), 489–501.

Rybtsov, S., Ivanovs, A., Zhao, S., & Medvinsky, A. (2016). Concealed expansion of immature precursors underpins acute burst of adult HSC activity in foetal liver. Development, *143*(8), 1284–1289.

Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., et al. (2011). Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. Journal of Experimental Medicine, *208*(6), 1305–1315.

Satoh, Y., Yokota, T., Sudo, T., Kondo, M., Lai, A., Kincade, P. W., Kouro, T., et al. (2013). The Satb1 protein directs hematopoietic stem cell differentiation toward lymphoid lineages. Immunity, *38*(6), 1105–1115.

Sawamiphak, S., Kontarakis, Z., & Stainier, D. Y. (2014). Interferon gamma Signaling Positively Regulates Hematopoietic Stem Cell Emergence. Developmental Cell, *31*(5), 640–653.

Schjerven, H., Ayongaba, E. F., Aghajanirefah, A., McLaughlin, J., Cheng, D., Geng, H., Boyd, J. R., et al. (2017). Genetic analysis of Ikaros target genes and tumor suppressor function in BCR-ABL1 +pre–B ALL. Journal of Experimental Medicine, *41*, jem.20160049.

Schlaeger, T. M. (2005). Tie2Cre-mediated gene ablation defines the stem-cell leukemia gene (SCL/tal1)-dependent window during hematopoietic stem-cell development. Blood, *105*(10), 3871–3874.

Schmidl, C., Rendeiro, A. F., Sheffield, N. C., & Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nature Methods, *12*(10), 963–965.

Schütte, J., Wang, H., Antoniou, S., Jarratt, A., Wilson, N. K., Riepsaame, J., Calero-Nieto, F. J., et al. (2016). An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. eLife, *5*, e11469.

Scott, E., Simon, M., Anastasi, J., & Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science, *265*(5178), 1573–1577.

Shalaby, F., Ho, J., Stanford, W. L., Fischer, K.-D., Schuh, A. C., Schwartz, L., Bernstein, A., et al. (1997). A Requirement for Flk1 in Primitive and Definitive Hematopoiesis and Vasculogenesis. Cell, *89*(6), 981–990.

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*, 62–66.

Shannon, P., Markiel, A., Ozier, O., Baliga N.S., Wang, J.T., et al., (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 13, 2498-504.

Shi, X., Richard, J., Zirbes, K. M., Gong, W., Lin, G., Kyba, M., Thomson, J. A., et al. (2014.). Cooperative interaction of Etv2 and Gata2 regulates the development of endothelial and hematopoietic lineages. Developmental Biology, *389*(2), 208–218.

Shivdasani, R. A., Mayer, E. L., & Orkin, S. H. (1995). Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature, *373*, 432–434.

Silacci, P., Mazzolai, L., Gauci, C., Stergiopulos, N., Yin, H. L., & Hayoz, D. (2004). Gelsolin superfamily proteins: key regulators of cellular functions. Cellular and Molecular Life Sciences CMLS, *61*(19), 2614–2623.

Silver, L., & Palis, J. (1997). Initiation of Murine Embryonic Erythropoiesis: A Spatial Analysis. Blood, *89*(4), 1154.

Smith, R. A., & Glomski, C. A. (1982). “Hemogenic endothelium” of the embryonic aorta: Does it exist? Developmental & Comparative Immunology, *6*(2), 359–368.

Song, H.-D., Sun, X.-J., Deng, M., Zhang, G.-W., Zhou, Y., Wu, X.-Y., Sheng, Y., et al. (2004). Hematopoietic gene expression profile in zebrafish kidney marrow. PNAS, *101*(46), 16240–16245.

Sood, R., English, M. A., Belele, C. L., Jin, H., Bishop, K., Haskins, R., McKinney, M. C., et al. (2010). Development of multilineage adult hematopoiesis in the zebrafish with a runx1 truncation mutation. Blood, *115*(14), 2806–2809.

Souilhol, C., Gonneau, C., Lendinez, J. G., Batsivari, A., Rybtsov, S., Wilson, H., Morgado-Palacin, L., Hills, D., Taoudi, S., Antonchuk, J., Zhao, S., & Medvinsky, A. (n.d.). Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells, *7*, 10784 EP –. doi:Article

Souilhol, C., Gonneau, C., Lendinez, J. G., Batsivari, A., Rybtsov, S., Wilson, H., Morgado-Palacin, L., Hills, D., Taoudi, S., Antonchuk, J., Zhao, S., & Medvinsky, A. (2016a). Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells. Nature Communications, *7*, 10784.

Souilhol, C., Lendinez, J. G., Rybtsov, S., Murphy, F., Wilson, H., Hills, D., Batsivari, A., et al. (2016b). Developing HSCs become Notch independent by the end of maturation in the AGM region. Blood, *128*(12), 1567–1577.

Souroullas, G. P., Salmon, J. M., Sablitzky, F., Curtis, D. J., & Goodell, M. A. (2009). Adult Hematopoietic Stem and Progenitor Cells require either Lyl1 or Scl for survival. Cell Stem Cell, *4*(2), 180–186.

Speck, N. A., & Baltimore, D. (1987). Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. Molecular and Cellular Biology, *7*(3), 1101–1110.

Stark, R., & Brown, G., (2011). DiffBind: Differential binding analysis of ChIP-seq data. Bioconductor. Available online at: http://bioconductor.org/packages/release/bioc/html/DiffBind.html.

Sugimura, R., Jha, D. K., Han, A., Soria-Valles, C., da Rocha, E. L., Lu, Y.-F., Goettel, J. A., et al. (2017). Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature*, *545*(7655), 432–438.

Sumanas, S., Gomez, G., Zhao, Y., Park, C., Choi, K., & Lin, S. (2008). Interplay among Etsrp/ER71, Scl, and Alk8 signaling controls endothelial and myeloid cell formation. Blood, *111*(9), 4500–4510.

Sumanas, S., Jorniak, T., & Lin, S. (2005). Identification of novel vascular endothelial–specific genes by the microarray analysis of the zebrafish cloche mutants. Blood, *106*(2), 534–541.

Sun, H. Q., Yamamoto, M., Mejillano, M., & Yin, H. L. (1999). Gelsolin, a Multifunctional Actin Regulatory Protein. J Biological Chemistry, *274*(47), 33179–33182.

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PloS one, *6*(7), e21800.

Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K. D., Yoshimoto, M., Yoder, M. C., Frampton, J., Schroeder, T., Enver, T., Göttgens, B., & de Bruijn, M. F. T. R. (2013). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. Nature Communications, *4*, 2924.

Szabo, E., Rampalli, S., Risueño, R. M., Schnerch, A., Mitchell, R., Fiebig-Comyn, A., Levadoux-Martin, M., et al. (2010). Direct conversion of human fibroblasts to multilineage blood progenitors. Nature, *468*(7323), 521–526.

Tahirov, T. H., Inoue-Bungo, T., Morii, H., Fujikawa, A., Sasaki, M., Kimura, K., Shiina, M., et al. (2001). Structural Analyses of DNA Recognition by the AML1/Runx-1 Runt Domain and Its Allosteric Control by CBFβ2. Cell, *104*(5), 755–767.

Taoudi, S., Gonneau, C., Moore, K., Sheridan, J. M., Blackburn, C. C., Taylor, E., & Medvinsky, A. (2008). Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-Cadherin+CD45+ pre-definitive HSCs. Cell Stem Cell, *3*(1), 99–108.

Thambyrajah, R., Mazan, M., Patel, R., Moignard, V., Stefanska, M., Marinopoulou, E., Li, Y., et al. (201). GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. Nature Cell Biology, *18*, 21–32.

Tober, J., Koniski, A., McGrath, K. E., Vemishetti, R., Emerson, R., de Mesy-Bentley, K. K. L., Waugh, R., et al. (2007). The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. Blood, *109*(4), 1433–1441.

Tober, J., Yzaguirre, A. D., Piwarzyk, E., & Speck, N. A. (2013). Distinct temporal requirements for Runx1 in hematopoietic progenitors and stem cells. Development, *140*(18), 3765–3776.

Vanhaesebroek, B., Stephens, L., & Hawkins, P. (2012). PI3K signaling: the path to discovery and understanding. Nature Reviews Molecular Cell Biology, 13, 195–203.

van Meeteren, L. A., & Dijke, ten, P. (2012). Regulation of endothelial cell plasticity by TGF-β. Cell and Tissue Research, *347*(1), 177–186.

Vargel, Ö., Zhang, Y., Kosim, K., Ganter, K., Foehr, S., Mardenborough, Y., Shvartsman, M., et al. (n.d.). Activation of the TGFβ pathway impairs endothelial to

haematopoietic transition. Scientific Reports, *6*, 21518.

Wang, J. H., Nichogiannopoulou, A., Wu, L., Sun, L., Sharpe, A. H., Bigby, M., & Georgopoulos, K. (1996a). Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. Immunity, *5*(6), 537–549.

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A. H., & Speck, N. A. (1996b). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. PNAS, *93*(8), 3444–3449.

Wareing, S., Mazan, A., Pearson, S., Göttgens, B., Lacaud, G., & Kouskoff, V. (2012). The Flk1-Cre-Mediated Deletion of ETV2 Defines Its Narrow Temporal Requirement During Embryonic Hematopoietic Development. Stem cells, *30*(7), 1521–1531.

Wilkinson, R. N., Koudijs, M. J., Patient, R. K., Ingham, P. W., Schulte-Merker, S., & van Eeden, F. J. M. (2012). Hedgehog signaling via a calcitonin receptor-like receptor can induce arterial differentiation independently of VEGF signaling in zebrafish. Blood, *120*(2), 477–488.

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. Developmental Cell, *16*(6), 909–916.

Wilkinson, A.C., Kawata, V.K., Schutte, J., Gao, X., Antoniou, S. et al. (2014). Single-cell analysis of regulatory network perturbations using enhancer-targeting TALEs suggests novel roles for PU.1 during haematopoietic specification. Development, 141(20), 4018–4030.

Will, B., Vogler, T. O., Bartholdy, B., Garrett-Bakelman, F., Mayer, J., Barreyro, L., Pandolfi, A., et al. (2013). Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. Nature Immunology, *14*(5), 437–445.

Williams, C., Kim, S. H., Ni, T. T., Mitchell, L., Ro, H., Penn, J. S., Baldwin, S. H., et al. (2010). Hedgehog signaling induces arterial endothelial cell formation by repressing venous cell fate. Developmental biology, *341*(1), 196–204.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., Offner, S., et al. (2008). Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. Cell, *135*(6), 1118–1129.

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P. M., Kinston, S., et al. (2010a). Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators.).. *Cell stem cell*, *7*(4), 532–544.

Wilson, N. K., Timms, R. T., Kinston, S. J., Cheng, Y.-H., Oram, S. H., Landry, J.-R., Mullender, J., Ottersbach, K., et al. (2010b). Gfi1 Expression Is Controlled by Five Distinct Regulatory Regions Spread over 100 Kilobases, with Scl/Tal1, Gata2, PU.1, Erg, Meis1, and Runx1 Acting as Upstream Regulators in Early Hematopoietic Cells. *Molecular and Cellular Biology*, *30*(15), 3853–3863.

Woll, P. S., Grzywacz, B., Tian, X., Marcus, R. K., Knorr, D. A., Verneris, M. R., & Kaufman, D. S. (2009). Human embryonic stem cells differentiate into a homogeneous population of natural killer cells with potent in vivo antitumor activity. Blood, *113*(24), 6094–6101.

Yamaguchi, T. P., Dumont, D. J., Conlon, R. A., Breitman, M. L., & Rossant, J. (1993). flk-1, an flt-related receptor tyrosine kinase is an early marker for endothelial cell precursors. Development, *118*(2), 489–498.

Yokomizo, T., Ogawa, M., Osato, M., Kanno, T., Yoshida, H., Fujimoto, T., Fraser, S., et al. (2001). Requirement of Runx1/AML1/PEBP2αB for the generation of haematopoietic cells from endothelial cells. Genes to Cells, *6*(1), 13–23.

Yu, M., Mazor, T., Huang, H., Huang, H.-T., Kathrein, K., Woo, A., Chouinard, C., et al. (2012). Direct Recruitment of Polycomb Repressive Complex 1 (PRC1) to Chromatin by Core Binding Transcription Factors. Molecular Cell, *45*(3), 330–343.

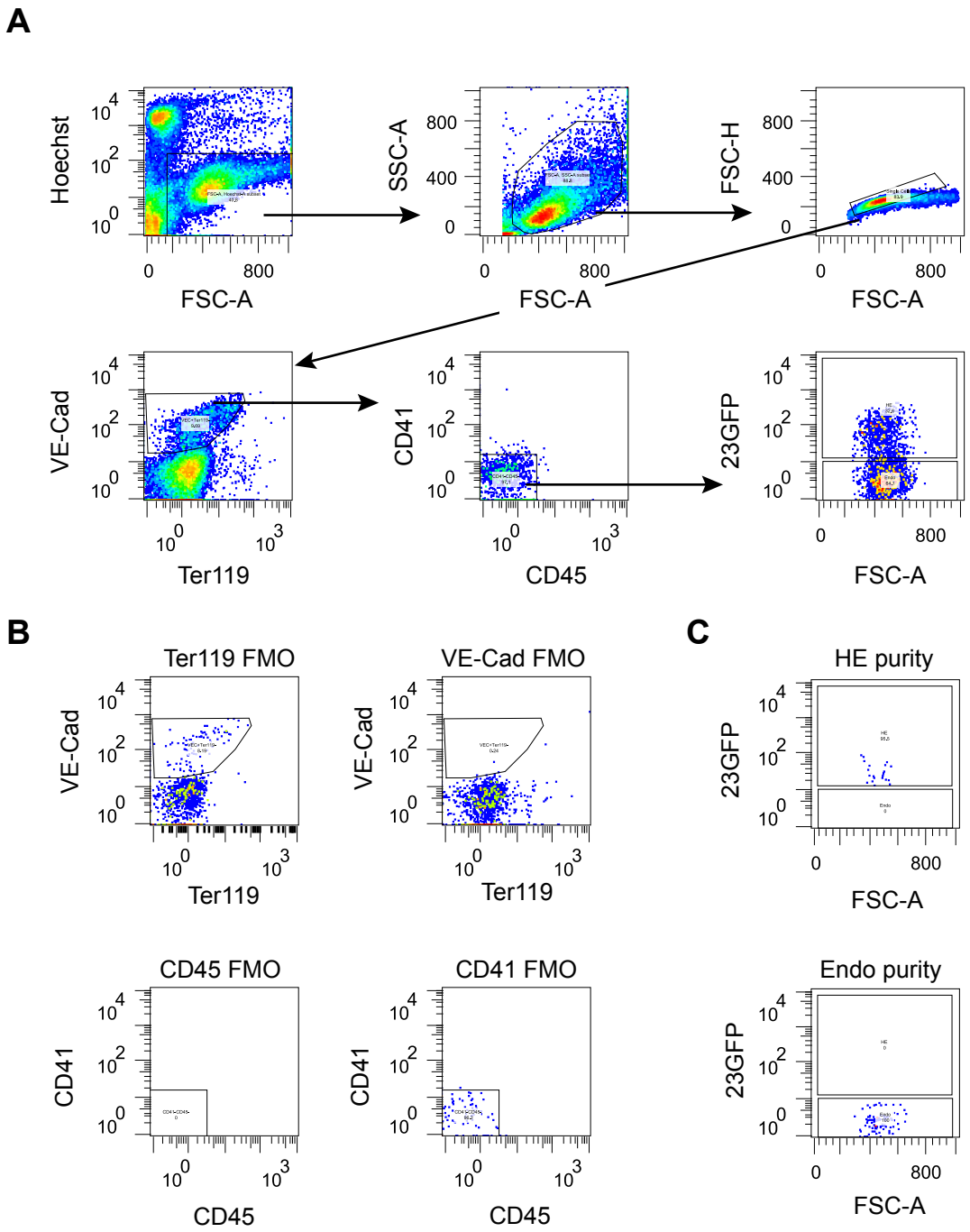
Zhang, C. C., Kaba, M., Ge, G., Xie, K., Tong, W., Hug, C., & Lodish, H. F. (2006). Angiopoietin-like proteins stimulate ex vivo expansion of hematopoietic stem cells, Nature Medicine, *12*, 240–245.

Zhong, T. P., Childs, S., Leu, J. P., & Fishman, M. C. (2001). Gridlock signalling pathway fashions the first embryonic artery. *Nature*, *414*, 216–220.

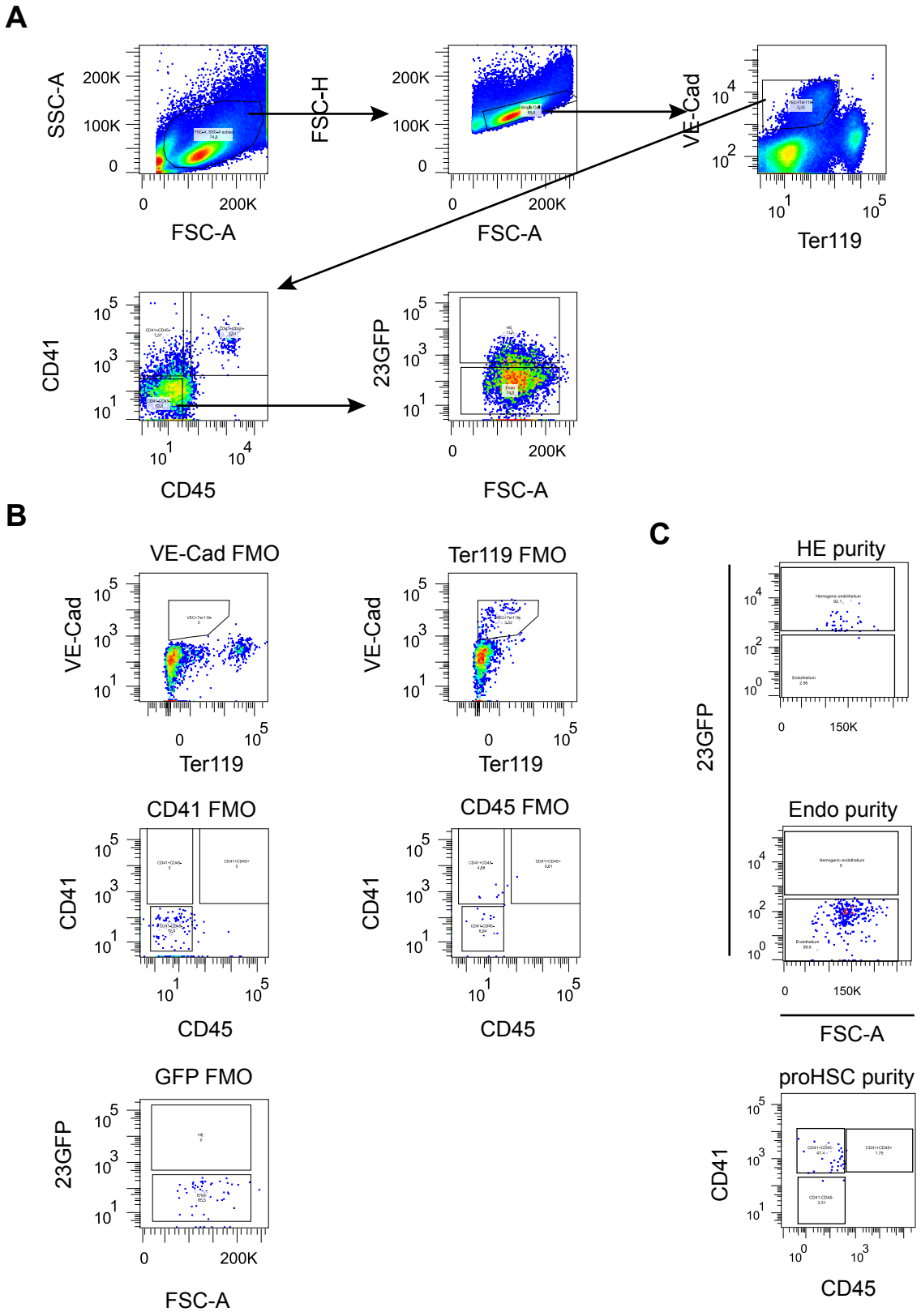
Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., Li, Z., et al. (2016). Tracing haematopoietic stem cell formation at single-cell resolution. Nature *533*, 487–492.

Zovein, A. C., Hofmann, J. J., Lynch, M., French, W. J., Turlo, K. A., Yang, Y., Becker, M. S., et al. (2008). Fate Tracing Reveals the Endothelial Origin of Hematopoietic Stem Cells. Cell stem cell, *3*(6), 625–636.

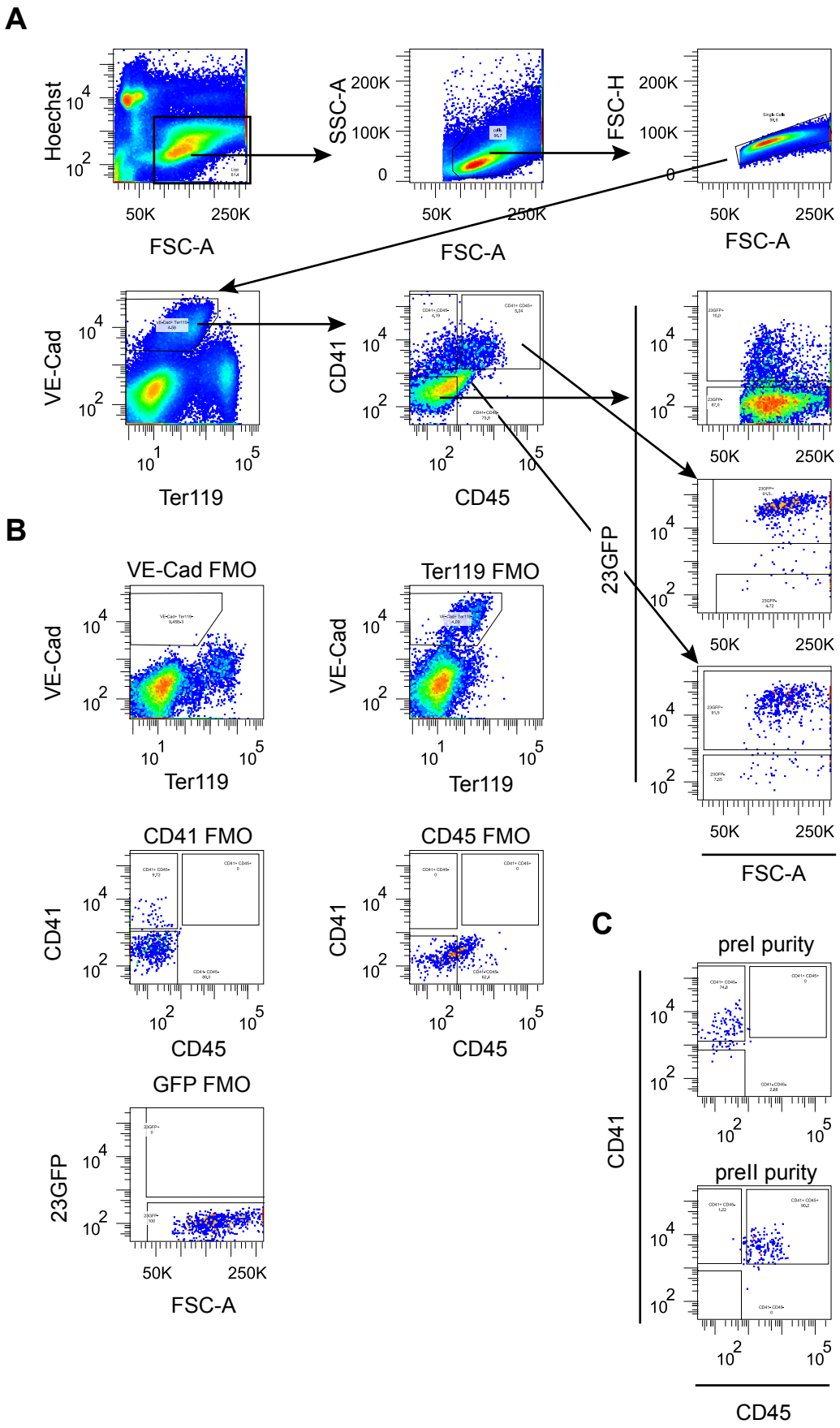
Appendix



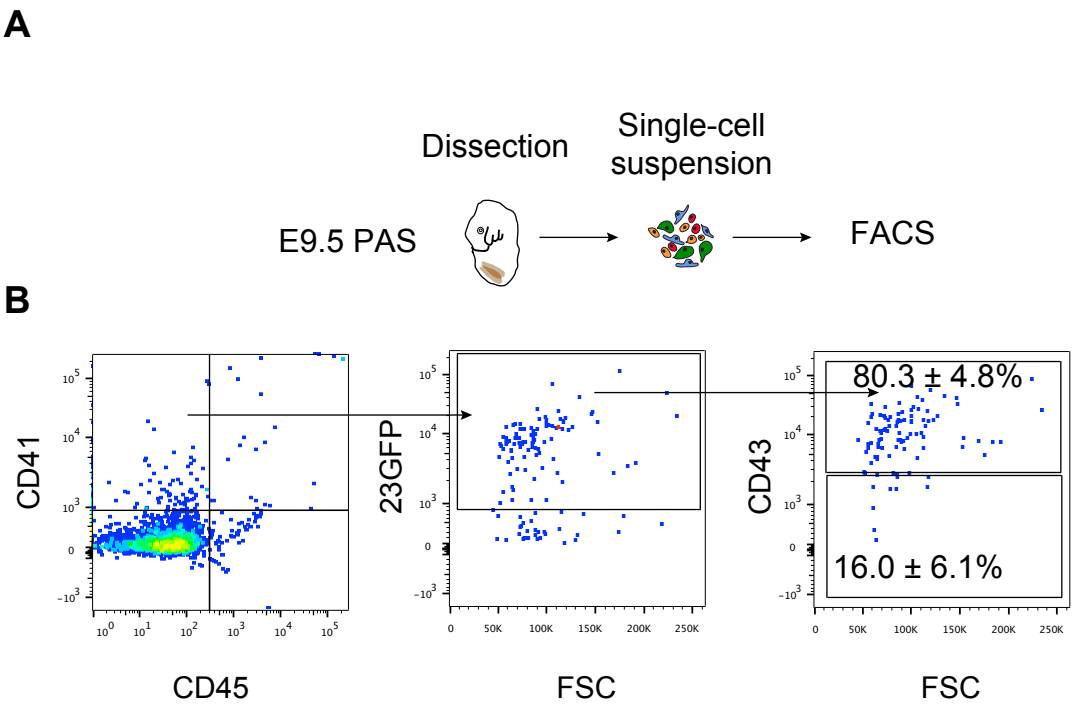
Appendix Figure 3.1: Control plots and purity checks for sorts at E8.5. Representative gates for live, single, Ter119/VEC and CD41/CD45 plots for E8.5. Gates were based on fluorescence minus one (FMO) control for all five fluorochromes. (A) Sorting strategy. (B) FMO controls. (C) Puirty checks done on HE and Endo populations.



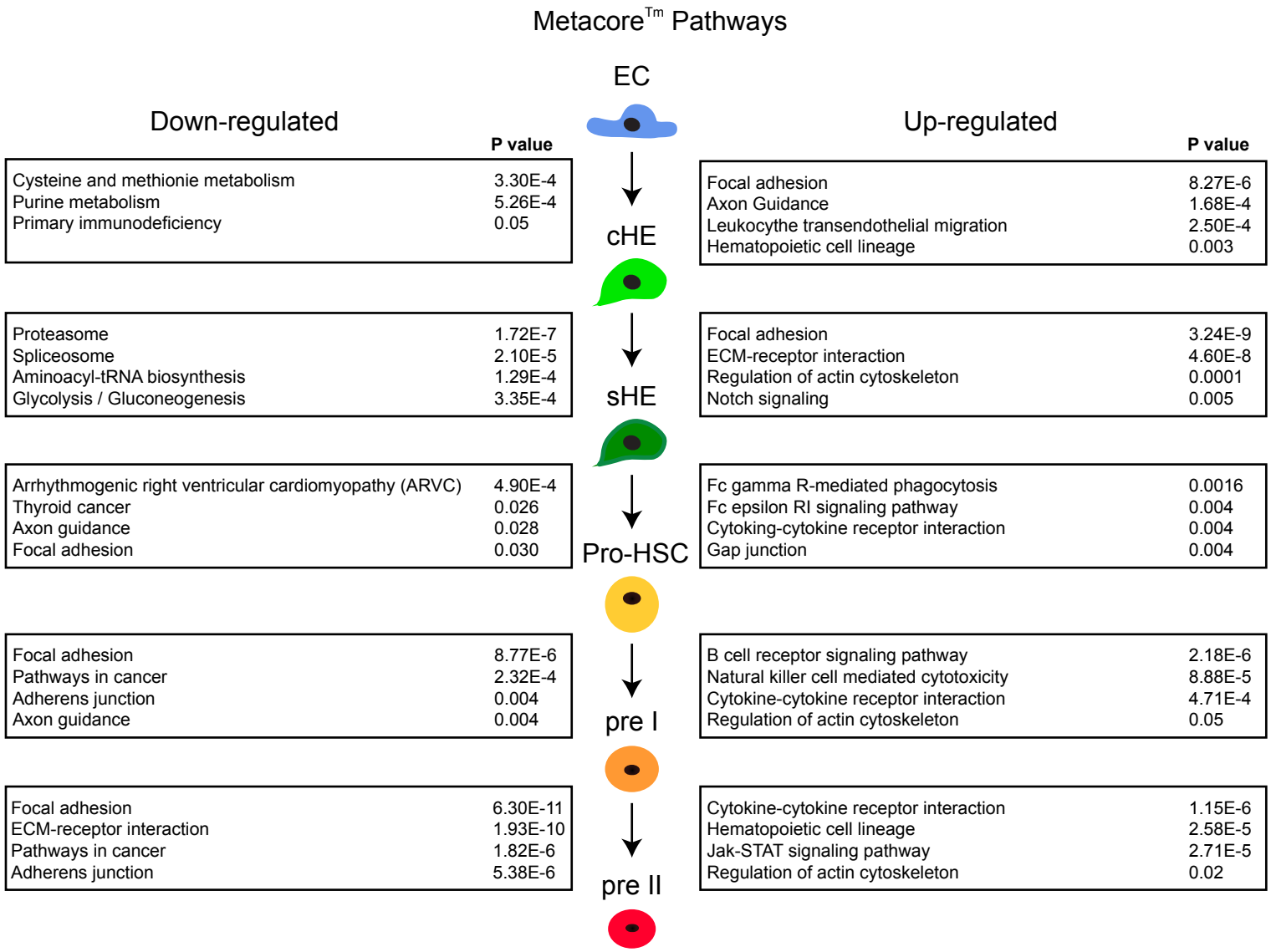
Appendix Figure 3.2: Control plots and purity checks for sorts at E9.5. Representative sort gates and reanalysis for E9.5 PAS + VU. Gates for live, single, Ter119/VEC and CD41/CD45 plots for E9.5 are shown. Representative reanalysis percentages are shown. Gates were based on fluorescence minus one (FMO) control for all five fluorochromes.



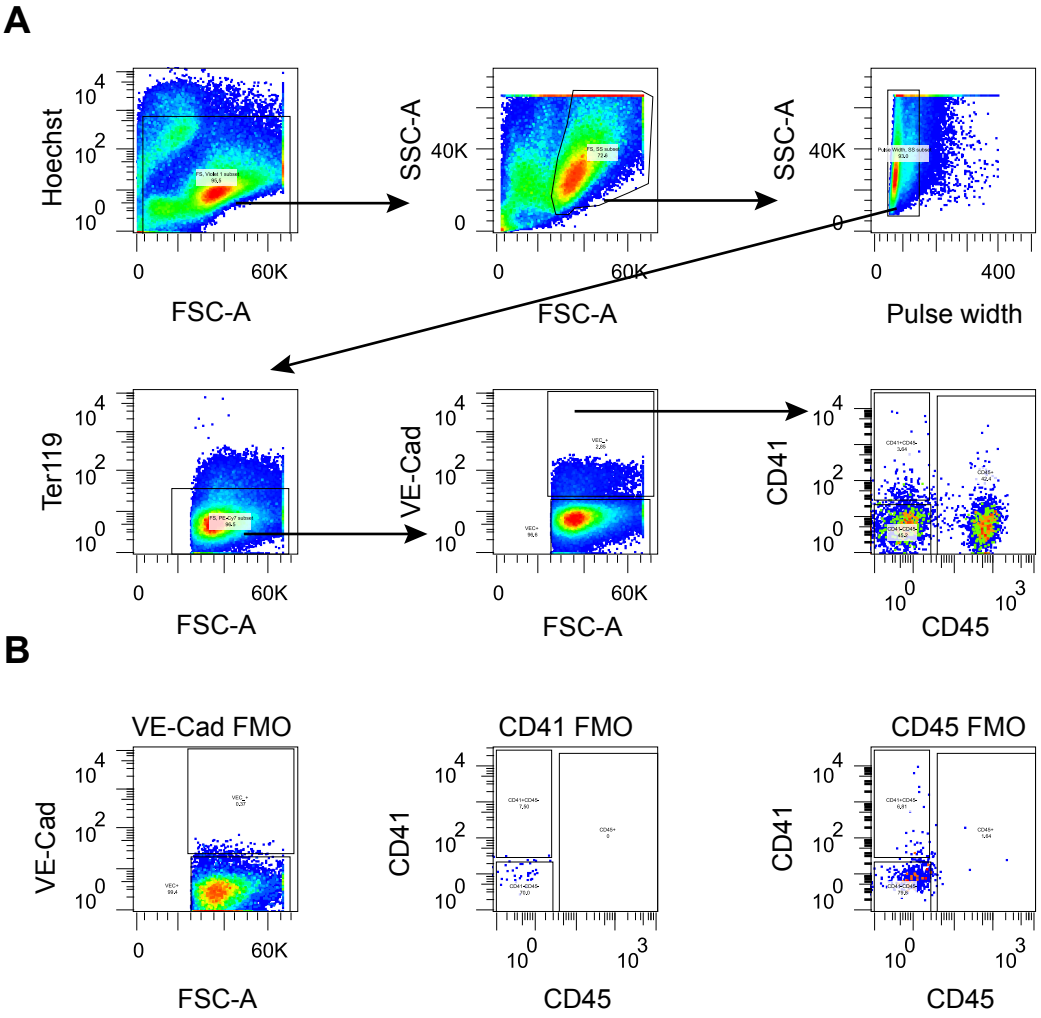
Appendix Figure 3.3: Control plots and purity checks for sorts at E10.5. Representative sort gates and reanalysis for E10.5 AGM + VU. Gates for live, single, Ter119/VEC and CD41/CD45 plots for E10.5 are shown. Representative re-analysis percentages are shown. Gates were based on fluorescence minus one (FMO) control for all five fluorochromes.



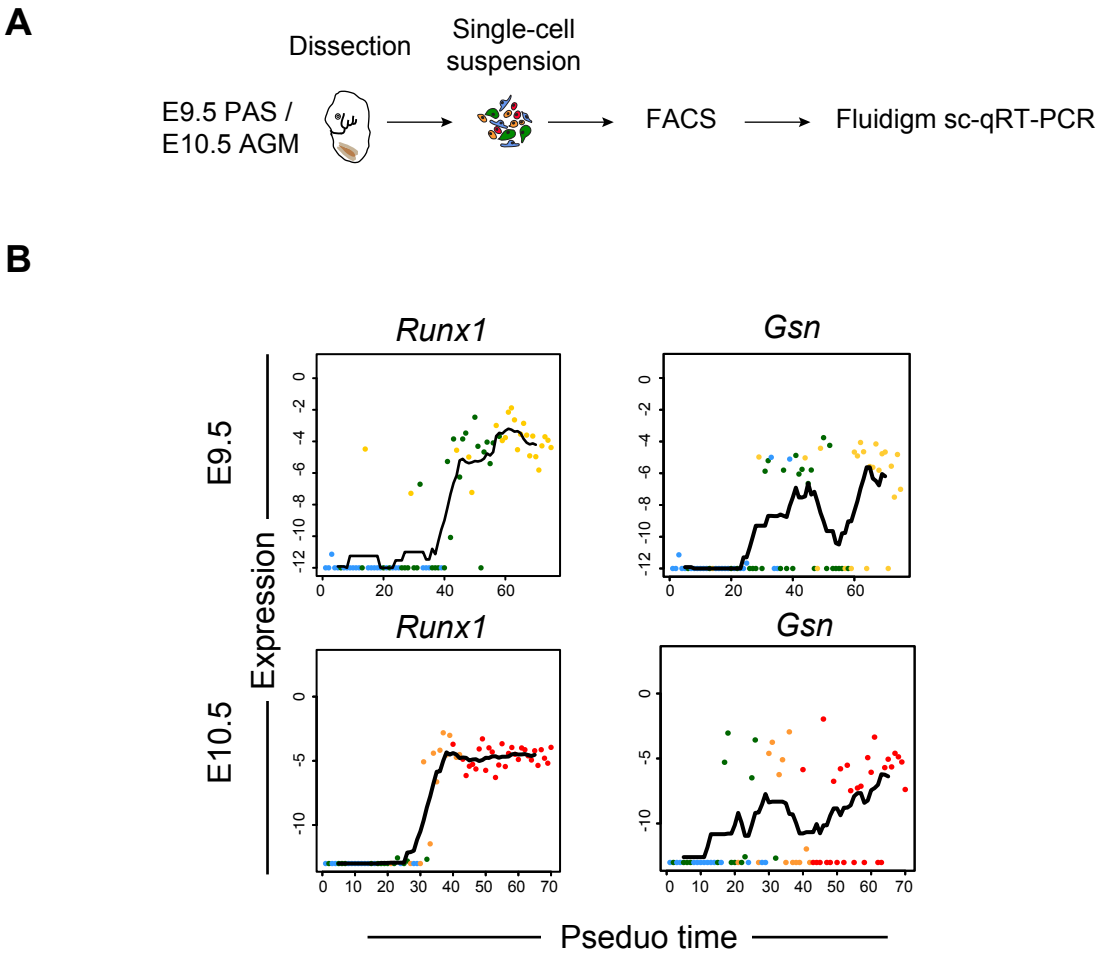
Appendix Figure 3.4: Examination of E9.5 CD41+ cells for phenotypic proHSC. (A) Schematic for analyzing E9.5 23GFP embryos. (B) Flow cytometry profile showing percent of E9.5 CD41+ cells that were also negative for CD43 (16.0± 6.1%). Work was done by Emmanuele Azzoni.



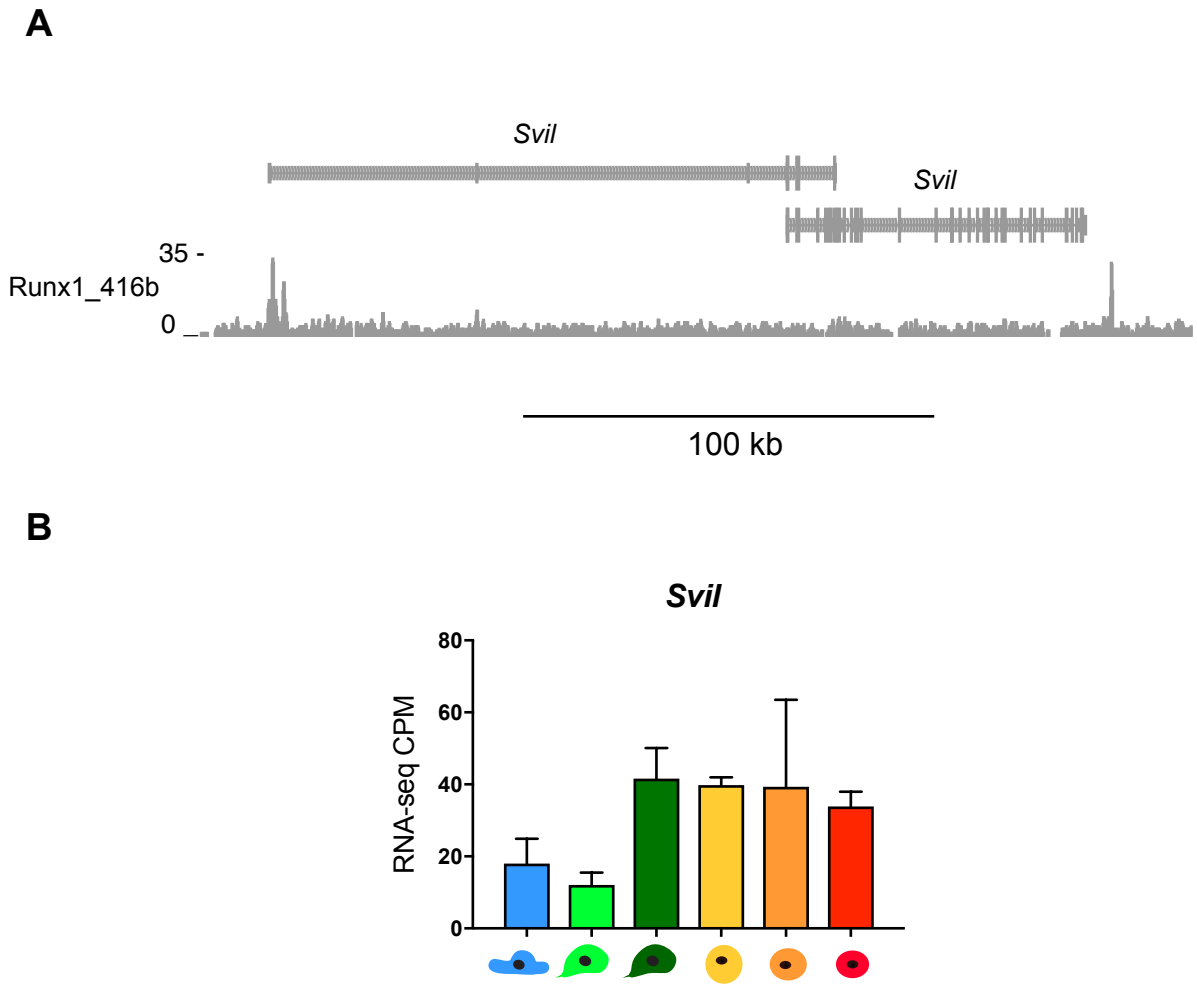
Appendix Figure 3.5: Summary of Metacore™ pathways enriched in up-regulated and down-regulated genes for each transition. The Top 4 pathways for each transition are listed in order of significance by on -log10 Pvalues.



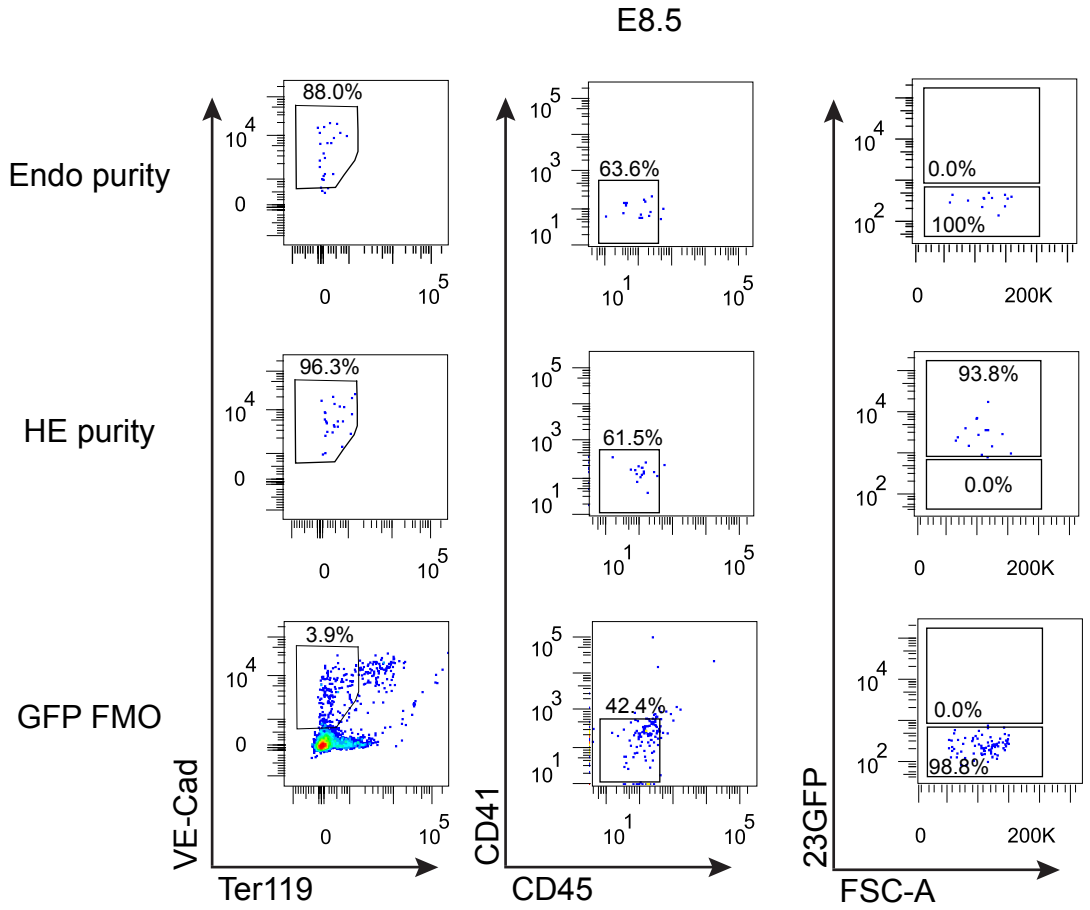
Appendix Figure 3.6: Gating strategy and control plots for mESC differentiations treated with CytoD. (A) Representative gates for live, single, Ter119, VEC and CD41/CD45 plots for mESC. (B) Gates were based on fluorescence minus one (FMO) control for all five fluorochromes.



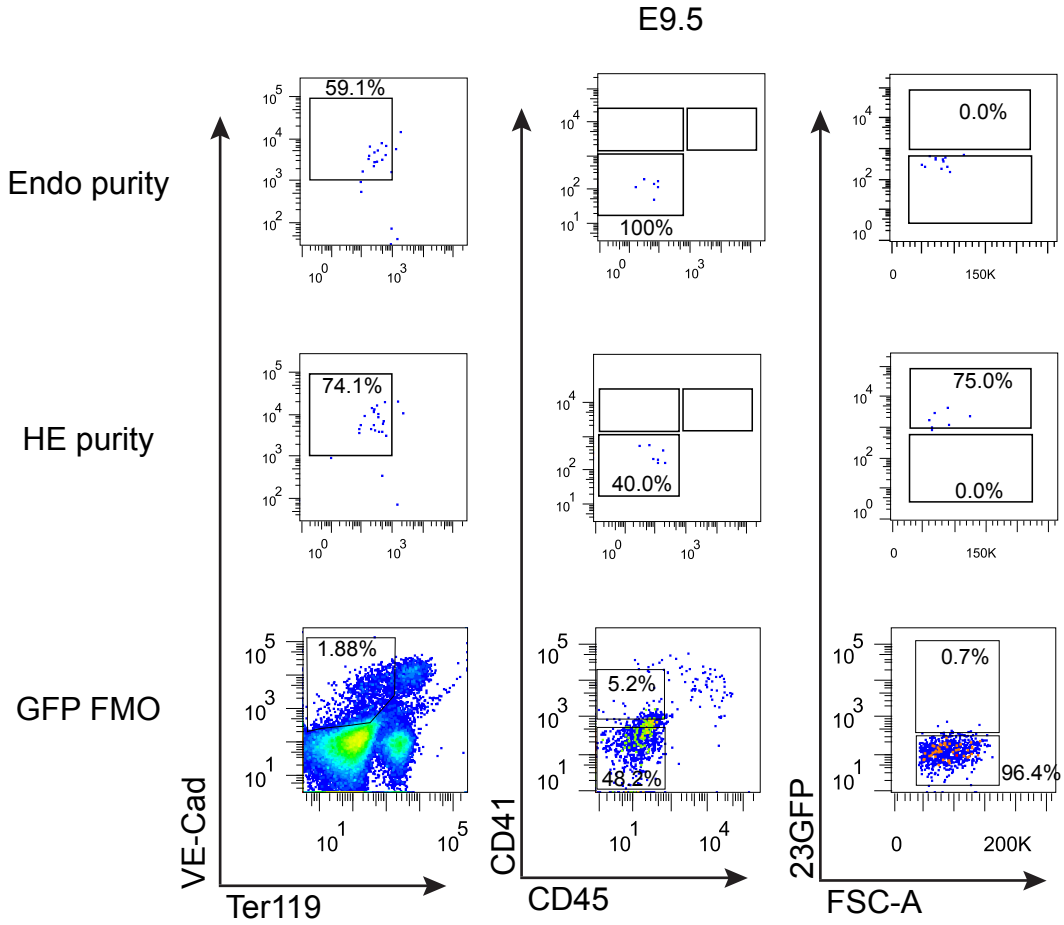
Appendix Figure 3.7: Single cell qRT-PCR for Runx1 and Gelsolin at E9.5 and E10.5. (A) Schematic of single cell qRT-PCR experiments. (B) Pseudo time plots for Runx1 and Gsn at E9.5 and E9.5. Solid line represents moving average with each dot representing expression for one cell. Flow plots and other controls are discussed in Chapter 6.



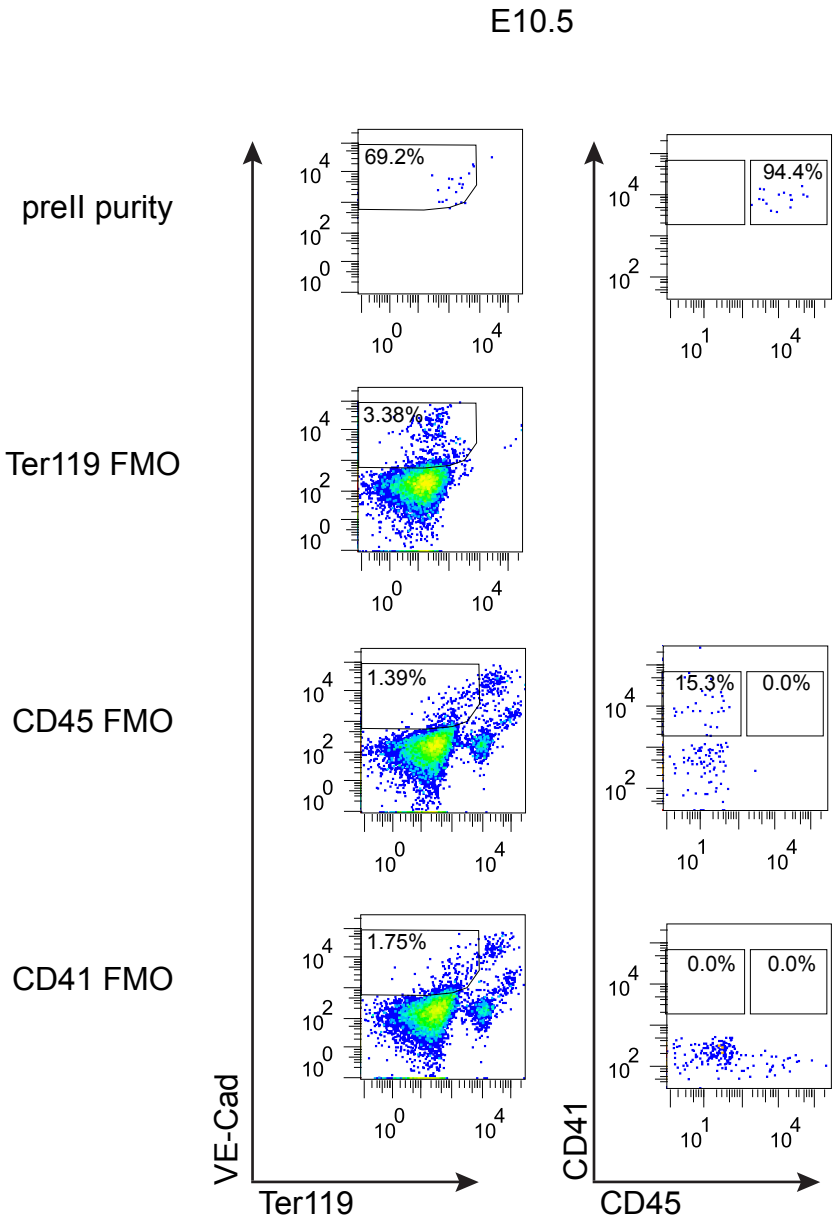
Appendix Figure 3.8: Gelsolin family member Supervilin. (A) Runx1 ChIP-seq at Svil locus (Chr18:4,904,116 - 5,149,496). Two major isoforms are shown in gene track. (B) RNA-seq CPM values for Svil during in vivo EHT.



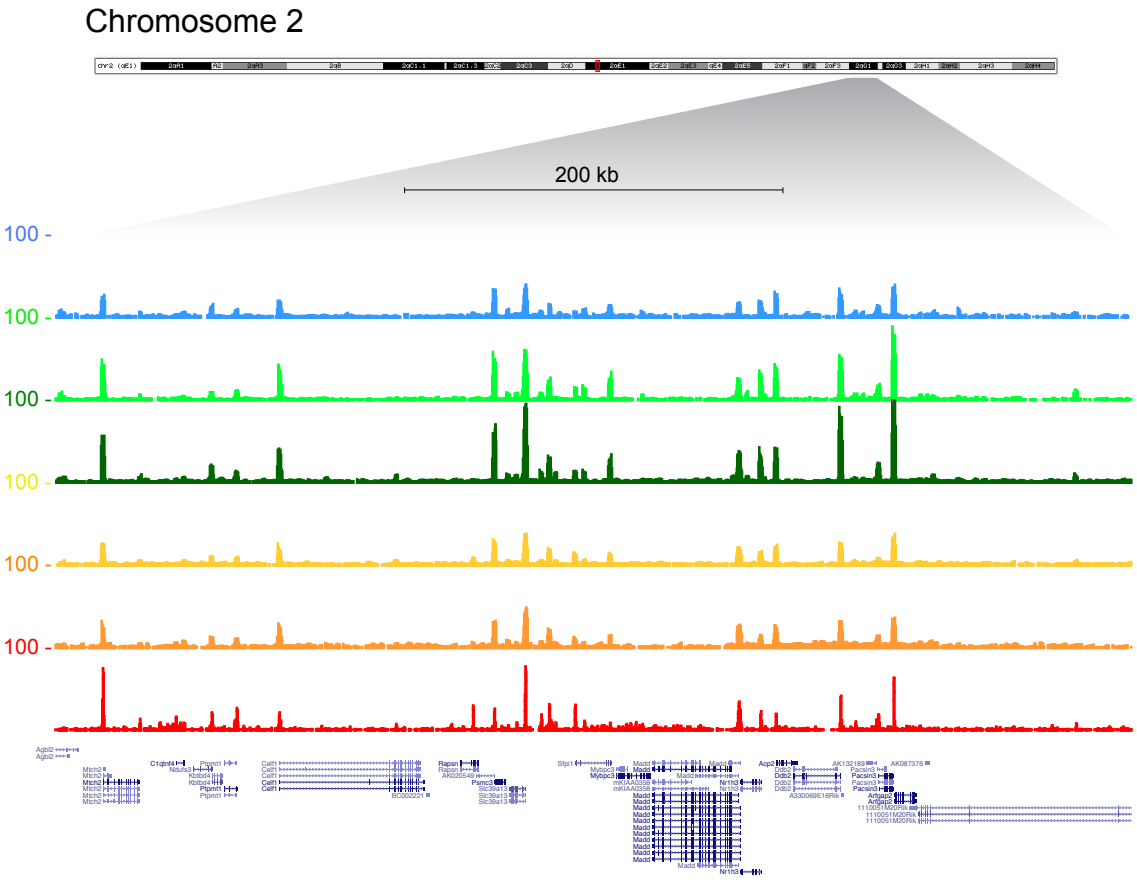
Appendix Figure 4.1: Control plots and purity checks for sorts at E8.5. Representative gates for live, single, Ter119/VEC and CD41/CD45 plots for E8.5. Gates were based on fluorescence minus one (FMO) controls for all five fluorochromes.



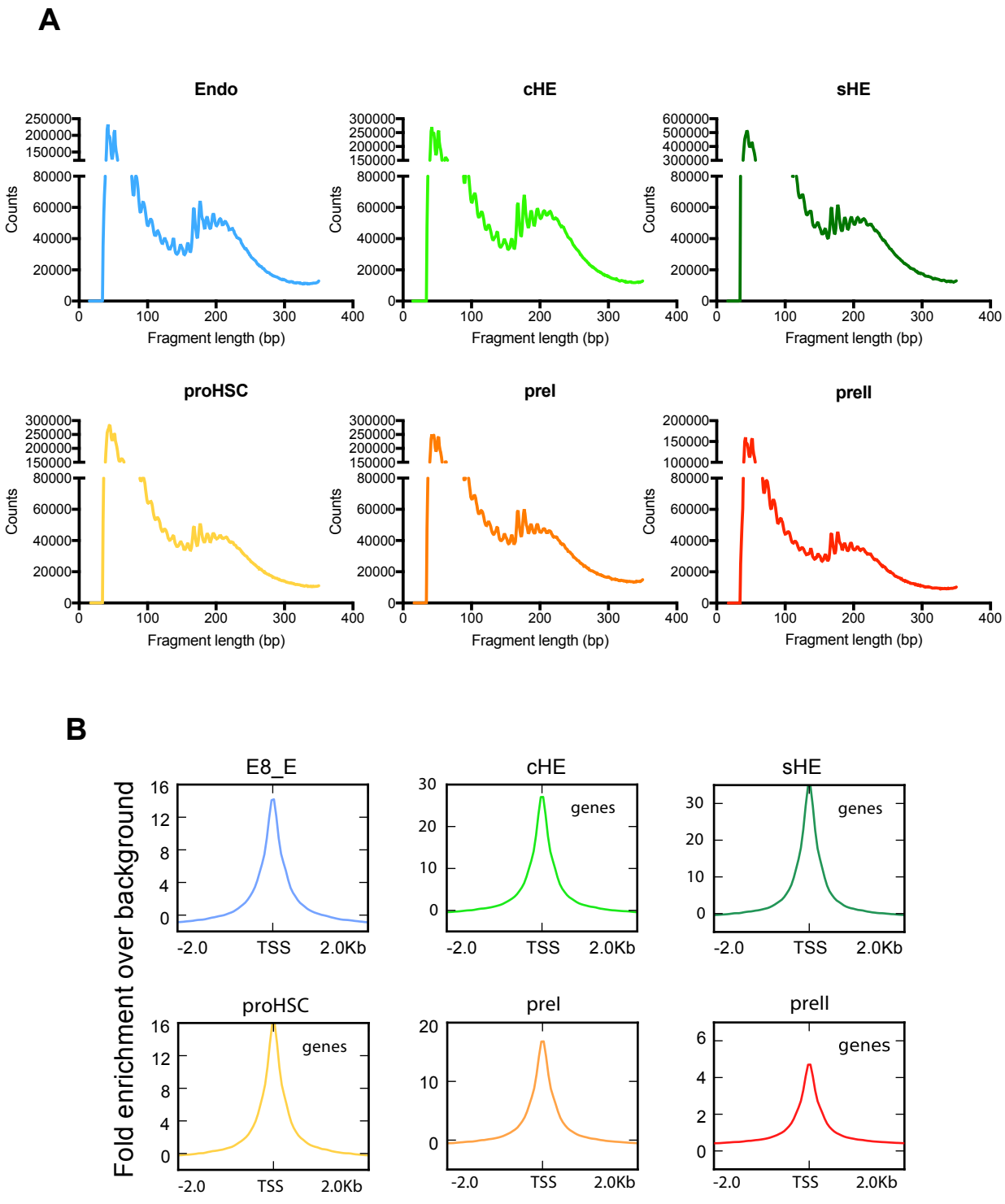
Appendix Figure 4.2: Control plots and purity checks for sorts at E9.5. Representative gates for live, single, Ter119/VEC and CD41/CD45 plots for E9.5. Gates were based on fluorescence minus one (FMO) controls for all five fluorochromes.



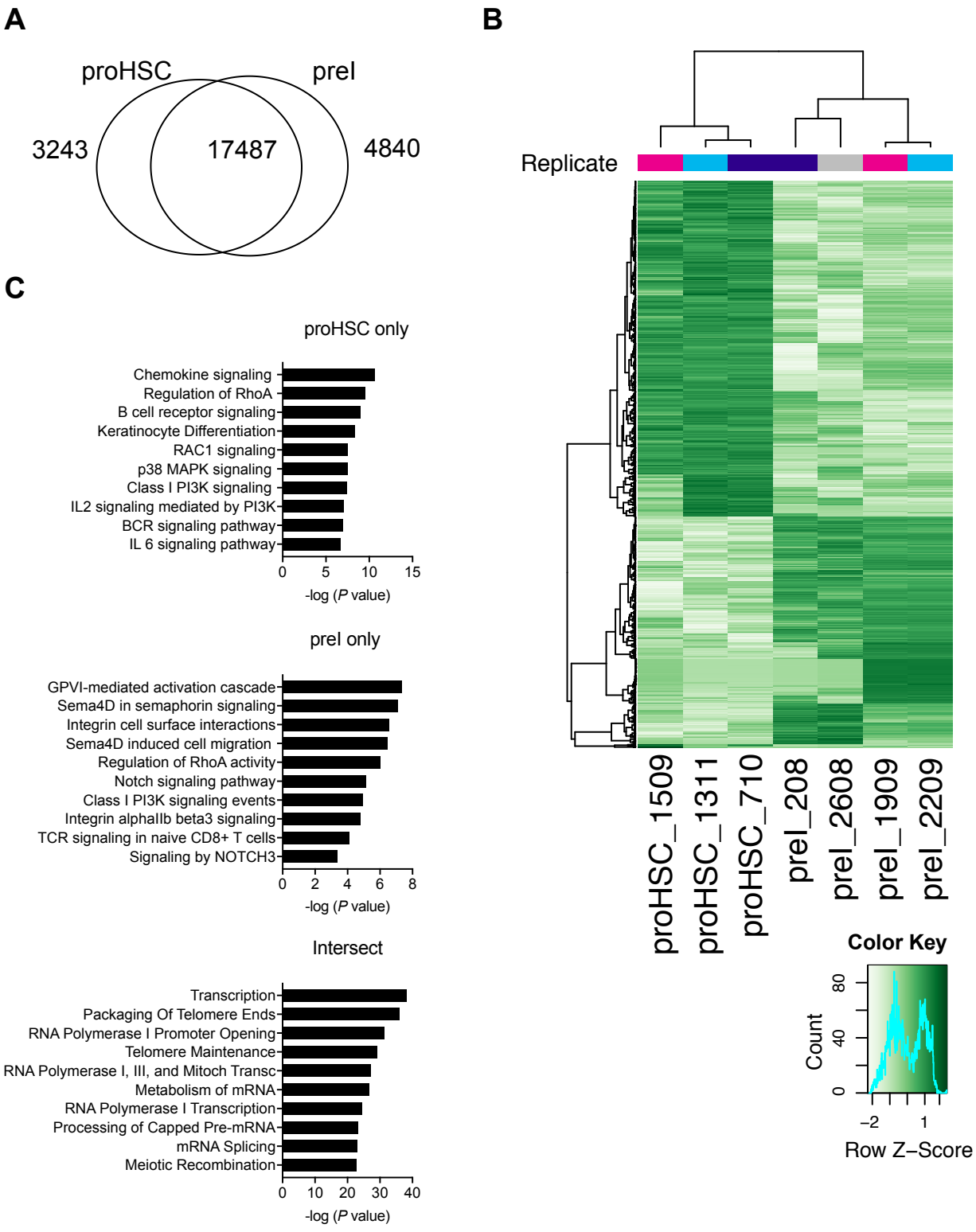
Appendix Figure 4.3: Control plots and purity checks for sorts at E10.5. Representative gates for live, single, Ter119/VEC and CD41/CD45 plots for E10.5. Gates were based on fluores cence minus one (FMO) controls for all five fluorochromes.



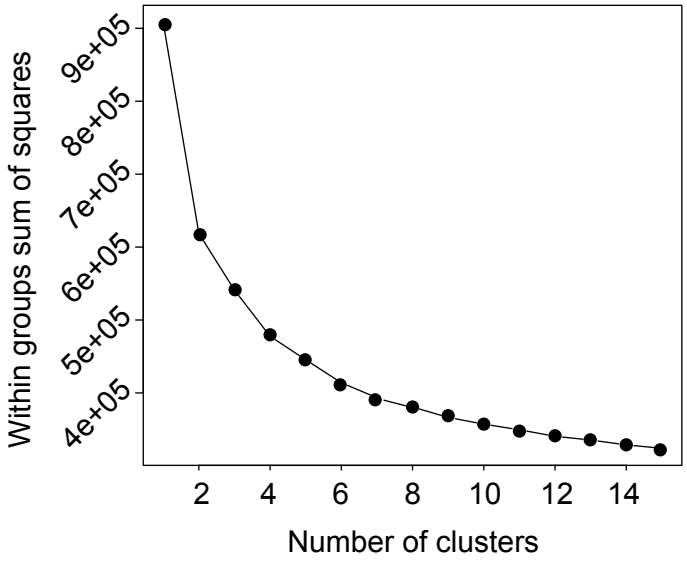
Appenix Figure 4.4: UCSC genome browser tracks showing coverage and quality of data.



Appendix Figure 4.5: ATAC-seq quality control. (A) Representative plot of aggregate signal around TSSs for each of the six populations sampled. **(B)** Representavie insert size distribution for each of the populations sampled.



Appendix Figure 4.6: Analysis of proHSC to prel transition. (A) Venn diagram shwoing intersect of all elements from each population. **(B)** Heatmap showing differentially accessible elements. (C) GREAT analysis of three groups from A.



Appendix Figure 4.7: Cumulative distribution function plot for determing k-means clusters.

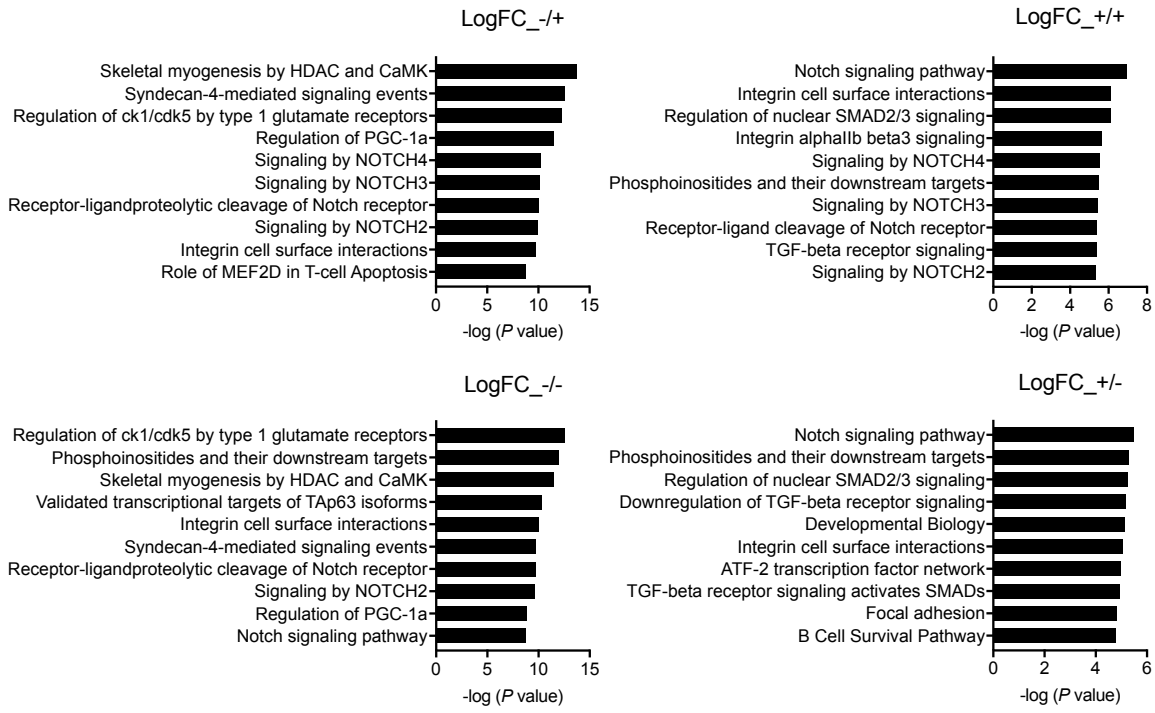


Figure Appendix 5.1: Integration of expression and chromatin profiles for cHE to sHE transition. Top 10 MsigDB pathways represented as histograms showing -log Pvalue.

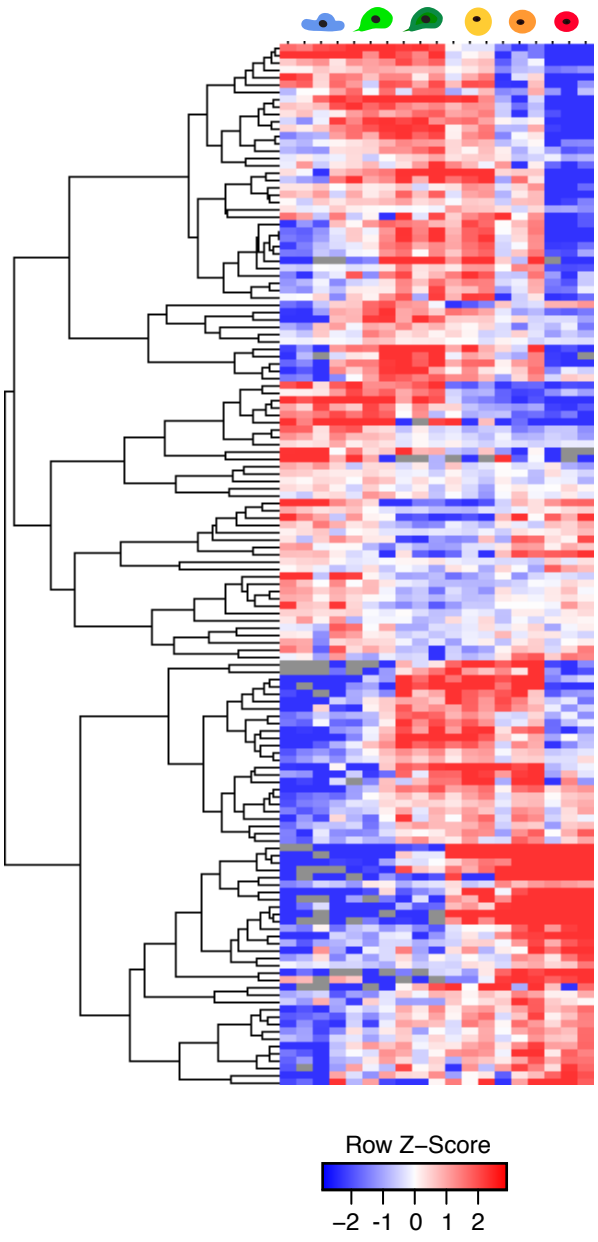


Figure Appendix 5.2: Differentially expressed genes from PI3K-AKT signaling pathway.

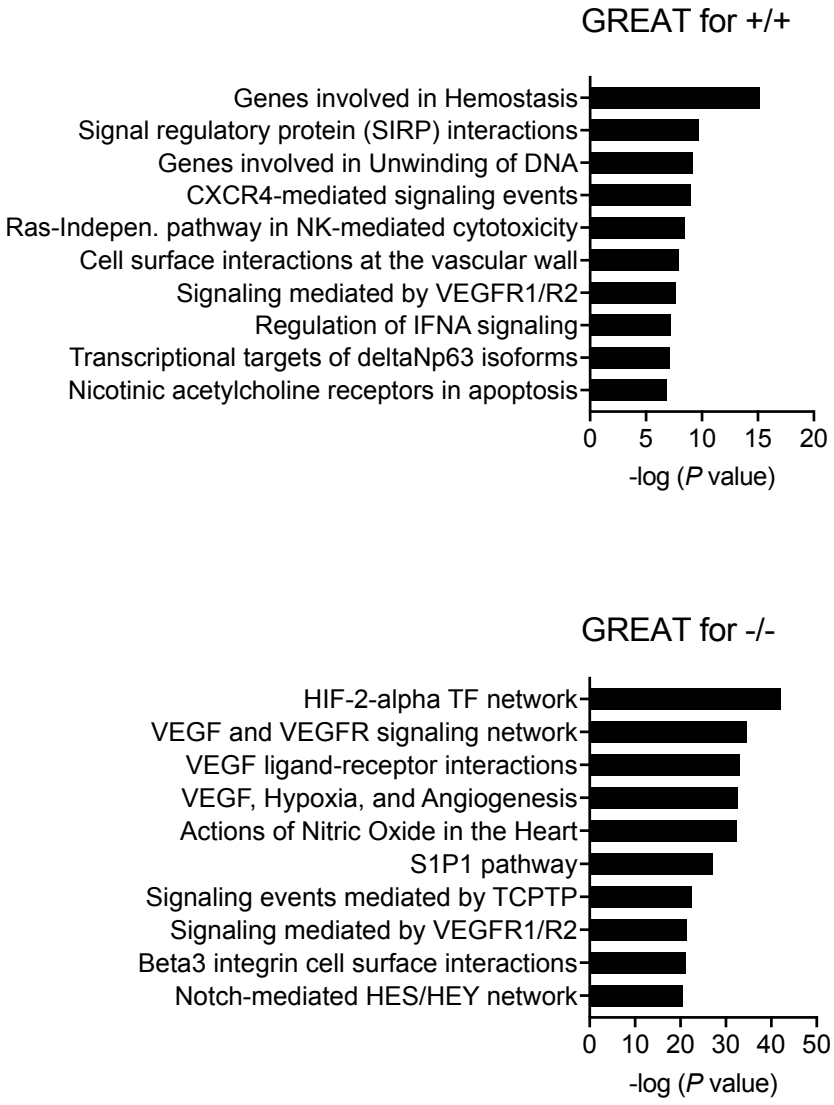
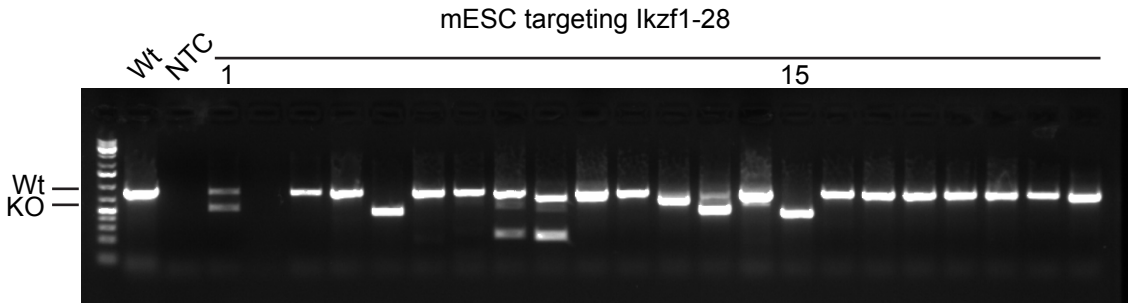


Figure Appendix 5.3: Great analysis for integrated elements in sHE to proHSC transition. Top 10 MsigDB pathways represented as histograms showing -log Pvalue.

A

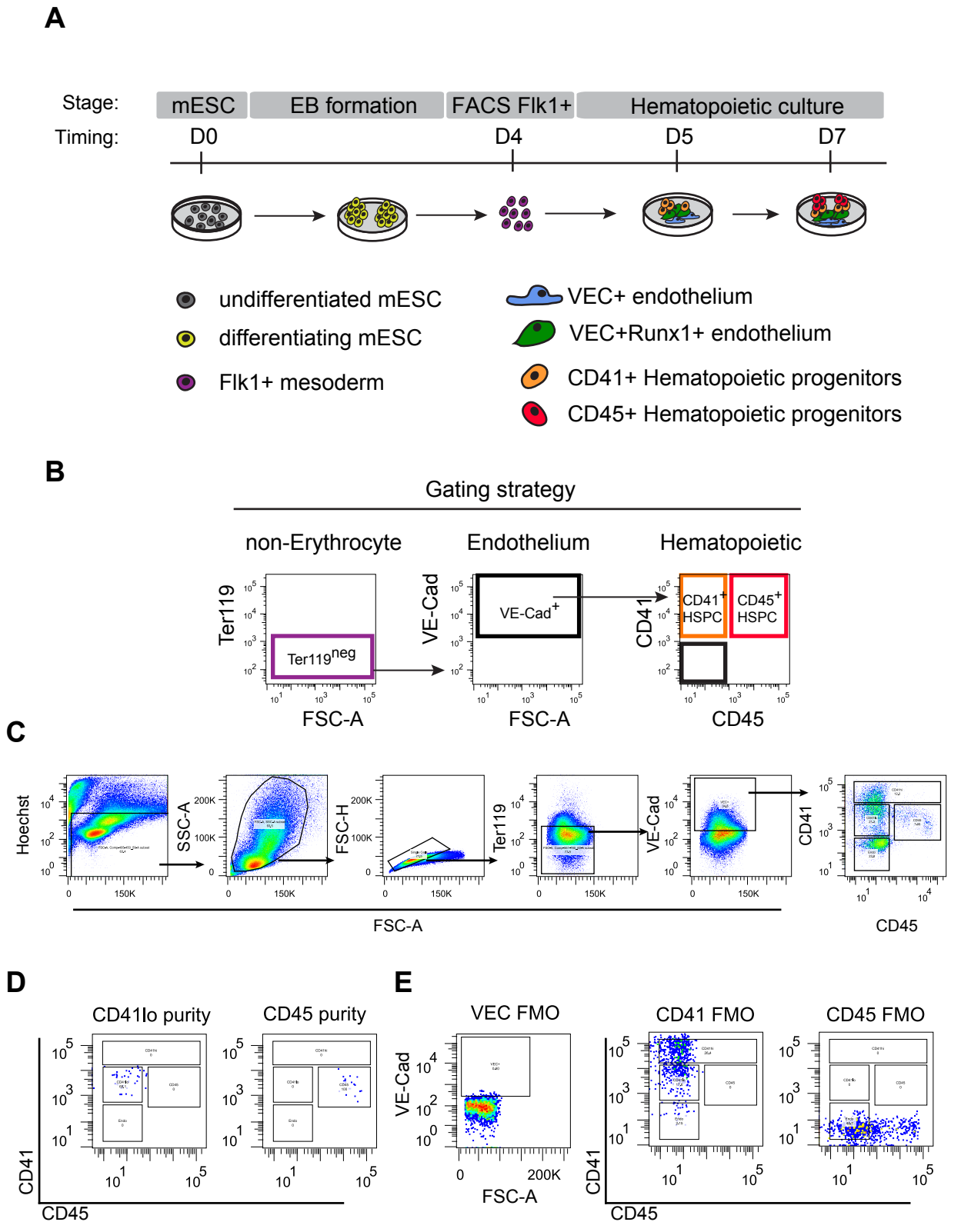


B

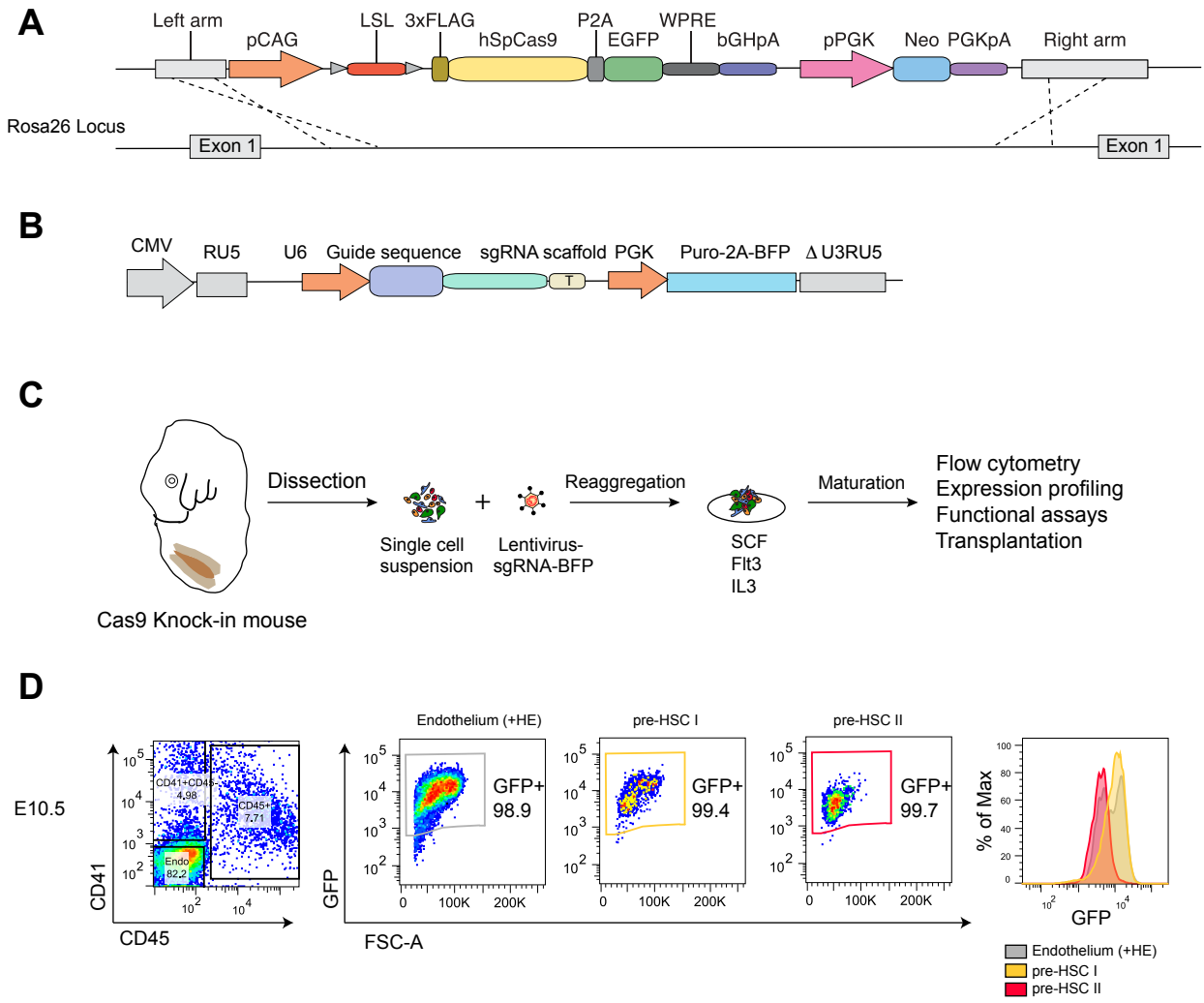
Ikzf1-28$^{-/-}$ clone 15

```
Wt        gcacactatagggcataaaattctagcacacagccctgggaatgtgttctattgaaacatttgctcacaactgaatgggaaagcaggaggccaaactgcccttgggtcagggtaaacactgctctttcctggct
              |||
Clone 15  gca----------------------------------------------------------------------------------------------------------------------------
Wt        ccagttgagtcatgtaaaagaaggtgattaatgagacattaaatggtgtgtatgttttaaggacacagaggacaggggctctcagaaccgcaggagcctgtggtttctgaaatactggtgggtggcttcctgtag
Clone 15  ----------------------------------------------------------------------------------------------------------------------------
Wt        ttgcaaggcatgctggggagatgagataggacaggccccacctaactcccttcatacggattctggcagtttggctcctgcagcctggtctgacctgctgctcccatgcagttagactggggaagcagagcag
Clone 15  ----------------------------------------------------------------------------------------------------------------------------
Wt        ctagctatcagccagacaagtaatagacatctaataaggtcacgtggccagaaacgcagaagcccacagaccagtttctgggggcagagatcggaagccattcactttattttgtgcctgtagacagtgactt
Clone 15  ----------------------------------------------------------------------------------------------------------------------------
Wt        taatgttatgtaaatgtagcccttaattcctaatgtctggaaatgtgtgtgtcagggctgtggcagctcaggttgtggatg
                                                                                |||||||||||
Clone 15  ----------------------------------------------------------------ctcangttgtggat
```

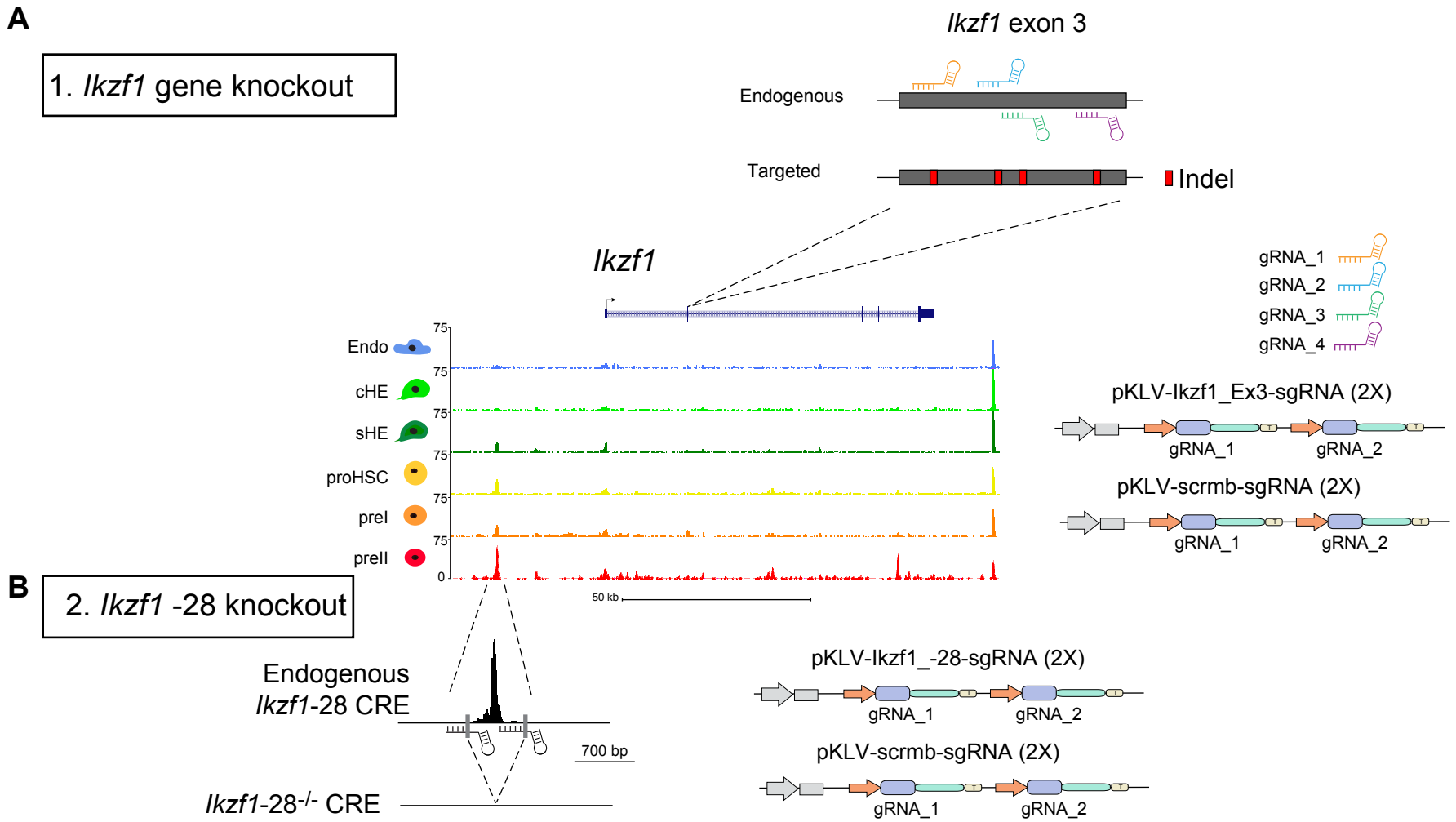
Appendix Figure 6.1: Generation of mESC harboring a Ikzf1-28 cis-element deletion. (A) Gel of clones targeted with CRISPR/Cas9 containing gRNA flanking Ikzf1-28. Clone 1 was identified as heterzygous using sanger sequencing to confirm both wt and -28 deleted allele. (B) Sanger sequencing track confirmed deletion of -28 for clone 15 from gel in (A) with "-" indicateing a deleted base pair.



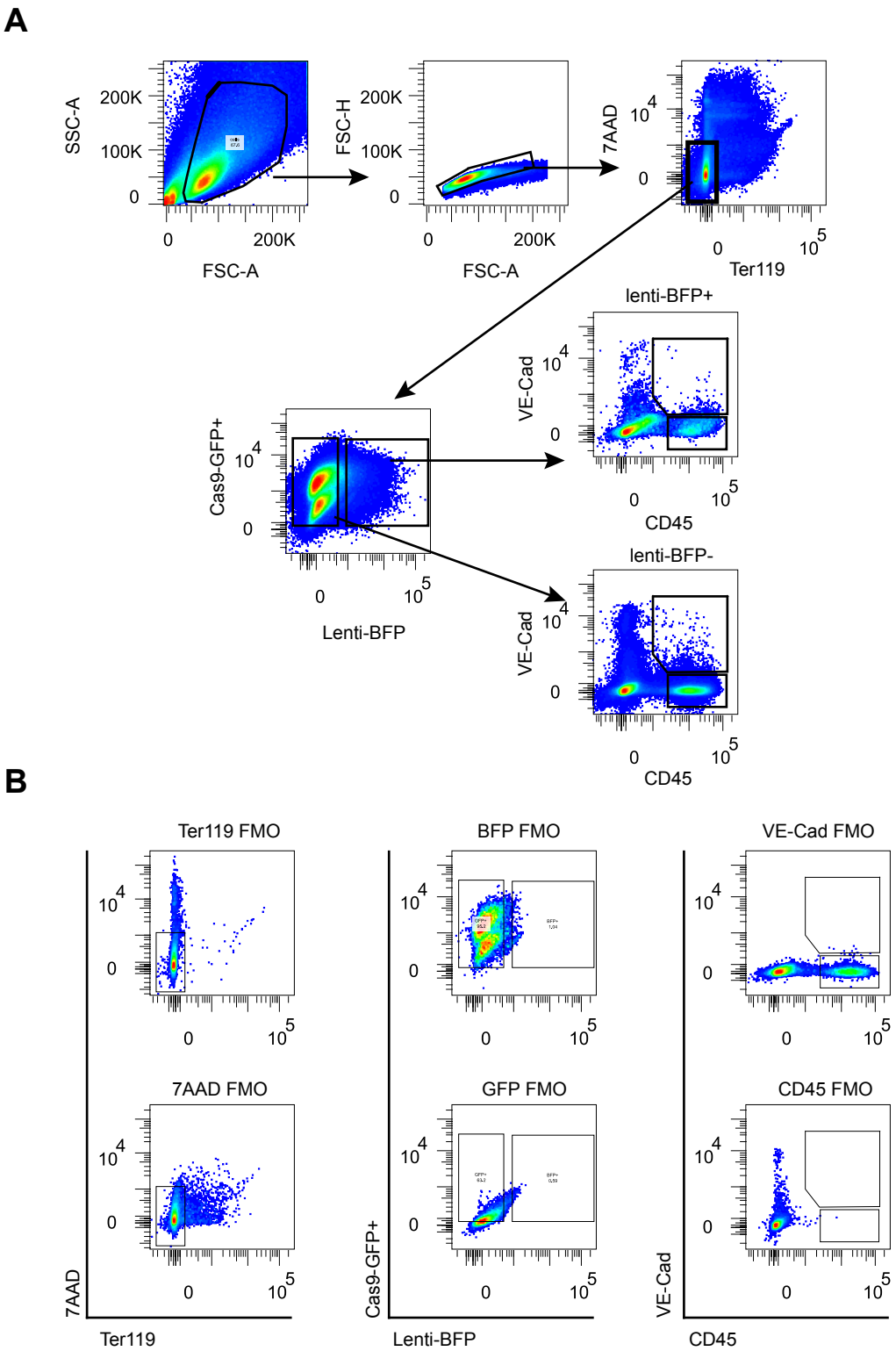
Appendix Figure 6.2: Characterization of mESC harboring a Ikzf1-28 cis-element deletion. (A) Schematic of mESC differentiation based on Pearson et al., 2015. Flk1+ cells at day 4 were sorted and replated onto gelatin in presence of cytokines. (B) Gating strategy for mESC differentiations. (C) Flow cytometry profile for Ikzf1-28-/- mESC. (D) Purity checks for CD41lo and CD45 sorted populations used for RNA analysis. (E) FMO controls used to draw gates in C.



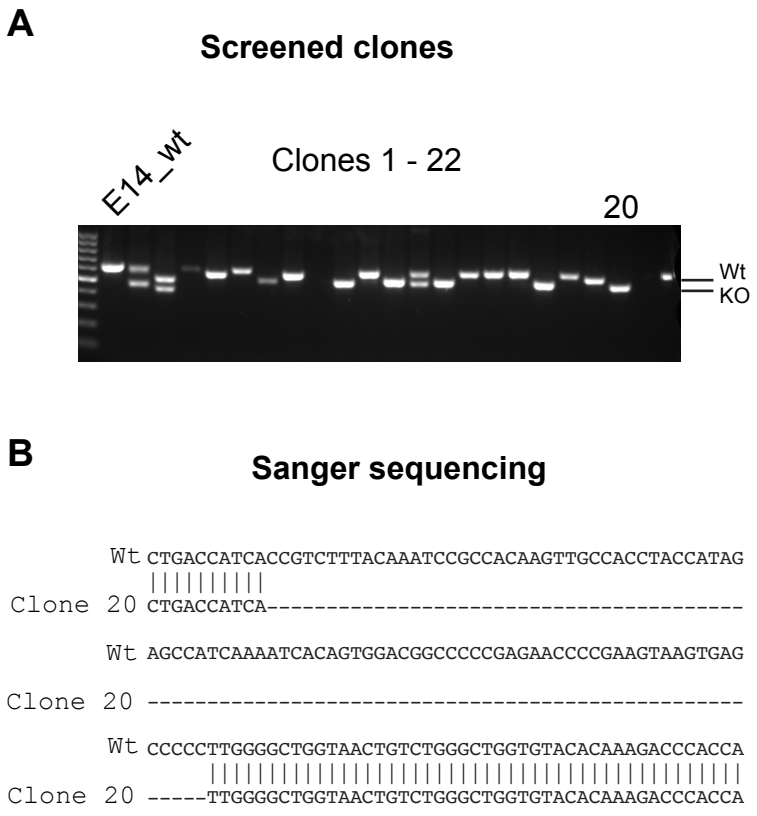
Appendix Figure 6.3: Characterization of Cas9 mouse used for ex vivo genetic perturbations. (A) Schematic of Cas9-EGFP transgene inserted into Rosa26 locus (Platt et al., 2014). (B) gRNA expression plasmid was obtained from addgene (plasmid 50946). Plasmid was modified to express two gRNA, each with own U6 promoter. (C) Experimental outline for ex vivo perturbation assay. Caudal portion of Cas9-GFP embryos was dissected, disassoiated into single cells, infected with lentivirus expressing gRNAs, re-aggregated with OP9 cells and cultured for 7 days on a PVDF membrane at the gas-liquid interface. (D) Flow cytometry profiles of Cas9-GFP E10.5 embroys to assess expression levels of Cas9-GFP in hematopoietic cells during EHT.



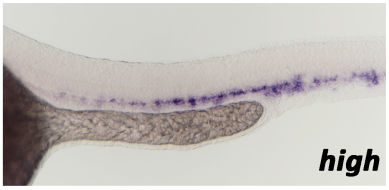
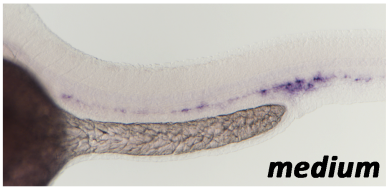
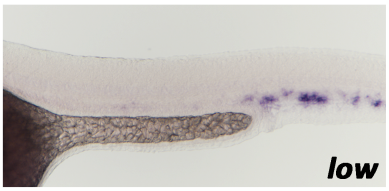
Appendix Figure 6.4: CRISPR/Cas9 targeting strategies for Ikzf1 gene knockout and -28 knockout. UCSC snapshot of AGM ATAC-seq for the Ikzf1 locus. (A) Schematic looking closer at exon 3 for creating gene knockout. Four different gRNAs were designed to target exon 3. Each is represented in a different color. gRNAs were cloned in tandem into an expression plasmid harboring Wt Cas9. (B) Schematic looking closer at -28 cis-element. 2 sets of 2 gRNA were designed with each pair containing gRNAs that flanked the -28 element. gRNAs were cloned in tandem into an expresison plasmid harboring Wt Cas9.



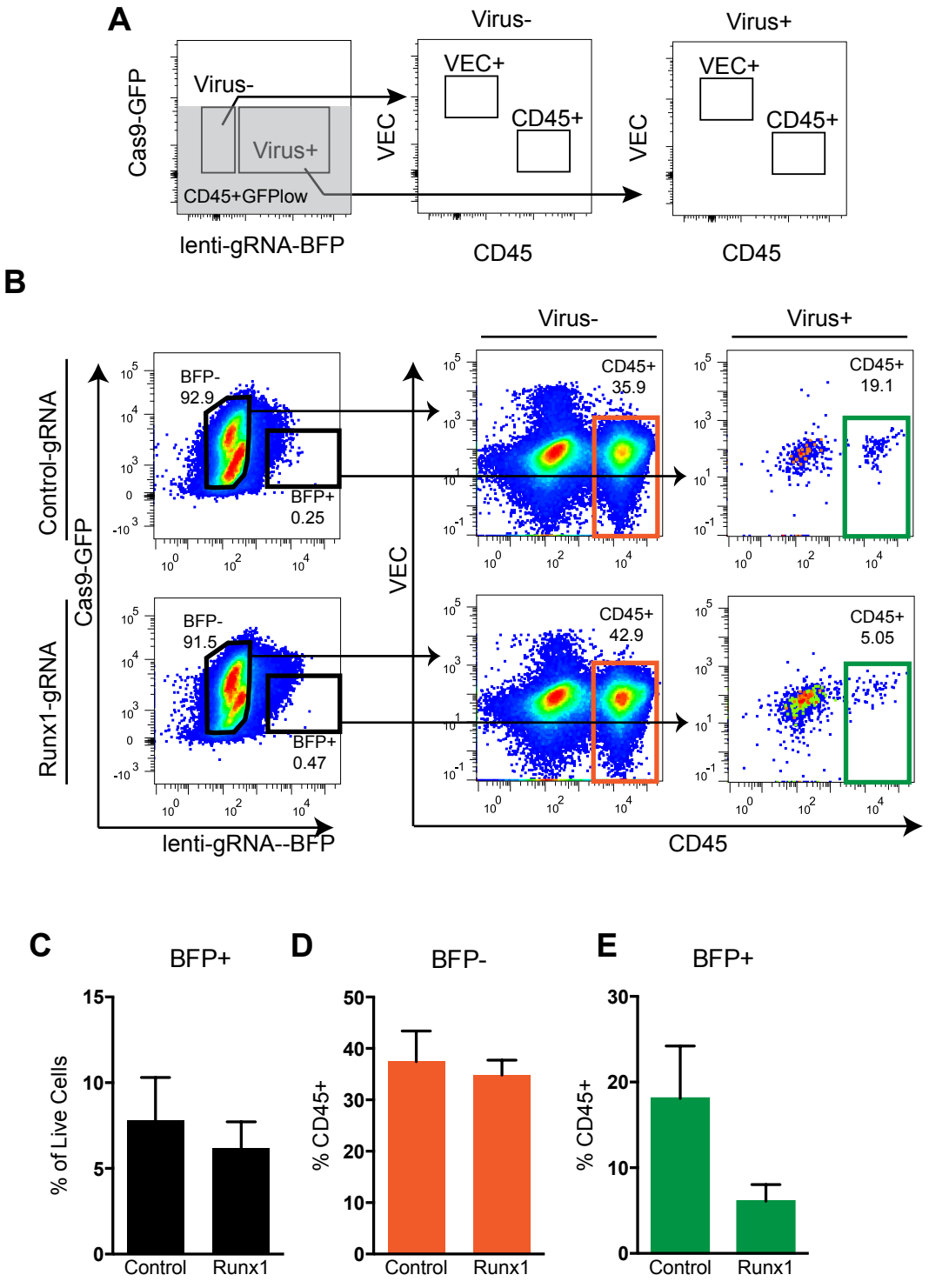
Appendix Figure 6.5: Flow controls for ex vivo targeting of Ikzf1-28. (A) Flow cytometry profile for d7 Re-aggregate cultures. (B) FMOs used to draw gates in A.



Appendix Figure 6.6: Generation and characterization of Runx1-/- mESC. (A) gDNA PCR used to confirmed targeted clones. (B) Sanger sequencing track confirmed deletion of key amino acids from Runx1 exon 4.

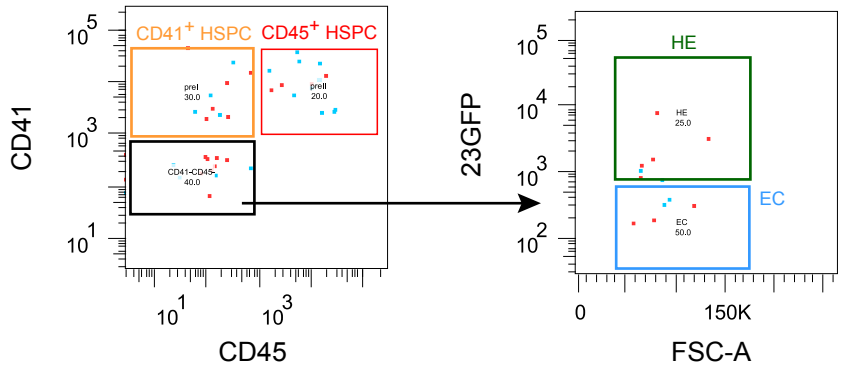
***runx1*$^{W84X/+}$ InX** **33 hpf** **ikzf1 ISH**	***runx1*$^{+/+}$**	***runx1*$^{W84X/+}$**	***runx1*$^{W84X/W84X}$**
	**30 %**	**31 %**	**/**
	**70 %**	**58 %**	**/**
	**/**	**11 %**	**100 %**

Appendix Figure 6.7: Ikzf1 in situ in zebrafish at 33 hpf reveals dose-dependent affect of Runx1 on Ikzf1 expression. Genotypes were confirmed by PCR after fixation and staging based on expression. Work was done by Florian Bonkhofer (Patient Lab).

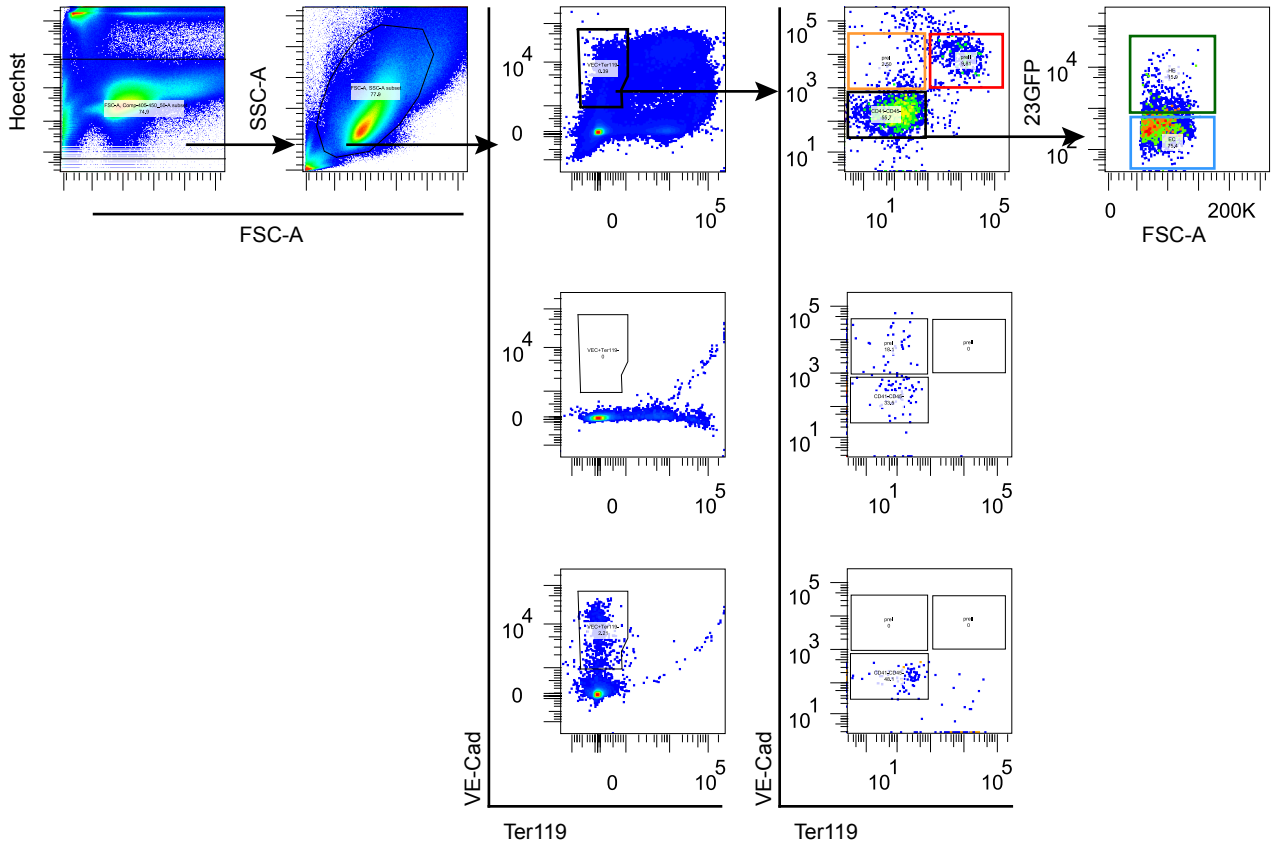


Appendix Figure 6.8: Characterization of ex vivo targeting of Runx1. (A) Gating and sorting strategy. (B) Flow cytometry profiles for control gRNA and Runx-gRNA. (C-E) Quantification of flow cytometry profiles in B.

A



B



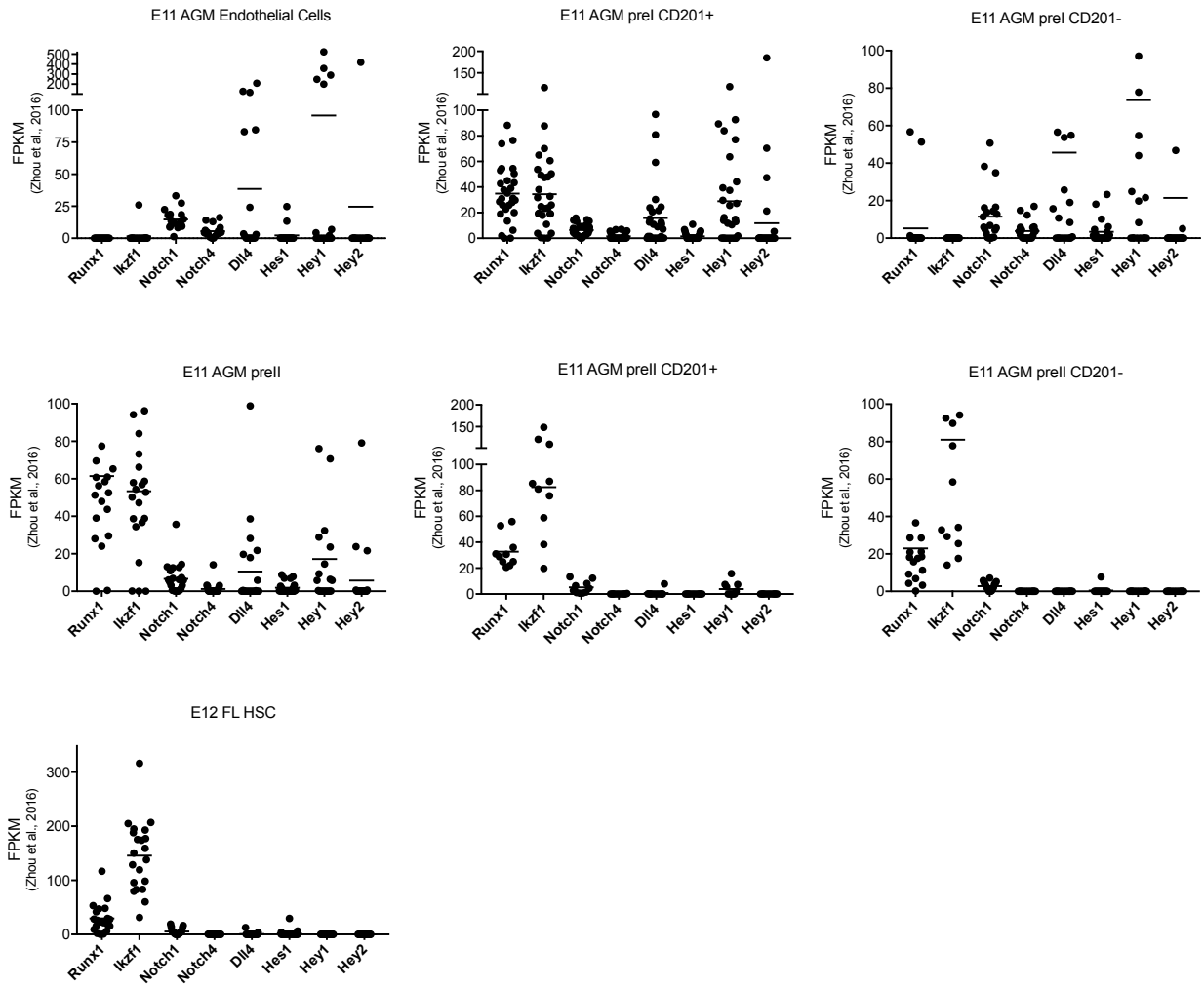
Appendix Figure 6.9: Flow controls for E10.5 single cell qRT-PCR analysis. (A) Position of all sorted cells captured using index sequencing. (B) FMOs used to draw gates in (A).

A

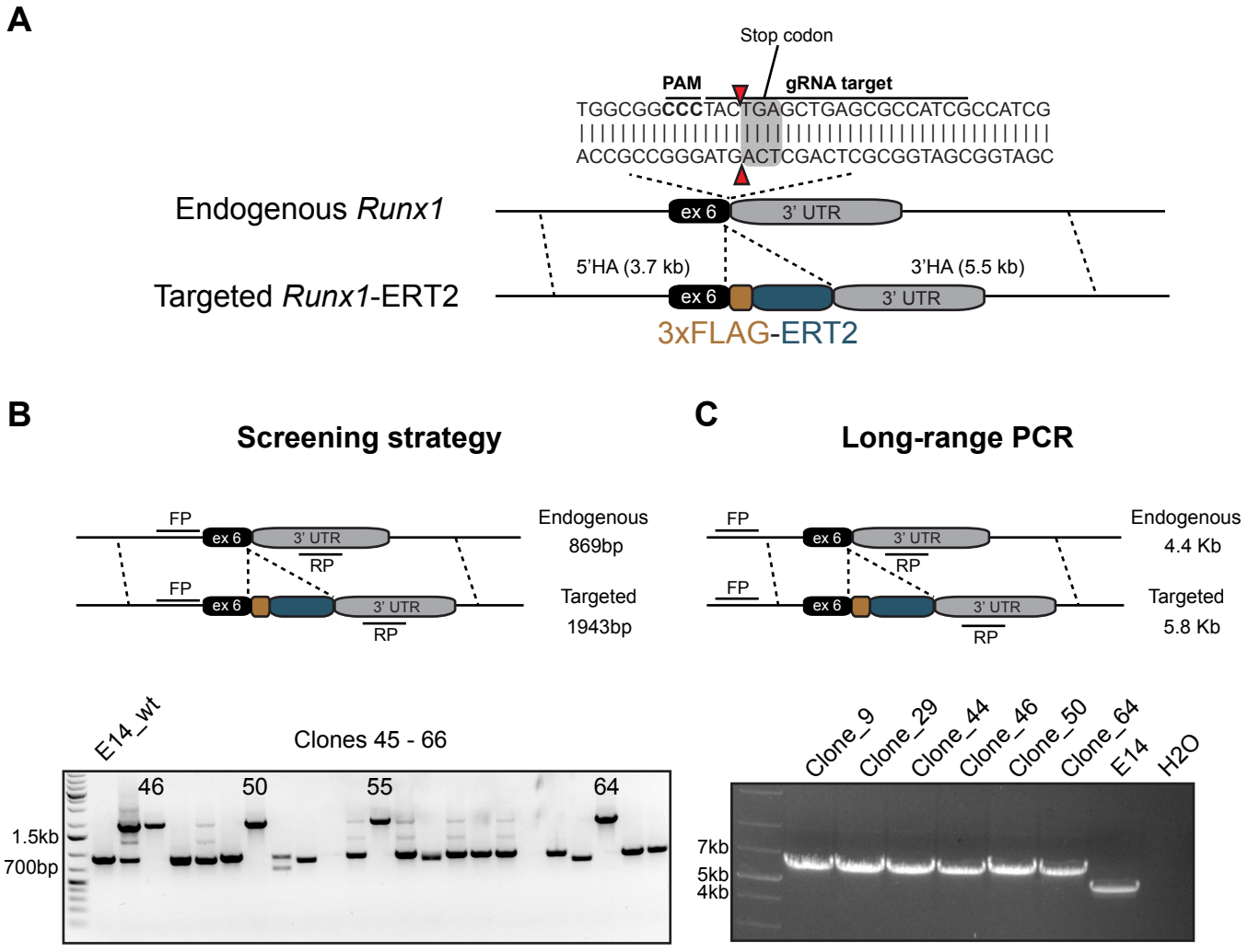
Stage	Tissue	Cell Type	Cell surface profile							
			Ter119	CD31	VE-Cad	CD41	CD45	cKit	CD201	CD43
E11	AGM	Endothelial cell	(-)	(+)	(+)	(-)	(-)	NA	NA	(-)
E11	AGM	Functional preHSC Type I	(-)	(+)	NA	(+)	(-)	(+)	(+)	NA
E11	AGM	Non-functional preHSC Type I	(-)	(+)	NA	(+)	(-)	(+)	(-)	NA
E11	AGM	preHSC Type II	(-)	(+)	NA	(+)	(+)	(+)	(+/-)	NA
E11	AGM	Functional preHSC Type II	(-)	(+)	NA	(+)	(+)	(+)	(+)	NA
E11	AGM	Non-functional preHSC Type II	(-)	(+)	NA	(+)	(+)	(+)	(-)	NA
E12	Fetal Liver (FL)	dHSC (Lin-Sca1+Mac1lowCD201	Lin-Sca1+Mac1lowCD201+							

*Non-functional = does not cotain LT-HSC

B



Appendix Figure 6.10: sc-RNA-seq of EHT at E11 from Zhou et al., 2016 supports Runx1-Ikzf1-Notch regulatory interaction. (A) Table with information regarding populations from Zhou et al., 2016. (B) FPKM for single cells for genes listed on the x-axis.

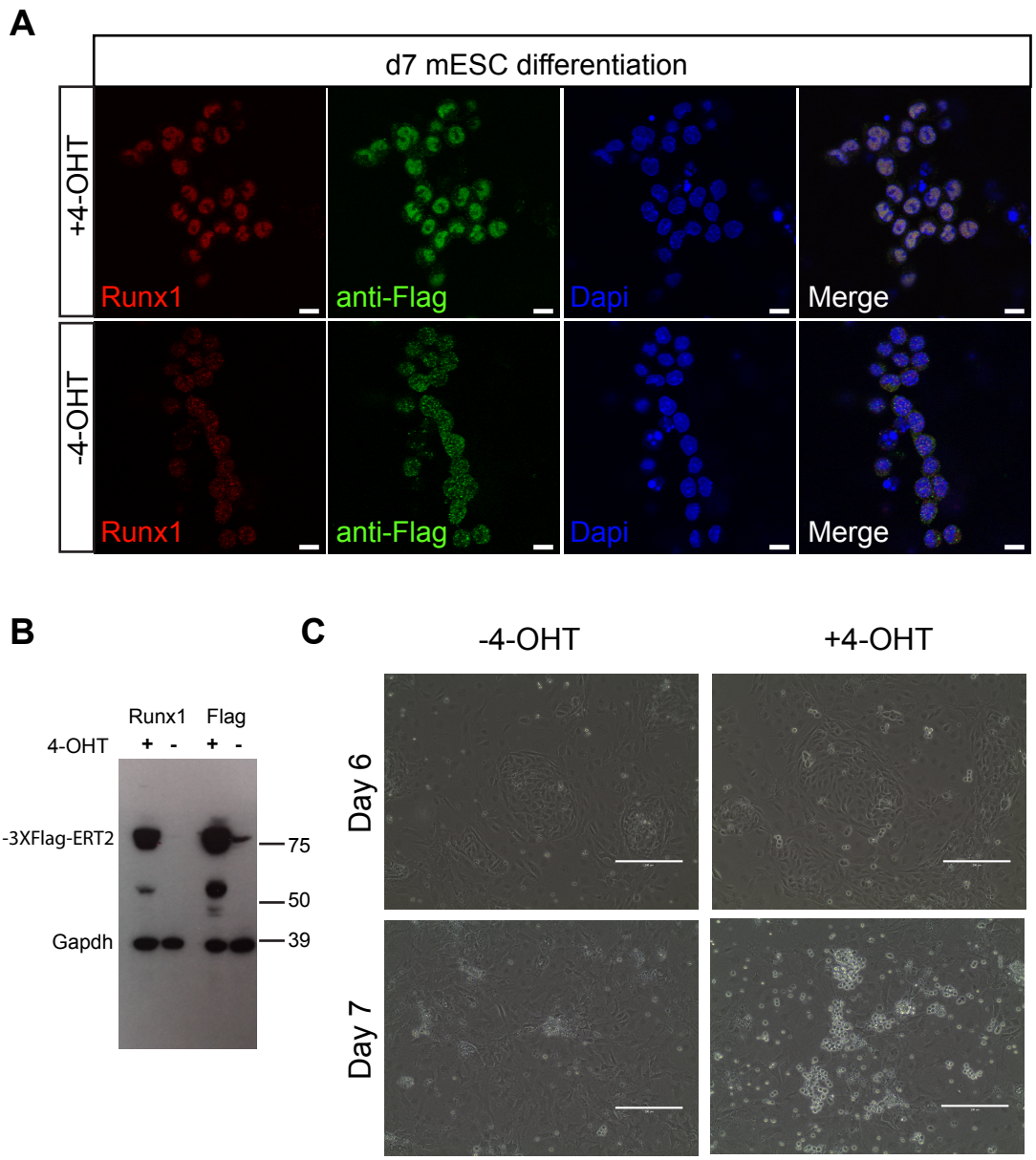


D

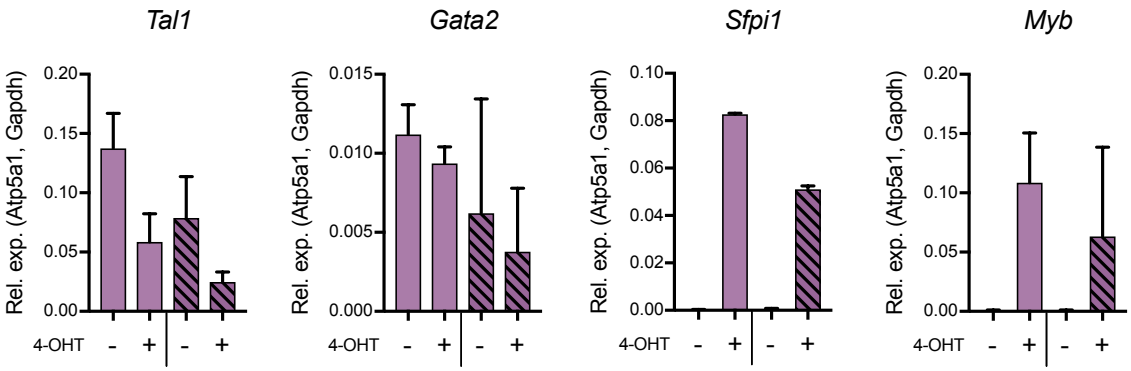
Colony screening statistics

Total Clones picked:	72
Clones expaned:	66
No. targeted:	34 / 66 (51.5%)
No. heteozygous:	22 / 66 (36.4%)
No. homozygous:	10 / 66 (15.2%)

Appendix Figure 6.11: Targeting ERt2 domain to endogenous *Runx1*. (A) Schematic showing location of gRNA used to cut immediately preceding the Runx1 stop codon. Homology arms (HA) were used to insert a 3X-FLAG tag. (B) PCR screening strategy to idenify positive mESC clones. (C) Long range (LR) PCR with primer pairs outside homology arms were used to confirm correctly targeted clones. (D) Efficiency of targeting events. Used L755,507 NHEJ-inhibitor for targeting.



Appendix Figure 6.12: Characterization of Runx1-ERt2 line. (A) IF for Runx1 and Flag tag showing colocalization of Runx1 and Flag. Also showed Runx1 sequestered to cytoplasm in the absence of 4-OHT. (B) Western blot confirming subcellular localization of Runx1 in the absence of 4-OHT. Gapdh used as loading contol. (C) Bright field images from EVOS microscope showing blood formation (round refractory clusters) present only in the 4-OHT treated.



Appendix Figure 6.13: qRT-PCR of Runx1-ERt2::Ikzf1-/- for EHT related genes.